

**TESIS DOCTORAL**

**2016**

**RECOMMENDER SYSTEMS IN SOCIAL SETTINGS:  
PROPOSAL, DEVELOPMENT AND TESTING IN REAL  
SCENARIOS.**

**ANGEL CASTELLANOS GONZÁLEZ**

**MSc IN COMPUTER SCIENCE, UNED**

**DOCTORAL PROGRAMME IN INTELLIGENT  
SYSTEMS**

**ANA GARCÍA SERRANO**

**JUAN CIGARRÁN RECUERO**



**TESIS DOCTORAL**

**2016**

**RECOMMENDER SYSTEMS IN SOCIAL SETTINGS:  
PROPOSAL, DEVELOPMENT AND TESTING IN REAL  
SCENARIOS.**

**ANGEL CASTELLANOS GONZÁLEZ**

**MSc IN COMPUTER SCIENCE, UNED**

**DOCTORAL PROGRAMME IN INTELLIGENT  
SYSTEMS**

**ANA GARCÍA SERRANO**

**JUAN CIGARRÁN RECUERO**





This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

2016 Ángel Castellanos González



**“Das Beste oder nichts”**

*Gottlieb Daimler*





# Acknowledgements

---

I would like to express the deepest gratitude to my advisors Ana García Serrano and Juan Cigarrán Recuero. Without the hundreds of hours they spent on me, their guidance and their valuable help this dissertation would not have been possible. I do especially appreciate the careful and extensive reviews Ana has done to every piece of my works, as well as its expert supervision and its uncountable advices and ideas. I also thank the patience of Juan helping me to turn all the scribbles in my whiteboard into valuable ideas for this work.

I can only thank Prof. Ernesto William de Luca and the whole Information Science Department of the University Of Applied Sciences Of Potsdam who hosted me during more than 7 months. The result of those two research stays, the discussions and all the knowledge acquired have had a significant impact in this dissertation.

I have to thank all my co-authors – Xaro Benavent, Ruben Granados, Thebin Lee or David Hernandez Aranda among others - who have represented and indispensable collaboration during all these years. I cannot forget all my former and current colleagues in the NLP & IR Group and the hours I spent with them during the writing of this dissertation.

Although they have not directly collaborated with my research, I want to remark here all the help, motivation and support I found in my family, friends and, of course, my girlfriend. Finally, the most important, my thanks to my parents. They told me many years ago that I should study and it is what I have done for all this time. I could never have imagined how important their support was going to be. Most of what I am I owe to them.

I thank all people who have been fighting for free public education.

**Institutional Acknowledgements.** The research presented in this thesis was partially supported by the Regional Government of Madrid under Research Network MA2VIRMR (S2009/TIC-1542), for Spanish Ministry of Science and Innovation under project BUSCAMEDIA (CEN-20091026) project HOLOPEDIA (TIN2010-21128-CO2) and project VOXPOPULI (TIN2013-47090-C3-1-P) and by UNED under a PhD grant (FPI-UNED 2014).

# Index

---

ACKNOWLEDGEMENTS .....	IV
INDEX .....	VI
RESUMEN .....	VIII
ABSTRACT .....	XII
LIST OF TABLES .....	XIV
LIST OF FIGURES .....	2
INTRODUCTION .....	2
1.1    MOTIVATION .....	3
1.2    SCOPE OF THIS WORK .....	11
1.3    OBJECTIVES .....	12
1.4    CONTRIBUTIONS .....	13
1.5    METHODOLOGY .....	13
STATE OF THE ART .....	15
2.1    KNOWLEDGE-BASED REPRESENTATIONS .....	16
2.2    TOPIC DETECTION .....	20
2.2.1.    Classification .....	20
2.2.2.    Clustering .....	21
2.2.3.    Probabilistic .....	22
2.2.4.    Graph-based .....	23
2.2.5.    Formal Concept Analysis .....	24
2.2.6.    Deep Learning for Topic Modelling .....	26
2.3    RECOMMENDER SYSTEMS .....	31
2.3.1.    Algorithms and Evaluation .....	33
2.3.2.    Specific Application Domains .....	43
2.3.3.    Related Topics .....	58
2.4    DISCUSSION .....	67
FORMAL CONCEPT ANALYSIS .....	70
3.1    FCA AT A GLANCE .....	71
3.2    STABILITY .....	75

3.3	COMPLEXITY REDUCTION .....	76
	CONCEPT-BASED MODELLING .....	80
4.1	APPLICATION SCENARIO 1: TOPIC DETECTION .....	82
4.1.1.	The RepLab 2013 Campaign.....	83
4.1.2.	Evaluation Strategy .....	86
4.1.3.	Topic Annotation by means of FCA .....	93
4.1.4.	Other Topic Annotation Approaches .....	101
4.1.5.	Experimentation.....	106
4.1.6.	Discussion.....	134
4.2	APPLICATION SCENARIO 2: IMAGE DIVERSIFICATION .....	138
4.2.1.	Test bed: Retrieving Diverse Social Images Task at MediaEval.....	139
4.2.2.	FCA-based Work Proposal .....	141
4.2.3.	Experimentation.....	147
4.2.4.	Discussion.....	154
	LOD-BASED REPRESENTATION ENHANCEMENT.....	156
5.1	KNOWLEDGE-BASED REPRESENTATION PROPOSAL.....	158
5.1.1.	DBpedia Modelling.....	159
5.1.2.	EWN Modelling .....	164
5.2	APPLICATION SCENARIO: IMPROVING THE TOPIC DETECTION .....	167
5.3	DISCUSSION .....	177
	RECOMMENDATION: FROM FORMAL CONCEPTS TO USER PREFERENCES	
	180	
6.1	FCA FOR RECOMMENDATION: PRELIMINARY APPROACH.....	182
6.1.1.	Recommendation Approach .....	182
6.1.2.	Application Scenarios .....	183
6.2	TOWARDS A COMMON CONCEPT-BASED REPRESENTATION SPACE .....	193
6.2.1.	Recommendation Approach .....	196
6.2.2.	Application Scenarios.....	198
	CONCLUSIONS .....	258
7.1	MAIN OUTCOMES .....	261
7.2	FUTURE DIRECTIONS.....	261

# Resumen

---

Desde los trabajos más tempranos en Sistemas de Recomendación, el principal objetivo de esta línea de investigación es el de asistir a los usuarios en el descubrimiento de contenidos relevantes entre la abrumadora cantidad de datos disponibles en la web. Los sistemas de recomendación fueron concebidos en los 90 con el auge de Internet y el incremento de los datos disponibles que ello conllevó. Hoy en día, con la explosión de los contenidos generados por los usuarios en el contexto de la Web 2.0, la necesidad de sistemas de recomendación es la misma que en los años 90, sino más, pero los problemas relacionados que deben abordar estos sistemas son más y más complejos.

Este contexto de los contenidos generados por usuarios y la web social afecta directamente al rendimiento de los sistemas de recomendación, siendo uno de los problemas más acuciantes el modelado con precisión de las preferencias de los usuarios. Los trabajos iniciales en el área principalmente abordaba este aspecto desde el punto de vista de los sistemas de filtro colaborativo; sin embargo, el uso de rasgos basado en el contenido de los ítems está cada día más extendido. De entre estos sistemas basados en el contenido de los ítems, la mayoría de los trabajos propuestos en la literatura normalmente dependen del modelado de usuarios e ítems por separado: los perfiles de usuario son analizados y modelados de acuerdo a rasgos basados en el contenido para luego buscar los ítems más relacionados con este modelo. Esta metodología introduce el problema de la separación entre usuarios e ítems; esto es, la separación entre sus ambos espacios de representación.

Para superar este problema, esta tesis propone un espacio común de representación para recomendación. Conceptualmente, modelar las dos dimensiones en conjunto parece ser el método más sensato. En particular, esta tesis propone un modelado conceptual de usuarios-ítems basado en conceptos a través de la aplicación del Análisis de Conceptos Formales (ACF). Nuestra hipótesis principal es que la abstracción basada en conceptos de los perfiles de usuarios e ítems que ACF genera facilitará la mejor identificación de relaciones entre los usuarios y los ítems, las cuales pueden ser entendidas como preferencias de usuario. Por lo tanto, usuarios e ítems serán representados en un espacio común mediante las preferencias de usuario descubiertas (en la forma de *conceptos formales*), organizadas jerárquicamente de un modo natural de acuerdo a su especificidad. De esta manera, se espera superar el problema de la separación entre usuarios e ítems, mejorando de este modo el proceso de recomendación.

De cara a probar nuestra hipótesis de trabajo, hemos aislado el proceso de la evaluación del rendimiento de nuestra propuesta. La razón de ello es la de primero evaluar el rendimiento de ACF para la representación de datos, para luego evaluar esta representación cuando se aplique a la recomendación de contenidos. Para ello, hemos aplicado nuestro modelado basado en ACF a dos escenarios independientes de la tarea de recomendación (Topic Detection @ Replab 2013 e Image Diversification @ MediaEVAL 2014 and 2015). La evaluación de ACF en estos escenarios prueba la idoneidad general de este modelado, logrando resultados en el estado del arte para ambos escenarios. Esta evaluación también prueba que, al contrario que otras propuestas en la literatura, nuestro sistema se ve a penas afectado por los diferentes parámetros relacionados con su funcionamiento. Finalmente, hemos abordado una extensiva comparación, en relación a la calidad de las representaciones generadas, con otras conocidas metodologías para la representación de datos (Hierarchical Agglomerative Clustering y Latent Dirichlet Allocation). Como es probado por esta comparación, la representación basada en ACF tiene más calidad y presenta un comportamiento más homogéneo que el resto de metodologías.

En una etapa posterior, hemos extendido esta metodología mediante la integración de rasgos semánticos relacionados con el contenido de los ítems. No solo este modelo logra mejorar la etapa de modelado, sino que también posibilita una representación de más alto nivel y más abstracta, la cual resulta en modelos más compactos y ligeros. Este aspecto facilita abordar los retos relacionados con la aplicación de nuestra propuesta a escenarios sociales (Topic Detection @ Replab 2013).

Hemos finalmente aplicado nuestro modelo FCA para crear un espacio de representación común para la recomendación de contenidos. En primer lugar, hemos llevado a cabo una aproximación preliminar para probar la idoneidad de nuestra propuesta en escenarios de recomendación sociales (NEWSREEL 2014 y ESWC LOD-RecSys 2014). Del análisis de los resultados de esta experimentación preliminar, hemos refinado nuestra propuesta para crear un espacio común de recomendación. La evaluación de esta propuesta, llevada a cabo en diferentes escenarios sociales (UMAP 2011 Dataset y ESWC LOD-RecSys 2015), hemos analizado también los diferentes aspectos envueltos en el proceso de recomendación, probando que, cuando están disponibles, el uso de modelos basados en rasgos semánticos de alto nivel conlleva una recomendación más precisa que cuando el texto en bruto es utilizado. Hemos confirmado también que, como ya dicho previamente por otros trabajos experimentales, en entornos sociales, los sistemas que aplican rasgos basados en contenido mejoran a los sistemas basados en filtro colaborativo.

Finalmente, este análisis extensivo demuestra que el buen rendimiento de nuestro modelo para la representación de datos permanece cuando es aplicado a la tarea de recomendación. En particular, nuestro espacio de representación común basado en ACF mejora el rendimiento de otros sistemas de recomendación reportados en la literatura como estado del arte para la tarea.



# Abstract

---

Since the earlier works in recommender systems, the main aim of this research area is to assist users in the finding of relevant content among the overwhelming amount of data available on the Web. Recommender systems research interest started in the 90s with the rise of the Internet and the increasing of available data that it entailed. Nowadays, with the explosion of user-generated content in the context of the Web 2.0, the necessity of recommender systems is the same than in the 90s, but the related problems that they have to face are more challenging every day.

This context of user-generated content and social web hinders the implementation of recommender systems, being one of the most acute the accurate modelling of user preferences. The initial works on the literature mainly addressed this issue from the perspective of Collaborative Filtering systems; however, the use of Content-based features is becoming more widespread. Among these Content-based systems, most of the works in the literature usually rely on the modelling of user and item dimensions by separate: user profiles are analyzed and modelled according to their Content-based features to then find the items that are most closely related to this model. This methodology introduces the problem of the user-item gap; i.e., the gap between both representation spaces.

To overcome this problem, this thesis proposes a common representation space for recommendation. The modelling of both dimensions together in a common representation space appears to be, conceptually, the most sensible choice. In particular, we propose on a concept-based user-item modelling generated through the application of Formal Concept Analysis (FCA). Our main hypothesis is that the concept-based abstraction of user and item profiles that FCA generates will facilitate the better identification of user-item relationships, which can be understood by user preferences. Therefore, users and items will be represented in a common space by means of the unfolding user preferences (in the form of *formal concepts*), hierarchically organized in a natural way according to this specificity. In this way, it is expected to overcome the user-item gap problem, thus improving the recommendation process.

In order to test our claim, we have isolated the evaluation of the performance of our proposal. The rationale is to firstly evaluate the performance of FCA for data representation to then evaluate this representation when applied for the recommendation task. To that end, we have applied the proposed FCA-modelling to two different scenarios

independently of the recommendation task (Topic Detection @ Replab 2013 and Image Diversification @ MediaEVAL 2014 and 2015). The evaluation of FCA in these scenarios proves its overall suitability, achieving state-of-the-art results for both scenarios. This evaluation proves as well that, in contrast to other proposals in the literature, our system is barely affected by the different parameters related to its operation. Finally, we have addressed an extensive comparison to other well-known data representation methodologies (namely, Hierarchical Agglomerative Clustering and Latent Dirichlet Allocation) in relation to the quality of the generated representations. As proven by this comparison, the FCA-based representation has more quality and presents a more homogeneous behaviour than the rest of methodologies.

In a later step, we have extended this modelling by integrating semantic features related to the item content. Not only does this enhanced model improve the modelling step, but it also enables a higher-level and more abstract representation, which results in lighter and more compact model. This aspect facilitates the overcoming of the challenges related to the application of our proposal to social-based real scenarios (i.e., Topic Detection @ Replab 2013).

We have finally applied our FCA-based model to the recommendation task. We have firstly conducted a preliminary experimentation to prove the suitability of our proposal in social-based recommendation scenarios (NEWSREEL 2014 and ESWC LOD-RecSys 2014). From the analysis of the outcome of this preliminary experimentation, we have refined our FCA-based recommendation approach to create a common representation space for recommendation. Throughout its evaluation carried out in different social-based scenarios (UMAP 2011 Dataset and ESWC LOD-RecSys 2015), we have analysed the different aspect involved in the recommendation process, proving that, when available, higher-level semantic features entails more accurate recommendations than when raw textual descriptions are applied. We have confirmed as well that, as stated by other experimental works in the literature, in these social-based environments, systems using Content-based features outperform Collaborative Filtering systems

Finally, this extensive analysis confirms our initial hypothesis in regards to our proposal. The high performance of our model for data representation remains when applied to the recommendation task. In particular, our FCA-based common representation space outperforms other recommender systems reported in the literature for the addressed tasks.

# List of Tables

---

TABLE 3.1 – EXAMPLE OF FORMAL CONTEXT .....	71
TABLE 3.2 – EXAMPLE OF THE SET OF GENERATED FORMAL CONCEPTS .....	72
TABLE 4.1 – ANALYSIS OF THE NUMBER OF TWEETS BY DOMAIN.....	85
TABLE 4.2 – STATISTICS ON TOPIC DETECTION TASK.....	86
TABLE 4.3 – CLUSTERING NOTATION .....	88
TABLE 4.4 – FORMAL CONTEXT OF A SET OF TWEETS AND THEIR CORRESPONDING SET OF ATTRIBUTES .....	94
TABLE 4.5 – ORIGINAL FORMAL CONTEXT TO APPLY THE BALANCED ALGORITHM .....	97
TABLE 4.6 – BALANCED ALGORITHM. FIRST ITERATION .....	98
TABLE 4.7 – BALANCED ALGORITHM SECOND ITERATION .....	99
TABLE 4.8 – BALANCED ALGORITHM THIRD ITERATION.....	99
TABLE 4.9 – EXAMPLE OF TWEET REPRESENTATIONS .....	108
TABLE 4.10 – F-MEASURE RESULTS OBTAINED BY THE DIFFERENT TWEET REPRESENTATIONS.....	109
TABLE 4.11 – IMPACT OF LOWER AND UPPER THRESHOLD VALUES ON THE FINAL CONCEPT LATTICE.....	111
TABLE 4.12 – RESULTS USING A 0.5% AND A 0.1% LOWER THRESHOLD.....	115
TABLE 4.13 – ATTRIBUTE REDUCTION ALGORITHM THRESHOLD ANALYSIS.....	116
TABLE 4.14 – RELIABILITY, SENSITIVITY AND F-MEASURE VALUES VARYING THE STABILITY VALUE FROM 0.2 TO 0.9 ON A SETUP OF 0.5%, 1% AND 5% LOWER THRESHOLD VALUES .....	117
TABLE 4.15 – RELIABILITY, SENSITIVITY AND F-MEASURE VALUES FOR DIFFERENT CLUSTER SELECTION SETUPS .....	121
TABLE 4.16 – RELIABILITY, SENSITIVITY AND F-MEASURE VALUES FOR DIFFERENT TOPIC ADAPTATION EXPERIMENTS .....	122
TABLE 4.17 – EXAMPLES OF TOPICS GENERATED BY THE DIFFERENT APPROACHES ..	123

TABLE 4.18 – HAC-BASED INTERNAL RESULTS .....	124
TABLE 4.19 – FCA-BASED INTERNAL RESULTS.....	125
TABLE 4.20 – LDA-BASED INTERNAL RESULTS .....	125
TABLE 4.21 – BASELINE INTERNAL RESULTS .....	129
TABLE 4.22 – BASELINE EXTERNAL RESULTS.....	130
TABLE 4.23 – HAC RESULTS.....	131
TABLE 4.24 – FCA-BASED RESULTS.....	132
TABLE 4.25 – LDA-BASED RESULTS .....	133
TABLE 4.26 – SUMMARY OF THE COMPARISON AMONG HAC, FCA AND LDA.....	137
TABLE 4.27 – FIGURES ABOUT MEDIAEVAL 2014 AND 2015 DATASET .....	140
TABLE 4.28 – KLD-BASED REPRESENTATION EXAMPLE .....	142
TABLE 4.29 – OFFICIAL RESULTS OF OUR APPROACHES FOR RETRIEVING DIVERSE SOCIAL IMAGES TASK .....	148
TABLE 4.30 – OFFICIAL RESULTS OF OUR APPROACHES FOR RETRIEVING DIVERSE SOCIAL IMAGES TASK 2014 — EXTENDED RESULTS .....	151
TABLE 4.31 – OFFICIAL RESULTS OF OUR APPROACHES FOR RETRIEVING DIVERSE SOCIAL IMAGES TASK 2015 — EXTENDED RESULTS .....	151
TABLE 5.1 – FCA DBPEDIA KNOWLEDGE ORGANIZATION STATISTICS.....	161
TABLE 5.2 – EXAMPLE OF DATA REPRESENTATION.....	163
TABLE 5.3 – FCA EWN KNOWLEDGE ORGANIZATION STATISTICS.....	164
TABLE 5.4 – EXAMPLE OF DATA REPRESENTATION USING THE EWN REPRESENTATION .....	167
TABLE 5.5 – TWEET REPRESENTATION EXAMPLE.....	170
TABLE 5.6 – TOPIC DETECTION RESULTS.....	171
TABLE 6.1 –FEATURE DISTRIBUTION ANALYSIS .....	230
TABLE 6.2 – DBPEDIA MODEL STATISTICS.....	246

TABLE 6.3 – OFFICIAL RESULTS FOR TOP-N RECOMMENDATION TASK COMPARED TO THE BASELINES.....	253
--	-----



# List of Figures

---

FIGURE 2.1 – TAXONOMY OF HYBRID SYSTEMS (EXTRACTED FROM [AGGARWAL, 2016]) .....	38
FIGURE 2.2 – LINKED OPEN DATA CLOUD (HTTP://LOD-CLOUD.NET/)	45
FIGURE 2.3 – EXAMPLE OF RANK-2 MATRIX FACTORIZATION .....	60
FIGURE 3.1 – NEXT NEIGHBOURS ALGORITHM.....	74
FIGURE 3.2 – EXAMPLE OF CONCEPT LATTICE REPRESENTATION .....	75
FIGURE 4.1 – CONCEPT LATTICE CORRESPONDING TO THE FORMAL CONTEXT OF TABLE 4.5 .....	95
FIGURE 4.2 – THE SAME CONCEPT LATTICE AS IN FIGURE 4.1 BUT WITH QUANTITATIVE INFORMATION ON OBJECT COUNT .....	96
FIGURE 4.3 - FINAL FORMAL CONTEXT, FILTERED BY THE BALANCED ALGORITHM AND ITS CORRESPONDING CONCEPT LATTICE .....	100
FIGURE 4.4 – EXAMPLE OF <i>DENDROGRAM</i> .....	103
FIGURE 4.5 – GRAPHICAL MODEL REPRESENTATION OF LDA .....	104
FIGURE 4.6 – IMPACT OF THE LOWER AND UPPER THRESHOLD ON THE FINAL PERFORMANCE .....	112
FIGURE 4.7 – TERM DISTRIBUTION AND THRESHOLDS .....	113
FIGURE 4.8 – RELIABILITY, SENSITIVITY AND F-MEASURE VALUES FOR DIFFERENT LOWER THRESHOLDS USING A 25% UPPER THRESHOLD .....	114
FIGURE 4.9 – RELIABILITY, SENSITIVITY AND F-MEASURE VALUES FOR 0.1%, 0.5%, 1% AND 5% LOWER THRESHOLD VALUES.....	115
FIGURE 4.10 – F-MEASURE VALUES VARYING THE STABILITY VALUES FROM 0.2 TO 0.9 .....	118
FIGURE 4.11 – DISTRIBUTION OF F-MEASURE VALUES IN TERMS OF STABILITY VALUES .....	119

FIGURE 4.12 – COMPARISON OF THE DUNN INDEX RESULTS FOR HAC, LDA AND FCA .....	127
FIGURE 4.13 – COMPARISON OF THE DAVIES-BOULDIN INDEX RESULTS FOR HAC, LDA AND FCA .....	127
FIGURE 4.14 – COMPARISON OF THE SILHOUETTE COEFFICIENT RESULTS FOR HAC, LDA AND FCA .....	128
FIGURE 4.15 – COMPARISON OF THE CALINSKI-HARABASZ INDEX RESULTS FOR HAC, LDA AND FCA.....	128
FIGURE 4.16 – HAC-BASED RESULTS FOR THE DIFFERENT SIMILARITY THRESHOLDS.	132
FIGURE 4.17 – FCA-BASED RESULTS FOR THE DIFFERENT STABILITY THRESHOLDS...	133
FIGURE 4.18 – LDA-BASED RESULTS FOR THE DIFFERENT CONFIGURATIONS.....	134
FIGURE 4.19 – CONCEPT LATTICE FOR THE CHRYSLER BUILDING IMAGES .....	144
FIGURE 4.20 – EXAMPLE OF FORMAL CONCEPT IMAGES (WITH INTENSION: ARTDECO, NATIONAL HISTORICAL LANDMARK) .....	144
FIGURE 4.21 – EXAMPLE OF THE ZEROS-INDUCED INDEX.....	145
FIGURE 4.22 – EXAMPLE OF A DIVERSIFIED RESULT LIST FOR THE CHRYSLER BUILDING IMAGES.....	146
FIGURE 4.23 – OFFICIAL RESULTS OF ALL THE TEXT-BASED SYSTEMS FOR RETRIEVING DIVERSE SOCIAL IMAGES TASK 2014 .....	149
FIGURE 4.24 – OFFICIAL RESULTS OF ALL THE TEXT-BASED SYSTEMS FOR RETRIEVING DIVERSE SOCIAL IMAGES TASK 2015 .....	150
FIGURE 4.25 –EXTENDED RESULTS: PRECISION BASED RESULTS .....	153
FIGURE 4.26 – EXTENDED RESULTS: CLUSTER RECALL BASED RESULTS.....	153
FIGURE 4.27 – EXTENDED RESULTS: CLUSTER RECALL BASED RESULTS.....	154
FIGURE 5.1 – FCA-BASED LOD MODELLING .....	162
FIGURE 5.2 – EXAMPLE OF EUROWORDNET HIERARCHY .....	165
FIGURE 5.3 – FCA-BASED EUROWORDNET MODEL .....	166
FIGURE 5.4 – TOPIC DETECTION RESULTS .....	173

FIGURE 5.5 – TOPIC DETECTION RESULTS (RELIABILITY, SENSITIVITY AND F-MEASURE FROM LEFT TO RIGHT).....	175
FIGURE 6.1 – EXAMPLE OF CHILDREN, PARENT AND SIBLING CONCEPTS OF A TARGET CONCEPT.....	183
FIGURE 6.2 – RESULTS FOR THE BASELINE 1 – MOST NOVEL ITEMS.....	187
FIGURE 6.3 – RESULTS FOR THE BASELINE 2 - TOP SCORED ITEMS.....	187
FIGURE 6.4 – RESULT OF THE APPROACH 1: FCA BASED RECOMMENDATION – MOST NOVEL.....	188
FIGURE 6.5 – RESULTS OF THE APPROACH 2: FCA-BASED RECOMMENDATION – TOP SCORED.....	189
<b>FIGURE 6.6</b> – ESWC2014 INTEREST PREDICTION RESULTS.....	192
<b>FIGURE 6.7</b> – EXAMPLE OF A USER PREFERENCE IN THE FORM OF A FORMAL CONCEPT. .....	194
FIGURE 6.8 – EXAMPLE OF THE CONCEPT-BASED COMMON REPRESENTATION.....	196
FIGURE 6.9 – RECOMMENDATION ALGORITHM.....	197
FIGURE 6.10 – PRECISION-BASED OVERALL RESULTS.....	216
FIGURE 6.11 – RECALL-BASED OVERALL RESULTS.....	217
FIGURE 6.12 – F-MEASURE-BASED OVERALL RESULTS.....	218
<b>FIGURE 6.13</b> – SUCCESS@K-BASED OVERALL RESULTS.....	220
<b>FIGURE 6.14</b> – OTHER RANKING-BASED OVERALL RESULTS.....	221
<b>FIGURE 6.15</b> – CONTENT-BASED OVERALL RESULTS (AVERAGED OVER THE DIFFERENT FEATURES).....	225
<b>FIGURE 6.16</b> – OVERALL COLLABORATIVE FILTERING RESULTS.....	229
<b>FIGURE 6.17</b> –RESULTS ACCORDING TO THE DIFFERENT FEATURES.....	233
<b>FIGURE 6.18</b> –RESULTS ACCORDING TO THE DIFFERENT ITEMS (AVERAGED OVER THE DIFFERENT APPROACHES AND FEATURES).....	237
<b>FIGURE 6.19</b> –RESULTS ACCORDING TO THE DIFFERENT K-VALUES.....	240

<b>FIGURE 6.20</b> –RESULTS ACCORDING TO THE FORMAL CONTEXT REDUCTION THRESHOLD .....	241
FIGURE 6.21 – EXAMPLE OF THE DBPEDIA MOVIE DOMAIN.....	247
FIGURE 6.22 – EXAMPLE OF THE COMMON REPRESENTATION SPACE.....	248
FIGURE 6.23 – OFFICIAL RESULTS FOR TASK 1 .....	255
FIGURE 6.24 – OFFICIAL RESULTS FOR TASK 2 .....	256



# Part I

## BASIS AND METHODOLOGIES

*“You don't write because you want to say something; you write because you've got something to say”*

F. Scott Fitzgerald, Novelist

*“The aim and meaning of Formal Concept Analysis as mathematical theory of concepts and concept hierarchies is to support the rational communication of humans by mathematically developing appropriate conceptual structures which can be logically activated.”*

Rudolf Wile, Mathematician and FCA co-inventor

# 1

## Introduction

---

*This chapter motivates the research conducted along the development of this thesis and gives the big-picture of the work that has been carried out.*

### Contents

---

1.1	MOTIVATION.....	3
1.2	SCOPE OF THIS WORK.....	11
1.3	OBJECTIVES.....	12
1.4	CONTRIBUTIONS.....	13
1.5	METHODOLOGY.....	13

---

## 1.1 Motivation

*As long as the centuries continue to unfold, the number of books will grow continually, and one can predict that a time will come when it will be almost as difficult to learn anything from books as from the direct study of the whole universe. It will be almost as convenient to search for some bit of truth concealed in nature as it will be to find it hidden away in an immense multitude of bound volumes.*

— Denis Diderot, "Encyclopédie" (1755)

This cite from the French philosopher Denis Diderot is more than 250 year old; however, it accurately expresses a XXI-century problem: information overload.

**Information Overload** Information overload refers to the difficulty that a lay user may have in finding relevant information among the huge amount of data currently accessible on the Web. In the era of user-generated content, social networks, and big-data, it is almost impossible for a user to digest the whole amount of available data. Instead of producing a benefit, choice, with its implications of freedom, autonomy and self-determination, can become excessive. As Schwartz exposes in his book "The Paradox of Choice", more is sometimes less [Schwartz, 2004].

**Recommender Systems** Recommender systems (RS) were conceived in the mid 90's to overcome the information overload problem by offering relevant recommendations that fulfil user preferences [Goldberg et al., 1992; Resnick and Varian, 1997; Resnick et al., 1994]. Nowadays that the social web have led to the so-called information explosion, the need for RS is more acute than ever, while they have to face more and harder challenges. In counterpart, this huge amount of available data has made easier to provide feedback about products or entities to improve the operation of recommender systems.

**Rating Prediction** Many systems in the literature [Delimitrou and Kozyrakis, 2013] address the recommendation task from the point of view of predicting the rating/s that a given user would give to an item/s. To that end, the previous user activity, in the form of item ratings, is analysed in order to infer the ratings for other items. This "predictive" way to address the recommendation task may be very useful when we want to predict the user

interest in a given item. Nevertheless, this predictive approach does not fit so well in social-based scenarios. In some practical settings, the goal is not necessarily to look for specific ratings. Rather, it is more interesting to discover the most relevant items for a particular user. Let us imagine a Twitter user expecting to get recommendations about interesting users, tweets, hashtags or lists. In this case, as in many others, it is more interesting to directly find such relevant items by applying a ranking version of the problem (i.e., find the ranking of more relevant items for a given user in a given context) [Aggarwal, 2016].

In this scenario (commonly referred as top-N recommendation), the task is not focused on predicting the user interest in a given item, but on offering a set of items in some degree interesting for the user. When state-of-the-art systems are applied to this specific recommendation task, their performance is far from the performance in the predictive task. In this regard, [Lü et al., 2012] review the different top-N recommendation algorithms and their performance when applied to well-known state-of-the-art datasets (Movielens and Netflix), which is far from being satisfactory: precision is in the level of hundredths; i.e., 9 out of 10 recommendations are wrong. Top-N  
Recommendation

This top-N recommendation environment is gaining momentum thanks to the emergence of new recommendation scenarios, such as Twitter or Facebook, where no user ratings, neither explicit user feedback, are available and where the huge amount of potentially-recommendable items limits the application of common recommendation approaches. The recommendation task must be therefore based on the accurate inference of the user preferences. In this sense, the scenario in which this research is framed is the user preference modelling for the top-N recommendation task, applied to the context of social-based environments.

User preference modelling is a key aspect in recommendation. Regardless of the predictive vs. top-N scenario, the main rationale of recommender systems is a rather simple observation: individuals tend to like in the future similar items to those liked in the present. Consequently, the recommendation task is highly dependent on the modelling of current preferences in order to predict future recommendations. In this sense, RS have traditionally followed two methodologies to accomplish this modelling: Collaborative Filtering (CF), which tries to group users with similar preferences together [Koren and Bell, 2011], and Content-based (CB), which tries to find items similar to the user preferences; i.e., similar items to those already consumed by the users [Lops et al., 2011]. In more detail, CF approaches are mainly focused on analysing the user-item matrix (i.e., the matrix including the users in the rows, the items in the columns, and the relationship

between them in the cells) to find groups of users that have consumed a similar group of items. The basic assumption is that users that have shared preferences in the past are likely to do it in the future. On the other hand, CB approaches rely on the content of the items. The process is based on finding items which content may cover the user preferences.

Applying CF approaches is usually challenging because user-item matrix is very sparse. This problem is especially acute in Big Data and social environments, such as Twitter or Facebook, where the number of documents (i.e., tweets, posts...) is even larger. In this regard, Content-based approaches are preferable because textual content is often more informative than raw user-item relationships. In addition, textual content enables richer and less sparse low-dimensional representations based on high-level features (i.e., topics, classes, semantic concepts, etc.). On the other hand, Content-based systems present the problem of how to link these richer item representations to the user profiles.

**User-Item Gap** Related to this latter problem is the gap between the item and the user preferences representations: how to bring the user and the item dimensions together. From the theoretical point of view, recommendation systems try to link the user space (i.e., profiles of the users to be offered recommendations) and the item space (i.e., representation of the items to be recommended). In this sense, Content-based systems usually model both by separate and, afterwards, they try to link the representations in both spaces.

**Common Representation Space** In contrast, this work proposes a common representation space for users and items. To that end, we have taken advantage of the CB recommendation rationale: user profiles in a CB-RS are an aggregation of the items already consumed/liked/purchased by the users. For instance, if a user has consumed two items  $Item_1 = \{Feature_a, Feature_b\}$  and  $Item_2 = \{Feature_b, Feature_c\}$ , they will be represented as  $User = \{Feature_a, Feature_b, Feature_c\}$ . Consequently, users as well as items can be represented by means of the item features. The intuition is that this common representation will better capture user-item similarities to be then exploited in the recommendation process.

Item features might potentially be any kind of data related to the items [Lops et al., 2011; Pazzani and Billsus, 2007]. Nevertheless, in recommender systems, as in other Information Retrieval or Natural Language Processing scenarios, the most commonly applied methodology is based on taking the textual information related to the items. However, due to the natural language ambiguity, textual features are not the best choice for data representation [Gattiker, 2014; Jeon et al., 2013]. While it can be enough when dealing with traditional datasets (news reports, articles, web pages), the new environments of the Web 2.0 (social networks or microblogging) entail new challenges

for the data representation approaches that potentially limit the performance of the traditional algorithms [Anta et al., 2012]. Consequently, more complex representations and models should be proposed [Pazzani and Billsus, 2007].

At this point, the application of knowledge-based resources appears as suitable solution for the aforementioned problem. Knowledge-based resources (e.g., linguistic-based as WordNet or semantic-based as DBpedia) offer a large amount of formalized data, as well as the relationships between these data, allowing the reasoning to infer and extract new knowledge. For instance, DBpedia provides a large amount of data, extracted from Wikipedia, in the form of a knowledge base. It allows the application of Semantic Web techniques to identify interesting information for content enrichment and modelling [Auer et al., 2007]. Similarly, WordNet [Fellbaum, 1998] or EuroWordNet [Vossen, 2004] include information about the human language in terms of cognitive synonyms (synsets) related according to different lexical relationships to be applied in a wide range of tasks dependent on content modelling. By using this information, knowledge-driven representations based on semantic and linguistic concepts can be generated.

The simplest way to use DBpedia (or other resources) is based on gathering its information for content enrichment [Di Noia et al., 2012b; Heitmann and Hayes, 2010; Luo et al., 2014; Musto et al., 2012; Passant, 2010a]. In the same way, works like that in [Hassanzadeh and Consens, 2009] propose the utilization of these semantic relations to perform a semantic-based item description. Some other interesting approaches not in context of recommender systems are [Damljanovic et al., 2012; Fernández et al., 2011].

However, DBpedia, like other Linked Open Data (LOD) resources, does not have a well-defined structure such as, for instance, an ontology. Consequently, the identification of the most interesting content to describe user and items is not a straightforward process [Berners-Lee et al., 2009; Di Noia et al., 2012a; Yao and Van Durme, 2014]. The main problem is that DBpedia structure is in the way of a general domain ontology. Therefore, the characteristics related to specific domains are not covered in detail, leading to similar problems to those cited in [De Luca and Nürnberger, 2006] (the lack of expressivity and the lack of conceptual description of datasets) in the practical application for data modelling and representation tasks [Jain et al., 2010]. The problem is similar when linguistic resources are considered. For instance, the EuroWordNet structure presents some limitations such as the "expressivity lack" [Gangemi et al., 2001] and the existence of terms categorized in too generic domains [De Luca and Nürnberger, 2005], entailing some problems when applied to specific tasks [De Luca and Nürnberger, 2006].

A more sophisticated application of these resources is based on a new organization of the knowledge they contain. In other words, the building of an extra-layer on top of their structure, which organizes the data in a better way [Bentivogli et al., 2004; Chen and Chen, 2012; Kent, 2003]. In this regard, the (re)organization of such resources might facilitate their application for the recommendation process. It might enable the identification of their most valuable information by taking advantage of the inferred knowledge-based structure [Castellanos et al., 2014a]. By means of these inferred features, it is possible to describe users and items in the same representation space, avoiding the problems inherent to other item representations.

Formal Concept  
Analysis

To that end, we propose the application of Formal Concept Analysis (FCA), a mathematical theory of concept formation proposed by [Ganter and Wille, 1997]. FCA may be seen as a biclustering technique which creates "rigid" binary biclusters (i.e., *formal concepts*) relating the space of the objects (i.e., user and items in this scenario) to the space of the attributes (i.e., DBpedia or EWN features). In addition, the identified *formal concepts* will be hierarchically organized in a lattice representation, according to the latent structure of the knowledge-based features (i.e., inferred from the relationships between features and entries in the original resources, namely DBpedia or ENW). This organization is inspired by the way humans order concepts in *subconcept-superconcept* hierarchies; e.g., a *car* is a subconcept of *vehicle*.

Our hypothesis relies in the fact that the already proven FCA performance for content organization applied to WOD-resources is likely to represent better the knowledge included in these resources through their modelling and organization [Kirchberg et al., 2012; Tanase, 2015]. Consequently, it might create a more abstract "concept-based layer", valuable to understand relationships, inherent structures, implications or dependencies among the data of these resources. This hypothesis has been proven in this thesis applied to two different data representation scenarios (Replab 2013 and MediaEVAL 2014 and 2015), where FCA was able to improve the representations created by two other state-of-the-art proposals: Hierarchical Agglomerative Clustering and Latent Dirichlet Allocation.

The powerful and automatic organization provided by FCA has been previously applied in the state of the art. The work presented herein differs from those, in two aspects. First, instead of taking only a bunch of data related to a specific environment (e.g. search results [Alam and Napoli, 2014], web data [Kirchberg et al., 2012], or experimental corpus [Falk and Gardent, 2014]), the FCA-based representation proposed in this thesis has been created by taking all the available information into account (i.e. the whole

amount of data contained in DBpedia or EuroWordNet). Secondly, the aim of our work and its application scenario are also different. We intend to apply the obtained representation to a specific task. In contrast, other works only use the resultant representation to conduct an experimental analysis into it. For instance, [Kirchberg et al., 2012] proposed the application of FCA to compute concepts from the Semantic Web. Particularly, their work focuses on the feasibility of the formal context generation and the FCA computation of large Semantic Web resources. However, it does not propose any specific application scenario for the generated FCA models, nor does it delve in the advantages that this kind of model might offer to such application scenarios. Likewise, the works presented in [Priss, 1998] to formalize WordNet mathematically or in and [Alam et al., Unpublished results] to RDF completion only focus on the FCA computation and the analysis of the FCA-based model, without paying attention to the practical application of that model.

Following the FCA theory, the knowledge-based common representation space can be interpreted as a bipartite graph, partitioned into objects  $O$  (users and items) and concepts  $F$  (the concept-based representations of the items). The bipartite graph can be interpreted as a formal context, and a set of formal concepts  $(A, B)$  can be inferred, where  $A$  is the set of users and/or items sharing the concept set  $B$  (i.e. users/items in  $A$  are described by the concepts in  $B$ ). The set of formal concepts can be therefore understood as the set of user preferences inferred from the user profiles and the items fulfilling these preferences. As stated by [Nenova et al., 2013], the generated formal concepts can be interpreted as the set of latent factors  $\mathcal{F} \subseteq \mathfrak{B}(X, Y, I)$  describing the data in the formal context. In fact, as also proved by [Nenova et al., 2013], these factors are the optimal decomposition of the formal context. In addition, the set of user preferences and the related items are hierarchically organized in a recommendation lattice. Recommendation will be then straightforward by recommending to a user the items associated to their preferences.

In this scenario, Formal Concept Analysis appears as a suitable solution, given that:

1. It is based on a well-defined mathematical theory (i.e., the adjective “formal” is meant to emphasize that we are dealing with mathematical notions).
2. It does not require prior information (i.e., it is an unsupervised approach).
3. The organization is based on a lattice structure, which provides a richer representation than a simple hierarchy, that better explores correlations, similarities, anomalies or even inconsistencies in the data structures [Carpineto and Romano, 2004].

4. It offers a readable representation of the resultant structure, facilitating its navigation and understanding.

Regarding the creation of a common user-item representation space, the approach proposed in this thesis has similarities with some works in the literature. For instance, [Shoval et al., 2008] propose an ontology-based user-item representation or [Cantador, 2008] presents a semantic-based common representation layer (also based on an ontology). Other alike proposals are those presented in [Huang and Bian, 2015], where authors propose the application of FCA to link two previously generated ontologies, those in [Yan et al., 2012] and in [Zheng et al., 2015] that propose a graph-based common representation, the works of [Singh and Gordon, 2008] and [McAuley and Leskovec, 2013] based on latent factor models; [Agarwal et al., 2011b] and [Ning and Karypis, 2012] that propose regression models, or the more recent proposal of [Wu et al., 2016a], applying Artificial Neural Networks (ANN). The main differences of the proposal presented in this work and these other are:

- The FCA-based representation is automatically inferred from the data. It does not need to be previously defined, for instance, by means of an ontology [Shoval et al., 2008]. In fact, this problem is related to the ontology definition itself: Ontologies are “predefined” formalisms to model knowledge domains [Cimiano et al., 2004]. On the other hand, FCA has need neither to predefine any data structure, nor any data relationship to model such knowledge domains.
- Other approaches, like the ontology-based ones [Cantador, 2008], do not create a unique common representation space, but a tripartite representation connecting users and items based on concepts relating them. Consequently, it needs to link items and users to concepts in the ontology (e.g., a user likes  $item_1$  and  $item_2$  because they are "pop\_music" of the "1960s" played by an artist from "UK"), whereas our approach automatically relates user and items.
- While the graph representations may provide an interlinked structure similar to that of FCA, the FCA-based model provides a lattice representation that hierarchically organizes the inferred user-item groups. It offers a coarse- and fine-grained representation (i.e., the most generic user preferences are at the top of the lattice structure whereas the most specific ones are at the bottom). [Zheng et al., 2015] speak about a hierarchical graph representation; however, their hierarchy is only based on three different predefined representation layers (user-, item-, and Content-based factors). In contrast, the hierarchical representation presented in this work is automatically inferred for the whole set of input data.

- In the work presented by [Huang and Bian, 2015], FCA is also proposed to create a representation for both users and items. However, they do not use FCA to infer this representation from the data, but apply FCA to link the concepts contained in two previously generated ontologies. In other words, they do not infer a unique user and item representation; they create both separately and then propose an automatic way of linking them. Consequently, the problems related to the manual construction of ontologies remain. Furthermore, they have selected the attributes related to the ontology concepts manually. In this sense, our proposal does not use FCA to link previously generated user and item representations, but we apply it to infer a data structure automatically, making explicit the users preferences and the items related to them. In addition, unlike the work of [Huang and Bian, 2015], the attributes to represent users and items are automatically identified and selected.
- Latent Factor Models, such as those presented by [McAuley and Leskovec, 2013] and [Singh and Gordon, 2008] try to factorize the information in the user and item dimension to create a common space. However, they take the user-item and item-attribute matrix by separated and we consider a unique user/item-attribute matrix. In addition, as proven by [Nenova et al., 2013], the concept lattice generated by FCA is an optimal factorization of the input matrix. Furthermore, Latent Factor Models do not capture the latent hierarchy in the user preferences (i.e., users that like comedy movies > users that like `Braindead`), as FCA does. This problem is shared with the regression models [Agarwal et al., 2011b; Ning and Karypis, 2012].
- One of the most novel research lines in this field is the use of ANN to create user-item representations. ANNs are able to automatically derive data representations from the input data. In this line, [Wu et al., 2016a] propose a Denoising Autoencoder as basis for this representation. This model is expected to represent the user-item relationships in a low-dimension common space of  $K$  factors. Following a rationale similar to that applied in this thesis, each one of these factors might be also considered as user preferences. Nevertheless, it presents some drawbacks:
  - It proposes a flat representation; however, user preferences are inherently hierarchical.
  - The  $K$  number factors in which the representation is based is manually settled. In contrast, the FCA-based structure is not parameterized and

the number of user preferences (i.e., formal concepts) will be automatically inferred from the input data.

These issues also appears in those works proposing common representation spaces based on embeddings of items and users [Moore et al., 2013; Wu et al., 2013], which, in addition, use to be based on shallow representations (e.g. check-ins), instead of on higher-level textual or semantic representations [Feng et al., 2015; Ozsoy, 2016].

Unlike other research fields (e.g. Information Retrieval), the recommendation task, especially when applied to social scenarios, lacks for a properly defined evaluation framework (i.e., standard datasets, evaluation setups, etc.). Consequently, we have developed an extensive experimental framework in order to evaluate our proposal in several social-based recommendation scenarios (NEWSREEL 2014, ESWC LOD-RecSys 2014 and 2015 and UMAP 2011). This evaluation confirms our previous hypothesis (i.e., FCA-based common representation space is able to accurately identify user preferences and the related recommendations). In particular, our FCA-based common representation space outperforms other recommender systems reported in the literature as state-of-the-art for the addressed tasks.

## 1.2 Scope of this Work

---

The scope of the work described in this Thesis is twofold: data representation and content recommendation. First, we investigate how to develop a meaningful concept-based model to represent data, independently on the addressed task. Specifically, this work focuses in the concept-based modelling of large amount of formalized data. The evaluation of this representation is carried out by measuring its performance in specific data-representation tasks. This evaluation setup – learning a good representation on a task A and then using it on a task B – is a general tactic, broadly applied in several contexts [Luong et al., 2013].

Thereafter, the developed (and evaluated) FCA-based data representation is applied to recommendation. This data representation enables the creation of a common space to represent users and items in the recommendation context. It is expected to enhance the recommendation process by reducing the problems derived from the gap between users and item representations. The evaluation of the proposed recommendation approach is performed in different recommendation scenarios. This latter evaluation is especially

important. In the recommendation field, there is no agreement regarding what is the best recommendation proposal (e.g., Content-based vs. Collaborative Filtering) or algorithm. In this regard, as proved by [Beel et al., 2016], even small changes in algorithm parameterization, the recommendation scenario or the evaluation setup have an impact on recommendation effectiveness. Therefore, an evaluation focused on the desired environment seems to be necessary, in order to compare the recommendation proposal presented in this thesis to several other state-of-the-art systems.

For the evaluation of recommender systems, different dimensions can be considered: novelty, diversity, etc. [Aggarwal, 2016]. The study of these dimensions is with no doubts an interesting side of the recommender system evaluation and it has attracted the interest of many researches in this field. Nevertheless, we are going to evaluate the recommender systems in terms of the relevance of the recommended items, which in turn is the methodology most widely applied in the literature.

## 1.3 Objectives

---

**Design and evaluation of a concept-based model to represent data from large information repositories and streams of social data.** This objective aims to create a fully automatic process to gather, process and model the input data in terms of high-level concept representations.

**Create a common recommendation space for users and items based on FCA lattices.** This objective pursues the proposal of a methodology that allows the representation of the two dimensions in the recommendation context — users and items — by applying Formal Concept Analysis

**Effectively address the top-N recommendation task.** This objective pursues to propose a methodology that addresses the recommendation task by actually inferring the user preferences and not only predict the outcomes of the system.

**Develop a recommendation algorithm that takes advantage of the relationships in the FCA lattice.** This objective pursues the formulation of a recommendation algorithm that may take advantage of the data explicitly represented in the FCA-based common representation space

**Apply the developed proposal to real Content-based recommendation scenarios.** This objective intends to evaluate our proposal in an evaluation scenario that

may replicate, to the extent possible, the scenario of a real recommendation task in social scenarios.

**Implement an experimental platform to evaluate the recommendation proposal.** This objective aims to implement a platform to allow a fair extensive evaluation of our proposal in comparison to other algorithms in order to frame its performance in the state of the art.

## 1.4 Contributions

---

The work carried out in this thesis succeeded in:

- Carrying out an extensive evaluation of the FCA performance in regards to the detection of thematically similar topics in social data.
- Providing a well-defined mathematical scenario for user-item modelling
- Proposing a novel application of Formal Concept Analysis
- Developing a recommendation algorithm for Content-based recommendation
- Deploying an experimental framework for real recommendation scenarios in which apply and evaluate the proposed approach by comparing it to other state of the art approaches

## 1.5 Methodology

---

- Literature review to understand the problems related to the recommendation task, especially in the sense of the data representation
- Review of the state-of-the-art systems applied to the task in order to have the big picture of the proposed solutions
- Thinking of possible solutions for the related problems, by focusing in solving the representation task to solve the recommendation task.
- Based on the review of the literature and the possible solutions, proposal of a new paradigm for data representation based on a conceptual modelling.
- Evaluation of the proposed modelling in data representation task.
- Based on the evaluation results, refinement of the proposed approach by applying a knowledge-driven representation.
- Application of the data representation for the recommendation task.

- Evaluation of the recommendation results by means of state-of-the-art configurations for the evaluation of recommender systems.
- Thorough results analysis to understand the involved issues.
- Publication of partial and total results in international conferences and workshops, as well as, impact-factor journals and inclusion of the gathered feedback to improve our proposal.
- Summarizing of the developed work and the drawn conclusions for its publication by means of this dissertation.

# 2

## State of the Art

---

*This chapter summarizes the state of the art in regards to the research fields addressed by the thesis*

### Content

---

2.1	KNOWLEDGE-BASED REPRESENTATIONS.....	16
2.2	TOPIC DETECTION .....	20
2.2.1.	Classification .....	20
2.2.2.	Clustering.....	21
2.2.3.	Probabilistic .....	22
2.2.4.	Graph-based.....	23
2.2.5.	Formal Concept Analysis.....	24
2.2.6.	Deep Learning for Topic Modelling.....	26
2.3	RECOMMENDER SYSTEMS.....	31
2.3.1.	Algorithms and Evaluation .....	33
2.3.2.	Specific Application Domains.....	43
2.3.3.	Related Topics .....	58
2.4	DISCUSSION .....	67

---

This chapter summarizes the state of the art of the fields related to scope of this work. This review does not pretend to be exhaustive, but it mainly aim to pay attention to the recent works that are more related to the research conducted herein (i.e., especially those carried out since the beginning of this work in 2013). We start focusing on the development of knowledge representations in section 2.1, as that proposed as basis of our common representation space. Section 2.2 includes novel approaches for the detection of the topics addressed in a stream of data. The detection of thematically similar topics is addressed as a previous step for the Content-based recommendation process. The detected topics will then serve to infer similar items to be offered as recommendations. Next, we discuss in section 2.3 the works conducted in the recommendation field, which is the application field of this thesis. Finally, in section 2.4, we summarize the lessons learned from the review of the state of the art and the implications they have in this thesis.

## 2.1 Knowledge-based Representations

---

Knowledge, as it is described in [Gradmann, 2010], is Information plus Context. Therefore, knowledge organization is based on the organization of the information belonging to a given context, according to the nature of the information and to the specific aspects involved in its context.

Knowledge organization may be addressed in a manual or automatic fashion. In the first case, it is usually conducted by experts, involving a great cost. On the other hand, automatic approaches facilitate the creation and utilization of these organizations. This is because automatic methodologies have become more popular.

The former are related to the ontologies, the latter to the conceptual representations. Ontologies are definitions of a knowledge domain, offered as hierarchies of concepts, which are described by means of attributes and linked by relationships between them (e.g., WordNet [Fellbaum, 1998]). As it has been argued in psychology research [Loftus and Scheff, 1971], the knowledge is organized in the human brain as a hierarchical structure. Therefore, ontologies try to replicate this knowledge organization in a formal way. A classification of ontologies according to their generality, complexity, or the information they represent can be consulted in [Cantador, 2008]. Ontologies are closely

related to the Semantic Web or Web 3.0. Web 3.0 intends to represent all the content on the Internet in a way it can be easily processed by the automatic systems. In this context, ontologies play a crucial role in the normalization of the data representations, categorizing the data and allowing the reasoning upon them. The final objective is to build a new Web organized in a set of categorized nodes (including the Web content), interlinked to others through the properties defined by shared ontologies. In the recommendation field, more “relaxed” representations, such as taxonomies, have also tried to apply hierarchical structures for item description [Zhang et al., 2014].

The formal representation provided by the ontologies allows the reasoning upon the data in order to infer new relationships or knowledge. It works well in reduced domains (e.g., medical domain); however, it is usually hard to apply to general domains or volatile scenarios. In this context a less strict representation, allowing fast changing and adaptation, seems more appropriate. In this regard, in contrast to ontological representations, conceptual representation also tries to create a more abstract representation of the content, by means of the concepts it address. The rationale is the same than with ontologies, but the creation of this representation follows a different approach. Conceptual representations are created dynamically by analysing the contents to represent, inferring their latent relationships and organization. It leads to a more dynamical representation, easy to change and more related to the contents (i.e., it is directly inferred from them).

Conceptual representations have been generated by following different methodologies. The analysis of special content features to infer the concepts from them is proposed by [Abel et al., 2011; Tao et al., 2012]. Similarly, the analysis of word senses for taxonomy extraction is presented at [Meijer et al., 2014]. The detection of the latent concepts has been also carried out by means of statistical methodologies [Ramage et al., 2010], algebraic factorization [Diaz-Aviles et al., 2012; Ticha et al., 2014] or, like in this work, the application of formal concept analysis [Valverde-Albacete and Peláez-Moreno, 2007].

By using statistical techniques, such as LSA or LDA, the conceptual representation is automatically derived from the data itself, based on the term frequency and distribution along the data. In this regard, hierarchical clustering and LDA are commonly viewed as desirable knowledge organization techniques [Hu et al., 2014]. A problem common to all these probabilistic methods is that the resulting representation does not have an evident interpretation for users [Zheng et al., 2015]. It reflects latent concepts but without a clear representation to a defined concept (i.e., in the sense that it may be understood for

a person). Instead, they are mathematical abstractions of the relationships between terms and documents that create these latent concepts.

In this context, FCA provides a mathematical framework in which to carry out this process, which has been extensively applied in the literature of content organization [Carpineto and Romano, 2004; Kim and Compton, 2004; Poelmans et al., 2013; Priss, 2000; Rahman and Chow, 2010; Tian, 2006]. In this regard, FCA also creates the latent conceptual representation, but the identified concepts correspond to concepts in the real world. In fact, the concepts are described by the combination of features that define them. FCA can be seen as a way to automatically generate ontology-like representations. According to [Gruber, 1995], ontologies are considered as an “explicit, [formal,] specification of a [shared] conceptualization [of a domain of interest]”. In other words, ontologies serve as models of a given domain based on some pre-defined formalism to describe concepts and relationships related to the data. Although theoretically FCA is not a model of a knowledge domain itself (i.e., Concept Lattices may be understood as way to structure a knowledge domain based on given data to mainly support their exploration), in their practical application it may serve as a data-based model to be applied in specific tasks. In fact, although theoretically speaking ontologies are constructed to describe a complete domain, in the vast majority of the cases they are simply made with a specific task in mind [Touzi et al., 2013].

FCA provides a set of benefits: 1) it does not require from any background information about the contents to organize them; 2) it provides an organization based on a lattice structure, rather than a simple hierarchy. The lattice represents a formalism that better explores correlations, similarities, anomalies or even inconsistencies in the data structures [Carpineto and Romano, 2004]; and 3) it offers an easy-readable representation for the users of the resultant structure, facilitating its navigation and understanding. A detailed review of the application of FCA for knowledge representation is conducted by [Poelmans et al., 2013] and by [Codocedo and Napoli, 2015]. Some other interesting researches on this subject are the works of [Tian, 2006] for building a hierarchical menu for content organization; those in [Priss, 2000] and in [Carpineto and Romano, 2004] for Information Retrieval systems; the research presented in [Alam and Napoli, 2014] on modeling SPARQL Query Results; or the work in [Kim and Compton, 2004] focused on document organization for Knowledge Acquisition.

These works mainly base their operation in the textual content of the documents. The reason is that text represents the main indicative of the document content, and the content is usually the desired feature to organize the documents. Although there are also

other types of organization (such as for example temporal-based), they are out of the scope of this work. Nevertheless, as it has been proposed in several works [McCallum, 2005; Ahern et al., 2007; Gattiker, 2014; Jeon et al., 2013], raw or unstructured text (i.e. natural language based text) is not the best information source given its lack of structure and organization. Some works have dealt with unstructured text [Hotho et al., 2002; Kuznetsov et al., 2007]; however, the main conclusion is that some kind analysis should be conducted to extract meaningful organizations, structures or patterns from the raw data. In this regard, projects such as Linked Open Data (LOD) provide a valuable source of semantically enriched and contextualized data, which has been widely addressed as a basis for knowledge-based representations [De Luca, 2013; Schandl and Blumauer, 2010; Shiri, 2014; Shiri, 2014; Summers et al., 2008].

However, in spite of the semantically-based structure of LOD, sometimes it is not clear how the LOD should be used, or what kind of data is more suitable to contain valuable information. A more advanced application of LOD would be based on the building of an extra-information layer on top of its structure. This layer should include some kind of knowledge-based structure, which could organize the data in a more “intelligent” way than the raw semantic relationships. At this point, the application of FCA jointly with LOD or other similar resources (ontologies [Cimiano et al., 2004], linguistic resources like Wordnet [Zhang et al., 2007; Falk and Gardent, 2014], or semantic-based data representations like DBpedia [Castellanos et al., 2014b; Kirchberg et al., 2012]) has been proposed as a solution in order to better represent content or documents. These applications intend to offer an extra knowledge about the data, more than the textual contents, moving then from Content Organization to Knowledge Organization.

This extra-knowledge represents a valuable information to create a more “intelligent” data representation, which will lead on a more “intelligent” data organization. In the same way that FCA applied over the textual contents provides a “rich” content organization, it is reasonable to think that applied over Knowledge-based contents it will provide a more “intelligent” Knowledge Organization. As it is postulated in [Kirchberg et al., 2012], the hypothesis is that FCA applied over LOD could create a more abstract “Concept Layer”, valuable to understand relationships, inherent data structures, implications or dependencies.

---

## 2.2 Topic Detection

---

The Topic Detection task is related to the ever-increasing need in analysing large corpora of data in order to understand, organize and summarize their content. In particular, it focuses on the discovering of groups of similar contents sharing an underlying common semantic theme (i.e. a common topic) [Mcauliffe and Blei, 2008].

The research into Topic Detection started several years ago, mainly motivated by the interest in the management the information contained in data streams. One of the first forums to focus on this area was the Topic Detection and Tracking (TDT), held within TREC [Fiscus and Doddington, 2002]. TDT pursued discovering and threading together topically related contents in streams of data, such as broadcast news. The works proposed within the scope of the TDT task, as well as in other works in the state of the art, have proven to be relatively satisfactory for Topic Detection in textual contents [Allan et al., 1998]. However, from these seed works, the focus has moved to Social Network data sources, and especially Twitter. In the state of the art, different methodologies have been proposed. In what follows, we explain the most noteworthy, as well as the more novel approaches proposed up to date.

### 2.2.1. Classification

---

The first approaches conducted for Topic Detection successfully applied classification techniques in different scenarios [Becker et al., 2011; Bengel et al., 2004; Kumaran and Allan, 2004; Wayne, 2000]. However, despite the extensive application of classification techniques to other sources, dealing with tweets involves some considerations that potentially limit the performance of the traditional classification algorithms. Some of them are explained in [Anta et al., 2012]: existence of special signs (i.e. abbreviations, emoticons or hashtags), use of slang, shortness of the contents (limited by Twitter), or spelling mistakes.

Notwithstanding the above considerations, classification-based approaches have been proposed in the state of the art. In [Sriram et al., 2010], the authors propose a classification algorithm to categorize tweets (e.g. news, opinions, personal messages, events, and deals). This algorithm uses Twitter-related metadata on the tweet's authors (e.g. name, information in the profile) to adapt classification methods to Twitter data.

In the work of [Phan et al., 2008] their authors also try to cope with classification drawbacks. In this case, they use external sources (Wikipedia and MEDLINE) to expand the tweet contents to increase the features (terms) available for the classification algorithm. Other solutions proposed also include tweet tokenization [Laboreiro et al., 2010], stemming, spelling analysis and use of dictionaries [Agarwal et al., 2011a].

Some novel approaches for topic detection based on classification methodology are shown in TASS (Workshop on Sentiment Analysis at SEPLN) [Díaz Esteban et al., 2013]. TASS is an experimental evaluation Workshop focusing on sentiment and reputation analysis. Within the scope of this forum, the authors of [Cordobés et al., 2013] propose a classification technique based on graph analysis. Their rationale is that any text can be represented as a graph and topics can be extracted from the structure of this graph. On the other hand, the authors of [Pla and Hurtado, 2013] apply a classic SVM (Support Vector Machines) approach, based on Twitter-dependant features and Tweet contents lexically and morphologically expanded. For more detail on other approaches recently proposed in this sense, as well as for a detailed analysis of the topic modelling from the point of view of supervised systems, please refer to [Blei and McAuliffe, 2010]

### 2.2.2. Clustering

---

In addition to the Twitter-dependant problems, which are not exclusive to classification, classification algorithms present another problem that limits their field of application: it is a supervised methodology. It means that the algorithms need to be trained with an annotated dataset to be able to classify new contents. This methodology has a high performance in classifying contents according to the features seen in the training set. Nevertheless, if new contents present new features, these will not be taken into account for the classification process. In the field of Topic Detection it means that if new topics, unseen in the training set, appear, they will not be detected.

A solution proposed for this problem is the application of clustering techniques. Clustering is based on an unsupervised methodology. Thus, it does not need a training set to compute the categorization, and consequently the topic discovery will not be restricted to the training data. In the Topic Detection context, several works make use of clustering techniques. [Phuvipadawat and Murata, 2010] present a typical clustering approach, based on a TF-IDF representation. [Sankaranarayanan et al., 2009] present the Tweet Stand system to cluster tweets on trending news. [Vakali et al., 2012] propose

a clustering framework (called Cloud4Trends) that includes information on the context of the tweet authors. The interesting aspect of this latter work is that clustering-based topic detection is only used as a previous step for the later trending detection. That is, contents are divided in topics, which are monitored in order to detect the appearance of trending content.

In the context of the RepLab Campaign [Amigó et al., 2013a], some works have been presented focusing on the application of topic detection approaches in a real environment. The UAMCLyR group proposes an approach based on a novel term-selection methodology for the clustering algorithm. Term selection is based on the *diversification* and *unification* concepts proposed in [Zipf, 1949]. UNED ORM group proposes different clustering techniques based on Latent Dirichlet Allocation (LDA) [Blei et al., 2003]. In this latter work, the authors propose a tweet expansion with Wikipedia contents to refine the operation of the clustering algorithm.

### 2.2.3. Probabilistic

---

In recent, probabilistic methods have emerged as almost a standard for the Topic Detection problem. The application of probabilistic techniques, mostly LDA, also known as probabilistic Topic Modelling (pTM), try to find the subjacent semantic latent space to group together (in topics) content, according to the shared latent space. Some of the most interesting works following this approach are [AlSumait et al., 2008; Anthes, 2010; Guo et al., 2013; Huang et al., 2012; Godin et al., 2013]

Although clustering and especially probabilistic techniques have been broadly applied to Topic Detection, there are still some open questions. One of the most important is: How many topics are there? and consequently, how many clusters can be generated? [Guo et al., 2013]. This problem has been addressed by means of: the analysis of the kernel matrix for clustering algorithms [Honarkhah and Caers, 2010]; the so-called Hierarchical Dirichlet Processes (HDP) for LDA algorithms [Paisley et al., 2012; Teh et al., 2006]; the application of supervised versions of LDA (Labelled LDA [Ramage et al., 2009] and its application for Twitter, TwitterLDA [Quercia et al., 2012]) and the use of Bayesian Inference for probabilistic algorithms [Cheng et al., 2015]. In this regard, [Petkos et al., 2014] present a clustering proposal based on Soft Frequent Pattern Mining (SFPM), a less strict version of traditional pattern mining algorithms that does not require all terms in a set to occur together frequently. In spite of the fact that these kinds of techniques

propose some interesting solutions and they may achieve good results, they require some supervision (i.e. training process, parametrization or previous analysis of the data) and their application is not straightforward.

As well as the number of topics problem, there is another limitation related to these methodologies: How can the systems take prior knowledge on the topics into account? That is, if there are some previously annotated training data, does the running of the systems have to be fixed by the data in the training set?, or do they have to show a certain degree of adaptability?

#### 2.2.4. Graph-based

---

Because of these problems, clustering and probabilistic techniques such as LDA have started to give way to other proposals. In this sense, the work in [Berrocal et al., 2013] in the context of Replab 2013, or the work in [Cataldi et al., 2010] apply novel techniques based on graph analysis. The graph analysis method uses the relationships between tweet terms to construct a graph-based representation of the data. These relationships are mainly term-based; however, some other works propose graphs of short phrases where tweets are connected by edges representing lexical inclusion [Leskovec et al., 2009]. The graph structure is then used to find topics in it. In this regard, the authors of [Berrocal et al., 2013] use the densest communities to be taken as topics, where density is the ratio between all the possible relationships and those that actually exist. Similarly, [Cataldi et al., 2010] apply the concept of *content energy* to find the emergent keywords, which are susceptible to defining new topics. Finally, the authors of [Sayyadi et al., 2009] use the *betweenness centrality* in the graph to carry out the clustering process.

Graph-based approaches are capable of coping with the topic detection related problems (i.e. topic adaptation and the need of parametrization). This kind of methodology does not restrict the number of topics to be detected, as happens with clustering, classification, or probabilistic algorithms (i.e. with the K-number of clusters to be generated, the number of classes learnt in the training process or the K parameter in LDA). In this way, the number of topics will only be based on the features of the data and, consequently, it will be adaptable to the different data. Graph based methodology also addresses the problem of detecting new topics, but taking the prior knowledge available into account: the graph can be generated with the prior knowledge and then, the new knowledge acquired over time is easily includable in the graph. This kind of techniques also offers a

graphical representation of the data relationship and makes the structure of the detected topics explicit.

### 2.2.5. Formal Concept Analysis

---

Just like graph-based approaches, the FCA for topic detection avoids the problems of the selection of the number of topics together with the problem of adaptability. We propose the application of Formal Concept Analysis as a way of organizing contents in a lattice structure. The generated lattice will automatically organize the contents into topics, based on the relationships between contents and their. FCA does not limit the number of generated topics and, as happens with graph-based methods; it is only dependent on the data features. With respect to the adaptability problem, the lattice can be constructed based on prior knowledge. Then, the new content will be included in this lattice by generating new topics if they appear or by including the content in the already existing topics.

This FCA-based proposal is somewhat similar to the graph-based approaches: it also provides a graph-like representation (the concept lattice) to be used for topic detection, being able to cope with the related problems in the same way as graph-based methodologies. Nevertheless, while graph representations are based on relationships between the data that have to be previously defined (e.g., term co-occurrence), FCA is able to automatically derive these relationships from the input data (see section 3). As a result, the input data are organized according to a set of formal concepts that groups similar data together [Ganter and Wille, 1997]. In contrast, graph-based approaches need to perform a later processing of the graph structure in order to identify such data groups (e.g., community detection). In this regard, the formal concept detection performed by FCA is a deterministic and exhaustive process, which cannot be assured in the case of graph-based representations. In addition, FCA theory defines a partial order relationship of the formal concepts that results in the construction of a concept lattice, a data representation richer than that provided by the graphs, that better explores correlations, similarities, anomalies or even inconsistencies in the data structures [Carpineto and Romano, 2004].

FCA has previously been proposed for topic detection [Geng et al., 2008; Ren et al., 2011], as we do in this work. In more detail, the work in [Geng et al., 2008] present a topic detection system that applies FCA to group mails together according to their

content and the temporal and social aspects related to them (i.e., the participants, sender and recipients). These features are weighted by using fuzzy membership functions that represent how likely an object is to belong to a given concept. In [Ren et al., 2011], a similar proposal is applied to cluster news stories gathered from the Internet, which are described by the terms appearing in their title. An extension of these works proposed in [Maio et al., 2016] relies on the use of Fuzzy FCA in order to take into account temporal aspects to model Twitter data according to semantic-based topics (i.e., those based on the Wikipedia concepts related to the tweets).

Some of the aforementioned works propose a similar idea to that presented in this paper: apply FCA to a series of documents described by a series of features for the detection of topics. In this regard, although the theoretical framework is similar, the application framework of those works is different. The dataset in [Ren et al., 2011] is not large enough (51 news stories) to draw any general conclusions. In addition, we propose a structured evaluation to measure the quality of the topics and their performance when applied to a real environment, as well as an extensive analysis of these results. In contrast, the evaluation in [Geng et al., 2008] is based on a manual analysis of the generated topics and it is only focused on the formal concepts related to the concept `trip`. Consequently, the conclusions can hardly be extrapolated to other contexts or datasets.

A similar issue is related to the work in [Maio et al., 2016] or in [Petkos et al., 2014]. [Maio et al., 2016] apply real-time Twitter data and the evaluation is better structured and more significant in terms of the evaluation set size than those proposed in the previously cited works. Their authors also propose a comparison to other state-of-the-art approaches — LDA (Latent Dirichlet Allocation), GDTM (Gaussian Decay Topic Model) and DTM (Decay Topic Model) — as we do. Nevertheless, the evaluation in [Maio et al., 2016] is based on some aspects defined by the authors (e.g., novelty, text-, hashtag- and concept-based coverage) and it is related to only 4 predefined topics: `Facebook IPO`, `Obamacare`, `Japan Earthquake` and `BP Oil Spill`. Consequently, the conclusions derived from this evaluation can only be understood in the context of this setting. In the same sense, the authors of [Petkos et al., 2014] propose an evaluation only focused on the analysis of four pre-selected topics according to the topic coverage and the precision of the keywords related to the topics. In contrast, we propose a general purpose evaluation based on the scenario proposed by the Replab Campaign, which tries to imitate a real environment for a Topic Detection system.

---

### 2.2.6. Deep Learning for Topic Modelling

---

Deep learning refers to the branch of machine learning in charge of training deep artificial neural network (ANN) architectures to discover high-level abstract features from raw data. In a sense, we can think of Deep Learning as a way to carry out feature engineering in an automatic way. To that end, Deep Learning algorithms build complex concepts out of simpler features in a hierarchical way. A layer including higher-level representations is built on top the output of another layer of simpler representations.

Although not Artificial Neural Network, neither Deep Learning are novel research areas, they have recently attracted a lot of attention because of their impressive performance in many tasks. For an extensive review of the deep learning techniques and their historical development, please refer to the recently published book of one of the fathers of the field, Yoshua Bengio, [Bengio and Courville, 2016]. Deep Learning has been successfully applied for perception tasks, such as image detection, speech recognition and even text understanding. However, tasks involving inference or deduction seem to be beyond the capability of conventional deep learning methods [Wang and Yeung, 2016].

In the natural language processing field, much of the works involving deep learning are focused on the learning of word vector representations through neural language models: Mikolov’s Word Embeddings [Mikolov et al., 2013], GloVe [Pennington et al., 2014] and others [Bengio et al., 2003; Yih et al., 2011].

The performance of deep learning for modelling data has been extensively proven in the literature. In this regard, its application to the scope of this work (i.e., the automatic modelling of textual data in a hierarchical conceptual based structure) appears as an interesting idea. As exposed by [Salakhutdinov et al., 2013]: “The ability to automatically learn in multiple layers allows deep models to construct sophisticated domain-specific features without the need to rely on precise human-crafted input representations”. Such features can be thus applied to create deep (i.e., hierarchical) models based on “high-level” conceptual representations.

Driven by these advantage, some works have proposed neural network-based topic models such as [Cao et al., 2015; Das et al., 2015], which use distributed representations of words to improve topic semantics.

Other methodologies are based on the use of Convolutional Neural Networks (CNN). CNNs are a kind of Artificial Neural Networks that creates several convolution layers

that relate regions of the input data to neurons in the output, identifying local connections in the input data. To that end, it relies in the use of the convolution operation applied over the input data to compute the output [LeCun et al., 1989]. One of the main strengths of CNN is that they are able to automatically learn a filter based on the task you want to carry out. The best example is CNN applied to Image Classification [Krizhevsky et al., 2012]. Each layer of the CNN is able to detect, through the application of the convolution operation, low-level image features (i.e., the first layer may detect edges in the image from its raw pixels). The features detected in some layer are used as input of the next layer. In this way, the network is able to learn higher-level features at each layer, creating a hierarchical model of the input data (i.e., edges from first layer may enable the detection of simple shapes, which can allow the detection of facial shapes in the next layer...). The last layer in CNN is then a classifier that applies these high-layer features.

CNN models have subsequently been shown to be effective for NLP: semantic parsing [Yih et al., Baltimore, Maryland, USA, June 23-25 2014], search query retrieval [Shen et al., 2014], sentence modelling [Blunsom et al., 2014], and other traditional NLP tasks [dos Santos and Gatti, 2014; Johnson and Zhang, 2015; Kim, 2014; Wang et al., 2015b]. However, unlike in the Image Classification scenario, it does not theoretically fit in the NLP scenario. The two rationales of CNN are Local Invariance and Local Compositionality that make intuitive sense for images, do not for NLP. In images, close pixels are commonly related which is not always true for words. Consequently, the CNN operation based on the detection of changes between image regions does not make so much sense for textual contents. In the same way, the compositional aspect that is obvious for images, it is not for words.

The main idea of CNN applied for textual representation and topic modelling is to make use of the 1-D text structure (word order) of document data, so that each unit in the convolution layer responds to a small region of a document. Usually, in CNN studies on text, the first layer of the network converts words in sentences to word vectors by table lookup. The word vectors are either trained as part of CNN training, or fixed to those learned by some other method e.g., [Mikolov et al., 2013]. In contrast, the proposal in [Johnson and Zhang, 2014] and in [Johnson and Zhang, 2015], instead of using low dimensional word vectors, directly applies CNN to high-dimensional text data for text classification.

In this sense, although some works have tried to adapt the CNN rationale for NLP scenario (e.g., [Johnson and Zhang, 2014; Kim, 2014]), some other neural network models

such as Recurrent Neural Networks (RNN) appear to make more sense for resembling the way we process the language. RNN model sequential information by taking into account the previous states to a given one [Mikolov et al., 2010]. For instance, if you want to predict the next word in a sentence, you better know which words came before it. These models allow overcoming the well-known theoretical problems of *bag-of-words* based models for the topic detection task (e.g., *the department chair couches offers* and *the chair department offers couches* have very different topics, although they have exactly the same bag of words [Wallach, 2006]). The work of Tian et al [Tian et al., 2016] makes use of different types of RNN (namely Long Short Term Memories [Hochreiter and Schmidhuber, 1997] and Gated Recurrent Units [Cho et al., 2014]) to address the topic detection problem. Their model assumes the words in the same sentence to belong to the same topic and the generation of a word to rely on the previous words in the sentence.

Other models have been also proposed for this task. [Glorot et al., 2011a] propose a supervised approach to train autoencoders on documents represented as bag of words. Autoencoders learn representations in a more reduced dimension than that of the input data. This representation can be seen as a more abstract description of the input data that groups together similar input entries according to their features. This more abstract description can be seen as the topics addressed by the data [Mirowski et al., 2010]. To allow hierarchical topic representations, several autoencoders can be stacked in layers (deep autoencoders) using the output of one layer as the input of the next one. Autoencoders are especially interesting because they operate in an unsupervised way. The input data is represented in the lower dimension representation. If this representation is accurate, it should be able to recreate the input data again when activated (i.e., the more similar the output to the input, the better the representation). This idea is applied to train the network by optimizing the loss function that minimizes this difference between input and output data.

Another unsupervised approach is followed in the work of [Larochelle and Lauly, 2012]. The authors present an unsupervised approach for Topic Modelling based on a Neural Autoregressive Distribution Estimation (NADE), a neural generative model inspired by the Replication Softmax [Larochelle and Murray, 2011]. Related to this latter work, Restricted Boltzmann Machines (RBM) may appear as a suitable approach for the Topic Modelling (i.e., Replicated Softmax is a generalization of RBM models). RBM models applied for topic modelling present some similarities with FCA, at least from the basic theoretical point of view. RBM models are undirected hierarchical (in the case of Deep architectures such as Deep Belief Networks) graphical models with binary observed (i.e.,

as FCA binary attributes) and latent variables (i.e., as formal concepts) organized in a bipartite graph (i.e., FCA is a way to bi-cluster bipartite graphs). Deep RBM architectures are known as Deep Belief Networks (DBN). DBN are stacked RBM where the visible layer of the bottom RBM is fed in with data, and the hidden layer of the bottom RBM is served as the visible layer of the second RBM [Hinton et al., 2006].

Because of their representativeness power, their ability to carry out non-linear dimensionality reduction and their ability to create high-order representations of the input data by means of an unsupervised process, DBN have been proposed for Topic Modelling. For instance, the work of [Maaloe et al., 2015] proposes a DBN model able to outperform other state-of-the-art topic modelling algorithm (based on LDA). In the same direction, but applied to the modelling of image representations, is the work of [Salakhutdinov et al., 2013]. Their proposal is not exactly unsupervised, they use very few training examples as a human would do (i.e., for human learners just one or a few examples are often sufficient to grasp a new category and make meaningful generalizations to novel instances). What they propose is a methodology to learn in an automatic fashion a hierarchical model based on high-level features that capture correlations among low-level features. Similar to this proposal is that proposed by [Wan et al., 2012]. They also propose a hybrid model integrating a neural net to provide a low-dimensional embedding for the input data and a hierarchical topic modelling approach to capture the subsequent topic distribution.

Despite of the unquestionable interest in Deep Learning and the proven performance of Artificial Neural Networks for many NLP tasks, there remain a number of concerns about them. One is that it can be quite challenging to understand what a neural network is really doing. If one trains it well, it achieves high quality results, but it is challenging to understand how it is doing so. If the network fails, it is hard to understand what went wrong [Lipton, 2015].

This latter aspect is special important given the complexity in the training of deep architectures. An extensive review of the issues related to the training process is presented in [Bengio, 2012]. This work can give an idea of the great number of parameters, hyper-parameters and configuration details that have to be taken into account to carry out the training (e.g., learning rate). As Bengio details: “deeper neural networks are more difficult to train than shallow one [...] there is a greater chance of missing out on better minima”.

In the same sense, [Levy et al., 2015] defend, focused in the context of Word Embeddings, that the performance improvement reported in the literature for different algorithms mostly relies in the hyper-parameter optimization. They prove this claim by applying the

same configuration to that applied for Word Embeddings computation to “old-school” algorithms (i.e., count-based distributional models) for word representation. The experimentation shows that the results for both approaches are similar and the differences come from the parameter configuration.

The complexity problem is especially acute when dealing with textual representations. Although many works have been proposed in regards to the automatic modelling and representation, most of them are focused on visual content (e.g., images, handwritten characters, human motion capturing). In contrast, textual representation highly increases the complexity in the training of deep architectures. In this regard, as explained by data scientist Will Stanton in a presentation prepared for the 2015 machine learning "Ski Hackathon", each hidden layer and each feature means more parameters to train, and human-generated text has a near-infinite number of features and data. Furthermore, the performance of ANN when of short texts are involved (e.g., Twitter) is not clear [dos Santos and Gatti, 2014; Johnson and Zhang, 2015]. In general, ANNs are not advisable in scenarios where a large collection of data is not available. This is because ANNs are highly noise sensitive, which might lead to overfitting when data size is small. In this sense, the use of pre-trained word embeddings would make sense to mitigate this issue, although this point is yet to be confirmed by means of a clear an extensive experimentation.

Other aspect to be taken into account is that the great heap of DL is based on the high performance of ANN in supervised scenarios. Unsupervised approaches have not still focused the attention of the researchers. Consequently, the performance of ANN for such tasks is yet to be proved. Although some works in the literature claim to be unsupervised or can be applied to unsupervised environments, they require in fact of a training process to adjust the parameters of the model (e.g., [Maaloe et al., 2015] or [Larochelle and Lauly, 2012]). For instance taking into account Mikolov’s Word2Vec, [Levy and Goldberg, 2014] have concluded that much of the technique performance comes from tuning algorithmic elements, such as sized context windows. In fact, as stated before, other approaches (Pointwise Mutual Information) achieve similar results than Word2Vec, or even better when their parameters are also optimized. A suitable solution to cope with this issue might be in the direction of the idea presented by [Maaloe et al., 2015] for the fine-tuning step but applied to the training process: transform the network into a Deep Autoencoder by replicating and mirroring the input and hidden layers and attaching them to the output of the DBN. In this way, the error estimation and the parameter

optimization could be done by comparing the normalized input data to the output of the network.

Finally, to the best of our knowledge, unsupervised hierarchical modelling has not been addressed by means of Deep Learning approaches. We do believe that some models, such as deep autoencoders, CNN or RBM (and most specifically Deep Belief Networks), appears as promising starting points for future works in this direction [Maaloe et al., 2015]. Deep autoencoders allows the automatization of the feature engineering by means of the automatic generation of complex features on top of simpler features, leading to a hierarchical structure. In this hierarchy, deeper layers are useful to extract topic representations from the input data (i.e., the complex features in these deeper layers put similar content together in a hierarchical representations). In a similar way, Deep Belief Networks also allows the automatic generation of hierarchical representation based on the input features. Finally, CNN have been proven to be useful in generating hierarchical representations of visual features, which can be also applied to textual features.

To sum up, the training of deep neural nets is expensive, complicated and it relies on many factors whose impact in the final performance of the model is not known a priori, being dependant on specific details of the task to be address or on the dataset applied for the experimentation.

## 2.3 Recommender Systems

---

Recommender Systems (RS) started in the early and mid-90s [Balabanovic and Shoham, 1997; Resnick et al., 1994] as a solution for the increasing information overload problem. RS base their operation on collecting information on the user preferences for a set of items. In origin, it was successfully applied for the recommendation of items in e-commerce sites (e.g., news, webpages, books, movies, products...), but nowadays the number of application fields have exploded: news, jokes, movies, applications, websites, travel destinations [Bidart et al., 2014; Huang and Bian, 2009; Yuan et al., 2013].

In last years, the work conducted in this area is huge and goes beyond those presented in this state-of-the-art. For instance, context-aware recommendation tries to consider the different contexts in which a user may interact with a recommender system (time of day, season, mood...). In any case, we decided to only focus on the special aspects related to this thesis.

For more details on the recommendation systems field, in [Rao, 2010] or more recently in [Park et al., 2012] and in [Lu et al., 2015], it is listed a wide set of RS as well as their application fields, both academics and business. For a more detailed information, the book of [Jannach et al., 2010b] offers an introductory view on recommender system, whereas the book of [Ricci et al., 2011] and more recently the book of [Aggarwal, 2016] discuss several advanced aspects.

RS research is a wide field and it involves several related fields, such as: Machine Learning, Information Retrieval, Natural Language Processing or Data Mining. Given the variety of application fields and the different research fields involved, RS has to face several well-known problems: scalability, proactivity, privacy, diversity, information acquisition, information integration, or cross-domain and cross-system integration. Some of them are exposed and addressed in [Ricci et al., 2011].

The literature in recommender systems have been traditionally organized according to two main typologies [Pazzani, 1999]: Collaborative Filtering based Recommenders, Content-based Recommenders. Although these methodologies form the fundamental pillars of research in recommender systems, driven by the appearance of new application fields and new kinds of information, specialized methods have recently been designed. For instance: time, location, social or demographic information [Porcel et al., 2012]. More detail about these typologies can be found in the next sub-sections.

The organization of this section intends to organize the recommendation literature according to these important topics. In particular, as proposed in [Aggarwal, 2016], the section is organized into three main categories:

- **Algorithms and Evaluation:** This section details the different methodologies and algorithms that have been commonly proposed in the literature of recommender systems. In particular, it introduces the two main methodologies applied in this sense: Collaborative Filtering in section 2.3.1.1 and Content-based recommendation in section 2.3.1.2. Section 2.3.1.3 presents the different proposal to hybridize both previous methodologies to overcome their individual limitations. Finally, section 2.3.1.4 details the works in the literature that studies the formal evaluation of recommender systems.
- **Specific Application Domains:** This section presents different specific applications scenarios that are particularly related to the scope of this thesis. In particular, we have studied the state of the art of: semantic-based, social, news and twitter recommendation.

- **Related Topics:** Finally, this section presents some topics related to the scope of the work addressed in this thesis: Graph-based recommender systems, Matrix Decomposition applied to recommender systems, other works proposing common representation spaces for recommendation and, finally, FCA applied to the recommendation task.

### 2.3.1. Algorithms and Evaluation

This subsection discusses the fundamental algorithms and methodologies into which recommender systems are organized. In more detail, section 2.3.1.1 details Collaborative Filtering methodologies, section 2.3.1.2 the recommender systems applying Content-based features, section 2.3.1.3 addresses the hybridization of different kinds of recommender systems and, finally, in section 2.3.1.4 the specific aspects on the evaluation of recommender systems.

#### 2.3.1.1. Collaborative Filtering Based Recommenders

The operation of this type of Recommenders are based on grouping together similar users, according to their past decisions [Linden et al., 2003; Kim et al., 2011]. The rationale is that if two users have liked the same set of items in the past, they will be likely to like the same set of items in the future. Based on that rationale, when two users have been considered as similar, if one of them consumes a new item, it will be recommended to the other user.

From a technical point of view, this task is closely related to missing value analysis. The system has an incidence user-item matrix, which includes the past interactions between users and items, and it should infer the missing values in this matrix based on the observed ones. In the context of the recommendation task, this operation is especially challenging because the user-items matrix use to be very large and very sparse [Koren, 2008].

Neighbourhood  
based CF

In the literature, two main types of CF methods have been defined. Firstly, neighbourhood-based CF algorithms (a.k.a. memory-based) are based on creating neighbourhoods of similar users –user equally rating the same items– (User-based CF) or items (Item-based CF) –items equally rated by the same users–. The recommendation task therefore relies on the generation of such neighbourhoods by following a process that can be seen as a generalization of nearest neighbour classifiers or k-means clustering. Thereupon, this methodology is highly dependent on how user or item similarity is

defined. The metrics proposed in the literature go from basic metrics like the cosine similarity [Devi and Venkatesh, 2009] to more sophisticated state-of-the-art similarities like BM25 [Parra-Santander and Brusilovsky, 2010]. However, most of these metrics do not take into account the length of the user-rating vectors (i.e., two users sharing the same 10 ratings are more likely to have similar tastes than two users sharing only a couple of ratings) [Ma et al., 2007]. To overcome this limitation, these measures may be applied in combination to Jaccard Similarity (or other measures that take into account the vector overlapping) [Candillier et al., 2008]. For more details, an extensive compilation of similarity measures in the context of Collaborative Filtering recommendation is included at [Bobadilla et al., 2013] and at [Pirasteh et al., 2015].

In particular, this latter work proposes an asymmetric user similarity measure intended to work in *cold-start* situations. The rationale of this measure is to distinguish between the impact that the target user has in his neighbourhood and the impact that the neighbourhood has in the user. In this regard, the work in [Jin et al., 2004] applies a similar idea but on the item side. They define a weighting scheme in order to capture the importance of each item for the recommendation process. Finally, external sources can be applied to calculate this similarity. In this sense, [Cilibrasi and Vitanyi, 2007] propose the Normalized Google Distance to measure the term similarity according to their co-occurrence on the Web, which has proven to achieve successful results in the recommendation field [Jack and Duclayee, 2008].

On the other hand, the so-called Model-based CF bases their operation on the generation of a summarized model of the input data that is created up front during the training process and it is thus applied to infer new recommendations. Different modelling techniques, most of them based on data classification models [Billsus and Pazzani, 1998], has been studied in the literature, such as: Rule-based systems [Shyu et al., 2005], or naïve Bayes classifiers [Miyahara and Pazzani, 2000]. Neural Networks has been also proposed for the recommendation field, especially in last years spurred by the interest of the research community in ANNs and Deep Learning. For instance, a model Restricted Boltzmann Machines is proposed by [Salakhutdinov et al., 2007], where the hidden units correspond to items and the user ratings of the items results in the activation of the visible units. RBMs have proven to achieve similar performance than other state-of-the-art models like latent factor models for scenarios such as the Netflix Prize [Bennett and Lanning, 2007].

Although model-based methods are usually seen as more refined and sophisticated methods, the complexity that relies on the creation of the recommendation model might

limit their suitability for big-data or real time scenarios. In contrast, the implementation of neighbourhood-based algorithms is lighter, given that the recommendation is an instance-based process, where no model has to be created up front.

A more recent research line is based on the so-called group-aware Collaborative Filtering [Ji and Shen, 2015]. These systems divide the large CF user-item matrix into some smaller subgroups (i.e., sub-matrices). The recommendation process will be then individually performed on every subgroup by, for instance, applying a CF algorithm on every sub-matrix. However, by following this methodology each user/item can only be assigned to a single subgroup, assuming that users do not have multiple interest, which is far to be true. In order to address this problem, different methodologies proposing overlapping co-clusters have been presented in [Xu et al., 2012; Zhang et al., 2013], and more recently in [Wu et al., 2016b]

Collaborative Filtering systems present a series of drawbacks. They need a critical mass of interactions in order to be able to find user similarities to enable the recommendation process. If these interactions are not available, it appears the *cold-start* problem; that is, when new users appear and the system has no information about them, the system is not able to provide recommendations. Many works in the literature have been tried to cope with this problem. Some of the most noteworthy are recompiled in [Schein et al., 2002], [Pirasteh et al., 2015]; and [Adomavicius and Tuzhilin, 2005]. The complexity of this kind of systems is often referred in the literature [Koren and Bell, 2011; Su and Khoshgoftaar, 2009]. Another problem of this kind of systems is related to the temporal dimension. Recommendations are inferred by past user interactions. However, the preferences that had related users in the past may not be valid in the present. For instance, two users may have been interested in some past event (e.g., the last Soccer World Cup) but they are not related anymore by similar tastes.

### 2.3.1.2. Content Based Recommenders

These systems are based on the content associated to the items to be recommended. The basic principle that underlies this methodology is that a user interested in a given content, like for instance action movies, is more likely to be interested in another action movie rather than in a romantic comedy movie. Therefore, the recommendations offered to the users are items similar to those already consumed by them [Lops et al., 2011].

In order to offer such recommendations, unlike Collaborative Filtering systems, these systems do not take the ratings of other users, but they largely rely on the user's own ratings. This is both an advantage and a disadvantage. Given that they are based on the

item content, the recommendations they enable are based on rich Content-based relationships discovered between items and user profiles [Ignatov and Kuznetsov, 2009]. Furthermore, when new items appear, they are easily recommended: they are offered to users that like the kind of content related to the item content. In contrast, being that Content-based systems only rely on the item content, they tend to overspecialize the recommendations by always recommending items with similar attributes [Zhang et al., 2002].

Because of these advantages and disadvantages, Content-based systems have been largely, but not exclusively, applied to domains where rich and informative item representations can be extracted, like for instance News Recommendation [Kompan and Bieliková, 2010]. Consequently, text classification and information retrieval are the most widely used techniques for developing this kind of systems [Pazzani and Billsus, 2007].

In brief, Content-based systems applying classification based their operation on classify items with a “similar” Content-based representation (i.e., features or keywords describing the items) together. This item similarity is based on some metric depending on the task and the item representations, being the cosine similarity the most widely applied. For more details on these metrics, please refer to [Spertus et al., 2005]. The recommendation is thus carried out by offering the items classified together with the items in the user profiles. This method is known as nearest neighbour classification and it is one of the simplest Content-based recommendation methodologies

Following the same rationale, more sophisticated algorithms have been proposed, such as: Naïve Bayes, applied by [Degemmis et al., 2007] to create Content-based neighbourhoods based on WordNet-enhanced user profiles or by [Semeraro et al., 2009] to exploit the knowledge stored in machine-readable dictionaries; and Rule-based Systems, like the one presented by [Abel et al., 2008] for Online Discussion Forums.

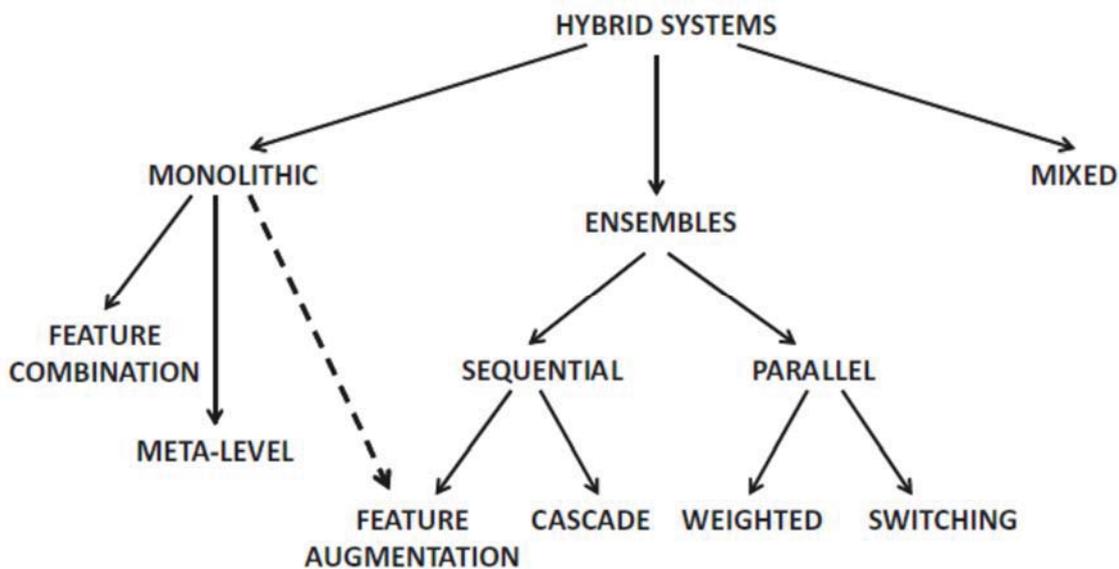
Regression-based models are especially useful for dealing with various types of ratings (binary, numerical, etc.) [Park and Chu, 2009]. In this regard, in the case of binary rating, Support Vector Machines are a commonly applied methodology because they are highly resistant to overfitting and results in high-performing systems [Wang et al., 2015c]. In scenarios of high-dimensional data (e.g., large rich-text descriptions) neural networks appears as an alternative given that they can be used to learn patterns among large sets of features. Related to this latter field, the work of [Ozsoy, 2016] applies Word Embedding to represent items in a Content-based recommender system; [Zhang et al.,

[2016] proposed a more advanced model to create Item Embeddings based on the knowledge data related to them.

### 2.3.1.3. Hybrid Recommenders

The two previous sections detailed the two main types of recommender systems: Content-based and Collaborative Filtering. As stated, both present different strengths and drawbacks when working on isolation. In this sense, Hybrid systems pursue the combination of both kind of systems in order to make use of all the knowledge available in different data sources (e.g., the user interactions or the content of the items). In this way it is attempted to leverage the complementary advantages of these systems, thus avoiding recommender system related problems like, for instance, cold start [Pereira and Hruschka, 2015; Tejada-Lorente et al., 2014; Wang and Wang, 2014].

One crucial aspect is how to carry out this hybridization; that is, how to combine the operation of different the different systems. In this sense, Figure 2.1 shows a classification of these hybridization methodologies proposed by [Aggarwal, 2016], according to the seven mechanisms combination proposed by [Burke, 2007]: 1) weighted, that combines the scores of different systems into an unified score [Mobasher et al., 2004]; 2) mixed [Smyth and Cotter, 2000]; 3) switching, that switches between different systems according to the current needs (e.g., a knowledge-based systems in the early phases to avoid *cold-start* problems and a CB or CF in later phases) [Billsus and Pazzani, 2000]; 4) feature combination of different sources into an unique system; 5) feature augmentation that uses the output of a system to create the features of the following system [Sullivan et al., 2004]; 6) cascade, in which a recommender system refines the recommendations given by another [Burke, 2002] and; 7) meta-level [Pazzani and Billsus, 1997] that shares the model used by one recommender system to another system.



**Figure 2.1** – Taxonomy of Hybrid Systems (extracted from [Aggarwal, 2016])

Higher level-categorization have been also proposed, like that in the figure proposed by [Aggarwal, 2016] or the classification of pipelined —one system is concatenated to the output of another system— and parallel systems —systems work in parallel— proposed by [Jannach et al., 2010a].

Recently hybrid systems, especially ensemble systems, have attracted a lot of attention because the winning systems of the Netflix Prize were systems of this kind [Koren, 2009a]. Hybrid systems applying weighted models are also commonly applied in the state of the art [Jahrer et al., 2010; Bar et al., 2012]. Finally, some other noteworthy examples of hybrid recommender systems are: [Bedi and Vashisth, 2014; Cacheda et al., 2011; Christensen and Schiaffino, 2014; Vozalis and Margaritis, 2004; Wang et al., 2006].

To sum up, hybrid systems allow the integration of different types of recommender systems, as well as of multiple data sources, thus improving the performance of individual systems. In this sense, these systems are the state-of-the-art in many scenarios, such as in the Netflix Challenge.

#### 2.3.1.4. Evaluation

There are three main evaluation paradigms for evaluating recommender systems: user studies, offline evaluations and online evaluations.

In brief, user studies evaluate recommender systems by asking test subjects to interact with them to perform specific tasks and give their feedback about the quality of the recommendations. User studies

- Online evaluations Online evaluations are similar to user studies, but in this case, the users are real users of fully developed recommender systems.
- Offline evaluations Finally, offline evaluations are based on testing recommendation algorithms' performance on historical data by comparing system outputs to actual user interactions (i.e., contained in the historical data).

User studies are the most desirable evaluation methodology because they collect information about actual user interactions with the system in a controlled evaluation environment. Nevertheless, they are difficult and expensive to carry out, since they involve the recruiting of a large number of users. Furthermore, the results are not comparable to other results outside of the study and they are difficult to replicate (i.e., the same evaluation configuration, the same kind of users and the same system operations should be replicated). Consequently, the conclusions drawn from this evaluation are difficult to extrapolate. Similar problems appear when applying online evaluation methodologies, although they allow for easier comparison to other algorithms [Kohavi et al., 2009]. Because of these technical problems, and although offline systems present the disadvantage that they do not actually evaluate the performance of recommender systems in the future (i.e., their evaluation is based on past user interactions), the offline evaluation paradigm is, by far, the most common methodology. In this sense, many standardized frameworks and evaluation measures are available in the state-of-the-art, which have allowed the development of many experimental works, comparing several recommendation methodologies and systems.

These evaluation methodologies have had to adapt to new scenarios, like, for instance, Twitter, where traditional methodologies are not suitable. In this regard, the problem of how to evaluate a Twitter-based recommender has been commonly posed in the literature, spurred by the growing interest in this research area (see section 2.3.2.4). To evaluate whether a given tweet is interesting for a given user is not a straightforward process. At this point it is important to remark that, unlike traditional recommender systems with explicit ratings (e.g., dislike/like or 1-5 ratings), in this scenario ratings are implicit and unary. That is, users do not explicitly rate an item; instead their interest by an item (tweet) have to be implicitly inferred from their behaviour (e.g., a user publishes some content, shares a URL or retweets other users' tweets). Moreover, this "ratings" has to be considered as unary: if a user has not interacted with some item (tweet), it could mean that he does not like it or he does not know it. Consequently, only positive user feedback can be taken into account. Formally defined, the rating of a user  $u_i$  by a content  $c_j$  is defined as

$$r_{i,j} \in \{0,1\} = \begin{cases} 1 & u_i \text{ likes } c_j \\ 0 & \text{otherwise } (u_i \text{ does not like or not know } c_j) \end{cases}$$

The main method to evaluate this interest is by means of a user study, like the ones in [Ramage et al., 2010] or in [Phelan et al., 2011]. However, the great time, resources and effort it takes, makes its application infeasible for large data. Another easier solution proposed in the state of the art is to infer evaluations from the implicit user feedback. In the context of Twitter it means that if a user has tweeted or retweeted some content it can be considered as interesting for them. This assumption has been applied in many works [Ramage et al., 2010], demonstrating its suitability in comparison to user studies.

In what follows, some of the most noteworthy evaluation metrics, which have been traditionally considered in the literature, are detailed. In this regard, one important aspect to be considered is that the recommender system evaluation is not based on a single criterion [McNee et al., 2006]. Although accuracy-based metrics are the most widely applied measure (in fact, state-of-the-art algorithms are commonly defined according to this criteria), some other aspects have been also considered in the literature: coverage, novelty [Konstan et al., 2006], *trustability* [Cramer et al., 2008], serendipity (i.e., lucky discovery) [Ge et al., 2010], diversity [Castells et al., 2014], robustness [Mobasher et al., 2007] or scalability [Takács et al., 2009]. For a more detailed discussion on the recommender system evaluation, as well as the limitations they present, refer to the corresponding chapter in [Aggarwal, 2016], the book chapter of [Shani and Gunawardana, 2011] or the survey presented by [Lü et al., 2012].

## Evaluation Metrics

One important issue in the evaluation process is to decide which metric is the most appropriate for the recommendation scenario. This aspect is related to the kind of criteria to pay attention in the evaluation process. This section is only focused on those metrics applied to measure the accuracy of a recommender system, because this is the aspect that it is considered in this work to evaluate our proposal. In particular, the metrics commonly proposed in the literature are divided into different types, which are detailed in what follows.

### *Precision-based Metrics*

This kind of metrics, as well as the evaluation scenario, is derived from the information retrieval field. To that end, they measure the quality of a system by comparing the

recommended items to a ground-truth including the items that are actually interesting for the target users. Among these metrics, it is included:

- **Precision (P):** It measures the probability that a recommended item fulfils the user's preferences.
- **Recall (R):** Recall (R) measures the ratio of recommended items which results to be relevant to the total number of relevant items.
- **F-Measure:** The F1-measure (F) of a system is the harmonic mean of its precision and recall.
- **Area under the ROC Curve (AUC):** The area under this curve (AUC) is a measure of how good is a system by comparing the true positive and the false positive rate.

### *Error-based Metrics*

The aforementioned accuracy metrics are focused on measure whether a recommended item is relevant (i.e., binary relevance commonly applied in Information Retrieval). In contrast, error metrics measure the quality of the systems from the point of view of the error produced in the rating prediction. In more detail, let  $r_{uj}$  be the rating of a user  $u$  to the item  $j$  (contained in the test set) and  $\hat{r}_{uj}$  the rating predicted by the system, the error is given by  $e_{uj} = \hat{r}_{uj} - r_{uj}$ . Among this type of measures, there are two notable metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

- **RMSE:** This metric represents the sample standard deviation of the differences between predicted ratings and the real ratings. The formulation of this metric is as follows:

$$RMSE = \sqrt{\frac{\sum_{(u,j) \in E} (e_{uj})^2}{|E|}}$$

Being  $E$  the set of entries on which the evaluation is performed and  $e_{uj}$  is the prediction error on the entry  $(u, j)$ . One important characteristic of this metric is that it tends to disproportionately penalize large errors because errors are squared in the formulation. Consequently, it is more affected by outliers or large errors.

- **MAE:** The mean absolute error is the average of the absolute error  $e_{uj}$  according to the following formulation:

$$MAE = \frac{\sum_{(u,j) \in E} |e_{uj}|}{|E|}$$

### *Rank-based Metrics*

These metrics try to measure not only how accurate the recommendations are, but also how well they are ordered by the recommender system. Some of the most noteworthy metrics in this category are:

- **Success@K:** This metrics stands for the mean probability that a relevant items appears within the top-K position in the ranking.
- **Mean Reciprocal Rank (MRR):** MRR, as defined by [Chakrabarti et al., 2008], refers to the inverse position of the first relevant item in the ranking:
- **Mean Average Precision (MAP):** MAP measures the mean of the Average Precision (AP) for each recommendation list, where the *AP* is equal to the average for the precision at each “seen” relevant item in the recommendation list
- **Normalized Discount Cumulative Gain (NDCG):** NDCG evaluates a recommendation list by measuring how much the overall quality of a given ranking improves by the appearance of a document with a given relevance (that offered by the recommendation algorithm) in a given ranking position.

### **Which measure is the best?**

In the state of the art, error metrics such as MAE or RMSE have concentrated the interest of researchers. Take for instance the Netflix Prize<sup>1</sup>, which has been the most extensively cited and reviewed recommendation challenge in the last years. Error metrics are intended to test how similar are the predicted ratings in comparison to the ratings actually set by the users. While error metrics are useful to evaluate recommender systems according to their predictive power, they “are not a natural fit for evaluating the top-N recommendation task” [Cremonesi et al., 2010].

The traditional recommendation scenario is based on predicting the rating given by a user to an item (i.e., to fill the empty cells in the user-item rating matrix). In this scenario, error measures are a good indicator of the system performance: the lower the error, the better the rating prediction, and the better the recommender system. However, the goal of Top-N recommendation is to offer the list of most appealing items to the user (i.e., to create a ranked list of items). Although a ranked item list may be inferred from

---

<sup>1</sup> <http://www.netflixprize.com/>

the predicted ratings by sorting the items according to them, recommendation algorithms optimized for error metrics do not necessarily perform well for this task.

In this regard, [Cremonesi et al., 2010] conduct an extensive analysis of different recommenders systems, from basic non-personalized models to the more sophisticated latent factor models. This experimentation found that there was no relation between error metrics and accuracy metrics. In other words, systems achieving a high performance in terms of error measures do not achieve the same performance in terms of accuracy. Based on these results, they claim that a re-evaluation of optimization goals for top-N systems, focusing more on accuracy-based metrics, is needed.

This latter aspect is especially interesting. Recommendation task, when rating prediction is considered, has achieved a significant degree of development in terms of error measures. State-of-the-art algorithms are able to accurately approximate user ratings. As Netflix recently claimed [Fiegerman, 2013], they are able to predict future user ratings better than the users themselves. Nevertheless, this classic scenario has given way to new scenarios like, for instance, Twitter and other social networks, where no user ratings, neither explicit user feedback are available.

These new scenarios entail new challenges with new requirements wherein traditional recommendations proposals, based on rating prediction and error-based evaluations, barely fit (e.g., huge amount of data, no explicit ratings, no explicit user feedback). In this regard, top-N recommendation and accuracy-based evaluations appear as a more suitable solution, more adapted to these new scenarios. In this regard, Precision, Recall, and the F-measure of both, are by far the most widely used metrics to evaluate top-N recommender systems. AUC is also often used to carry out this evaluation. However, they are not able to evaluate the recommendations as a ranking. In this context, ranking metrics such as NDCG, MAP or MRR are preferable to evaluate recommender systems.

### **2.3.2. Specific Application Domains**

---

This section delves into the literature addressing the specific aspects related to different application scenarios and contexts. It does not pretend to be an exhaustive analysis, but it only focuses on those aspects related to the work presented in this thesis.

### 2.3.2.1. Semantic-based Recommendation

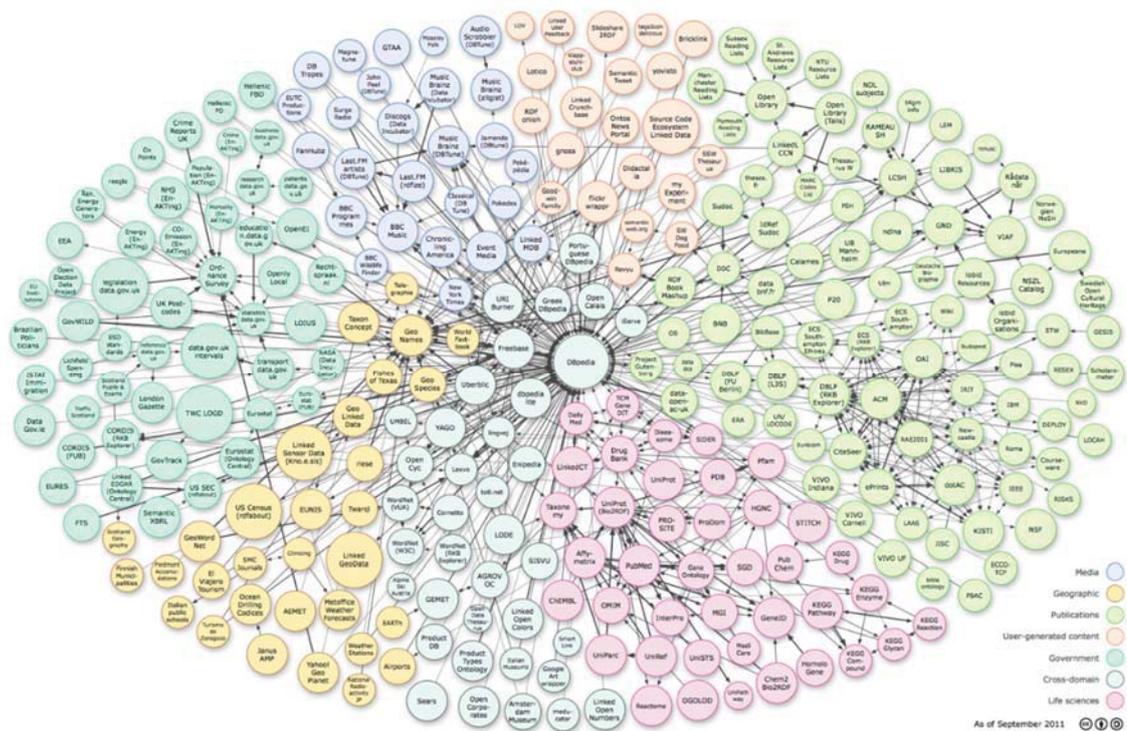
This new type of recommender systems tries to take advantage of semantic-based information in order to better characterize and easily manage the information about the items to be recommended.

Semantic-driven recommendation techniques infer item and user relationships by the interlinking structure included in the semantic resources. This typology has gained significant attention due to its high performance [Bobadilla et al., 2013] and the ability to limit the classical problems related to the recommender systems, such as overspecialization or cold-start situations. In this regard, in [Peis et al., 2008] several semantic recommenders trying to cope with these problems are studied

### Linked Open Data

Over the last years, the amount of semantic information has exponentially grown, driven by the initiatives of data publishing by following the Linked Data principles. This information covers a wide range of fields, such as: geographical locations, people, companies, movies, books, time information, medical information, genes, or drugs among others. The whole amount of public-access information and the links between different initiatives create the so-called *Linked Open Data Cloud*, organized around the DBpedia initiative (see Figure 2.2).

This kind of information has been broadly applied in the recommendation field (e.g. [Blanco-Fernández et al., 2008; Yang, 2010]). This increasing interest has also led to the emergence of specific challenges like for instance the Linked Open Data-enabled Recommender Systems challenge, held at the European Semantic Web Conference (ESWC 2014).



**Figure 2.2** – Linked Open Data Cloud (<http://lod-cloud.net/>)

Linked Open Data  
Cloud

The huge amount of interlinked information contained in the *Linked Open Data Cloud*, can be effectively explored to tackle some recommendation problems [Khrouf and Troncy, 2013], or simply to improve the performance of recommender systems by generating a better item representation through the semantic information. The information available on the Internet has been designed to be readable only by humans. Therefore, computer systems cannot process, nor interpret this information. However, the semantic-based data allow the representation of the information in a way that can be easily processed by the Recommender Systems.

The utilization of LOD for recommendation can be done by only gathering the information related to the items to create richer item representations. This rationale has been followed by works like the one presented in [Passant, 2010b], where the authors propose the utilization of LOD information to create a representation of the artists in a music recommender system; or the work in [Musto et al., 2012], where an enrichment approach for information in music playlists extracted from Facebook is proposed. The utilization of LOD for Content-based recommendation has been also proposed in [Di Noia et al., 2012a], in [Peska and Vojtas, 2015] and in [Di Noia et al., 2012b]. All of these works did not consider the hierarchical structure of the DBpedia categories (derived from Wikipedia) and apply the DBpedia graph (DBpedia entities, their related

features and the links among them) as a flat structure. In contrast, the work presented by [Cheekula et al., 2015] takes advantage of the subjacent hierarchy in the DBpedia categories to infer a taxonomy of the classes. The generated taxonomy is then used to recommend new entities by applying a spreading activation algorithm.

Semantic information has been also proposed to create hybrid recommenders as in [Ostuni et al., 2013], where the ontological knowledge contained in DBpedia is used to create a Content-based recommender to be executed jointly with a Collaborative Filtering system. An interesting point of this work is how the authors solve the integration problem between the Collaborative Filtering and the Content-based data representation. They propose a data model based on a graph representation, including users, item, entities describing the items and the different relations between them. They later use this graph representation to find the most interesting paths connecting users and unknown items through other users with whom the target user share some item or through some properties contained in the consumed items.

Other hybrid proposal is detailed in [Ticha et al., 2014] wherein the authors propose a user modelling based on the semantic features of the consumed items (movie features extracted from IMDB and Rotten Tomatoes) and a CF-based algorithm that relates users based on the semantic profiles created in the previous step. An interesting issue addressed in this work is the splitting of semantic features in order to create two user models: one based on Dependent semantic features (very variable number of values like, for instance, *actors* or *directors*) and another on Non-Dependent ones (very few values like *genre* or *country*). Hybrid LOD-enabled recommendation is also presented in the work in [Luo et al., 2014] that proposes a hybrid user modelling by integrating the object user model (i.e., users are described by means of the LOD features of the consulted items) and the predicate user model (i.e., the modelling of the user related history) in a movie recommender system.

Also in [Khrouf and Troncy, 2013] a hybrid system is proposed for event recommendation. To that end, LOD are used to create an event description to be used in a Content-based system, complementing the operation of a Collaborative Filtering system using social information about the users. A similar idea is applied in the work in [Meymandpour and Davis, 2015], where LOD is used to create the *item neighbourhood* (i.e., to find similarities between items) in a Collaborative Filtering system. Other related works are presented in [Basile et al., 2014] and in [Ristoski et al., 2014] in the context of the Linked Open Data-enabled Recommender Systems Challenge. Both works are based on the combination of different recommenders including the semantic information extracted from DBpedia.

### *Feature Extraction*

It exists two main ways to extract features from the Linked Open Data databases:

- Use the Uniform Resource Identifier (URI): The URI directly connects the content or entity to get the features to the related Linked Open Data features and, consequently, it is the most desirable way to conduct the extraction. However, it is not normally possible to have access to the URI related to the contents.
- Entity Linking: When it is not possible to have the URI, some entity linking approach has to be applied in order to find the resource/s wherein the features are included. The problem of this methodology is that some mistakes may happen in the resource identification (i.e. there are more than one resources related to “Michael Jordan”). In this sense, information retrieval and disambiguation techniques should be applied to refine the operation.

### *Similarity*

Similarity measures are one of the most important issues in the RS operation, especially in the Content-based RS. By means of the similarity measures, the system can set the closeness degree between two items (using the content of the items) and consequently, can offer similar items as recommendations.

Most similarity measures come from the Information Retrieval field; for instance: cosine similarity, BM25, Pearson Correlation, Jaccard Similarity or Manhattan Similarity. All of these measures can be also applied for dealing with semantic information. However, some works have proposed more sophisticated similarity measures, especially adapted to deal with semantic information. [\[Di Noia et al., 2012b\]](#) propose an approach to find similarity between items by exploring the RDF graph. This approach is based on an adaptation of Vector Space Model (VSM) to include the information in the format of RDF triples.

### *Related Problems*

On the other hand, with the utilization of these interlinked data new problems arise, mostly related to the integration of all of this knowledge in the recommender systems operation. The information (related features) contained in Linked Open Data databases is huge, making necessary the selection of the most interesting features. This selection can be carried out manually, by identifying the most interesting features to retrieve, or automatically by applying some data mining technique.

## Ontologies

Ontologies offer a formal representation of a knowledge domain. This formalization facilitates the managing of the data belonging to the modelled domain, as well as the reasoning upon the data in order to infer new relationships or data structures. This inferred knowledge can be then applied to offer new recommendations. For instance, a domain ontology about cinema, including information about movies, actors, directors, genres, dates..., may allow offering recommendations such as: movies of the actor X (who the user likes) about the IIWW (a topic interesting for the user), but not directed by Y (who the user really dislikes), released between 1980 and 1990. Some works in the literature making use of ontologies for the recommendation process are: [Anand et al., 2007]; [Cantador et al., 2008a]; [Chen et al., 2014]; [Mobasher et al., 2004].

Ontologies have been also proposed to enable the diversification of the result list, like in the work of [Bedi and Richa, 2015], where spreading activation on the concepts of a domain ontology is applied. Ontologies can be also applied to the user information. In this sense, ontologies especially developed to represent personal data, such as the Friend-Of-A-Friend (FOAF) ontology, have attracted the interest of the researchers like, for instance, the authors of [Sabucedo et al., 2014] who propose the use of FOAF to model users in a recommender for e-Government services.

### 2.3.2.2. Social Recommendation

The utilization of Social Information in Recommender Systems has become a trendy field within the RS research. As the Web 2.0 has developed, the amount of social information available has exponentially grown (e.g. user-related information, social network based information such as followers or friends, tags, post, blogs...). Everyday billions of users interact with online social networks, generating a huge amount of data [Cheung et al., 2011]. The hypothesis in which these systems rely upon is that the recommendation process has an inherent social dimension. For instance, one user likes an item because the item content, but also because the interaction with their social environment (family, friends, co-workers...). Furthermore, users likely show interest by an item when a number of users in their social environment did [Mossel and Roch, 2007]. Hence, the use of this social information can be useful to replicate this process.

Related to this growing in the use of Social Information in RS, the interest in Content-based RS has also grown. Content-based RS allows including this social information in the recommendation process [Arazy et al., 2009]. As it was pointed out in [Bonhard et al., 2007], the operation of recommender systems in environments suffering from

information overloading has been not as good as expected, especially in social networks. In this regard, there is a need to integration of the social information in the recommendation process. From the first seminal works in this area, it has been proven that social information in Content-based RS can significantly improve the performance of the RS [Ma et al., 2008; Ma et al., 2009; Ma et al., 2011; Tan et al., 2011] and the quality of the predictions [Arazy et al., 2009; Carrer-Neto et al., 2012]. In these works and in the state of the art in general, Social Information has been used as contextual information in order to mitigate the sparsity problem (i.e. to try to fill the information gaps in the user-item matrix). The overspecialization problem related to Content-based recommenders is also addressed by means of Social Information [Alexandridis et al., 2013; Hu and Pu, 2011; Ugander et al., 2012]. Academic recommendation has also benefited from the application of social information, as for example the Content-based recommender proposed in [Rohani et al., 2014]

Some other researches have applied social information with other minor objectives: 1) propose or generate new types of RS, or 2) identify relationships between social information and collaborative process. Some of these new works are exposed in [Bobadilla et al., 2013].

Nevertheless, Content-based RS is not the only application field of social information. Collaborative Filtering RS have also benefited from the use of Social Information. In [Pham et al., 2011] a Collaborative Filtering algorithm based on the hierarchical clustering of the social information is proposed. Similarly, clustering-based approaches for Collaborative Filtering RS using social information are presented in [Pitsilis et al., 2011] and in [Alexandridis et al., 2013]. On the other hand, [Bidart et al., 2014] propose a graph-based recommendation to suggest cities to visit in the context of an e-tourism system (TripAdvisor). The authors of [Colace et al., 2015] also propose a graph-based recommendation, integrating information from several data sources and taking different features into account, such as: preferences, opinions, behaviours and user feedback. Some other methodologies in this field are summarized in the survey of [Yang et al., 2014].

Besides its individual application in Collaborative Filtering or Content-based systems, the hybridization of both kind of systems by means of social information has been carried out. In [Carrer-Neto et al., 2012], their authors propose a system that puts together social and semantic based information. Semantic-based information is used to develop a Content-based recommender system that offer as recommendations items similar to the user preferences according to their semantic features. Social Information is then used in a Collaborative Filtering system that gathers the ratings of similar users according to the

social information to create a second recommendation list. Finally, both list are fused together leading on the final recommendation list. A similar idea is followed in the e-Government recommender presented in [Sabucedo et al., 2014]. Herein, the items and users are described by means of ontology-based models and social information (folksonomies) to enable a Content-based recommendation which result list is ranked by combining it with a Collaborative Filtering algorithm.

Social information is also closely related to trust and reputation [O'Donovan and Smyth, 2005; Victor et al., 2011]. The idea is to use this social information to try to set the trust (or reputation) of:

- The trust degree of a given user, either through: 1) implicit information gathered from social networks, as in [Golbeck, 2006] or in [Park et al., 2007], where the concept “you are what you consume” is proposed. It bases its recommendation in a Bayesian modelling of the item features and user interactions to represent users by means of the features of the items they like. 2) Explicit ratings of the rest of the users [Yuan et al., 2010].
- The trust degree of a given item, obtained by means of user feedback [JÃ, sang et al., 2007] or analysing user-item interactions [Kitisin and Neuman, 2006].

More information on specific ways to obtain and apply social information in RS, especially developed recommender algorithms, new fields of application of RS driven by social information and some other related issues can be consulted in [Bobadilla et al., 2013] in the related section. [Zhou et al., 2012] includes a more specific review of the state of the art of social recommendation. Other recommended readings related to social recommendation can be found in [Esslimani et al., 2011; Golbeck and Hansen, 2011] and [Guy et al., 2009].

### 2.3.2.3. News Recommendation

News reading has experimented a vast change from its beginning. The first jump happened with the change from physical newspapers to the digital formats. As long as Internet has grown, every newspaper or news agency offers its content via web. Furthermore, an increasing number of media contents are only offered via online. The number of contents available has become too large and, consequently, platforms like Google News<sup>2</sup>, Yahoo! News<sup>3</sup>, or Digg has appeared to aggregate and resume these

---

<sup>2</sup> <https://news.google.com/>

<sup>3</sup> <http://news.yahoo.com/>

contents. However, these platforms were also insufficient because of the huge amount of information to process. In this sense, recommendation algorithms have been proposed to address the information overload in these system, like the one presented in [Das et al., 2007] dealing with Google News, or the one in [Lerman, 2007], applied to Digg.

The problem is ever increasing, driven by the appearance of self-generated content (the so-called *information explosion*). Each user has become a potential journalist, being able to create, share or spread news stories from their computer, mobile phone, or tablet. In this sense, the Web 2.0 and social platforms like Facebook<sup>4</sup>, Twitter<sup>5</sup> or Flipboard<sup>6</sup>, among others, have moved and often replaced the traditional mass media. Such was the boom that even mainstream media have been forced to offer their contents in these social platforms. Therefore, in the current scenario where the amount of contents can be overwhelming to the users, the challenge of news recommender systems is to help users find news articles interesting to read, among the overall available contents.

News recommendation is not a new research field. It has been mainly studied from the point of view of Content-based recommendation [Cantador et al., 2008b; Kompan and Bieliková, 2010; Phelan et al., 2011], although not exclusively [Saranya and Sadhasivam, 2012]. However, the current context presents a challenging scenario [Liu et al., 2010a]. First, news recommendation cannot be considered as equal than other recommendation domains. In news recommendation, users are looking for novel contents. For instance, in a music recommendation scenario there is no problem in recommending an item already known for the user (a user can be interested on listening a sing more than one time). In contrast, it does not make sense to recommend a news story already known and read by the user.

Other issues related to news recommendation to be taken into account are reflected in [Tavakolifard et al., 2013] and in [Garcin and Faltings, 2013]. Some of the most important are: 1) trendy news should have a high relevance regardless of the degree of relatedness to the user profile. Related to the previous one: 2) freshness also represents an important aspect. Usually the most novel news has to be deemed as more relevant than older ones. 3) In news recommendation, user preferences are commonly event-oriented; that is, the preference in a specific topic or group of news are only due to an

---

<sup>4</sup> <https://www.facebook.com/>

<sup>5</sup> <https://twitter.com/>

<sup>6</sup> <https://flipboard.com/>

ongoing event (e.g. a peak in the readings about politics because the proximity to the presidential elections).

Experimental workshops like the International News Recommender Systems Workshop and Challenge 2013 (NRS 2013) [Gulla et al., 2013], held in the ACM Conferences Series in Recommender Systems 2013 (RecSys 2013); the News Recommendation Evaluation Lab (CLEF-NEWSREEL), as well as contests, such as the pLista Contest, have focused the attention of the researchers in news recommendation. Their leitmotif has been the development of news recommender systems capable to work online in a real environment. Even best performing algorithms in a theoretical environment can be useless or inefficient in a real environment. In this sense, such workshops and challenges not only offer the possibility to have access to a real recommendation scenario, but also evaluate the algorithms according to metrics adapted to this context. Some novel works done within this context have been [Said, A. Bellogín, A. and de Vries, A., 2013] or [Garcin and Faltings, 2013]. The former proposes an infrastructure adapted to the real-time recommendation scenario and testes different state of the art algorithms. The later uses a recommendation infrastructure to demonstrate the validity of different recommendation approaches, such as Context-Tree recommendation, Collaborative Filtering, and a Content-based approach.

Social networks and the information that they provide are focusing most of the works related to this area [Abel et al., 2011; Tao et al., 2012]. Twitter provides a huge data repository for many tasks, news recommendation among them. Given the abundance of news-related content published by Twitter users, it appears as the ideal information source for news recommendation. As it is posed by the analysis in [Kwak et al., 2010], about 85% of the post in Twitter are about headlines or persistent news. In addition to the vast amount of information, the immediateness is other of the main advantages of Twitter. In many cases, news appear in Twitter before in any other source, even news agencies [De Francisci Morales et al., 2012]. In fact, as it is said in [Zhao and Rosson, 2009], users tends to immediately share real time information about on-going events.

The main rationale of these approaches is to take advantage of the information about users contained in Twitter to model their preferences. Twitter provides not only Content-based data (e.g., tweet contents, news shared in the tweets) but also social-based information (e.g., followers, folowees) and contextual information (e.g., geo-position or date of the tweets) about the users. The Twitter-based news recommendation also poses a number of challenges related to the Twitter environment (e.g., large volume of tweets, scalability issues, tweets include jargon and slang, shortness of the tweets content) and

some other related to the specific news recommendation environment (e.g., news are highly dynamic and have a short life cycle).

The works in the literature for Twitter-based news recommendation go from simple TF-IDF-based approaches [Phelan et al., 2011] to more sophisticated systems applying classification [O'Banion et al., 2012], topic modelling [Michelson and Macskassy, 2010], context trees [Garcin et al., 2013] or learning approaches [De Francisci Morales et al., 2012]. Some approaches have intended to take into account special Twitter issues, such as hashtags [Abel et al., 2011] or trending topics [Asur et al., 2011].

#### 2.3.2.4. Twitter Recommendation

Twitter<sup>7</sup> is a microblogging service in which users share public short text messages (up to 140 characters) about, mostly, whatever. Answering the question “What’s happening?” has turned to be a social phenomenon and Twitter has rapidly grown to become one of the most important sites of Web 2.0 with more than 500 million users posting hundreds of millions of tweets per day. Users in Twitter may also follow (and be followed by) other users in order to subscribe to their tweets. So, Twitter can be considered as a huge social network (or an “interest network” as Twitter defines it [Gupta et al., 2013]) wherein users interchange public information, opinions, and content among them. This vast amount of public social information has attracted the interest of the researchers in several areas, such as web mining, social sciences, marketing, opinion mining, or recommender systems.

Focusing on recommender systems, the conducted researches have tried to address the overload problem; that is, the users are overwhelmed by the amount of available information in Twitter. According to what has been reported in [Qu and Liu, 2011], Twitter users follow on average 80 other users, leading to hundreds or thousands of tweets per day. Although the information overloading problem is the classical scenario in the recommendation field, in Twitter it is even more problematic. Twitter presents the tweets in chronological order, so interesting tweets may be covered up by new tweets, not so interesting.

In general, all researches have intended to take advantage of the huge amount of public information in Twitter to improve the recommendation process. If you have more information, it is rational to think that you will infer better user preferences. However,

---

<sup>7</sup> <https://twitter.com/>

the main benefit is also the main problem; dealing with such amount of information poses a challenge to the state-of-the-art recommender systems.

The conducted works go from the most basic bag-of-words approaches [Chen et al., 2010] to other more sophisticated techniques including specific Twitter contextual and social features. These latter have proved to be of great interest for the recommendation process. Some of them are: the user of real-time information through the Twitter API [Diaz-Aviles et al., 2012], the geographical information available about the users, social relationships among users [Armentano et al., 2012; Yan et al., 2012], existence of user-generated thematically-based annotations (hashtags) or URLs included in the tweets which are likely to be related to the tweet content [Huang et al., 2010; Laniado and Mika, 2010].

Other researchers focus on other interesting Twitter-related issue: the temporal dimension. As Twitter allows the monitoring of a specific topic, set of contents or user activity during a time span, it is possible to study how they evolve and spread along the time. In this sense, the work in [Abel et al., 2011], studies how to integrate this information to improve the recommendation process and how user preferences relate to the temporal dimension in Twitter.

In general, most of the works conducted in Twitter Recommendation have been focused on filter out the Twitter stream (i.e., all the tweets shared by the users whom a user is following) in order to recommend the most interesting tweets in this stream. In these sense, there are three user behaviours especially interesting for the recommendation process [Chen et al., 2012]: follow another user, publish a tweet and retweet another user tweet. Retweet mechanism allow users to share tweets from their stream (i.e., coming from the users they follow) which they have considered interesting. From a recommendation-based point of view, retweets are an active user-feedback whereby a user explicitly sets their interest in a content. Some works focused on studying the impact of retweets in the recommendation process are presented in [Zaman et al., 2010].

However, some other works have addressed the problem from different points of view. For instance, [Lin et al., 2013] apply Twitter data to recommend mobile applications; [Wang et al., 2013a] propose a system to recommend users to be referenced in the tweets; [Gupta et al., 2013] propose the recommendation of related users to follow, based on the analysis of the whole Twitter graph (i.e., the users in Twitter plus the follower/followee relationships).

The recommendation of users in Twitter has become a hot topic in the Twitter-based Recommendation. Some interesting works are presented in [Garcia and Amatriain, 2010; Hannon et al., 2011; Kwak et al., 2010], or [Weng et al., 2010]. Especially interesting is this latter, where TwitterRank, an extension of the Google's PageRank algorithm, is presented. It measures the influence of a user by analysing their link structure. Although it is out of the scope of this dissertation, some of the conclusions learnt from these works can be applicable to the recommendation of tweets as a way to weight influential users and, consequently, favour their tweets in the recommendation process.

Regarding the typology of the applied recommendation approaches, both Collaborative Filtering and Content-based have been proposed. In addition, more sophisticated methodologies applying some of the special Twitter features have been proposed. In what follows, a review of the most novel works classified by the type of recommendation is presented.

### Collaborative Filtering Recommendation

Collaborative Filtering appears as the obvious methodology to try to exploit the links among users (i.e., following relationships) [Abel et al., 2011; Hannon et al., 2010; Ramage et al., 2010; Sun et al., 2009; Uysal and Croft, 2011; Yan et al., 2012].

However, because of the high volatility of Twitter information, traditional CF algorithms present some problems when they have to infer novel recommendations. Some solutions proposed have to do with the inclusion of temporal aspects in the Collaborative Filtering process. The most basic methodology would be to split the input information, considering only the very recent content. It poses a problem: the restriction of the amount of input information may not be suitable to capture all the interesting data patterns or users preferences [Muthukrishnan, 2005]. In this sense, more elaborated algorithms are needed. For instance, in [Diaz-Aviles et al., 2012] it is presented a strategy for updating CF-based models by applying active learning techniques upon randomly selected input data coming from the Twitter stream in real time.

### Content-based Recommendation

Several works have been proposed following this approach for Twitter recommendation [Chen et al., 2010; Naveed et al., 2011; Ramage et al., 2010]. However, given the special characteristics of the tweets (e.g., shortness, noise, informal language or typos and grammatical mistakes), the application of such techniques is not straightforward and it might lead to inaccurate recommendations [Liu et al., 2011]. To mitigate these problems,

the processing of the tweet content is a commonly applied methodology [Abel et al., 2011; Chen et al., 2012; Kapanipathi et al., 2011; Stankovic et al., 2010].

An important factor to discover user interests is the Named Entities included in the tweets. Entities are, in general, a good signal to set the topic of some content (e.g., tweets including Madonna are likely to be about music whereas tweets including BMW about automotive). Entity-based approaches, like the one presented in [Michelson and Macskassy, 2010], therefore aim to discover topics of interest for the users based on the named entities and references in the tweets. In the analysis conducted in [Abel et al., 2011], their authors conclude that the entity-based profiles outperform other approaches including hashtags or topic-based profiles.

The managing of special Twitter features, mainly references and hashtags, has been specially studied in the literature. Namely, hashtags have attracted the interest of researchers in the recommendation fields. They were thought to be topic-based annotations created by the users (i.e., a sort of folksonomy), becoming especially important because their widespread application, mainly due to: 1) Twitter identifies the highest published hashtags and offers them as “Trending Topics” and 2) they have been used as publicity for events, shows, mainstream media, TV shows, sport events, etc. In consequence, its analysis, managing and processing might lead to a better understanding and identification of the content in the Twitter stream, allowing the more accurate recommendation of such content. In this regard, the authors of [Laniado and Mika, 2010] analyse and introduce metrics to characterize the hashtags, meanwhile in [Huang et al., 2010] it is analysed their temporal dimension.

On the other hands, dealing with hashtags entails some problems. In the same way than it happens when using folksonomies, the creation of hashtags is not restricted, leading to data redundancy. That is, users create different hashtags for the same event (e.g., #WorldCupBrazil, #WorldCup2014, #FifaWorldCup2014) and in different languages (e.g., #MundialBrasil, #Weltmeisterschaft, #CopadoMundo). One possible solution to this latter problem is proposed in the work in [Rahman et al., 2013]. Herein the authors propose an algorithm for hashtag recommendation; that is, to recommend hashtags wherewith to label a given tweet. In this way, it is expected to reduce the redundancy by recommending the same set of hashtags for similar tweets. They proposed a recommendation algorithm based on a graph (“Hashtag Graph”) based on the occurrence of a pair of hashtags in a given tweet. To infer new recommendations, the graph structure is navigate, offering the most similar hashtags, according to an own-

developed measure called “Weighted Tag-Similarity”. Other approaches to cope with the same problem are also presented in [Gassler et al., 2012; Zanardi and Capra, 2008].

### **Social and Contextual-based Recommendation**

This social and contextual information could be considered as a kind of Content-based information. However, given the special twitter framework, it should be considered as a different kind of information. In fact, a research line has appeared, including many proposals that address the recommendation process only from this point of view. All of these proposal share the same rationale: trying to infer user preferences from the social and contextual information related to the tweets users have consumed or shared and to the user profiles. It intends to reduce the sparsity problem (i.e., lack of enough information relating users and items). While sparsity is a challenging problem in recommendation, it is even harder to solve in environments such as Twitter. The huge amount of available information makes harder to find user-item relationships than in other more limited scenarios (e.g., movie recommendation, book recommendation and tourist recommendation). Work proposals like the one presented in [Cui et al., 2011] or the one in [Yang et al., 2011] delve into the managing of the social aspects to improve the recommendation process.

### **Hybrid Recommendation**

As with other kind of recommenders, those operating on Twitter suffer from topology-related drawbacks. Collaborative Filtering based recommender lose the information provided by the textual contents, while Content-based ones do not take into account other factors besides the content that may be representatives of the user preferences: tweet quality, author of the tweet or social information about the user among others.

A hybrid approach trying to integrate both content and social features is presented in [Chen et al., 2012]. In this work the authors integrate different factors (tweet contents, social information, tweet quality or authority of the publisher) to capture the personal interest of the users in a collaborative ranking recommendation algorithm.

One hybrid-based approach is presented in [Yan et al., 2012] where the authors propose a bipartite graph including users and items and a co-ranking algorithm to infer recommendations from the information in the graph.

---

### 2.3.3. Related Topics

---

This section presents some topics related to the issues addressed in this work and particularly to our FCA-based proposal. In brief, it includes recommender systems that make use of graphs to model user-item interactions (i.e., our FCA-based representation is a graph-like model of these interactions); matrix decomposition methodologies applied to the recommendation field (i.e., FCA can be seen as a matrix decomposition methodology); approaches proposing a common representation space for modelling user-item interactions (as we do in this work); and, finally, some works directly applying FCA for the recommendation process.

#### 2.3.3.1. Graph Based Recommendation

Graphs allow the representation of users and items, as well as the relationships between them, making explicit the inherent structure of these relationships. A graph-based representation can identify, for instance, the relationships between items, grouping together similar ones [Gori and Pucci, 2007]; or also the relationships between users, making possible the detection of communities (group of users sharing similar tastes or preferences) [Bidart et al., 2014]. This latter user-user relationship is especially interesting since, as it has been proved in the state of the art, it offers the best performance for Collaborative Filtering recommendation [Hernando et al., 2014]. In addition, one of the main advantages of graph-based methods is that they overcome the problem of the user-item sparsity: users (or items) do not need to share many ratings in order to be considered neighbours (i.e., to be considered as being related) as long as some path exists between them.

This graph-based representation can be also very useful for predicting item rating or creating recommendation rankings by the propagation of information throughout the graph. In this sense, [Adomavicius and Kwon, 2011] present a graph-based approach for improving the diversity in the recommended item set. [Fouss et al., 2005] and [Brand, 2005] try to gain insight into the graph structure information to offer more accurate recommendations. Especially interesting is the proposal in [Shi, 2013] wherein different recommendation related aspects (such as accuracy, diversity, similarity, and long tail) are addressed by means of a graph-based approach using what the author called cost-flow recommendation. This approach is based on navigating across the graph representation and offering as recommendations those items with a lowest cost, according to the four aforementioned aspects.

Besides, graph representation allows the visualization of the set of information and the set of relations linking the information. This feature is especially interesting for the recommendation field. This kind of visualization could really improve the user experience in their interaction with a recommender system. Nevertheless, only a few works have addressed the graph-based visualization for recommender systems. Some remarkable examples are the works presented in [Hernando et al., 2014] or in [Hernando et al., 2013]. These two works, and the majority of the ones in the state of the art, propose a graph visualization in the form of a hierarchical graph map [Abello, 2004]. The reason for this is that this type of representations facilitates the visualization of very large graphs, such the ones related to recommender systems.

### 2.3.3.2. Matrix Decomposition based Recommendation

This kind of techniques pursues to reduce the matrix dimensionality, without losing data representativeness, according to subjacent factors. The basic rationale behind this methodology is to exploit the fact that a significant number portions of the user-item matrix are highly correlated (i.e., there are highly correlated groups of users and items) according to a set of latent factors. In consequence, the data can be represented in a low-rank matrix, based on these latent factors.

Applied to the recommendation field, these techniques allow the decomposition of the original user-item matrix in these latent factors, according to the interactions between users and items. These factors include closely related user-item groups that are expected to enable accurate recommendations. In fact, these methodologies perform particularly well in the recommendation field given that:

- 1) User-item matrices are highly sparse; therefore, the expected reduction is larger than in other contexts.
- 2) There is a clear set of latent factors that have given rise to the data in the user-item matrix (the user preferences), expressed through the user-item interactions.

## Decomposition Methods

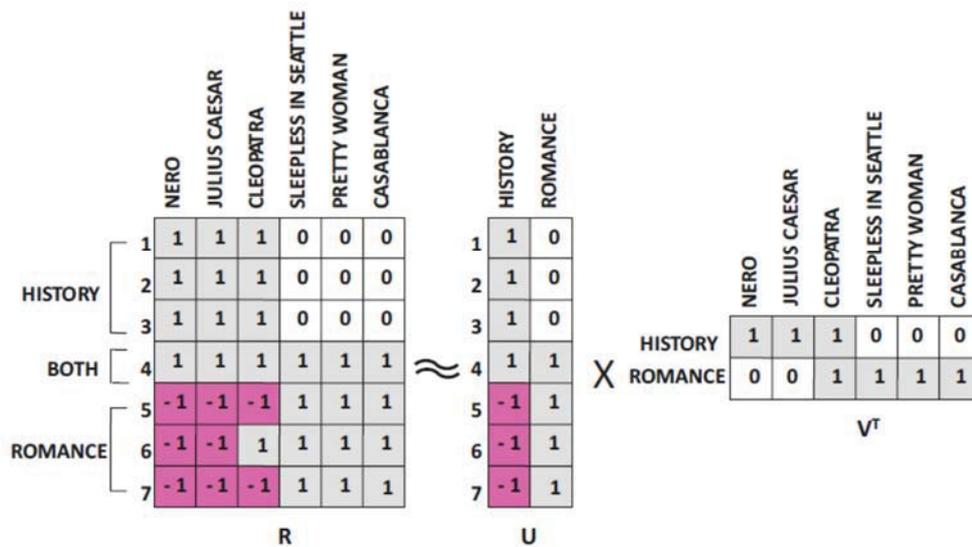
Many methods have been proposed for matrix decomposition; some of the most common are presented below, paying especial attention to those applied to the recommendation field.

### *SVD*

SVD (Singular Value Decomposition) [Eldén and Berry, 2008] is a method to decompose a rectangular matrix  $A \in \mathbb{R}^{m \times n}$  ( $m > n$ ) into the product of three matrices:

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V$$

In recommendation, rows of  $U$  and  $V$  can be interpreted as vectors of the user's and item's loyalty (attitude) to a certain topic (factor), while the singular values of  $\Sigma$  as the importance of the topic/factor among others. That is, given a user  $u_i$  the first row in  $U$  will be the interest of the user  $u$  for the implicit topic/factor in this row; the first row in  $V$  includes the items corresponding to this factor; and finally, the values of  $\Sigma$  the weight of the user interest for the corresponding factor. [Figure 2.3](#) extracted from the book of [\[Aggarwal, 2016\]](#) illustrates this point. This figure presents a rating matrix with 7 users and 6 items, where users show a clear tendency in their ratings, related to the genre of the movies (historic and romance). As a result, this matrix can be factorized into rank-2 factors, where the matrix  $U$  shows the interest of the users towards the genres and matrix  $V$  shows the relationship between these genres and the movies.



**Figure 2.3** – Example of rank-2 matrix factorization

Based on the idea of SVD, more refined methodologies have been proposed, as the well-known SVD++ [\[Koren, 2008\]](#) that have offered state-of-the-art results in the Netflix Challenge. SVD++ is an enhanced SVD that takes into account the implicit feedback information in the form of the set of items that the user has already rated. A temporal enhancement of SVD, called time-SVD++ has been proposed [\[Koren, 2009b\]](#). The idea behind this algorithm is to model the SVD parameters as a function of time. In particular, time-SVD++ assumes that the user and item biases as well as the user factors are functions of time.

### ***BMF***

Like SVD, BMF (Binary Matrix Factorization) is focused on the reduction of the matrix dimensionality, but applied over a binary matrix; where binary matrices represent the interest of an user by an item with a binary value (0 for uninterested/unseen and 1 for interested/seen).

BMF is a decomposition of the original binary matrix  $I \in \{0,1\}^{n \times m}$  into a Boolean Matrix product  $P \circ Q$  of binary matrices  $P \in \{0,1\}^{n \times k}$  and  $Q \in \{0,1\}^{k \times m}$  for the smallest possible value of  $k$ :

$$(P \circ Q)_{ij} = \bigvee_{l=1}^k P_{il} \cdot Q_{lj}$$

### ***NNF***

NNF (Non-Negative Matrix Factorization) refers to the specific matrix factorization methodology that applies to Non-Negative Matrices. NNF decomposes the original matrix  $V \in \mathbb{N}^{n \times m}$  into two matrices  $W \in \mathbb{N}^{m \times r}$  and  $H \in \mathbb{N}^{r \times n}$  (where  $r$  is the number of latent factors, also called aspects), such that:

$$W * H = V$$

This technique, as stated by [Aggarwal, 2016], facilitates the understanding of the user-item interactions, especially in cases in which the users have no mechanism to specify a dislike (i.e., unary ratings).

### ***FCA***

As already mentioned, FCA can be seen as a matrix decomposition methodology for binary data, that factorized the input matrix, in the form of a *formal context*, into a set of latent factors (i.e., *formal concepts*) that group closely objects according to their shared attributes. FCA guarantees that this factorization is unique (i.e., the same input data always results into the same formal concept set) because, as demonstrated by [Nenova et al., 2013], the factorization provided by FCA is the optimal factorization of the input matrix.

## **Recommendation Approaches**

Recommendation techniques based on this methodology mainly applies a Collaborative Filtering approach, that is, find the most similar users to a target user. Considering this, given the reduced matrices obtained through the application of the aforementioned techniques, the recommendation will be based on finding the most similar users (neighbourhood) according to the latent factors. Then, other factors related to the users

in the neighbourhood will be used to recommend new items, associated to these factors. The rationale is that the latent space (obtained through the decomposition of the user-item matrix) will better identify the implicit user-item relationships by means of the detected latent factors [Koren, 2008].

Some of the earliest uses of latent factor models were proposed as stand-alone methods for recommendation [Aggarwal and Parthasarathy, 2001; Sarwar et al., 2000], leading to state-of-the-art results. Because of this early proposals, different forms of matrix factorization has been proposed, including factory analysis [Canny, 2002], latent semantic models [Hofmann, 2004], NNF [Zhang et al., 2006]. After the popularization of these models by the Netflix Prize contest [Bell and Koren, 2007], later works presented more advanced models. In [Paterrek, 2007] the aspects related to the latent factor models are discussed (e.g., asymmetric factor model, biases) and they were proposed some of the basic innovations that were later combined to create state-of-the-art methodologies, like SVD++ [Koren, 2009a; Koren, 2008]. Other novel proposals in this regard are presented at [Devooght et al., 2015] and at [Jain and Dhillon, 2013].

The application scenarios for latent factor models include fields so much diverse as e-commerce [Schafer et al., 2001], movie recommendation [Pirasteh et al., 2015], or travel recommendation [Noulas et al., 2012] [Wang et al., 2013b]. They have been also applied to deal with Big-data environments as in the work in [Yu et al., 2014].

The study of the approaches addressing the recommendation task by applying FCA to factorize the user-item matrix are left for the next section focused on the FCA-based recommendation techniques.

### 2.3.3.3. Common Representation Space for Recommendation

The works in this regard are focused on fusing the representation spaces of users and items (i.e., the rating and content matrices). By sharing a common representation, it would be only necessary to find the closest items in the space to the representation of the user profiles.

Knowledge-based representations have been commonly proposed as basis for this common representation. An example is presented in [Shoval et al., 2008], wherein the authors propose a common ontology to represent user and items. In this work, it appears one of the problems to tackle when using ontology representation, the need to define a similarity measure to relate user and items. This latter work and others in the literature, e.g. [Cantador, 2008], present a similar idea: the creation a tripartite representation connecting users and items in common representation layer based on concepts relating

them. In this regard, the work presented by [Huang and Bian, 2015] proposes the application of FCA to relate users and items in a common representation space. In particular, they apply FCA to link two ontologies, one reflecting the tourists' preferences and one for the services offered by the tourism information providers. To that end, the concepts of both ontologies are taken as the objects of the formal context while the attributes are seven concepts manually extracted from the tourism literature. The final concept lattice may be seen as a formal representation of the users, described in the tourist ontology, and the items, described in the tourism providers' ontology and the relationships among them.

Graph-based representations have been also proposed in order to provide a common model for users and items. For instance, [Chen et al., 2013] propose a bipartite graph as a model to integrate both user and item representations. In the graph, users are related according to the links among them, while items are represented according to their content similarity.

In the context of hybrid systems, some techniques have been proposed in this sense. The rationale is to combine Collaborative Filtering systems, which operate over the user-item dimension, and Content-based systems, which operate over the item-attribute dimension, into a unique system that integrates the information of both dimensions into a common representation. One of the first attempts along this line is presented at [Basilico and Hofmann, 2004], where the authors propose a unified approach to integrate all the training information (user-item matrix and item attributes) into joint feature maps. The work proposed at [Singh and Gordon, 2008] present an analogous approach that applies a collective matrix factorization model to simultaneously factorize the user-item and item-attribute matrices into a common model. [McAuley and Leskovec, 2013] also presents a factorization model for combining review text and ratings. Regression-based models have been also proposed to create this latent factor models [Agarwal et al., 2011b] [Ning and Karypis, 2012]. Finally, proposals like those in [Moore et al., 2013], [Feng et al., 2015] or [Wu et al., 2013] propose the use of embeddings to represent users and items into a common latent space.

Concerning this latter proposal, the recent hype of Artificial Neural Networks (ANN) and Deep Learning has attracted the interest of researchers in the field of Recommender Systems (more information and an extensive review of Deep Learning literature can be consulted at section 2.2.6). ANNs have proven to learn complex latent representations of the input data. In this regard, some works have tried to take advantage of this ability to implement common representation spaces for users and items. The first attempt in

this direction is presented at [Salakhutdinov et al., 2007], where a Restricted Boltzmann Machine is applied for top-N Recommendation. [Iyyer et al., 2014] propose a neural network for question answering over paragraphs that models both, questions in the form of paragraphs and answers, in the same vector space. They thus expect to encourage questions representations to be near their correct answer representations and far away from incorrect answers in this common vector space. A similar idea is presented in [Socher et al., 2014]. Their authors present a model based on a multi-modal representation space to map the outputs of a convolutional network applied to detect visual objects in images and vector representations for sentences generated by a DT-RNN (Dependency Tree – Recurrent Neural Network). It is expected to allow the linking of images to sentences that may describe the image. Similar approaches for linking images and words in a common space are presented in [Socher and Fei-Fei, 2010] and in [Srivastava and Salakhutdinov, 2012]. More recently, Autoencoders have been proposed for predicting user ratings [Sedhain et al., 2015; Wang et al., 2015a]. Especially interesting is the approach presented by [Wu et al., 2016a] that proposes a new model for CF recommendation. The authors propose a Denoising Autoencoders to build a latent representation of users and items, which is able to outperform other state of the art recommenders for top-N recommendation.

#### 2.3.3.4. FCA-based Recommendation

The context of a recommender system can be interpreted as a bipartite graph partitioned into users ( $U$ ) and items ( $I$ ). The edges in this graph, of the form  $\rho = r(u, i)$ , establish the relation of interest of the user  $u$  by and item  $i$  weighted with a rating  $r$ . Following the FCA theory, the triple  $(U, I, \rho)$  can be interpreted as a *formal context* (or a *recommendation context*), according to the definition in section 3, which can be factorized into a set of FCA *formal concepts* including the set of users that have rated the same set of items.

In this regard, several works have addressed the recommendation problem from the point of view of FCA. In [Simovici et al., 2012], the authors apply the FCA basis to obtain user subsets sharing the same purchases. Then, they calculate the entropy of each subset in order to find the most suitable user sets to recommend a specific item. In [du Boucher-Ryan and Bridge, 2006] the authors propose a Collaborative Filtering approach that intends to take advantage of the structure of the lattice to find similarities between users according to the items with which they interact. To that end, two methods based on the *entry level concept* are proposed: one based on the entry level of an attribute and another

one on the entry level of a user. This latter is especially interesting since they pose a methodology to go over the lattice in order to find the users in the neighbourhood of the target user.

Association Rules can be also a valuable technique to generate recommendations. In this respect, in [Zhou et al., 2005] association rules are applied for web usage mining to detect navigational patterns. Given a web access sequence, conducted by a user in a session, the identified rules will be used to recommend new contents to be accessed by the user. Association rules have been also proposed to expand user profiles like. Some examples of this latter are presented in [Shaw et al., 2010] and in [Sobhanam and Mariappan, 2013]. In [Ignatov and Kuznetsov, 2009], and expanded in [Ignatov et al., 2012], the authors apply FCA to a Collaborative Filtering approach but instead of recommend items (as in the previous examples), their system recommends terminology for Internet Advertisement: given a company that has used some terms in the past for marketing campaigns; the system recommends new related terms. The recommendation is carried out by using association rules: the higher the confidence of association rule is the more probable that the consequent of the rule is recommended.

[Senatore and Pasi, 2013] presents other example of Collaborative Filtering approach based on FCA. The particularity of this approach is the application FCA on fuzzy data (i.e., instead using binary values, values are continuous in a [0-1] interval). To carry out the recommendation process the authors propose a basic Collaborative Filtering algorithm that recommends the items already seen by the users that share some item with the target user. Recommendations are then ranked according to the fuzzy values.

However, not only Collaborative Filtering based recommendation can be addressed through FCA; it can be also applied to Content-based recommendation. In this context, instead of considering the items to be recommended as attributes, the contents of these items are considered. [Ignatov et al., 2013] applies FCA to a crowdsourcing platform to represent users according to the content (mainly keywords) of the projects in this platform with which the users have already interacted. Their modelling proposal takes into account that the attributes can be multi-valued by using multi-valued *formal concepts* (more concretely *triadic concepts* [Wille, 1995]). Based on this modelling, they propose two different recommendation methodologies:

- Recommend similar users to a target user: The system look for users that have interacted with content similar to those related to the target user. To do this, the

intents of the formal concept in which the target user appears are used to find other users belonging some of these items.

- Recommend antagonist users (i.e., user that have interacted with the same set of contents but whose opinion about them is completely different). The system look for users sharing the same set of attributes (those in the same formal concept) and it calculates the distance between them, according to their interactions with these attributes (i.e., their opinion about the projects of the platform which are identified by these attributes).

In [Li and Murata, 2010] FCA is used to model item profiles and to construct the candidate recommendation set. To create the item profiles, a *formal context* is created taking into account the items and their metadata. This *formal context*, represented in a lattice, is used to infer relations between user and items FCA-based descriptions, reflected in the *formal concepts* of the *formal context*); recommending the items related to these descriptions (i.e., those that belong to the *formal context*). The work in [Maio et al., 2012] proposes a recommendation approach based on FCA for their application in an e-learning environment. More concretely, the authors applies Fuzzy Formal Concept Analysis (FFCA) to model RSS-feeds content. In this scenario, given a learning context of a user, the most similar concepts in the lattice (according to Wu and Palmer similarity) are recommended to the user. Other interesting approaches of FCA-based recommendation in e-learning scenarios are detailed in [Fang and Zheng, 2009] and in [Lau et al., 2008].

FCA and Recommendation have been also applied together in other scenarios. An interesting application is described in [Kashnitsky and Ignatov, 2014]. Herein, FCA is proposed to model a set of classifiers in a Multiple Classifier System according to their predictions and, given a new content to be classified; the lattice structure is used to select the proper classifier. In [Asmus et al., 2014] FCA has been applied in order to develop a system for recommending algorithms for black-box optimization.

Some of the main *affordances* and *challenges* as well as the future lines of the application of FCA to the Recommender Systems field are detailed in [Valverde-Albacete and Peláez-Moreno, 2013]. Although this work focuses on the relation of FCA and Information Retrieval, most of its conclusions are also applicable to the Recommender systems field.

## 2.4 Discussion

---

This section intended to summarize the aforementioned proposals in state of the art.

Section [2.1](#) in relation to Knowledge-based representations, presents two main types of representations: ontological and conceptual. Although ontological representations are preferable when dealing with well-defined domains (e.g., medical domain), their generation is difficult and expensive. Furthermore, their structure is very rigid, which does not allow the fast updating that social or online domains require. In contrast, conceptual representations, although not so formally defined as ontologies, do allow this updating. The organization of the knowledge contained in these representations is also analysed in this section. In general, this organization is useful, and in fact recommendable, when raw (unorganized) representations have been automatically generated, because they might include redundant or incorrect information (e.g., DBpedia ontology). In this scenario, the organization of the data belonging to them is expected to generate a more abstract and informative representation, thus improving its performance when applied to specific tasks. In this regard, Formal Concept Analysis (i.e., the methodology we use in this thesis) offers a series of advantages over some other traditionally applied methodologies.

The study in section [2.2](#) on topic modelling shows that probabilistic methodologies have become the state-of-the-art of the task. New techniques, in particular graphical models, have attracted the interest of researches in the area because of its proven performance in social domains and the easy interpretation of their results. In contrast, supervised techniques, which were widely applied in the early works in the area, have been relegated to very specific domains, mainly because the difficulty in obtaining annotated dataset to train these systems. Finally, Deep Learning, in spite of their impressive results for some tasks (e.g., POS Tagging, Image Classification, Machine Translation), presents some limitations when dealing with texts (the specific scenario of this work): they are mostly supervised methodologies and their application to rich-text representations are highly limited because of the complexity it entails. From this section, it can be extracted that unsupervised methodologies (as the one we proposed) are preferred, especially in fast changing environments like Twitter where no training data is usually available. In addition, the approaches resulting in hierarchical representations are better suited to capture the inherent hierarchical structure of textual content.

Regarding recommender systems, the section [2.3.1](#) presents the different algorithms and methodologies in each one of the three main topologies —Collaborative Filtering, Content-based and hybrid systems— and their evaluation. The main conclusion that can be drawn is that there is no such a thing like a “best recommender system”. Each type presents some advantages and disadvantages. On one hand, Content-based systems are easy to use, their recommendations are easily explainable to the users and they are user-independent (i.e., they exploit solely the ratings of the target user); but, in contrast, they tend to offer overspecialized recommendations. On the other hand, Collaborative Filtering systems offer state-of-the-art performance in some scenarios and they are not restricted to application domains where complex item descriptions are available. Nevertheless, among other problems already referred in section [2.3.1.2](#), they suffer from cold-start problems when new items appear in the dataset and they do not take advantage of item relationships, which may enhance the recommendation process (i.e., *The Godfather I* and *II* are similar items). Content-based are preferred in contexts where rich and informative representations are available, as the one applied in this thesis. The hybridization of both systems appears as a sensible approach in order to cover the disadvantages of each type with the advantages of the other. In this sense, the approach presented in this thesis try to address this issue by linking both, user interactions (the basis of CF models) and item representations (the basis of CB models) in a common representation space.

Section [2.3.1](#) studies different application domains for recommender systems, paying special attention to those in the scope of this work. From this study, it is important to remarks some issues:

- In social recommendation environments, and especially Twitter, it is difficult, if not impossible, to collect explicit ratings, as can be done in other domains like movie recommendation. In consequence, approaches working in these domains should be focused on dealing with implicit ratings (i.e., ratings that have been implicitly collected from the user activity [[Albanese et al., 2011](#)]). Furthermore, given the nature of this implicit scenario, ratings are mostly unary: it is possible to know what user likes but not what dislikes.
- News recommendation presents specific requirements related to the application domain. For instance, although freshness (i.e., recommendation of recently appeared items) and novelty in recommendations is a desirable characteristic in any recommendation methodology, it is much more important in the field of news recommendation. Users are rarely interested in old news reports, even though their content may be interesting for the users.

The use of higher-level features (e.g., semantic, contextual) provides more informative representations; that is, items are better described, thus enhancing the recommendation process. Furthermore, these higher-level features enable the generation of models based on abstract contents. This leads to more compact representations, which reduces the dimensionality and consequently the complexity of the operation of the recommender systems

As regards the evaluation, there are three evaluation paradigms. Although user studies or online evaluations may give a clearer insight into the improvement of the user experience provided by the recommender systems, they are difficult and expensive. For this reason, offline methodologies are the common way to evaluate recommender systems (i.e., this is the methodology applied in this work). Among the large set of evaluation methodologies related to this paradigm, for the context of top-N recommendation (the one addressed in this work) those based on accuracy-based metrics are preferable to error-based evaluations. Furthermore, those metrics taking into account the ranking of the recommended items offers a more accurate view of the overall system performance.

From this analysis of the recommendation field, we have derived the recommendation scenario we propose in Chapter [6](#). In particular, we propose the scenario of a Content-based recommender system in social environment, making use of semantic information. It is proposed as top-N recommendation task to be evaluated by means of accuracy metrics. The recommendation methodology we propose is based on the development of a common representation space, based on the representation proposed in Chapters [4](#) and [5](#).

# 3

## Formal Concept Analysis

---

*This chapter describes the theory related to Formal Concept Analysis.*

### Content

---

3.1	FCA AT A GLANCE .....	71
3.2	STABILITY .....	75
3.3	COMPLEXITY REDUCTION.....	76

---

Formal Concept Analysis (FCA) is a mathematical theory of concept formation [Belohlavek, 2008; Ganter and Wille, 1997; Wille, 1992; Wille, 2009] derived from lattice and ordered set theories. In brief, FCA studies how objects interacts with attributes and, from this interaction, how they can be hierarchically grouped together according to their common attributes.

In what follows, an introductory view to FCA is given in section 3.1. Section 3.2 defines the concept of the stability of a *formal concept*, which is important for our later experimentation. Finally, section 3.3 discusses about some techniques to minimize the impact of the FCA complexity in its implementation.

### 3.1 FCA at a Glance

Formal Context FCA provides a theoretical model to organize information represented in *formal contexts*. A *formal context* is defined as a set structure  $\mathbb{K} := (G, M, I)$ , where  $G$  is a set of (formal) *objects*,  $M$  a set of (formal) *attributes* and  $I$ , a *binary has-a* relationship between  $G$  and  $M$  ( $I \subseteq G \times M$ ), denoted by  $gIm$ , which is read as: the object  $g$  has the attribute  $m$ . An example of *Formal Context* can be seen in the Table 3.1, where there is a set of *objects* ( $G$ ), a set of *attributes* ( $M$ ) and a relation between them ( $I$ ), denoted by the crosses in the Table, each one representing that *object*  $g$  has the *attribute*  $m$ .

	<i>Attr</i> <sub>1</sub>	<i>Attr</i> <sub>2</sub>	<i>Attr</i> <sub>3</sub>	<i>Attr</i> <sub>4</sub>	<i>Attr</i> <sub>5</sub>
<i>Object</i> <sub>1</sub>	×				×
<i>Object</i> <sub>2</sub>	×	×	×		
<i>Object</i> <sub>3</sub>	×	×		×	
<i>Object</i> <sub>4</sub>					×

Table 3.1 – Example of Formal Context

Formal Concept The main construct of the theory is the *formal concept*. From the information in the *formal context*, a set of *formal concepts* can be generated. To define *formal concepts* it is needed the following *derivation operation*:

$$A \mapsto A^I := \{m \in M \mid gIm \text{ for all } g \in A\}$$

$$B \mapsto B^I := \{g \in G \mid gIm \text{ for all } m \in B\}$$

Where  $A$  is a set of objects ( $A \subseteq G$ ),  $B$  a set of attributes ( $B \subseteq M$ ) and  $I$  the *prime operator*. Applying the *prime operator* to  $A$  we obtain the set or those attributes that are present in *all* the objects belonging to  $A$ , denoted by  $A^I$ . Conversely, by applying the *primer operator* to  $B$  we obtain the set of objects that have *at least* the attributes given in  $B$ , denoted by  $B^I$ . This operation satisfies the following properties:

$$Z_1 \subseteq Z_2 \Rightarrow Z_1^I \supseteq Z_2^I$$

$$Z \subseteq Z^{II}$$

$$Z^{III} \subseteq Z^I$$

Thus, a *formal concept* is a pair  $(A, B)$  *iff*  $A \subseteq G$  is a set of objects and  $B \subseteq M$  is a set of attributes describing these objects; being that  $B$  is the *extension* of  $A$  ( $A = B'$ ) and, conversely,  $A$  is the *intension* of  $B$  ( $B = A'$ ) [Ganter et al., 2016]. A formal concept has the following properties:

- If an object  $a \in A$  is tagged with an attribute  $b$ , then  $b$  must be included in  $B$  (i.e.,  $B = A^I$  the intent of the formal concept includes all the attributes shared by the objects in the extent).
- Conversely, if an object  $a$  is tagged with all the attributes in  $B$ , then  $a$  must be included in  $A$  (i.e.,  $A = B^I$ : the extent of the formal concept includes all those objects filtered out by the intent).

To exemplify the generation of *formal concepts*, giving the *formal context* in the Table 3.1, the *formal concepts* in Table 3.2 are generated.

	Extent	Intent
$C_1$	{O1, O2, O3, O4}	$\emptyset$
$C_2$	{O1, O2, O3}	{Attr1}
$C_3$	{O1, O4}	{Attr5}
$C_4$	{O2, O3}	{Attr1, Attr2}
$C_5$	{O1}	{Attr1, Attr5}
$C_6$	{O2}	{Attr1, Attr2, Attr3}
$C_7$	{O4}	{Attr1, Attr2, Attr4}
$C_8$	$\emptyset$	{Attr1, Attr2, Attr4, Attr4, Attr5}

Table 3.2 – Example of the set of generated Formal Concepts

Subconcept-  
superconcept  
Relation

*Formal concepts* can be formally (partially) ordered in a *subconcept-superconcept relation* according to their extents. It is based on the natural way in which humans usually order concepts: a `car` is a subconcept of `vehicle` because every `car` is a `vehicle`. To that end, it is possible to define an order relation  $\leq$  on the formal concepts that orders them from the most generic to the most specific one as follows:

$$(A, B) \leq (C, D) : \Leftrightarrow A \subseteq C$$

where  $(C, D)$  is called a *super-concept* of  $(A, B)$  and, conversely,  $(A, B)$  is a *sub-concept* of  $(C, D)$  (i.e.,  $(A, B)$  is more specific than  $(C, D)$ ).

Concept Lattice

The order that results can be proven to be a *lattice*, which is called the *concept lattice*, denoted as  $\mathfrak{B}(G, M, I)$ , associated to the *formal context*. In *concept lattices*, two important types of *formal concepts*, needed for our later recommendation process, are *object concepts* and *attribute concepts*:

- The *object concept*, denoted as  $\gamma g$ , associated with an object  $g$  is the most specific concept (the smallest concept) including  $g$  in its extent. In order to construct it, it is necessary to include in its *intent* all the attributes of  $g$ , and to include in its *extent*, in addition to  $g$ , all those objects tagged exactly with the same attributes than  $g$  (i.e.,  $\gamma g := (\{g\}^I, \{g\}^I)$ ).
- Conversely, the *attribute concept*, denoted as  $\mu m$ , associated with the attribute  $m$  is the most generic concept including  $m$  in its intent. It can be constructed in a dual way to an object concept: (i) add all the objects tagged by  $m$  to the extent, and (ii) in addition to  $m$ , add all the attributes shared by those objects to the intent (i.e.,  $\mu m := (\{m\}^I, \{m\}^I)$ ).

The computation of all the formal concepts as well as the creation of the Concept Lattice by means of the *subconcept-superconcept-relation* in this thesis is conducted by following the *Next Neighbours* Algorithm, detailed in the [Figure 3.1](#). For more information about this algorithm, its complexity issues and some other related algorithms, please refer to [\[Carpineto and Romano, 2004\]](#).

Since *concept lattices* are ordered sets, they can be naturally displayed in terms of *Hasse diagrams* [\[Ganter and Wille, 1997\]](#). In a *Hasse diagram*:

- There is exactly one node for each *formal concept*.
- If  $C \subseteq C'$ , then  $C'$  is placed above  $C$  ( $C$  is a *sub-concept* of  $C'$  or  $C'$  is a *super-concept* of  $C$ ).

- If  $C \subseteq C'$  but there is no other intermediary concept  $C''$  such as  $C \subseteq C'' \subseteq C'$ , there is a line joining  $C$  and  $C'$ .

---

**Algorithm 2** Next Neighbours

---

**Input:** Context  $(G, M, I)$   
**Output:** The Concept Lattice  $= (C, E)$  of  $(G, M, I)$

```

1:  $C := (G, G')$ 
2:  $E := \emptyset$ 
3:  $currentLevel := (G, G')$ 
4: while  $currentLevel := \emptyset$  do
5:    $nextLevel := \emptyset$ 
6:   for each  $(X, Y) \in currentLevel$  do
7:      $lowerNeighbours := FindLowerNeighbours((X, Y))$ 
8:     for each  $(X_1, Y_1) \in LowerNeighbours$  do
9:       if  $(X_1, Y_1) \notin C$  then
10:         $C := C \cup \{(X_1, Y_1)\}$ 
11:         $nextLevel := nextLevel \cup \{(X_1, Y_1)\}$ 
12:       end if
13:       Add edge  $(X_1, Y_1) \rightarrow (X, Y)$  to  $E$ 
14:     end for
15:   end for
16:    $currentLevel := nextLevel$ 
17: end while
18:
19: function  $FindLowerNeighbours((X, Y))$ 
20: /* Returns the lower neighbours of a concept */
21:  $candidates := \emptyset$ 
22: for each  $m \in M \setminus Y$  do
23:    $X_1 := (Y \cup \{m\})'$ 
24:    $Y_1 := X_1'$ 
25:   if  $(X_1, Y_1) \notin candidates$  then
26:     $candidates := candidates \cup \{(X_1, Y_1)\}$ 
27:   end if
28: end for
29: return maximally general  $candidates$ 

```

---

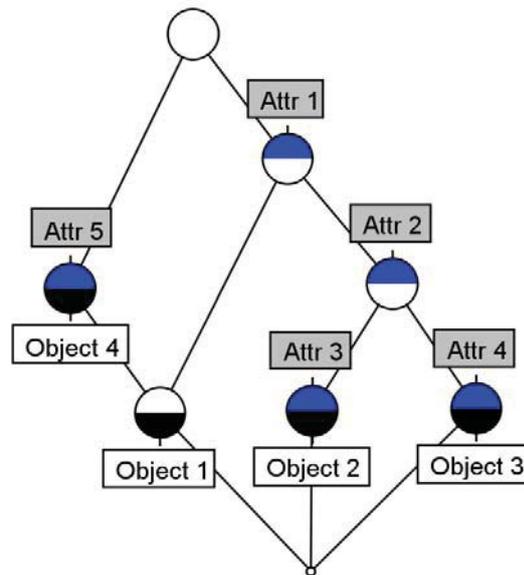
**Figure 3.1** – Next Neighbours Algorithm

In [Figure 3.2](#) it can be viewed an example of the *concept lattice representation* corresponding to the *Formal Context* in the [Table 3.1](#). In this figure, white labels refer to the entries (objects) and grey labels to the features (attributes). The node to which an object label is attached represents its *object concept* ( $\gamma g$ ) and it is denoted by a black semicircle; conversely, the node to which an attribute label is attached represents its *attribute concept* ( $\mu m$ ) and it is denoted by a blue semicircle. To avoid an overloaded representation, each *formal concept* is depicted with a minimal set of object and attribute labels. From this diagram, each *formal concept* can be easily reconstructed as follows:

- The *extent* includes all the objects depicted in the nodes on the paths leading from the target's *formal concept* to the bottom concept in the diagram. For example, the *extent* of the *formal concept* associated with the node marked as "Attr5" in [Figure 3.2](#) is {Object 1 and Object 4}.

- The *intent* includes all the attributes depicted in the nodes on the paths leading from the target's *formal concept* to the top node in the diagram. For example, in [Figure 3.2](#) the *intent* of the concept labelled with the film "Attr 1" is {Attr 1, Attr 2, Attr 4 and Attr 3}.

For more detail on how to read *concept lattices* displayed as Hasse Diagrams, please refer to the recently published book of [\[Ganter et al., 2016\]](#), in particular to section 1.3.



**Figure 3.2** – Example of Concept Lattice Representation

One of the problems associated with FCA is the handling of very large databases, mainly for two main reasons: the computational cost in order to calculate the concept lattice and the complexity of the generated concept lattice [\[Codocedo et al., 2011\]](#). Related to the latter, stability (in section [3.2](#)) can be helpful to improve the readability of the concepts by removing concepts from data. Related to the former problem, the solution focuses on the reduction of the complexity of the *formal context* as a previous step to the application of FCA to compute the *concept lattice*.

## 3.2 Stability

Stability is a technique to reduce the number of *formal concepts* in a given *concept lattice*; selecting “the most interesting” groups [\[Kuznetsov, 2007\]](#). Other ideas have been

also proposed for the readability improvement, like for example the building of the *iceberg lattices* [Jay et al., 2008; Kuznetsov et al., 2007]. An *iceberg lattice* includes only the upper part of the lattice; i.e., those concepts with extents comprising at least  $n\%$  of all objects. Nevertheless, the *iceberg lattices* could overlook small but interesting groups (i.e. exotic or emergent groups not present in a large number of objects). To address this particular problem, stability takes all the concepts in the lattice into account and not only the “top concepts” (as iceberg lattices technique does). Stability focuses on measuring the dependence between the *extent* of a concept and its *intent*. The formal definition of stability that we apply is the one proposed in [Roth et al., 2008]:

**Definition 1.** Let  $\mathbb{K} := (G, M, I)$  be a *formal context* and  $(A, B)$  be a *formal concept* belonging to  $\mathbb{K}$ . The stability index,  $\sigma$ , of  $(A, B)$  is defined as follows:

$$\sigma_i(A, B) = \frac{|\{C \subseteq A \mid C' = B\}|}{2^{|A|}}$$

where  $|A|$  is the number of objects in  $A$ , and  $C$  is each subset of  $A$  whose concept’s intent ( $C'$ ) is equal to the concept intent of  $A$ , that is,  $C' = B$ . Stability so-defined indicates “how much” the *intent* of a given concept depends on particular objects of the concept *extent* (*intensional stability*). The least dependant the intent of a concept, the more stable. An “instable” concept represents a concept with noisy data (i.e., if we remove some object/s of a concept, the concept will be seriously affected; given that the concept’s objects are not representative objects: there are noisy objects) [Cooper et al., 2010].

Although computing stability can seem an easy task, in fact this is a #P-complete problem [Kuznetsov, 2007]. Thus, some heuristic are necessary to easily compute the stability values of all the concepts in a lattice. In this sense, in [Kuznetsov et al., 2007] some properties are presented in order to facilitate the calculation of stability values, while in [Roth et al., 2008] an algorithm to calculate stability of the *concepts* of a *formal context* is described. It is this latter work (that of [Roth et al., 2008]) which we have applied in order to compute the stability in our experimentations.

### 3.3 Complexity Reduction

---

Although stability or iceberg lattices are useful to improve readability, detect interesting groups and even to prune lattices; these techniques are a process posterior to the lattice generation; so the complexity problem associated with this generation still remains. In this regard, the reduction of dimensionality of the *formal context* has been proposed as a

previous step to the application of FCA. The idea is that, if we reduce the amount of data (with the minimum loss of representativeness) the FCA algorithms will be easy to compute. Some of the most noteworthy works in this line are based on finding latent relations in the data in order to summarizing them according to these relations. [Codocedo et al., 2011], and earlier in [Gajdos et al., 2004] apply Latent Semantic Analysis (LSA) to compress a *formal context* into a smaller one (*reduced context*), by using as objects of the lattice the implicit dimensions identified by the application of LSA. [Snásel et al., 2007] apply Non-Negative Matrix Factorization (NMF). [Aswanikumar and Srinivas, 2010] propose a K-means clustering to reduce the data dimensionality. Finally, [Cheung and Vogel, 2005] present a data reduction based on looking for equivalence relations between objects.

Besides of data dimensionality reduction, the complexity problem can be addressed by taking into account background knowledge about the data, avoiding the need to discard some data (with its subsequent loss of information). For example, the authors of [Belohlavek and Vychodil, 2009] apply the background knowledge of a set of users to identify their most important priorities (*concept intents*) and to use only these priorities for the posterior FCA modelling.

Different to the former approaches, the authors of [Ignatov et al., 2013] propose a method to complexity reduction not based on the reduction/manipulation of the data but on modify the generation of *formal concepts*. More concretely, they propose an alternative approach based on the relaxation of the definition of *formal concepts* by using object-attribute (OA) biclusters. Basically, the difference between OA-biclusters and *formal concepts* is that in OA-biclusters not all the cells, which represents the relations that form a *formal concept*, must be filled (i.e., it is possible that some *objects* in the *formal concept* are not related with some of the *attributes* of the *formal concept*). By means of OA-biclusters is possible to generate broader concepts by sacrificing some density in these concepts (i.e., the amount of relations between objects and attributes that does not exist in the concepts). The minimum density required to form a *formal concept* is set by a  $\rho_{min}$  value, being a non-negative real number such that  $0 \leq \rho_{min} \leq 1$ ; that is,  $(A, B)$  will be a formal concept if their density value,  $\rho(A, B)$  is greater than the minimum density:  $\rho(A, B) \geq \rho_{min}$ .

# Part II

## MODELLING PROPOSAL

*“Unstructured content is stupid and old-fashioned. It’s costly, complex and does not generate a competitive advantage”*

Ann Mulhay, ex-CEO of Xerox



# 4

## Concept-based Modelling

*This chapter presents the experimentation conducted in regards to text modelling by applying Formal Concept Analysis*

### Content

---

CONCEPT-BASED MODELLING .....	80
4.1 APPLICATION SCENARIO 1: TOPIC DETECTION .....	82
4.1.1 The RepLab 2013 Campaign .....	83
4.1.1.1 Use Case .....	83
4.1.1.2 The RepLab 2013 Dataset.....	84
4.1.2 Evaluation Strategy .....	86
4.1.2.1 Internal vs. External Evaluation.....	86
4.1.2.2 Internal Evaluation Setup.....	87
4.1.2.3 External Evaluation Setup: Replab Evaluation Framework.....	92
4.1.3 Topic Annotation by means of FCA.....	93
4.1.3.1 Concept Lattices as Topic Representations.....	93
4.1.3.2 How to deal with big concept lattices?.....	96
4.1.4 Other Topic Annotation Approaches.....	101
4.1.4.1 HAC Clustering .....	101
4.1.4.2 Latent Dirichlet Allocation .....	104
4.1.4.3 Comparison of HAC and LDA vs. FCA.....	105
4.1.5 Experimentation .....	106
4.1.5.1 Content Pre-processing .....	107
4.1.5.2 FCA-based experimentation.....	110
4.1.5.3 Comparison to other Topic Annotation approaches.....	122
4.1.6 Discussion .....	134
4.2 APPLICATION SCENARIO 2: IMAGE DIVERSIFICATION .....	138
4.2.1 Test bed: Retrieving Diverse Social Images Task at MediaEval .....	139
4.2.1.1 Dataset.....	140
4.2.1.2 Evaluation Setup.....	140
4.2.2 FCA-based Work Proposal .....	141
4.2.2.1 Image Content Representation.....	141
4.2.2.2 FCA-based Modelling.....	143
4.2.2.3 HAC-based grouping.....	144
4.2.2.4 Result List Creation.....	146
4.2.3 Experimentation .....	147
4.2.4 Discussion .....	154

---

One important part of a recommender system is the representation of the content that the users have consumed, as well as the future items which are susceptible to be consumed. The intuition behind is that the better the items are represented, the better will be the recommendations inferred from this representation. Although these items come from different sources and may include several types of information, they are mainly based on textual representations (e.g., movie reviews, news reports, tweets, books, scientific papers...). Therefore, the scope of this thesis focuses on dealing with such textual representations.

The representation of textual content was one of the first lines in which researchers and practitioners in the field of Natural Language Processing and Information Retrieval focused their attention [Salton, 1971; Salton and McGill, 1983]. As stated in section 2, it has been pointed out many times in the literature that “raw” textual representations present some problems related to the ambiguity and lack of structure of text. Consequently, when working with textual representations some kind of analysis should be conducted to extract meaningful organization and structures of patterns from the raw text [Hotho et al., 2002; Kuznetsov et al., 2007]. In that context, the modelling of textual content was proposed trying to mitigate these problems by provided a more structured representation. Resources like taxonomies, or ontologies, or more recently probabilistic-based models or distributed representations have been implemented in this regard.

To cope with these problems, we propose a text modelling approach based on a concept-based representation generated through the application of Formal Concept Analysis. We intend to create a more abstract representation based on formal concepts, automatically inferred from the text. This more abstract representation is expected to better represent items described by textual content.

To prove this hypothesis, we present two application scenarios for the FCA-based representation. The rationale of this experimentation is to evaluate our representation proposal, which will be later applied for recommendation, independently of the recommendation task. By abstracting the evaluation process and only focusing on the representation step, we are going to have a measure of its actual performance in modelling and representing content instead of its accuracy when applied for recommendation.

The first of these two scenarios, presented at section 4.1, focuses on the detection of topic on a Twitter stream. The detection of topics is highly correlated to the data representation. An accurate representation of the tweet content entails a significant improvement in the topic detection process, inasmuch as the proposed topic detection

systems mainly rely on the clustering of the data representations to infer thematic-based topics. Furthermore, the task is related to the recommendation scenario proposed in this thesis. The detection of topics relies on the identification of thematically similar items and the recommendation methodology that we propose in subsequent sections is based on finding thematically related user-item groups. The scenario is also the same: data coming from social networks, Twitter in particular.

The second application scenario in section [4.2](#) is the diversification of images. Although this scenario might seem to be less related to our proposal, it is in fact similar. The image diversification is focused on the identification of similarities between image text-based descriptions. This scenario is also interesting because the textual content is coming from social data; i.e., social annotations made by Flickr users. These data are similar to the noisy, blurry and sparse data that can be found in the item representations in social-based recommendation scenarios, as those propose in this thesis.

To sum up, the two proposed application scenarios share a common nexus: the automatic (i.e., unsupervised) data representation of social contents. Throughout these scenarios, our representation proposal is evaluated in the detection of groups of similar content among social-based data. This is also what the recommendation algorithm based on this representation is expected to do. This evaluation proves that the FCA-based proposal developed in this thesis is able to accurately represent such content, improving the state of the art of the tasks in which it is applied. This improvement is expected to lead on a more accurate system when applied for recommendation in next sections.

## 4.1 Application Scenario 1: Topic Detection

---

Topic Detection refers to the finding of topics in data streams on a company, product, person or service. These topics will be useful, for instance, for identifying trending opinion streams, divide contents or users of interest groups, or warn of some risk to the entity's reputation, based on the appearance of a controversial topic.

Traditionally, classification and clustering techniques have been applied for Topic Detection. More recently, probabilistic techniques have been also attracted the interest of the research community. However, all of these techniques have drawbacks such as the need to fix the number of topics to be detected, or the problem of how to combine the

previous knowledge on topics with the detection of new topics in an adaptable way (see section 2.2 for more details). As a suitable solution to these drawbacks, we propose a novel methodology in the field of the Topic Detection: Formal Concept Analysis (FCA).

Applied to the Topic Detection problem, FCA allows the organization of objects into thematically similar *formal concepts*, in accordance with their shared attributes. In addition, these formal concepts are partially ordered by means of a generalization specialization relationship, capturing in this way the inherent hierarchy in the topics (i.e., a topic `sports` is more general than other focused on `soccer`). FCA is also able to deal with some other related problems: the need to know a priori the number of topics to be detected and the adaptability to new content (see section 2.2 for more details).

To delve into this matter, this proposal is evaluated in the context of a real-life topic detection task, the Twitter corpus from the RepLab 2013 campaign. It is a corpus of 143.000 tweets about 61 entities in different domains, manually annotated by experts. Evaluation metrics and scripts are provided by the organization, enabling the comparison of our results and the ones in the Topic Detection literature. In order to demonstrate the efficiency of the proposal, several experiments have been performed focused on testing: a) the impact of terminology selection as an input to the FCA based algorithm, b) the impact of concept selection as the outcome of our algorithm, and; c) the efficiency of the proposal to detect new and previously unseen topics (i.e., topic adaptation).

---

#### 4.1.1. The RepLab 2013 Campaign

---

This section introduces the scenario applied for the experimentation of our proposal: The Topic Detection Task @ Replab 2013. The RepLab 2013 Evaluation Campaign is one of the main international forums for experimentation and evaluation in the field of Online Reputation Management (ORM). One of its task is related to the Topic Detection problem in social networks. More specifically, the Topic Detection task focuses on the detection of topics related to a set of entities in a large Twitter collection. The outcome of this task could be used by entities (i.e. companies) to detect emerging topics that can affect their reputation, allowing an agile and appropriate response to them (e.g. avoid losing reputation by means of negative comments in social networks).

##### 4.1.1.1. Use Case

Suppose an expert in Online Reputation Management who is in charge of monitoring Twitter comments about a given entity (BMW). The expert is expected to review all the

tweets about the entity, previously filtered out, understand them and organize them in thematically related topics. The expected output of their work would be a report detailing this information. This report is useful in order to know the main topics being commented on by the people about the entity (e.g., thoughts about new car models, general opinions about the company, problems arising in real-time). Moreover, this information, which may be seen as an implicit user feedback, might be useful for the early detection of reputational alerts (e.g., customers that have bought a new BMW model experiencing some problems). To do so, the expert would need to review many tweets, which is a tedious process that is not always possible as the number of tweets about the entity grows. This context justifies the existence of automatic systems to carry out this topic detection.

In order to cover the expert needs and offer valuable input for later tasks, like, for instance, alert detection, Topics to be generated are expected to be:

- 1) Thematically similar, the idea is to find the different subjects that are being commented on by the users.
- 2) Cohesive in terms of intra- and inter-similarity, if two topics address a quite similar thematic they should be merged and if one topic includes subjects which are not closely related, it should be divided.
- 3) Both topic precision and coverage are important, the most number of topics addressed in the tweets should be detected with the best possible precision. Systems with low coverage might overlook an important topic and systems with low precision will lead to many erroneously classified tweets, which would be rather useless for the expert.

In addition, the kind of system proposed to fulfil the aforementioned task requirements should carry out the topic detection process in an automatic and unsupervised way, because no previous information about the topics to be detected is expected to be available. As our efforts focused on analysing the topic detection task itself, it has been considered that the unrelated tweets (i.e., tweet that are not actually related to the entity) have been filtered out as a previous step to the topic detection process.

#### **4.1.1.2. The RepLab 2013 Dataset**

This is a collection of tweets related to 61 entities crawled from the 1st June 2012 to the 31st Dec 2012 using the entity's canonical name as query (e.g. BMW). As well as the tweets, there are also other data on the entities: Wikipedia pages of the entities, homepages of the entities and contents of the Webs mentioned in the tweets (i.e. web pages corresponding to the shortened URLs).

Selected entities (61 in total) belong to four domains: automotive, banking, universities and music/artists. For each entity, at least 2,200 tweets were collected: the first 700 were used as training set, and the rest as a test set. The corpus also comprises additional background tweets for each entity (up to 50,000, with a large variability across entities). The corpus is in English and Spanish; however, the language is highly dependent on the entity (i.e. most of the tweets of the entity are usually in the same language, either Spanish or English). The balance between both languages depends on the availability of data for each of the entities included in the dataset. Some numbers on the dataset and the different domains are set out in the [Table 4.1](#), taken from the Overview of the Task [\[Amigó et al., 2013a\]](#)

Annotators have manually labelled the dataset. These annotators have been trained and guided by experts in ORM. Each tweet in the training and test sets are annotated with the identifier of the topic (cluster) to which the tweet belongs. Participants can only access the ground truth (i.e. topic identifier) of the training test.

	All	Automotive	Banking	University	Music/Artist
# Entities	61	20	11	10	20
# Training Tweets	45,679	15,123	7,774	6,960	15,822
# Test Tweets	96,848	31,785	16,621	14,944	33,498
# Total Tweets	142,527	46,908	24,395	21,904	49,320
# Tweets EN	113,544	38,614	16,305	20,342	38,283
# Tweets ES	28,893	8,294	8,090	1,562	11,037

**Table 4.1** – Analysis of the number of tweets by domain

Although a first approach to this task could suggest applying an automatic classification strategy, it is not possible as topics in the test dataset can be different to those detected and annotated in the training set. In other words, it could be possible to train a classifier using the training set but it would only accurately classify those tweets related to the trained topics. Hence, an unsupervised approach seems to be more appropriate for this task. We will apply Formal Concept Analysis from this perspective, but also taking into account all the previous knowledge gathered from the training set. In this way, our proposal will not only cluster new tweets related with older topics but will also discover new emerging topics not present in the training set. [Table 4.2](#) shows some statistics on the annotations by domain and collection.

	All	Automotive	Banking	University	Music
<b>Training Set</b>					
<b>Topic Number</b>	3,813	1,389	831	503	1,090
<b>Avg. Topics</b>	62.51	69.45	75.91	50.3	54.5
<b>Avg. Tweets per Topic</b>	14.40	12.36	11.35	17.57	16.53
<b>Test Set</b>					
<b>Topic Number</b>	5,757	1,959	1,121	1,035	1,642
<b>Avg. Topics</b>	94.38	97.95	101.91	103.5	82.1
<b>Avg. Tweets per Topic</b>	21.14	18.42	18.95	21.78	24.74
<b>All</b>					
<b>Topic Number</b>	9,570	3,348	1,952	1,538	2,732
<b>Avg. Topics</b>	156.89	167.4	177.45	153.8	136.6
<b>Avg. Tweets per Topic</b>	17.77	15.39	15.15	19.67	20.64

Table 4.2 – Statistics on Topic Detection Task

### 4.1.2. Evaluation Strategy

The evaluation proposed in the Replab Campaign is carried out by comparing the system results to the data contained in the gold standard of the task. Moreover, evaluation can be also carried out by considering only the internal data to analyse the quality of the generated clusters. In what follows both kinds of evaluation are explained in detail.

#### 4.1.2.1. Internal vs. External Evaluation

In clustering evaluation, the adjectives internal and external refer to the source of the data that are used to evaluate the results. The internal evaluation takes into account only the information obtained in the clustering results (cluster documents, distances between clusters, cluster in-between similarity...) to evaluate the goodness of a clustering structure [Liu et al., 2010b]. In contrast, the external evaluation takes data from an external source — a gold standard — to carry out the evaluation. However, the source of data is not the only difference; the objective of the evaluation is also different. The internal evaluation tries to measure the *quality* of the generated clusters, whatever *quality* means, while the external evaluation measures how the systems are able to offer results that may be suitable for a specific task, by comparing them to a given gold standard.

The RepLab experimental setup follows this latter methodology: the system results are evaluated by comparing them to the gold standard provided by the annotators. Although this kind of evaluation is broadly applied and their results may give a good insight into

the system performance, it might bias the evaluation result, thus favouring some types of approaches. If there is any bias in the gold standard, systems offering results with the same bias will be promoted by the evaluation. For instance, the annotation could have been made in a generic way (i.e. a few generic clusters with many tweets belonging to each one). Consequently, systems offering a very specific topic classification would obtain poor results, even though their topic representation may be accurate. To sum up, this kind of evaluation basically measures to what degree the results are similar to the gold-standard.

In the RepLab Evaluation Campaign, this problem has been minimized thanks to:

- 1) The gold standard being generated by human annotators, trained by experts, who carefully reviewed the annotation process, avoiding errors attributable to inexperience.
- 2) Each annotator worked with a different entity in the dataset; so, since the evaluation is based on the overall performance of the system throughout all the entities, the bias related to the annotator's criteria is minimized.
- 3) Each annotation has been created by aggregating the opinion of different annotators.
- 4) The aim of a topic detection system is to assist experts in the field. Therefore, nobody is better than the experts to set the guidelines on what is a good topic representation and what is not.

The final comment is that, due to the task definition, a gold standard for topic detection depends on the annotation criteria and the annotator's possible bias. Consequently, it cannot be taken as a general truth, as it could happen for instance in the Web People Search task, where you can check if a web page actually belongs to a given person. Thereupon, even though we can assume that a high-performing system in Replab is a high-performing system for the Topic Detection task, the internal evaluation framework is necessary in order to reflect the quality of the generated topic representations instead of just its adaptation to the gold standard.

#### 4.1.2.2. Internal Evaluation Setup

The internal evaluation proposed in this section is a comparison of the *quality* of the topic representation computed by three different approaches. To that end, several cluster quality metrics were applied to measure the cohesion of the generated clusters. Cohesion is a desirable characteristic that a good topic representation should have in the topic detection task (see section [4.1.1.1](#)). This kind of internal metrics has been already proven to be effective in discovering the inherent clustering structure of a dataset ([Rendón et al.](#),

[2011]. Consequently, a high-performing approach in terms of these measures is expected to identify this clustering structure, which in our task will mean the identification of the latent topic structure. To carry out this evaluation, we implemented four of the most noteworthy measures in the state of the art, which are explained below. To better understand the following formulations, notation referring to the clusters is detailed in Table 4.3.

$K$	Number of Clusters
$C_k$	Cluster $k$
$C_i^k$	Point $i$ within the Cluster $k$
$G^{\{k\}}$	Centroid of the Cluster $C_k$
$G$	Centroid of the Dataset
$n_k$	Cardinal of the Cluster $C_k$
$I_k$	Set of indices of the observations belonging to the Cluster $C_k$

Table 4.3 – Clustering notation

### Davies-Bouldin Index [Davies and Bouldin, 1979]

The Davies-Bouldin Index measures which algorithm generates the clusters with the lowest intra-cluster distance and the highest inter-cluster distances.

**Definition 2.** The Davies-Bouldin Index is computed by the formula in (1). The algorithm that produces the smallest value of this index is the best one.

$$DB_{index} = \frac{1}{K} \sum_{k=1}^K \max_{k' \neq k} \left( \frac{d_k + d_{k'}}{\Delta_{kk'}} \right) \quad (1)$$

$d_k$  being the mean distance of the points belonging to cluster  $C_k$  to the centroid  $G^{\{k\}}$ :

$$d_k = \frac{1}{n_k} \sum_{i \in I_k} \|C_i^{\{k\}} - G^{\{k\}}\| \quad (2)$$

and  $\Delta_{kk'}$  the distance between the centroids of clusters  $C_k$  and  $C_{k'}$ :

$$\Delta_{kk'} = d(G^{\{k\}}, G^{\{k'\}}) = \|G^{\{k'\}} - G^{\{k\}}\| \quad (3)$$

### Dunn Index [Dunn, 1974]

Dunn Index is similar to the Davies-Bouldin Index. This index favours the dense and well-separated clusters. The larger the value the better the clustering.

**Definition 3.** The calculation of the Dunn Index is carried out by means of the formulation in (4):

$$Dunn_{index} = \frac{d_{min}}{d_{max}} \quad (4)$$

where  $d_{max}$  is the largest of the  $D_k$  distances of each cluster:

$$d_{max} = \max_{1 \leq k \leq K} D_k \quad (5)$$

where  $D_k$  is the largest distance separating two points inside a cluster (also called the diameter of a cluster):

$$D_k = \max_{\substack{i, j \in I_k \\ i \neq j}} \|C_i^k - C_j^k\| \quad (6)$$

on the other hand,  $d_{min}$  is the smallest of the  $d_{kk'}$  distances:

$$d_{min} = \min_{k \neq k'} d_{kk'}$$

where  $d_{kk'}$  is the distance between clusters  $C_k$  and  $C_{k'}$ , measured by the distance between their closest points:

$$d_{kk'} = d(C^{\{k\}}, C^{\{k'\}}) = \min_{\substack{i \in I_k \\ j \in I_{k'}}} \|C_i^k - C_j^{k'}\| \quad (7)$$

### Silhouette Coefficient [Rousseeuw, 1987]

This coefficient is applied over each object in the collection. It compares the distance of an object to the elements in the same cluster and to the distance to elements in other clusters. The higher is the coefficient the better the object has been clustered. Its computation is conducted as follows:

**Definition 4.** To define the Silhouette Coefficient let us to consider each point  $i$  in a cluster  $k$  ( $C_i^k$ ) and  $a(i)$ , the within-cluster mean distance of the point  $C_i^k$  to the other points in the cluster:

$$a(i) = \frac{1}{n_k - 1} \sum_{i' \in I_k, i' \neq i} d(C_i^k, C_{i'}^k) \quad (8)$$

and also consider the mean distance of the point  $C_i^k$  to the rest of the points in the another cluster  $C_{k'}$ :

$$\mathfrak{d}(C_i^{\{k\}}, C_i^{\{k'\}}) = \frac{1}{n_{k'}} \sum_{i' \in I_{k'}} d(C_i^{\{k\}}, C_{i'}^{\{k'\}}) \quad (9)$$

and  $b(i)$  as the smallest of these mean distances:

$$b(i) = \min_{k \neq k'} \mathfrak{d}(C_i^k, C_{k'}) \quad (10)$$

then, the silhouette width of a point  $C_i^{\{k\}}$  is computed by applying the following formulation:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (11)$$

The value obtained is between -1 and 1. A value close to 1 indicates that the point belongs to the right cluster, while a value close to -1 indicates that the point has been wrongly clustered. The mean of the widths of each of the points inside a cluster denotes the silhouette of the cluster:

$$\mathfrak{s}_k = \frac{1}{n_k} \sum_{i \in I_k} s(i) \quad (12)$$

Finally, the overall silhouette coefficient is the mean of the mean cluster silhouettes:

$$Silhouette_{coefficient} = \frac{1}{K} \sum_{k=1}^K \mathfrak{s}_k \quad (13)$$

### The Calinski-Harabasz Index [\[Calinski and Harabasz, 1974\]](#)

The index is based in the ratio of between-cluster variance and within-cluster variance. The larger is the between-cluster variance and the smaller is the within-cluster variance, the better is the clustering quality. Consequently, a good clustering representation will maximize the value of this index.

**Definition 5.** The formulation for the computation of the Calinski-Harabasz Index is:

$$CH_{index} = \frac{N - K}{K - 1} * \frac{SS_B}{SS_W} \quad (14)$$

where  $SS_B$  is the overall between-cluster variance and  $SS_W$  the overall within-cluster variance,  $K$  the number of clusters and  $N$  the number of observations. The overall between-cluster variance,  $SS_B$ , is defined as:

$$SS_B = \sum_{k=1}^K n_k \|G^{\{k\}} - G\|^2 \quad (15)$$

where  $G^{\{k\}}$  is the centroid of the cluster  $k$  and  $G$  is the centroid of the whole dataset. On the other hand, the overall within-cluster variance,  $SS_W$  is defined as:

$$SS_W = \sum_{k=0}^K \sum_{i \in I_k} \|C_i^{\{k\}} - G^{\{k\}}\|^2 \quad (16)$$

being  $C_i^k$  the point  $i$  in the cluster  $k$ .

### Comparison of the Internal Metrics

All the above metrics are essentially a ratio of two criteria: compactness or intra-cluster similarity (points in the same cluster should be similar) and separation or inter-cluster similarity (the points in the different clusters should be dissimilar) [Liu et al., 2010b]. They only differ in how these two criteria are defined and combined [Van Craenendonck and Blockeel, 2015].

The Silhouette Coefficient defines the cluster compactness based on the pairwise distances between all points in the cluster, and separation based on pairwise distances between all points in the cluster and all points in the closest other cluster. It might represent a problem when dealing with datasets with subclusters (i.e., clusters that are close together): the cluster separation will achieve its maximum value when close subclusters are considered as one big cluster. In addition, unwanted behaviour takes place in presence of unbalanced clusters where this metric tends to score well (e.g., isolate one point as a cluster and grouping all other points together in a single cluster).

The Davies-Bouldin index defines compactness based on the distance of points in the cluster to its centroid, and separation based on distances between centroids. In other words, this index relies on the ratio of the within-cluster scatter, to the between-cluster separation. As separation is based on centroid distances, this index is able to handle subclusters better than, for instance, the Silhouette Coefficient [Liu et al., 2010b].

The Calinski-Harabasz Index defines the cluster validity based on the average between- and within-cluster sum of squares. This index is suitable when cluster shapes are more or less spherical and compact in their middle; therefore, it tends to prefer cluster solutions with clusters consisting of roughly the same number of objects.

Finally, the Dunn Index uses the minimum pairwise distance between objects in different clusters to measure separation and the maximum diameter among all clusters to measure compactness. When noise is introduced, the inter-cluster separation can decrease sharply since it only uses the minimum pairwise distance, rather than the average pairwise distance.

#### 4.1.2.3. External Evaluation Setup: RepLab Evaluation Framework

The RepLab evaluation framework is focused on comparing the system results to a manually generated gold standard. For this evaluation, two new measures were proposed by the organizers: Reliability and Sensitivity (similar to B-Cubed Precision and Recall respectively [Bagga and Baldwin, 1998]). These measures are presented and detailed in [Amigó et al., 2013b]. Briefly explained, Reliability ( $R$ ) is defined as the precision of binary relations predicted by the system with respect to those that derive from the gold standard and Sensitivity ( $S$ ) is similarly defined as the recall of these relationships. In more detail, Reliability and Sensitivity are defined as:

$$R(Entropy) = P_{j \in \mathcal{J}} (rel_{gold}(i, j) = rel_{sys}(i, j) \mid rel_{sys}(i, j)) \quad (17)$$

$$S(Entropy) = P_{j \in \mathcal{J}} (rel_{gold}(i, j) = rel_{sys}(i, j) \mid rel_{gold}(i, j)) \quad (18)$$

where  $\mathcal{J}$  is the set of tweets considered in the evaluation,  $rel_{gold}(i, j)$  shows that a tweet  $i$  and a tweet  $j$  belong to the same cluster in the gold standard and  $rel_{sys}(i, j)$  is analogous but applied to the system output. The final  $R$  and  $S$  measures are the average of the individual  $R(Entropy)$  and  $S(Entropy)$  for the 61 entities in the dataset.  $R$  and  $S$  are combined with the Micro-average F-measure: the final F-measure values (i.e., those shown in the results) indicate the average of the different F-measures  $F(Entropy)$  of the individual  $R(Entropy)$  and  $S(Entropy)$  values for each of the 61 entities in the dataset:

$$F = Avg(Entropy) = \frac{\sum_{Entropy=1}^{|Entities|} F(Entropy)}{|Entities|} = \frac{\sum_{Entropy=1}^{61} \frac{R(Entropy) * S(Entropy)}{R(Entropy) + S(Entropy)}}{61} \quad (19)$$

F-measure defined in this way penalizes the combined score if there is a low score according to any of both measures ( $R$  and  $S$ ). A good precision and coverage is one of the desired criteria for the generated topics (see section 4.1.1.1).

---

### 4.1.3. Topic Annotation by means of FCA

---

FCA can be seen as a powerful tool to structure and classify a set of resources represented through a set of attributes automatically (i.e. terms, metadata, semantics, etc.). Using simple words, FCA unfolds an *is-a* or *has-a* relationship into a complex set of formal concepts, related by a *generalization-specialization* relationship which facilitates the discovery of hidden relationships. Applied to the topic detection task, FCA deals with the main problems usually related to this task. For instance, FCA does not need to know a priori the number of topics and it behaves well in the adaptation to new topics or the selection of new features. More details on its application for the Topic Detection task are included in the following subsections.

#### 4.1.3.1. Concept Lattices as Topic Representations

FCA can be applied to our approach in a straightforward way by taking tweets as objects, and their terms as attributes. Therefore, all the tweets belonging to the same formal concept will share the same terminology that, in turn, means that these tweets are related to the same topic. In other words, we will consider each *formal concept* as a topic. By means of the order relationship, it is possible to define that a given formal concept is greater than (more generic), smaller than (more specific) or not comparable to another one.

With these assumptions, the resulting concept lattices will be shaped as follows. The upper part of the lattice will be made up of those concepts representing general topics of the set of tweets, whereas lower concepts will represent topics that are more specific. Due to the generalization-specialization structure of a concept lattice, it is possible that the same tweet will be categorized in more than one topic. This situation facilitates the discovery of new topics in the final concept lattice that are built as a combination of the upper concepts. The concept lattice also enables the identification of disjoint partitions in the data set, which are defined by a generic formal concept that includes the specific formal concepts related to them (i.e. those that are a specification of the generic one and have no relationship with data in other partitions) [Geng et al., 2008]. This latter property may be useful in detecting generic topics that split the data set at a given specificity level (e.g., sports, politics and so on), as well as the specified formal concepts (topics) that are related to these generic ones (e.g., soccer, basketball, etc. related to sports).

Table 4.4 illustrates this setup. The formal context is built using a set of tweets (i.e. 7 tweets) as objects and a set of terms (i.e. 8 terms) as attributes. The table shows the incidence relationship between objects and attributes. In our particular case, this incidence relationship means that the tweet has a specific term. For instance, *Tweet1* has *Term1* whereas *Tweet2* has Terms  $\{2, 4, 7\}$  and 8. Only unigrams have been used to represent the tweet (i.e., individual terms). Although word bi-grams has demonstrated its effectiveness for some tasks [Glorot et al., 2011b; Wang and Manning, 2012], the use of word n-grams with  $n > 1$  on topic categorization is not always effective (see, e.g., references in [Tan et al., 2002]).

	<i>Term</i> <sub>1</sub>	<i>Term</i> <sub>2</sub>	<i>Term</i> <sub>3</sub>	<i>Term</i> <sub>4</sub>	<i>Term</i> <sub>5</sub>	<i>Term</i> <sub>6</sub>	<i>Term</i> <sub>7</sub>	<i>Term</i> <sub>8</sub>
<i>Tweet</i> <sub>1</sub>	×							
<i>Tweet</i> <sub>2</sub>		×		×			×	×
<i>Tweet</i> <sub>3</sub>	×	×		×				×
<i>Tweet</i> <sub>4</sub>						×		
<i>Tweet</i> <sub>5</sub>	×		×		×		×	
<i>Tweet</i> <sub>6</sub>	×		×		×		×	
<i>Tweet</i> <sub>7</sub>		×		×			×	×

**Table 4.4** – Formal context of a set of tweets and their corresponding set of attributes

Figure 4.1 shows the concept lattice corresponding to the formal context of Table 4.4. The lattice contains nine formal concepts of which seven will be considered as topics. Figure 4.2 also depicts the object count of each concept, which can be read as how many tweets are related to the topic. For instance, the formal concept (i.e. topic) marked (1) contains four tweets (i.e. a 57% of the whole set of tweets) and Term7 describes it. Lower concepts represent topics that are more specific as well as those related to the generic ones. For instance, the formal concept marked (2) is a specialization of its two upper neighbours. In this particular case, this concept owns two tweets (i.e. 29%) and Term2, Term4, Term7 and Term8 describe it. It can be easily seen that this description is a combination of the descriptions of its upper neighbours. Finally, it is also remarkable how the concept lattice naturally depicts non-comparable areas that can be understood as disjoint topics. The formal concept marked (3) is an example. In this case, Tweet4 is isolated in a topic described by Term6 that is not related to any other formal concept of the lattice. On the other hand, the other formal concepts are more or less related to each other, which mean that they represent closer topics.

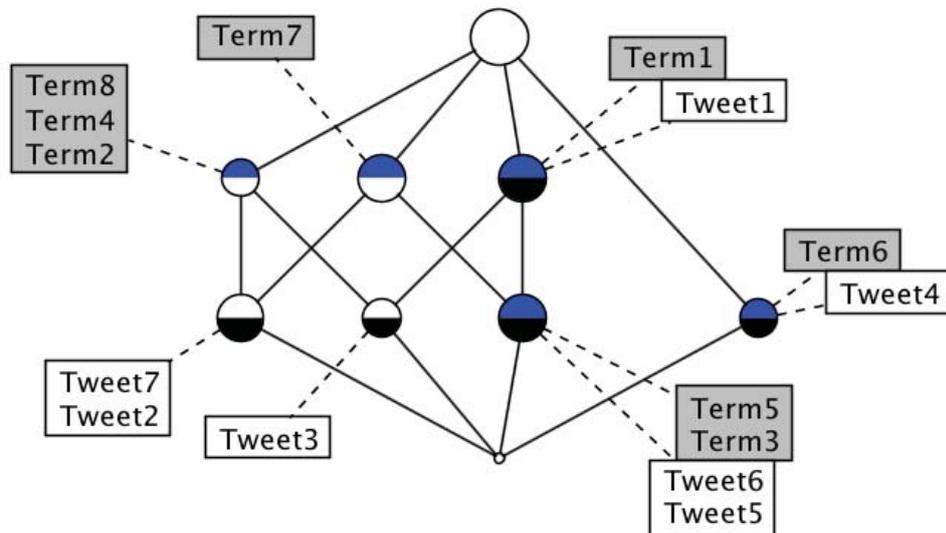
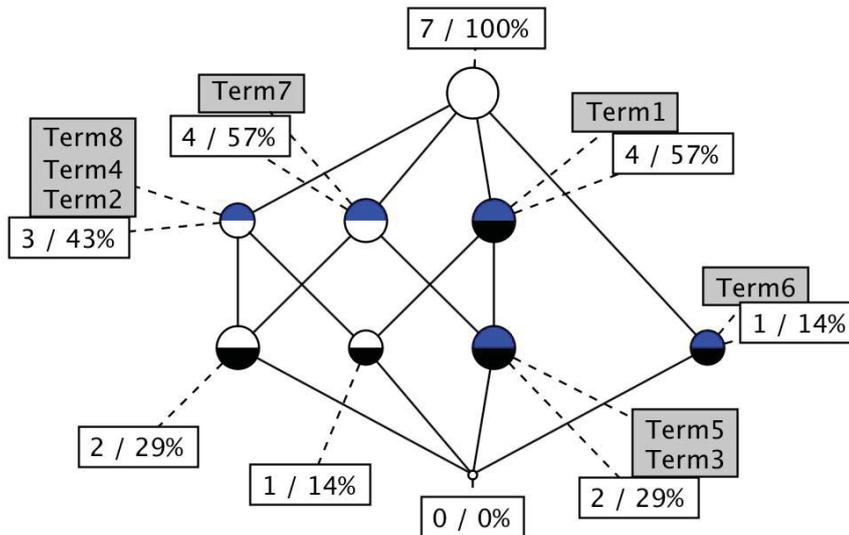


Figure 4.1 – Concept lattice corresponding to the formal context of Table 4.5



**Figure 4.2** – The same concept lattice as in [Figure 4.1](#) but with quantitative information on object count

#### 4.1.3.2. How to deal with big concept lattices?

FCA is a powerful mathematical theory dealing with unstructured data and discovering relationships and hidden information. However, its application is not straightforward when the formal context is too large and there are too many incidence relationships between objects and attributes. In this work, we propose an approach to reduce the final number of terms to represent the formal concept and introduce the idea of the formal concept stability to guide the selection of formal concepts (i.e. topics) once the lattice is built.

#### Term Selection Strategy

Although in the theoretical model all the tweet terms can be considered as attributes, in a real scenario this would generate an unmanageable concept lattice with a huge number of concepts. To solve this problem, we have applied a balanced selection strategy, presented and evaluated in [\[Cigarrán, 2008\]](#) to filter out these attributes according to their representativeness. This approach tries to maximize the distribution of the set of tweets in the lattice. The idea is to generate a balanced lattice with a low populated top concept. Therefore, our approach focuses on the selection of a set of terms able to cover (i.e. to describe) as many tweets as possible. To do so, the algorithm selects the most shared attributes, based on two frequency thresholds:

- a) A bottom frequency threshold used to discard those terms appearing in just a few tweets. These terms will not be considered in representing the tweets.

- b) An upper frequency threshold used to select a term as representative and therefore consider it to represent the tweets.

Once the high and low frequency terms have been selected or discarded respectively, the balanced selection algorithm is applied on the remaining terms. For each iteration, the algorithm selects the term with the highest frequency (i.e. appearing in most tweets) and then removes it from the list of selectable terms. Furthermore, all those tweets containing the selected term are also removed and the frequencies of the remaining terms are recalculated.

Table 4.5 to Table 4.8 illustrate how the balanced algorithm works. Table 4.5 shows the initial set of tweets and its corresponding terms. It also shows the document frequency of each term (i.e. number of tweets in which the term appears).

	$Term_1$	$Term_2$	$Term_3$	$Term_4$	$Term_5$	$Term_6$	$Term_7$	$Term_8$
$Tweet_1$	X	X						
$Tweet_2$	X	X	X					
$Tweet_3$	X		X		X			
$Tweet_4$	X		X			X		
$Tweet_5$		X				X		
$Tweet_6$		X			X			
$Tweet_7$		X			X		X	
$Tweet_8$			X	X	X			
$Tweet_9$				X				X
$Tweet_{10}$							X	X
<b>FREQ</b>	<b>4</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>4</b>	<b>2</b>	<b>2</b>	<b>2</b>

Table 4.5 – Original formal context to apply the balanced algorithm

Table 4.6 shows the first selection decision. In this case,  $Term_2$  is selected as it has the maximum frequency. Then  $Term_2$  is removed from the selectable set of terms as well as all the tweets in which  $Term_2$  appears (i.e.  $Tweet_1$ ,  $Tweet_2$ ,  $Tweet_5$ ,  $Tweet_6$  and  $Tweet_7$ ). The algorithm then iterates using the data shown in Table 4.7 where all the

frequencies are recalculated. In this case, Term3 is selected and Tweets 3, 4 and 8 are removed. In the final iteration, [Table 4.8](#), the algorithm selects Term8 and it finishes,  $D = \{\text{Term2}, \text{Term3}, \text{Term8}\}$  being the set of selected descriptors used to build the formal context.

	<i>Term<sub>1</sub></i>	<i>Term<sub>2</sub></i>	<i>Term<sub>3</sub></i>	<i>Term<sub>4</sub></i>	<i>Term<sub>5</sub></i>	<i>Term<sub>6</sub></i>	<i>Term<sub>7</sub></i>	<i>Term<sub>8</sub></i>
<i>Tweet<sub>1</sub></i>	X	X						
<i>Tweet<sub>2</sub></i>	X	X	X					
<i>Tweet<sub>3</sub></i>	X		X		X			
<i>Tweet<sub>4</sub></i>	X		X			X		
<i>Tweet<sub>5</sub></i>		X				X		
<i>Tweet<sub>6</sub></i>		X			X			
<i>Tweet<sub>7</sub></i>		X			X		X	
<i>Tweet<sub>8</sub></i>			X	X	X			
<i>Tweet<sub>9</sub></i>				X				X
<i>Tweet<sub>10</sub></i>							X	X
<b><i>FREQ</i></b>	<b>4</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>4</b>	<b>2</b>	<b>2</b>	<b>2</b>

Table 4.6 – Balanced algorithm. First iteration

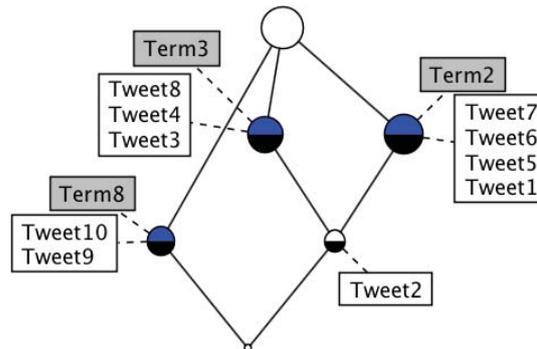
	$Term_1$	$Term_3$	$Term_4$	$Term_5$	$Term_6$	$Term_7$	$Term_8$
$Tweet_3$	X	X		X			
$Tweet_4$	X	X			X		
$Tweet_8$		X	X	X			
$Tweet_9$			X				X
$Tweet_{10}$						X	X
<b><i>FREQ</i></b>	<b>2</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>2</b>

Table 4.7 – Balanced algorithm second iteration

	$Term_1$	$Term_4$	$Term_5$	$Term_6$	$Term_7$	$Term_8$
$Tweet_9$		X				X
$Tweet_{10}$					X	X
<b><i>FREQ</i></b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>2</b>

Table 4.8 – Balanced algorithm third iteration

	$Term_2$	$Term_3$	$Term_8$
$Tweet_1$	X		
$Tweet_2$	X	X	
$Tweet_3$		X	
$Tweet_4$		X	
$Tweet_5$	X		
$Tweet_6$	X		
$Tweet_7$	X		
$Tweet_8$		X	
$Tweet_9$			X
$Tweet_{10}$			X



**Figure 4.3** - Final formal context, filtered by the balanced algorithm and its corresponding concept lattice

### How to select the best topics?

Although initially every formal concept is susceptible to representing a topic, the selection of all the generated concepts as topics can be overwhelming. Moreover, the application of the term selection strategy presented above, although it reduces the potential number of formal concepts, it might lead to a large number of topics. How can we deal with this situation? Is it possible to select a subset of formal concepts to obtain the most representative topics detected?

In this sense, a desirable characteristic of the concepts (i.e. topics) would be the cohesion between their objects (i.e. tweets). Otherwise, it would indicate that a concept is not

really a topic but a general grouping of different topics. To reflect how much each formal concept in the lattice fits with this requirement we propose the use of the stability concept presented at section 3.2. To recap, the stability of a *formal concept* indicates how much the concept intent depends on particular objects of the extent. In other words, the stability of a concept is the probability of preserving its intent after leaving out an arbitrary number of objects. Thus, a high stability value indicates that the concept represents a cohesive set of tweets or, equally, it can represent a proper topic.

#### 4.1.4. Other Topic Annotation Approaches

---

Our proposal to address the topic detection task is to apply Formal Concept Analysis to model the data, grouping together similar contents and finally selecting the most appropriate clusters. To put in context the results of our proposal, we have also implemented two common approaches for the Topic Detection Task: Hierarchical Agglomerative Clustering (HAC) and Latent Dirichlet Allocation (LDA)<sup>8</sup>. HAC has been demonstrated as the best performing approach in the past RepLab Campaign [Amigó et al., 2013a] as well as in more recent experimentations [Spina et al., 2014], while LDA is one of the most widely applied approaches in the state of the art of Topic Detection and Topic Modelling [Blei, 2012]. In the following, our FCA-based approach, the HAC and LDA approaches are detailed. Other approaches, e.g., graph-based (see section 2.2), have also been proposed for the Replab Campaign; however, none of them obtain valuable results for the topic detection task [Amigó et al., 2013a]. In what follows, we provide more details on HAC and LDA and their application for Topic Detection.

##### 4.1.4.1. HAC Clustering

Clustering refers to the set of methodologies applied to group a series of documents together in cohesive clusters in such a way that the documents in the same cluster are more similar (based on some closeness degree set by a distance measure) to each other than to the documents in other clusters. The basic clustering operation is based on a flat structure, that is, all the clusters belong to the same hierarchical level.

---

<sup>8</sup> HAC and LDA have been implemented with the help of the LingPipe text processing tool kit: <http://alias-i.com/lingpipe/index.html>

Nevertheless, the data to be clustered rarely have a flat hierarchy. For instance, topics are usually susceptible to being classified at different levels, depending on the granularity desired for the annotation. For instance in the tweet:

Berkeley Researchers Say #Carbon #Pollution Can Be Turned Into #Energy

The following topics, from more generic to more specific, can be identified:

University

Berkeley University

Berkeley University Research

Berkeley University Research in Energy

Hierarchical Clustering (HC) pursues to build a cluster representation that not only cluster the data, but also create a hierarchy of this representation. Within this methodology, Hierarchical Agglomerative Clustering or HAC seeks to create a clustering representation based on a hierarchical structure by applying an agglomerative approach [Manning et al., 2008]. The HAC computation starts by creating a cluster for each document to be clustered. Then, the most similar pair of clusters is merged in a new cluster. The merging decisions depend on the similarity between documents established by some measure: the next pair of clusters to merge is the one with the fewest distance between them. This process is iteratively repeated until only one cluster remains. Because of this computation a binary tree hierarchy (usually known as *dendrogram*) of the clusters is created. Figure 4.4, contained in [Manning et al., 2008], shows an example of a *dendrogram* obtained from the result of a HAC computation. It shows the clustering of 30 documents from Reuters RCV1. In the y-axis the similarity between clusters is exposed while the x-axis contain each of the documents to be clustered. Each merge between clusters, from the bottom to the top, can be reconstructed by checking the *dendrogram*.

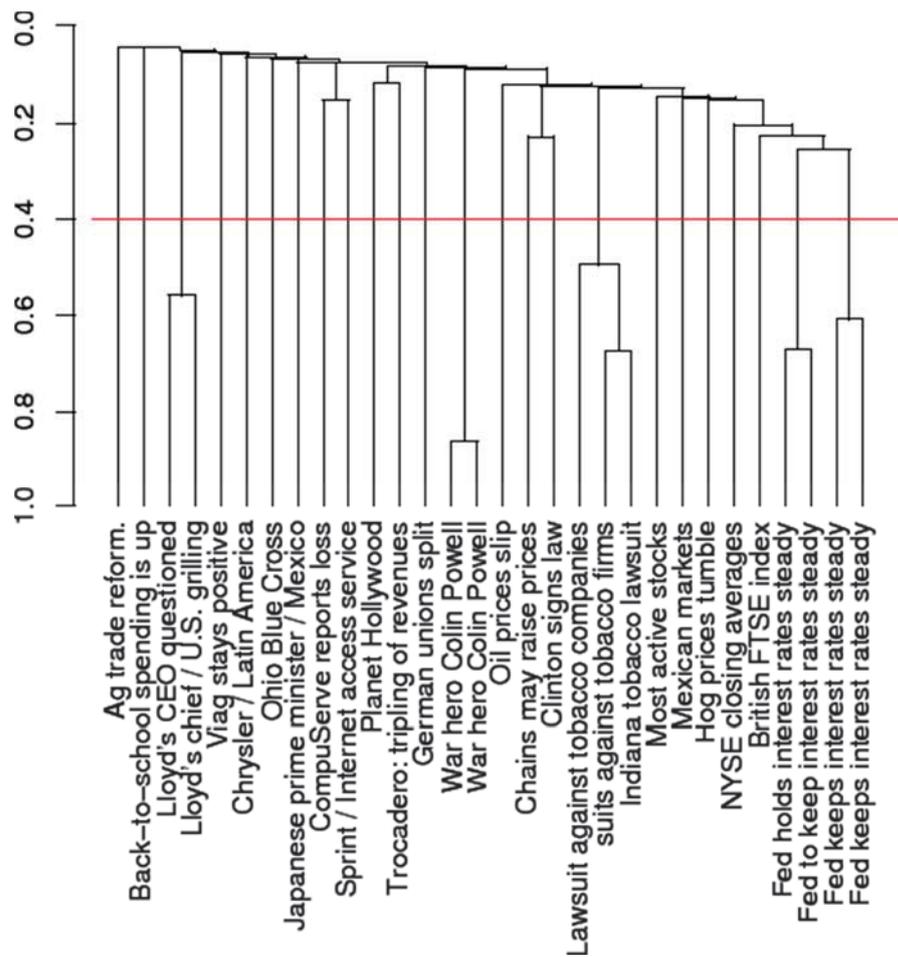


Figure 4.4 – Example of *dendrogram*

HAC offers a hierarchical representation; however, in many task, and specifically in the Topic Detection task, a flat data representation is needed. Therefore, a flat representation has to be inferred from the cluster hierarchy. Hence, once the hierarchy is built, a degree of granularity has to be chosen to generate the clustering result (i.e. from the most generic, only one cluster, to the most specific, one cluster for each document). Therefore, the final clustering representation will be the set of clusters at the selected hierarchy level. In [Figure 4.4](#), this level is represented by the horizontal line. This line represents the degree of granularity (i.e. the minimum similarity that any two documents should reach to belong to the same cluster). In the example, this level is equal to 0.4 and it generates 24 clusters.

The topic detection task can be understood as a clustering task: a set of documents should be classified in a set of unknown classes. Therefore, the application of HAC for Topic Detection is straightforward. In more detail, we have implemented a single-linking HAC algorithm applying the Jaccard Similarity as distance measure. Once the

*dendrogram* has been generated, several similarity levels for creating the final clustering representation have been applied to test the HAC performance with different granularities.

#### 4.1.4.2. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a generative probabilistic technique for modelling a text corpus through the underlying set of topics addressed in it [Blei et al., 2003]. In this sense, LDA proposes a model similar to Probabilistic Latent Semantic Analysis (pLSA) [Hofmann, 1999]. The main differences lie in that in LDA the topic distribution follows a Dirichlet distribution and that pLSA does not provide a probabilistic model at the level of documents.

The basic idea of LDA is that each document is represented as random mixture over latent topics (i.e., the intuition behind LDA is that documents are likely to exhibit multiple topics.), which are represented by a distribution over words. More in detail, for each document  $d$ , LDA assumes the following generative process:

- Document  $d$  has a number of words  $N$ , such as  $N$  follows a Poisson distribution:

$$N \sim \text{Poisson}(\xi)$$

- The document has a topic mixture (according to a Dirichlet distribution over a fixed set of  $K$  topics with a prior  $\alpha$ ):

$$\theta_d \sim \text{Dir}(\alpha)$$

- For each of the  $N$  words  $w_n$  in the document  $d$ :
  - Choose a topic  $z_n \sim \theta_d$
  - Choose a word  $w_n$  from  $p(w_n | z_n, \beta)$ , a multinomial probability conditioned on the topic  $z_n$ , and the prior  $\beta$ .

Figure 4.5 depicts the graphical model representation of this process where the boxes are “plates” representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.

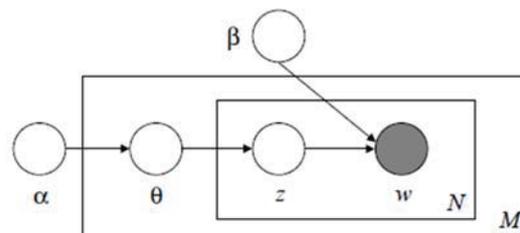


Figure 4.5 – Graphical model representation of LDA

Applying the generative process described before it is possible to describe each document by means of the set of hidden or latent topics. Then, the problem will be based on, using the observed documents, to infer the hidden topic structure. This can be thought of as “reversing” the generative process: What is the hidden structure that likely generated the observed collection? [Blei, 2012].

Therefore, the final topic modelling (the topics to be detected in the corpora and their relationship to the documents) will arise by computing the hidden structure that more likely generated the observed documents. It is important to highlight that the number of topics has to be fixed a priori (i.e. the number of topics is an input parameter for the LDA algorithm). It represents a significant drawback for the topic detection task. As was pointed out, the number of topics is not known a-priori, neither their distribution throughout the documents.

More details about the application of LDA for Topic detection, as well as the specific issues and related problems can be found at [Blei, 2012], at [AlSumait et al., 2008] and in their related works.

#### 4.1.4.3. Comparison of HAC and LDA vs. FCA

After reviewing the three approaches, FCA presents some advantages for the proposed task in comparison to HAC and LDA. The first one is that FCA does not limit the topic representation to one hierarchy level. As we said before, the topic representation rarely presents a flat hierarchy, but it is likely to include topics with different granularities. In this sense, HAC creates a multi-level representation just as FCA does; however, in order to create the final clustering partition, the *dendrogram* has to be cut at a specific granularity level. As regards LDA, the topic representation is a flat partition, losing the information that a hierarchical representation offers. In contrast, FCA creates a hierarchical representation and it selects topics from different levels in the hierarchy.

Another important aspect is related to the algorithm parametrization. The FCA computation may be completely unsupervised: it does not require any training process or any parameter to be implemented. In this work, we have modified the FCA implementation with some parameters to adapt the algorithm to different situations (e.g. topic selection thresholds or stability). Nonetheless, this parametrization has been only applied to test its influence in the system performance for the proposed task; that is, how these parameters affect the FCA performance in comparison to the other approaches. In other words, these parameters are not necessary for the FCA computation and it would work without them anyway. In contrast, although the HAC algorithm does not have to be parametrized, a similarity threshold for generating the final cluster partition has to

be selected. LDA also has to be parameterized to infer the latent topics: it is mandatory to set the number of topics to be detected previously. FCA has other advantages. For instance, unlike HAC and clustering in general, the clusters/topics are automatically labelled. FCA also offers a better representation — the concept lattice — of the results than the HAC-dendrogram and the flat representation provided by LDA.

As regards the complexity of the algorithms, all three have a high computational complexity, although FCA is the worst case when the number of attributes is very large. The algorithm we use to compute FCA is the Next Neighbours algorithm [Carpineto and Romano, 2004], which presents a complexity of  $\mathcal{O}(|G|^2|M||C|)$  being  $|C|$  the number of formal concepts that has been generated at each step of the algorithm,  $|G|$  the number of objects and  $|M|$  the number of attributes. The complexity of LDA for topic detection grows linearly with the number of topics  $K$ , the number of documents  $N$  and the number of words in the vocabulary  $V$ :  $\mathcal{O}(|K||N||V|)$ . Finally, the HAC's is  $\mathcal{O}(N^3)$ , where  $N$  is the number of documents.

Taking all the previous considerations into account, FCA seems to be more suitable approach for the Topic Detection task than those applied in the literature. In the following experimentation, we intend to test this point experimentally by means of a cluster quality evaluation, using some well-known metrics in the state of the art and an external evaluation, based on the evaluation environment proposed by the RepLab Campaign.

### 4.1.5. Experimentation

---

This section presents the experimental results applying the aforementioned scenario, Topic Detection Task @ RepLab 2013. In particular, section 4.1.5.2 includes a study of the FCA performance according to the different aspects related to the experimentation (e.g., term selection, stability, cluster selection and topic adaptation). This evaluation is conducted according to the RepLab Evaluation Framework proposed at section 4.1.2.3. On the other hand, section 4.1.5.3 details the comparison of the FCA results and those of the other two proposals (LDA and HAC), applying both, internal and external evaluations, to carry out this comparison.

Given that the topic detection is mainly based on the tweet content to be processed, an important step is the pre-processing of these data. Twitter has some special characteristics that make its application more challenging: the existence of special signs

(i.e. abbreviations, emoticons or hashtags), the use of slang, and the shortness of the content (limited by Twitter) or spelling mistakes.

To address some of these special characteristics and to test their impact on the final algorithms performance, a pre-processing methodology has been proposed to enrich the tweet content and to manage the special Twitter signals (e.g., references and hashtags). The pre-processing, detailed at section 4.1.5.1, is carried out in two steps: 1) enrichment of the content with external knowledge, and 2) managing the enriched data to create an accurate representation.

#### 4.1.5.1. Content Pre-processing

The textual content in the tweets is quite reduced due to the character limitation imposed by Twitter. In this context, content enrichment with external information appears as an indispensable step in the management of Twitter data and a valuable help in the topic detection process. To test that end, we have enriched the tweets in the RepLab dataset by means of Textalytics<sup>9</sup>, a Semantic API that integrates several linguistic and semantic tools for text processing. Tweets have been represented by means of the following features using the information in the dataset and the enrichment provided by Textalytics:

- **Text:** Textual content of the tweets. After removing stop-words and stemming the terms in the tweets, each term is considered as a feature to represent the tweet. Stop-words were removed in order to avoid the inclusion of uninformative terminology, while stemming has been done to increase the density of the relationships between tweets and content (i.e. car or cars are considered as the same term).
- **Hashtags:** Hashtags in the tweets. Hashtags are thematic-based tags used by the Twitter users to label the tweets (e.g. #worldcup2014, #oscars...).
- **References:** References to other user profiles (e.g. @username).
- **URLs:** URLs contained in the tweets. In case of a shortened URL, we have resolved it and used the complete URL.
- **Named Entities:** Named Entities identified by Textalytics.
- **Concepts:** General concepts detected by Textalytics.

---

<sup>9</sup> <https://textalytics.com>

- **Categories:** Category/ies identified by Textalytics in which the tweet could be classified, based on the IPTC categories<sup>10</sup>.

An example of these tweet representations is shown in [Table 4.9](#):

<b>Tweet</b>	Seen the brand new 2013 Bmw #m5 again in San Leandro! Fastest 4 door car in the world <a href="http://instagr.am/p/THbbcnsZi9/">http://instagr.am/p/THbbcnsZi9/</a>
<b>Text</b>	see, brand, 2013, bmw, #m5, san, leandro, fast, 4 door, car, world
<b>Hashtags</b>	#m5
<b>References</b>	-
<b>URLs</b>	<a href="http://instagr.am/p/THbbcnsZi9/">http://instagr.am/p/THbbcnsZi9/</a>
<b>NE</b>	BMW, San_Leandro
<b>Concepts</b>	BMW, car, world
<b>Categories</b>	Sport, motor_racing

**Table 4.9** – Example of Tweet Representations

After the enrichment process, different tweet representations may be created using the previous features separately and all together (ALL in [Table 4.10](#)). It is expected that the aggregation of all the features will lead to a more accurate representation considering that more information is taken into account. In this sense, we propose a refined way of integrating all the features contained in each tweet (ALL + Processing in [Table 4.10](#)). This process is carried out as follows:

1. Normalize named entities in each tweet by unifying them under common labels (e.g. BMW\_M3 and M3 are considered the same entity). The idea is to unify different lexical realizations of the same concept. To that end, an heuristic process has been automatically carried out, by applying the following steps:
  - a. As a previous step, the NE in the tweets are split into their constituent parts (e.g., BMW\_M3\_coupe is converted to {BMW, M3, coupe}) and their occurrence frequencies are identified.

<sup>10</sup> <http://cv.iptc.org/newscodes/subjectcode/>

Named Entity	<i>frequency</i>
{BMW, M3}	100
{BMW, M3, V8}	10
{M3}	60
{M3, V8}	1

- b. Given a tweet with named entities to be normalized, for each named entity the lexical realization with more frequency is selected. For instance, if a tweet has the named entity M3 or {M3, V8}, it will be represented as BMW\_M3.
2. Expand the features in the tweets with the identified hashtags: if there is a hashtag #M3, all the tweets with the term/concept/category/named entity M3, it will be represented with the hashtag #M3. This process is automatically carried out by looking for lexical matches between features and the text of the hashtags (i.e., without the symbol #). The objective is again to unify similar content, expressed as different terms.

The same process has been applied for the three topic detection methods (HAC, LDA and FCA). The results of the aforementioned representations (see Table 4.9) are shown in Table 4.10 in terms of Reliability, Sensitivity and the F-measure of both (see section 4.1.2). In order to just focus on the representations, Table 4.10 presents the results obtained by the best configuration for the three approaches: 10 clusters for LDA, stability equals to 0.2 for FCA and a similarity threshold of 0.8 for HAC.

	LDA			HAC			FCA		
	R	S	F(R,S)	R	S	F(R,S)	R	S	F(R,S)
<b>Text</b>	0.48	0.12	0.13	0.36	0.62	0.32	0.61	0.23	0.29
<b>Hashtags</b>	0.97	0.01	0.01	0.33	0.55	0.29	0.99	0.01	0.01
<b>References</b>	1.00	0.00	0.00	0.22	0.79	0.17	1.00	0.00	0.00
<b>URLs</b>	0.76	0.06	0.10	0.59	0.39	0.39	1.00	0.01	0.02
<b>Named Entities</b>	0.57	0.09	0.13	0.27	0.57	0.22	0.87	0.06	0.10
<b>Concepts</b>	0.71	0.06	0.10	0.27	0.49	0.21	0.90	0.04	0.07
<b>Categories</b>	0.94	0.01	0.02	0.23	0.66	0.17	0.94	0.02	0.04
<b>ALL</b>	0.76	0.04	0.07	0.33	0.65	0.29	0.64	0.18	0.25
<b>ALL + Processing</b>	0.45	0.21	0.15	0.38	0.60	0.41	0.41	0.32	0.33

**Table 4.10** – F-measure results obtained by the different Tweet Representations

The table shows that the strongest signal for detecting topics appears to be the textual content of the tweets: text is the best-performing feature. Even when all the features (including the text) are aggregated, their performance is worse than that obtained using

just the text. It is at this point where our proposal to aggregate the features proves its suitability. As can be seen in [Table 4.10](#), our proposal (**ALL + Processing**) achieves the best performance for all the algorithms (HAC, LDA and FCA). It demonstrates that the proposed extra features can improve the text-based results by offering other signals for topic detection, different from those offered by the text.

Consequently, for the following experimentation, the proposal including the entire tweet features aggregated according to the content representation presented before (**ALL + Processing**) is selected to represent the input data for all the experimental configurations.

#### 4.1.5.2. FCA-based experimentation

As said before, FCA has been applied for the detection of topics in a set of tweets. The results of FCA are detailed in the following sections. In particular, this analysis is focused on different aspects: the impact of the term selection, the impact of the stability in the selection of formal concepts, the impact of the cluster selection strategy and, finally, the adaptability of the topic detection to the appearance of new topics.

#### Analysing the impact of the Term Selection Strategy

Prior to the application of FCA, and to avoid huge concept lattices with low populated formal concepts (i.e. clusters or topics); we applied the term selection strategy as explained in section [4.1.3.2](#). To recall, the objective of this term selection is to remove those terms that can be considered uninteresting, due to their low frequencies, and retain those terms with high frequencies as representative of the formal context. To that end, the algorithm iteratively selects the best terminology to represent each tweet at least with one term but minimizing the number of terms finally selected to build the formal context.

This first experiment analyses the impact of this algorithm setup by selecting different values for the upper cutting threshold and the lower cutting threshold respectively. The values of these thresholds have a high impact on the final number of terms selected and consequently, the number of formal concepts (i.e. topics). Thus, a high lower cutting threshold value will select a small set terms and, conversely; a low lower threshold value will produce a higher set of selected terms. Remember that the lower threshold determines the frequency value below which the terms are not considered by the algorithm.

On the other hand, a high upper threshold will keep a small portion of terms in the formal context, whereas a low upper threshold will consider a higher number of terms to

be included in the formal context. Remember that the upper threshold determines the frequency value above which the terms are considered as part of the formal context.

The question here is to analyse the outcome of the proposed FCA approach by modifying these thresholds or, in other words, to check the correlation between the final numbers of attributes selected and the system performance. Although, it seems natural that the more attributes, the greater the system performance; we would like to experiment with this assumption in order to test whether there are a worthy increase in the performance that justifies increasing the algorithm computation time (i.e., if the threshold value is relaxed, the formal context will be bigger, increasing the FCA computation time).

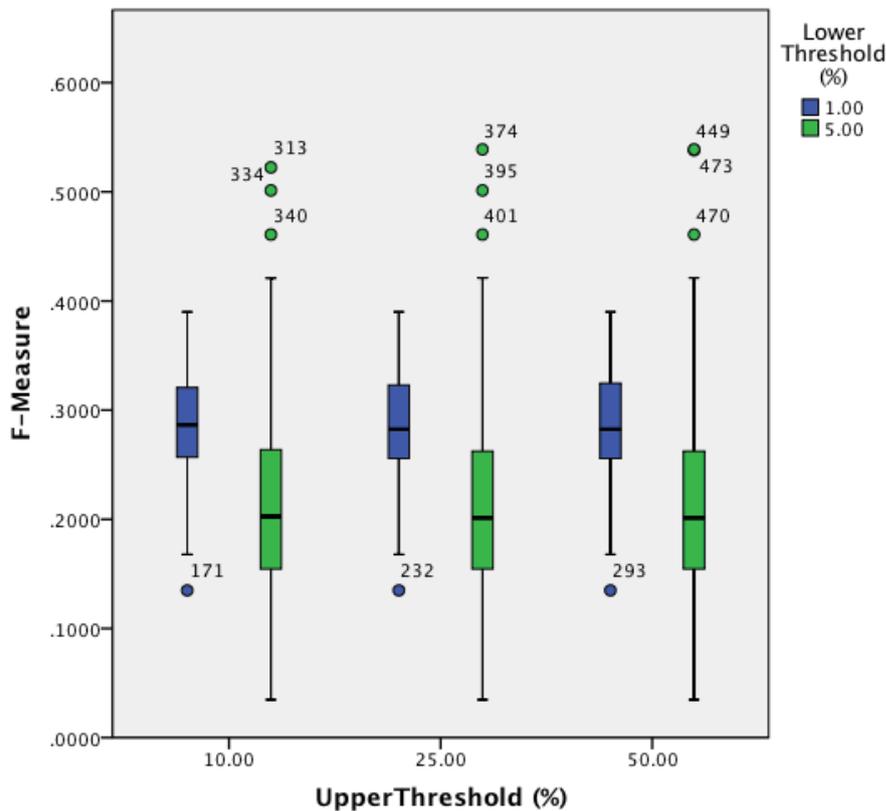
Table 4.11 shows the overall results, expressed in terms of Reliability, Sensitivity and F-measure, considering lower thresholds of 1% and 5 % and upper thresholds of 50%, 25% and 10%. For these experiments we selected a stability value of 0.9 (i.e. a discussion about which stability values to choose will be presented on the following section). It can be seen that the results, according to the F-measure, are better with the low lower threshold. In particular, the 1% and 50% configuration increases the F-measure value a 28.54% with respect to the configuration taking 5% as the lower threshold. These results confirm the initial intuition that the more information, the better. They also show that the increase in performance is worthy enough to justify the increase in the computation time of the FCA algorithm.

Lower Threshold	Upper Threshold	Reliability	Sensitivity	F(R,S)
1	50	0,3021	0,3343	<b>0,2882</b>
1	25	0,3029	0,3324	0,2878
1	10	0,3039	0,3311	0,2877
5	50	0,1678	0,6778	<b>0,2242</b>
5	25	0,1680	0,6746	0,2235
5	10	0,1685	0,6715	0,2236

**Table 4.11** – Impact of lower and upper threshold values on the final concept lattice

Figure 4.6 shows this scenario in more detail. The boxplot shows the median and quartiles of the upper threshold with respect to the F-measure grouped by the lower threshold. This plot can give an idea about the performance for each analysed entity. First, the experiments using 5% for the lower threshold show a higher dispersion of results due to bigger quartile ranges (for both Q1 and Q4 quartiles and interquartile range). These results explain a non-uniform behaviour. They also have more outliers. On the other

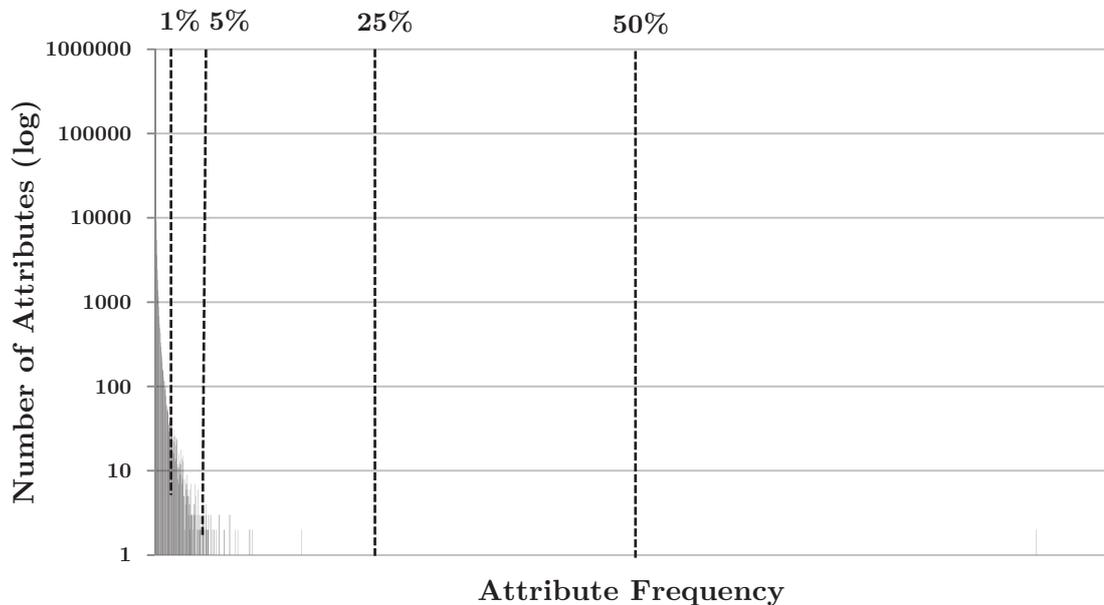
hand, a 1% lower threshold shows more uniform (i.e. less disperse) results, quite independent of the upper threshold selected value and with higher medians for all cases. These results indicate that a low lower threshold can improve the term selection algorithm, thus allowing it to manipulate more terminology and consequently, generate better formal contexts.



**Figure 4.6** – Impact of the lower and upper threshold on the final performance

These differences rely on the term distribution (i.e., a long tail shape as typically happens in the distribution of terms in a set of documents). This distribution can be seen in the histogram in [Figure 4.7](#). This figure shows the number of attributes (in the y-axis) with a given frequency (in the x-axis, which is ordered from frequency 1 at the left to the max frequency at the right) and the position in the histogram of the threshold values (vertical lines). All the terms in the left of the lower threshold values (i.e., with a frequency lower than that established by the threshold) are removed. Conversely, all the terms in the right of the upper threshold values are considered. As can be seen, there are a large number of terms with low frequency values and only a few terms with high frequencies. As the lower threshold removes the terms with frequencies below it, a high value of this threshold will delete too many terms that will not be considered by the term selection algorithm, resulting in a great loss of information (see the difference between 1% and 5%

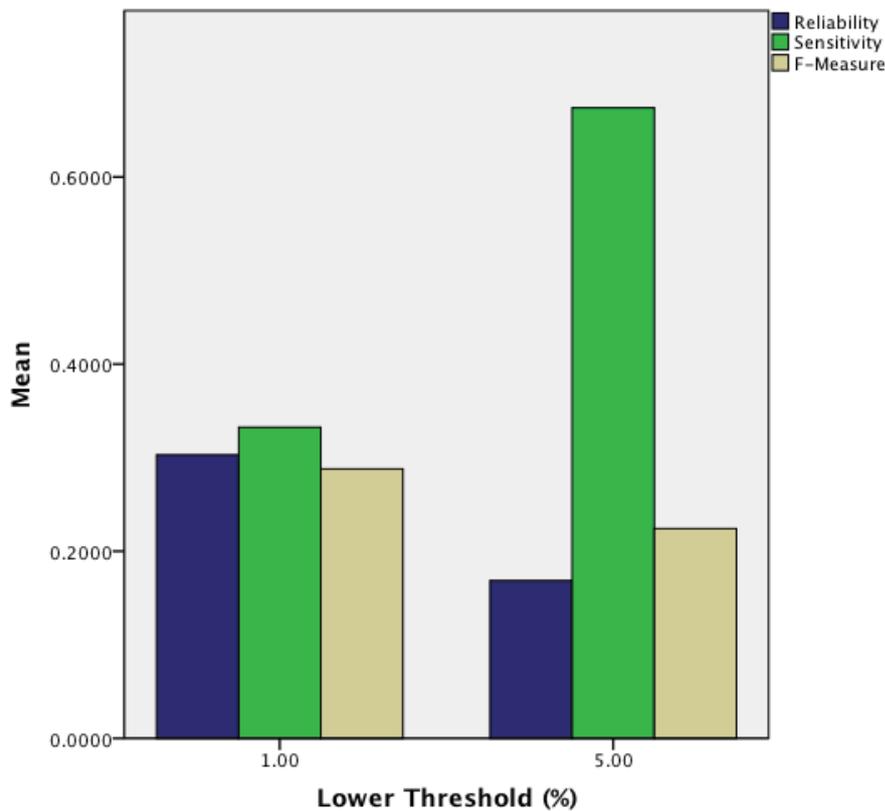
lines in the figure). This is also the reason why the changes in the upper threshold values do not affect the final F-measure results. In this particular case, the terms involved have high frequencies but there are not too many; so, their previous selection by the threshold does not affect to the final performance of the algorithm.



**Figure 4.7** – Term distribution and thresholds

Figure 4.8 shows the same output but in terms of reliability, sensitivity and F-measure for both lower thresholds. It can be seen that differences between reliability and sensitivity in the case of a 1% lower threshold are not quite as different as in case of 5%. In other words, the reliability of the system creating clusters is around 30% in the case of a 1% threshold, whereas it is around 16% in the case of a 5% threshold. This means that the precision of the set of clusters presented is more accurate by choosing a 1% threshold. On the other hand, a threshold of 1% produces a 33% sensitivity value, whereas a 5% threshold produces a 67% sensitivity value.

What does this mean? Is the 5% approach doing well? The answer is no. The meaning of sensitivity indicates the proportion of relationships covered by the system with respect to the gold standard. A hypothetical case in which the system would cluster all the tweets using only one cluster would produce a sensitivity value equal to 1. Therefore, the answer is that although the 5% system is capturing more relationships, it is not clustering them correctly because of the low reliability value. The balance of both measures is shown in the F-measure bars, where it can be clearly seen that the system with a 1% threshold performs better.



**Figure 4.8** – Reliability, Sensitivity and F-Measure values for different lower thresholds using a 25% upper threshold

However, to go further with this issue, we decided to experiment with even lower thresholds (0.1% and 0.5%), which barely reduce the attribute set. Although these values increase the computation time, it is worthy to experimenting with them for research purposes. [Table 4.12](#) shows the results of these experiments when choosing an upper threshold of 25%. While a lower threshold for values 1% and 5% seems to be better if the threshold is extremely low, the system performance starts to decrease again in terms of the F-Measure. This decrease is due to the reduction in the recall (i.e. sensitivity) values and it can be explained by the following: if we add more attributes, the system will generate more and, consequently, smaller clusters. Having smaller clusters means it will be easy for those tweets in the same cluster to be really related which, in turn, means a higher precision or reliability value. However, it is more difficult for the whole set of tweets that are related to be in a single cluster.

Lower Threshold	Upper Threshold	Reliability	Sensitivity	F(R,S)
5	25	0,1680	0,6746	0,2235
1	25	0,3029	0,3324	0,2878
0.5	25	0,3836	0,2412	0,2710
0.1	25	0,4075	0,2204	0,2671

Table 4.12 – Results using a 0.5% and a 0.1% lower threshold

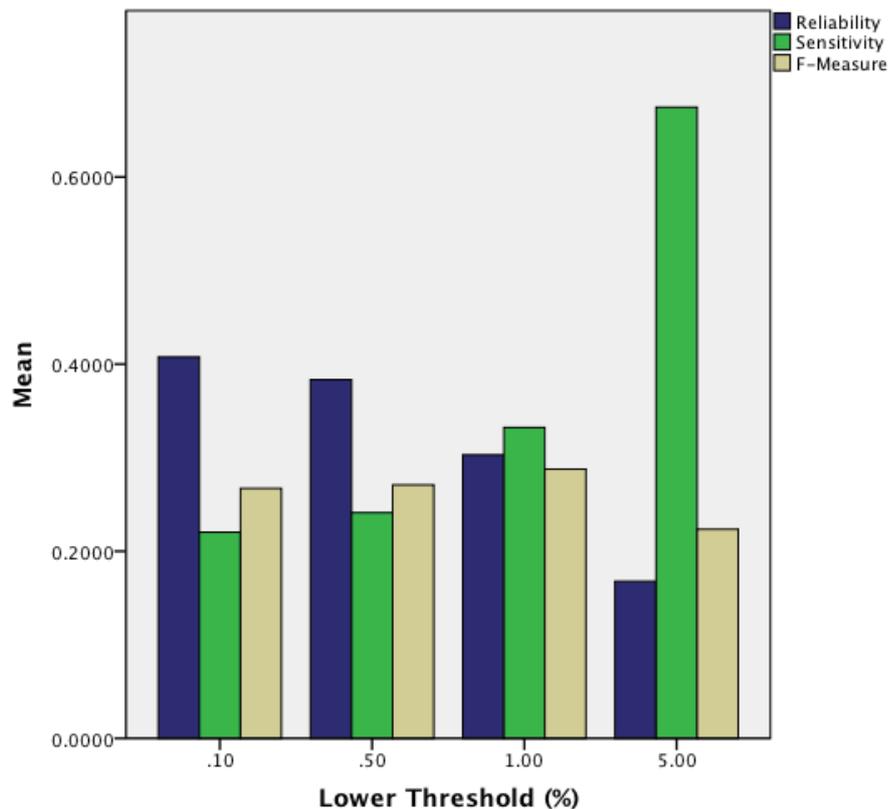


Figure 4.9 – Reliability, Sensitivity and F-Measure values for 0.1%, 0.5%, 1% and 5% lower threshold values

To sum up (Figure 4.9), the lower threshold is directly correlated to the Sensitivity value (the higher the threshold, the higher the sensitivity value) and inversely correlated to the Reliability value (the higher the threshold, the lower the reliability value). Considering the overall performance, the best results are obtained where reliability and sensitivity values are similar; that is, in those cases where the system does not present extreme behaviours. From these results, we can conclude that although more terms (attributes) mean more information it does not imply a better performance on our system, as the increase of information is penalized by the inclusion of noisy data.

Another point is that, although the lower threshold has a great influence on the results, the upper threshold does not seem to be as relevant. An explanation to this phenomenon is that a change in the upper threshold does not affect the number of selected attributes too much. In contrast, a small variation in the lower threshold does affect the final number of selected attributes too much. Because of this, the number of formal concepts created (i.e. the number of possible topics) will differ a lot, relying on the number of attributes selected (see [Table 4.13](#) for more information).

Hence it can be posited that only lower threshold of the selection algorithm has a real impact on the lattice computation. In this way, it can be concluded that, in general, more attributes (i.e. more information) are better. This enhancement in the performance is enough to justify the increase in the computation time that it entails. However the removal of the less frequent (and noisy) attributes make sense to improve general performance. In our case, the best performance is obtained when a lower threshold between 0.5 and 1 % is used.

Lower Threshold	Upper Threshold	Number of Generated Concepts	Average Concepts by Entity
1	50	29836	489
1	25	31384	514
1	10	32566	533
5	50	1100	18
5	25	1154	18
5	10	1258	20

**Table 4.13** – Attribute Reduction Algorithm Threshold Analysis

### Analysing the impact of the Stability-based Cluster Selection

Once the final attributes has been selected, the concept lattice is built. Although the attribute selection strategy produces a reduced concept lattice, depending on the final number of attributes selected, the final number of formal concepts (i.e. clusters) can be still large (see [Table 4.13](#)). For this reason, it might be appealing to filter out the concept set in order to reduce the final number of selected clusters. To than end, in section [3.2](#), we introduced the concept of stability as a feasible way of selecting the most suitable formal concepts to be taken as clusters or topics.

Thus, stability is also an important parameter to adapt the final system performance. The higher the stability value, the more restrictive the topic selection (i.e. fewer clusters are finally selected). In this sense, we experimented with different stability values to get

a better understanding of the influence of this parameter in the overall results. Table 4.14 and Figure 4.10 show the F-Measure values considering a fixed upper threshold of 25% and a 0.5%, 1% and 5% lower threshold respectively. In this scenario, we varied the stability value from 0.2 to 0.9. Results demonstrate that stability values have a significant impact on the final results, especially for the 1% lower threshold experiment.

The experiment with a 5% lower threshold shows that the stability value is not relevant. This is because the original number of clusters generated is already low, containing too many tweets per cluster. It makes that, although the cluster set is not fine grained, each of its clusters is quite stable (i.e., it is not affected by removing a small number of objects). This means that relaxing the stability value does not have a great impact on the final 5% results.

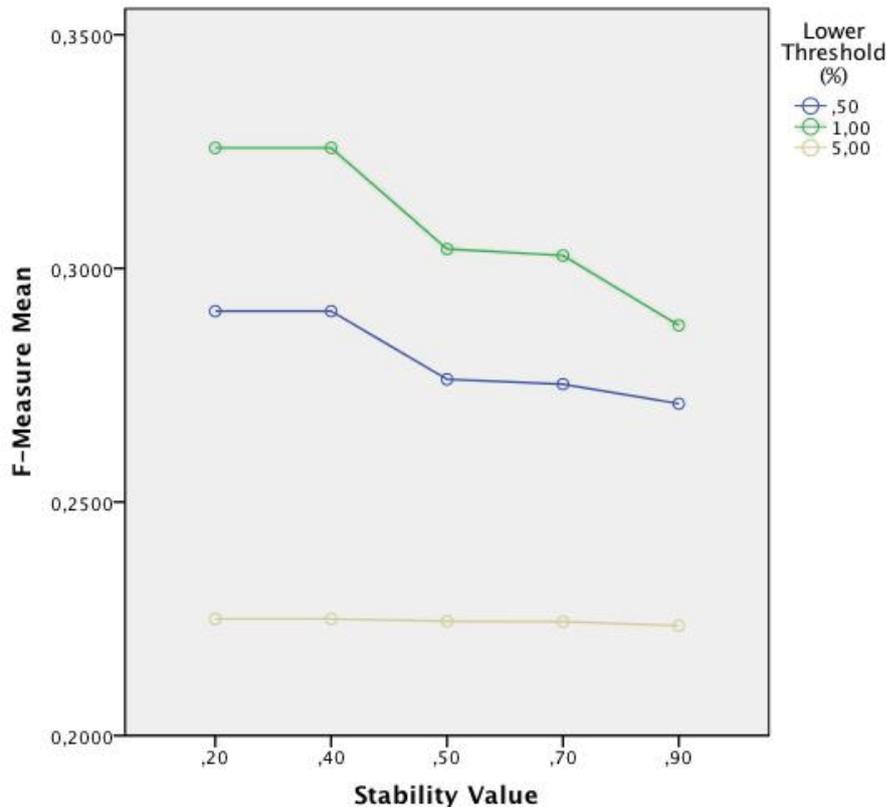
On the other hand, a 1% and 0.5 % lower threshold value generates a fine-grained cluster set. It means more clusters but less populated. Is this configuration more sensitive to changes in the stability value? The answer is yes, as can be seen in Figure 4.10.

Stability Value	Lower Threshold	Reliability	Sensitivity	F(R,S)
0.2	0.5%	0,6053	0,2099	0,2908
0.4	0.5%	0,6053	0,2099	0,2908
0.5	0.5%	0,4499	0,2245	0,2763
0.7	0.5%	0,4386	0,2265	0,2752
0.9	0.5%	0,3836	0,2412	0,2710
0.2	1%	<b>0,4090</b>	<b>0,3163</b>	<b>0,3258</b>
0.4	1%	<b>0,4090</b>	<b>0,3163</b>	<b>0,3258</b>
0.5	1%	0,3455	0,3228	0,3041
0.7	1%	0,3407	0,3236	0,3027
0.9	1%	0,3029	0,3324	0,2878
0.2	5%	0,1696	0,6744	0,2250
0.4	5%	0,1696	0,6744	0,2250
0.5	5%	0,1691	0,6744	0,2245
0.7	5%	0,1690	0,6744	0,2244
0.9	5%	0,1680	0,6745	0,2235
<b>Best RepLab</b>		<b>0,4624</b>	<b>0,3246</b>	<b>0,3252</b>

**Table 4.14** – Reliability, Sensitivity and F-Measure values varying the stability value from 0.2 to 0.9 on a setup of 0.5%, 1% and 5% lower threshold values

In general, low stability values produce the best F-Measure results in our experiment scenario. This means that less stable clusters match with the annotators' criterion that,

in turn, means that the annotators have preferred to group tweets into small clusters, not sharing too many terms. This effect can be confirmed as the stability value increases. For instance, changing the stability value from 0.2 to 0.9 produces a decrease of 11.7% and 6.8% of the F-Measure values for the 1% and 0.5% lower threshold experiments. Note that with a configuration of a 1% lower threshold and a 0.2 stability value we have improved the best RepLab 2013 result for this task.



**Figure 4.10** – F-Measure values varying the stability values from 0.2 to 0.9

Figure 4.11 shows a boxplot with the distribution of the 0.5%, 1% and 5% lower threshold samples for the different stability values. As happened with the analysis of the term selection strategy, the 1% approach shows less dispersed F-Measure values, which means a more homogenous behaviour. Finally, the stability value, as happens with the lower threshold value, is directly correlated with the Sensitivity value and inversely correlated with that of Reliability. This correlation is due to the reduction in the number of selected clusters/topics as the stability value increases. Not too many topics mean larger topics (i.e. topics with more tweets inside). Hence, the reliability value will be lower given that, more tweets will be put together, some of them unrelated. Inversely, putting more tweets together will increase the sensitivity value. In this sense, the best results are obtained

with the lowest stability values, that is, when smaller but more precise clusters are selected.

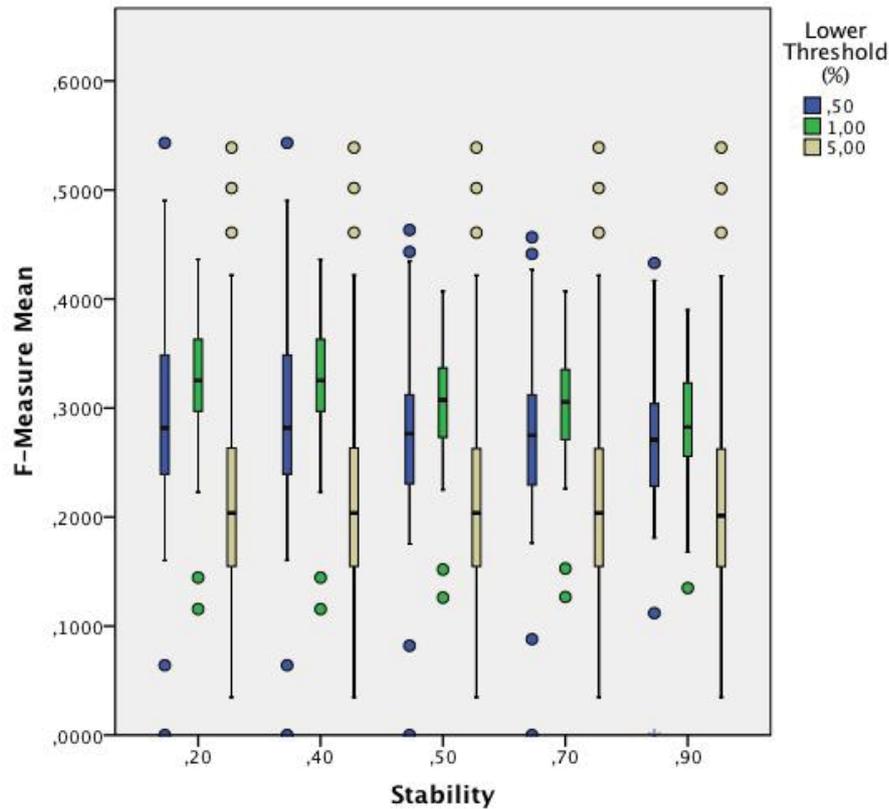


Figure 4.11 – Distribution of F-Measure values in terms of stability values

### Cluster Selection for Topic Annotation

The FCA nature and its generalization specialization relationship make it possible for an object to belong to several formal concepts at the same time. In the context of the Topic Detection task, this means that a tweet can belong to different topics. Although this feature makes FCA a powerful theory, in our particular case it was disadvantage, because the RepLab 2013 ground truth only annotated one tweet per topic, not allowing multi-topic classification.

To 1) overcome this problem and 2) to experiment with other methods of including a tweet in some of its possible topics, we experimented with different ways of selecting the candidate cluster:

- **Most Specific Concept Selection.** If any tweet appears in several formal concepts, it will be classified in its most specific cluster. This means that the tweet will belong to the most specialized topic and, as a consequence, to the smallest topic. This is the approach followed for the experiments presented until

now. The intuition behind this approach is that the most specialized clusters represent more specific and interesting topics (i.e. top clusters are no more than a joining of many tweets only related to each other based on general topics, such as “sports”, instead of “soccer” and “basketball”). It has to be taken into account that unstable clusters, which are also at the bottom of the lattice, have been previously removed in accordance with the stability-based concept selection strategy.

- **Most Stable Formal Concept Selection.** The previous selection method does not take into account any information on the clusters themselves. In this regard, it is not always the most specific clusters that will be the most suitable ones. This approach looks for those concepts with a higher stability value to be selected as topics. If some tweet appears in several formal concepts, it will be included in the concept (i.e. topic) with a higher stability value.
- **Multiple Selection.** In spite of selecting only one formal concept to be assigned as topic for the tweet, if some tweet appears in several formal concepts it will be included in all of them. With this approach, we allow some tweet to belong to different topics at the same time. By following this approach, the evaluation algorithm of the RepLab 2013 will penalize those badly chosen topics but we will be able to assure the success if one of the selected topics belongs to the ground truth.

The results obtained for each of these approaches are shown in [Table 4.15](#). According to these results, the best performance is obtained by the **Most Specific Concept Selection**. Although, the approach that allows tweets to be annotated with several concepts/topics obtains almost the same performance. The better performance of these approaches is due to the increase in the Reliability (precision-based) value; compared with the **Most Stable Concept Selection** approach. This behaviour could be explained because the **Most Stable Concept Selection** approach tends to classify tweets in bigger (and less precise) clusters. According to these results, the **Multiple Selection** and the **Most Specific Selection** approaches seem to be the most suitable ones, according to F measure-based values. Nevertheless, it is important to highlight that the **Multiple Selection** approach is also interesting not only for its results, but for the fact that allows tweets to belong to several topics. In some scenarios, this feature could be interesting, given that some content could involve different topics (e.g. some news report talking about politics, economics, and commerce).

Run	Reliability	Sensitivity	F(R,S)
Multiple selection	0,4058	0,3177	0,3256
Most Stable Concept Selection	0,2690	0,4374	0,3001
Most Specific Concept Selection	0,4090	0,3163	<b>0,3258</b>

**Table 4.15** – Reliability, Sensitivity and F-Measure values for different cluster selection setups

### Topic Adaptation

One possible advantage of the application of FCA over the classic clustering techniques is the topic adaptation of the technique. That is, Can FCA facilitate the detection of new topics based on information from past topics? To answer this question we have experimented with the performance of FCA to detect topics already seen in the training set and new topics that appear in the test set of the RepLab 2013 corpus. [Table 4.16](#) shows the results of these experiments, grouped by lower and upper threshold values and with a stability value set at 20%. For each group, the table shows the Reliability, Sensitivity and F-Measure values considering the full topic set (i.e., all), only the topics that have already appeared in the training set (i.e. seen) and, finally, considering only the topics which have not appeared in the training set (i.e. unseen).

The results confirm our intuition that FCA is a technique suitable for topic adaptation. For the different threshold configurations, FCA has almost the same performance for both the seen and unseen topics. The higher differences occur in the 5% threshold runs, probably due to the low amount of attributes (information). This represents one of the main advantages that we have postulated about our approach as compared to the clustering approaches, which are not able to take the previous knowledge in the training set into account as our FCA-based approach does.

One final remark is the overall Sensitivity values for the unseen topics, which are higher than for those already seen in the training set (and consequently the Reliability is lower). It seems to be counterintuitive; however, it can be explained by the absence of information in the unseen topics. As these topics have not appeared before, the system has not as much information as it has about the seen topics. This issue leads to bigger clusters, less precise (Reliability) but with a higher coverage (Sensitivity), as it also happens with the 5% approaches compared to the 1% (see previous section).

Run	Topics	Reliability	Sensitivity	F(R,S)
20% stab, 0.5-25% threshold	All	0.3836	0.2412	0.2710
20% stab, 0.5-25% threshold	Seen	0.6151	0.2115	<b>0.3014</b>
20% stab, 0.5-25% threshold	Unseen	0.4379	0.2616	0.2982
20% stab, 1.0-25% threshold	All	0.4090	0.3163	0.3258
20% stab, 1.0-25% threshold	Seen	0.5764	0.3018	<b>0.3730</b>
20% stab, 1.0-25% threshold	Unseen	0.4504	0.3302	0.3447
20% stab, 5.0-25% threshold	All	0.1696	0.6744	0.2250
20% stab, 5.0-25% threshold	Seen	0.3605	0.6735	<b>0.4062</b>
20% stab, 5.0-25% threshold	Unseen	0.1824	0.6796	0.2362

**Table 4.16** – Reliability, Sensitivity and F-Measure values for different topic adaptation experiments

#### 4.1.5.3. Comparison to other Topic Annotation approaches

This section presents the experimental comparison of our FCA-based approach to the other two aforementioned topic detection proposals — HAC and LDA — presented at section 4.1.4. This comparison is carried out in terms of internal and external evaluation, by applying the evaluation configuration presented at section 4.1.2. An example of the topics generated by each of the three approaches is shown in Table 4.17, which exemplifies some differences between the approaches.

HAC-based clustering is highly reliant on the exact similarity between terms. In other words, topics group together tweets whose textual content is quite similar. For instance, Topic 24 includes almost equal tweets about a new functionality of BMW cars (LTE Hotspot). Topic 133 and Topic 92 (about Bank of America) are more diverse, but still include quite similar tweets. Topic 133 is an interesting example since it takes advantage of the patterns for publishing tweets — tweets selling BMW cars in this particular case — to detect the topic. This behaviour can be adapted by reducing the similarity threshold; however, as is going to be shown late in this section, lower threshold values mean worse performance.

On the other hand, LDA-based topics include tweets, which at first sight are more textually diverse. The way in which LDA detects the topics (based on a latent representation space) allows the inclusion of textually different tweets. It enables the detection of more abstract topics than those detected by HAC. For instance, Topic 23 in the table includes tweets with information about releases of new BMW models, even though some of these tweets barely share terminology.

Approach	Entity	Tweet	Topic
HAC		BMW's Car Hotspot LTE means Bavarian Motor WiFi Mobilehttp://zite.to/V1TEuZ	
		BMW introduces LTE Car Hotspot with NFC functionality http://goo.gl/QM3r2	24
	BMW	TESTNEWS 1: BMW's Car Hotspot LTE means Bavarian Motor WiFi - The dedicated car phone may be a thing of the past,... http://fb.me/1mjwEzdmX	
		BMW's Car Hotspot LTE means Bavarian Motor WiFi: The dedicated car phone may be a thing of the past, but the i... http://engt.co/105XWYWM	
	Thomas A. is selling a 2002 BMW 540i in Dallas, TX with a Standard Ad.		
	Lauren B. is selling a 2009 BMW X5 in Port Arthur, TX with a Standard Ad.		133
	Shawn W. is selling a 2005 BMW X5 in Parker, TX with a Standard Ad.		
	<b>Bank of America</b> to add mobile payment service: Bank of America will jump on the mobile payments bandwagon soon with a service... http://dlvr.it/2TyZGr		92
	<b>America</b> Bank of America releases a mobile payment service to compete with Square by @tched		
	Watching 2011 Bmw X3 M Sport @ 2010 Paris Motor Show on http://www.sportsphotoart.com/2011-bmw-x3-m...		
	BMW 1 Series Coupe and Convertible Lifestyle Edition for 2013 Detroit Motor Show		
LDA	BMW	Auto: BMW Will Have Wi-Fi Built Into Cars: This Is HOT - http://tinyurl.com/br8c2of #IFWT	23
		BMW's Car Hotspot LTE means Bavarian Motor WiFi: The dedicated car phone may be a thing of the past, but the i... http://engt.co/105XWYWM	
		TESTNEWS 1: BMW's Car Hotspot LTE means Bavarian Motor WiFi - The dedicated car phone may be a thing of the past,... http://fb.me/1mjwEzdmX	
		BMW's Car Hotspot LTE means Bavarian Motor WiFi (bah... just get an iPhone and turn on 'personal hotspot') http://ip.it...	
FCA	BMW	BMW's Car Hotspot LTE means Bavarian Motor WiFi: The dedicated car phone may be a thing of the past, but the i... http://engt.co...	[BMW car, LTE, motor hotspot]
		BMW Car Hotspot LTE, un paso adelante en el uso de internet en el coche #motor #tecnologia @BMWESpana	
		TESTNEWS 1: BMW's Car Hotspot LTE means Bavarian Motor WiFi - The dedicated car phone may be a thing of the past,... http://fb.me...	
		BMW's Car Hotspot LTE means Bavarian Motor WiFi" good read	
	BMW's Car Hotspot LTE means Bavarian Motor WiFi http://engt.co/ThbuLR		
	Bank of America to get into mobile payments (bizjournals): Share With Friends:    Industry - Legal Servi... http://bit.ly/W63eSm #law		
	Bank of America Vs. Square - The Charlotte, N.C.-based institution became the first bank to offer its own mobile car... http://ow.ly/2tbQEQ		[bank, america mobile]
	Bank of America trials mobile QR code payments with Paydiant http://goo.gl/C7LnX		
	Bank Of America Spurns NFC For Image-Based Mobile Payment System, Now In		
	Testing http://ow.ly/e513I Do you think it's better than NFC?		

Table 4.17 – Examples of topics generated by the different approaches

Finally, FCA topics are also based on textual similarities. However, FCA generates topics that are more textually diverse and abstract than HAC, although more similar than LDA topics. For instance Topic [bmw, car, LTE, motor, hotspot] can be likened to HAC Topic 24 and Topic [bank, america, mobile] to HAC Topic 92. Nevertheless, FCA topics include more diverse tweets, also related to the topic, which are not included in the HAC topic (e.g., *BMW's Car Hotspot LTE means Bavarian Motor WiFi (bah... just get an iPhone and turn on 'personal hotspot') http://flip.it... or Bank of America to get into mobile payments (bizjournals): Share With Friends: Industry - Legal Servi... http://bit.ly/W63eSm \#law*). Another advantage of FCA with respect to LDA and HAC is also shown in Table 4.17: the automatic labelling of the topics with the terms that have been used to generate the topic (instead of a number ID), facilitating the understanding of its content.

These differences have some implications in the quality of the generated topics and their performance in the topic detection task, as is going to be seen in the following subsections.

### Internal Evaluation Results

Table 4.18 to Table 4.20 show the results obtained by HAC, FCA and LDA approaches according to the evaluation proposed in section 4.1.2.2. HAC and FCA tables also show the average number of clusters per entity and the average cluster size. As the number of clusters is the LDA threshold, this information is not included in the Table 4.20.

Similarity Threshold	Dunn Index	Davies-Bouldin Index	Silhouette Coefficient	Calinski Harabasz Index	Avg. Clusters	Avg. Cluster Size
0.1	0.0364	12.5626	-7.0057	0.0120	1100.48	7.8
0.2	0.0399	12.6760	-2.8023	0.0178	1075.85	8
0.3	0.0675	13.5150	-6.2656	0.0273	1046.18	8.2
0.4	0.0516	8.2813	-9.7287	0.0396	990.39	8.7
0.5	0.0525	8.2722	0.0015	0.0550	904.21	9.5
0.6	0.0542	8.2178	-0.0021	0.0680	814.49	10.5
0.7	0.0421	8.1700	-0.00354	0.0866	633.72	13.5
0.8	0.0361	8.0370	-0.0052	0.1171	293.3	29.2
0.9	0.0505	7.8201	-0.0017	0.1747	21.62	396.2

Table 4.18 – HAC-based Internal Results

Stability Threshold	Dunn Index	Davies-Bouldin Index	Silhouette Coefficient	Calinski-Harabasz Index	Avg. Clusters	Avg. Cluster Size
<b>0.1</b>	0.0389	2.2467	-0.0135	0.1467	514	17
<b>0.2</b>	0.0299	2.9158	-0.0116	0.1885	434.93	19
<b>0.3</b>	0.0299	2.9158	-0.0116	0.1885	447.87	19
<b>0.4</b>	0.0299	2.9158	-0.0116	0.1885	447.87	19
<b>0.5</b>	0.0205	3.7707	-0.0136	0.1940	447.87	19
<b>0.6</b>	0.0215	3.8549	-0.0133	0.2031	212.67	41
<b>0.7</b>	0.0215	3.8549	-0.0133	0.2031	201.57	43
<b>0.8</b>	0.0190	4.6852	-0.0304	0.2513	191.44	45
<b>0.9</b>	0.0190	4.6852	-0.0304	0.2513	135.97	63

Table 4.19 – FCA-based Internal Results

Number of Topics	Dunn Index	Davies-Bouldin Index	Silhouette Coefficient	Calinski-Harabasz Index
<b>10</b>	0.0002	29.2826	-0.0186	0.6406
<b>30</b>	0.0003	24.2439	-0.0200	0.4995
<b>50</b>	0.0003	24.1787	-0.0237	0.4609
<b>70</b>	0.0003	22.4080	-0.0395	0.4194
<b>90</b>	0.0004	23.0144	-0.0476	0.3945
<b>100</b>	0.0004	22.4877	-0.0526	0.3834
<b>200</b>	0.0005	20.9919	-0.0928	0.3132
<b>400</b>	0.0005	29.6324	-0.1733	0.2542
<b>500</b>	0.0006	38.3605	-0.2179	0.2391

Table 4.20 – LDA-based Internal Results

As a general comment, FCA offers the best results for most of the configurations (i.e. note that for the Davies-Bouldin Index a lower value means a better performance). The only exceptions are some values achieved by the HAC approach in terms of the Dunn Index and the LDA results for the Calinski-Harabasz Index.

As regards the HAC results, the best Silhouette Coefficient results are obtained for low values of the similarity threshold, when a lot of clusters are generated. As said before (see section 4.1.2.2), this index is sensitive to the presence of subclusters, which is the case when a low similarity threshold is selected and many clusters are generated. Consequently, these values may be due to this aspect and not to a well-performing clustering. In contrast, the best Silhouette Coefficient results are achieved by the FCA-based clustering for the stability values that also achieve the best results in the external evaluation (see section 4.1.2.3). The results of this index seem to indicate that adapting

the stability threshold not only affects the granularity of the generated clustering, but also the type of the generated clusters. In contrast to what happens with HAC Silhouette Coefficient results, which are correlated to the number of clusters, FCA results are related to the accuracy of the clustering (in terms of the external evaluation results). The best results are achieved, not when stability generates more clusters, but when stability generates clusters with the degree of cohesiveness required by the task. Consequently, in spite of HAC offering the best results according to this index, FCA appears to generate more interesting results.

Considering the Dunn Index results, HAC does offer better results than FCA for some threshold values, while for other values FCA results are better. One of the main drawbacks of the Dunn Index is that it is highly affected by noisy data. Therefore, the random-like HAC results (i.e., there is no a clear tendency in the results) in terms of this index might be related to this issue and, consequently, they are not given any interpretable information about HAC performance. In contrast, FCA does not present this behaviour. FCA applies the stability to generate the clustering, so the final results are based on the cohesiveness of the clustering. On the other hand, the HAC results are based on the granularity defined by the similarity threshold. While it may affect the cohesiveness of the clustering, it seems that the selection based on the stability of the *formal concepts* better captures this aspect than the HAC similarity threshold.

In addition, the FCA performance is more homogeneous (stable) than the other two approaches (see [Figure 4.12](#) to [Figure 4.15](#)) for all the measures and threshold values. In other words, the FCA-based approach provides a model less dependent on a specific parametrization. In our view, this is a profitable behaviour: you would prefer a stable topic detection approach (in terms of the quality of the clustering) that is not affected by the granularity of the clustering. An approach only performing well for some specific setups is not suitable for real environments where the specific details required to configure the algorithms are not known (e.g., granularity of the clustering, topic cardinality, topic distribution...).

Another interesting aspect to highlight is that for both HAC and LDA the internal evaluation results directly correlate with the granularity of the clustering (the more specific, the better), but for the FCA-based results this correlation is not so clear. We have already pointed out this aspect for the Silhouette Coefficient results. Equally, the Davies-Bouldin Index also achieves the best performance for the stability values of around

0.1 and 0.4 (for the values with the best performance in the external evaluation, see section External Evaluation).

The internal measures promote well-separated and cohesive clusters. Consequently, it does make sense that whereas the more specific the clustering, the better the internal results: a more specific clustering representation creates smaller and, in theory, more cohesive clusters. However, why is it different for some FCA values? To answer this question we have to remind ourselves that FCA clusters are selected by means of their stability value. A higher stability leads to a more generic clustering, given that fewer clusters are selected, but it also implies that the selected clusters are more cohesive, in terms of the definition of stability (see the formulation in section 3.2). Therefore, as the internal evaluation looks for cohesive clusters, a more stable clustering could also lead to a better clustering in terms of this evaluation, although it generates bigger clusters. It could be said that a more stable *formal concept* (as defined in this paper) might better represent a topic/cluster. However, this point cannot be fully confirmed for all the results, so no general conclusion can be obtained.

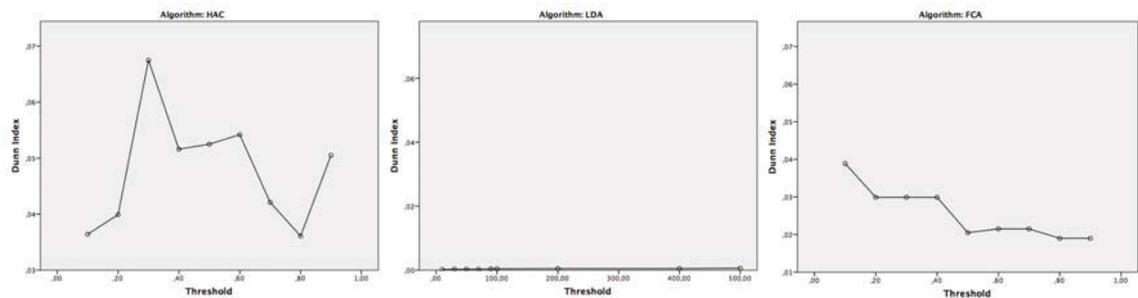


Figure 4.12 – Comparison of the Dunn Index results for HAC, LDA and FCA

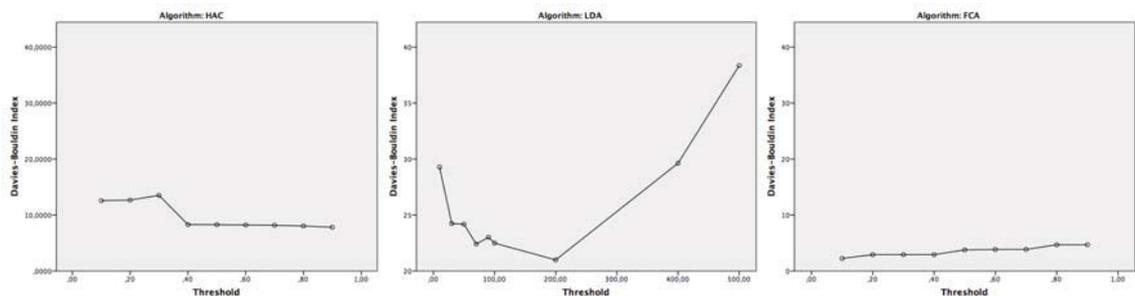
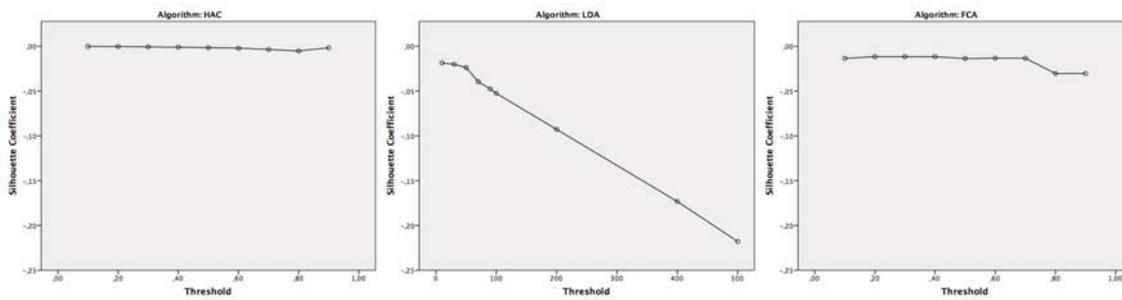
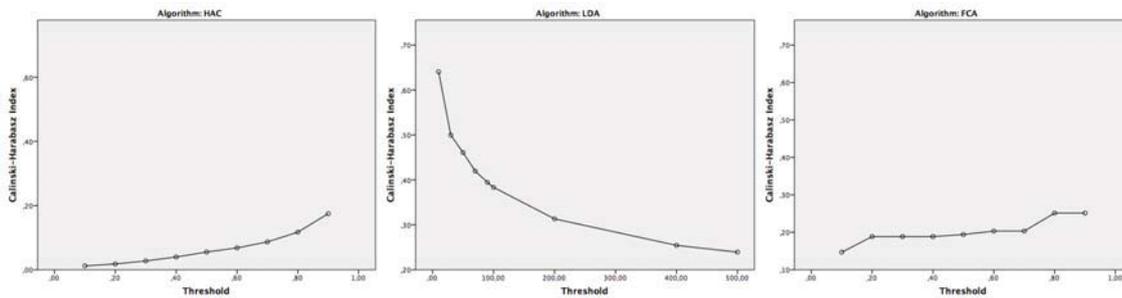


Figure 4.13 – Comparison of the Davies-Bouldin Index results for HAC, LDA and FCA



**Figure 4.14** – Comparison of the Silhouette Coefficient results for HAC, LDA and FCA



**Figure 4.15** – Comparison of the Calinski-Harabasz Index results for HAC, LDA and FCA

To sum up, this evaluation has proven FCA as a more suitable and reliable approach. FCA offers a high and stable performance for all the possible configurations in all the different metrics, outperforming the results of the other two approaches in most of the cases. The evaluation also showed that our topic selection strategy, based on the stability of the *formal concepts*, generates a more cohesive clustering, which is a desired feature for the topic representation, at least in this task (see section 4.1.1.1).

However, the cohesiveness of the clusters is not the only criteria to be taken into account. We seek to apply our proposal to a specific task: Topic Detection (see section 4.1). Creating cohesive clusters only meets one of the desired criteria of the topic detection algorithms. In fact, good results in this internal evaluation may not represent good results in terms of the performance for the task. In this sense, Table 4.21 shows that the RepLab Topic Detection gold standard does not achieve the best results in terms of this internal evaluation. If we only take cohesiveness into account, we can propose other baseline approaches focussing on this aspect. For instance, Table 4.21 presents the internal

evaluation results for the One-In-One (one cluster per tweet) and the Duplicated Tweets (One-In-One but considering duplicated tweets) approaches, which outperform the gold standard results for most of the measures. Nevertheless, these baseline approaches are rather useless at solving the topic detection task. In fact, as it is going to be proven in the next section, they offer a poor performance when applied to the task.

Approach	Dunn Index	Davies-Bouldin Index	Silhouette Coefficient	Calinski-Harabasz Index
Replab Gold Standard	0,0224	6,4707	-0,0357	0,3338
Duplicated Tweets	0,0125	4,7824	-0.002	0,0058
One-In-One	0,0048	17,4000	-0,0138	0,2117

**Table 4.21** – Baseline Internal Results

Finally, in order to test whether the good FCA performance in terms of cluster quality also results in a good topic detection approach in a real scenario, the following section presents the comparison of our approach to the other two approaches in terms of the external evaluation proposed by the RepLab Evaluation Campaign.

### External Evaluation Results

According to the internal evaluation presented above, FCA has demonstrated that it creates quality and cohesive clusters. In addition, FCA was less affected by the threshold configuration and offered more stable results than HAC and LDA. The aim of this section is to study how this cluster quality and cohesiveness are related to an accurate topic representation for the specific RepLab Topic Detection task.

The [Table 4.23](#), [Table 4.24](#) and [Table 4.25](#) and the [Figure 4.16](#), [Figure 4.17](#) and [Figure 4.18](#) show the results of HAC, FCA and LDA. The results are obtained by using the official RepLab 2013 dataset and evaluation tool. The results are expressed according to the different threshold configurations: stability for FCA, cluster similarity for HAC and number of Topics for LDA (i.e., the K value for the number of LDA topics tries to cover the range of feasible values for the number of topics per entity in the RepLab dataset, see [Table 4.2](#)). [Table 4.22](#) also includes the performance for different baselines in order to frame the results of the proposed systems and configurations:

- **One-In-One** clustering where one cluster is generated for each tweet
- **All-In-One** where all the tweets are included in a unique cluster

- **Duplicated Tweets** that are similar to One-In-One but including duplicated tweets (e.g., retweets) in the same cluster
- **Best Replab** that is the best-performing approach at Replab 2013 (in terms of F-measure).

Approach	Reliability	Sensitivity	F(R,S)
<b>Duplicated Tweets</b>	0.989	0.1	0.1712
<b>All In One</b>	0.0678	1,0000	0.121
<b>One-In-One</b>	1	0.386	0.0693
<b>Best RepLab</b>	0.4624	0.3246	0.3252

**Table 4.22** – Baseline External Results

The first point to remark is that, for some configurations, HAC results outperform FCA results and especially LDA results. Beyond the specific threshold values for these configurations, which are dependent on the collection, the application of HAC with Jaccard Distance as similarity measure offers the top-performing results for this task. These results confirm those obtained by other works, already mentioned in the state-of-the-art in section 2.2. As regards FCA, this evaluation also confirms its high performance, in the same level as the best results obtained for the task (see Table 4.24). In contrast, the average LDA performance is far cry from that of HAC and FCA.

A general behaviour for both systems (HAC and LDA) is that the best results are obtained by those configurations that maximize the number of clusters; that is, that generate more specific clusters. In the case of HAC, a higher similarity threshold means less number of cluster merging, given that only those which similarity value are higher than the threshold will be merged. Meanwhile, in FCA, a smaller stability value means a great number of formal concepts selected as topics: only those formal concepts with a similarity value lower than the threshold will not be considered as topics.

If we take a closer look at the FCA-based results, they are much more homogeneous than the HAC-based ones (i.e., they are not as dependent on the threshold, performing similarly for all the configurations).

The LDA-based results also present stable values; however, the performance of LDA is significantly lower than that of the FCA and HAC. Therefore, the FCA approach does not suffer from the granularity of the representation as HAC does, maintaining a good overall performance. This is a desirable characteristic of a topic detection system and a symptom of a good topic representation.

These external results go in the same direction as those in the internal evaluation. The high system performance and the homogeneous values of our FCA approach remain in accordance with both evaluations.

Therefore, we may conclude that FCA provides a high quality topic detection based on cohesive clusters, which also provides accurate results applied to the specific RepLab task. In contrast, the variance in HAC results seems to be related to a higher adaptation to the gold standard.

Similarity Threshold	Reliability	Sensitivity	F(R,S)
<b>0.1</b>	<b>0.9940</b>	0.1119	0.1912
<b>0.2</b>	0.9910	0.1170	0.1995
<b>0.3</b>	0.9830	0.1237	0.2098
<b>0.4</b>	0.9600	0.1403	0.2342
<b>0.5</b>	0.9130	0.1696	0.2710
<b>0.6</b>	0.8620	0.2079	0.3095
<b>0.7</b>	0.7510	0.3184	0.3707
<b>0.8</b>	0.3820	0.6044	<b>0.4072</b>
<b>0.9</b>	0.0930	<b>0.9612</b>	0.1632
<b>Best RepLab</b>	<b>0.4624</b>	<b>0.3246</b>	<b>0.3252</b>

Table 4.23 – HAC Results

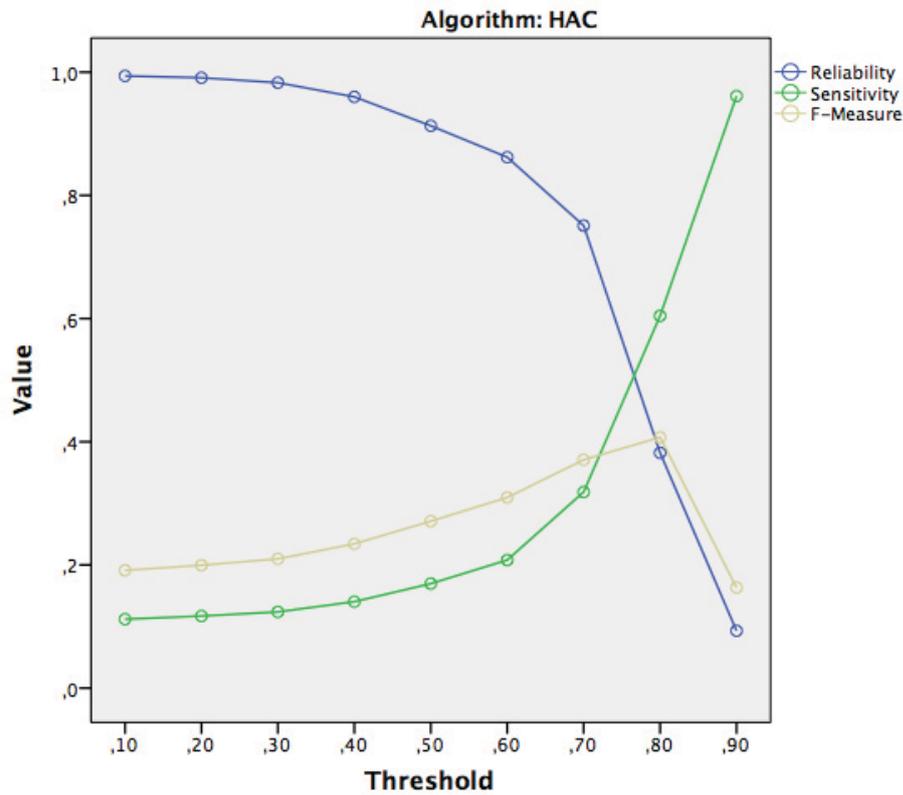


Figure 4.16 – HAC-based results for the different Similarity Thresholds.

Stability Threshold	Reliability	Sensitivity	F(R,S)
0.1	<b>0.5070</b>	0.2494	0.3208
0.2	0,4090	0,3163	<b>0,3258</b>
0.3	0,4090	0,3163	<b>0,3258</b>
0.4	0,4090	0,3163	<b>0,3258</b>
0.5	0,3455	0,3228	0,3041
0.6	0.3407	0.3236	0.3027
0.7	0,3407	0,3236	0,3027
0.8	0,3029	<b>0,3324</b>	0,2878
0.9	0,3029	<b>0,3324</b>	0,2878
<b>Best RepLab</b>	<b>0.4624</b>	<b>0.3246</b>	<b>0.3252</b>

Table 4.24 – FCA-based Results

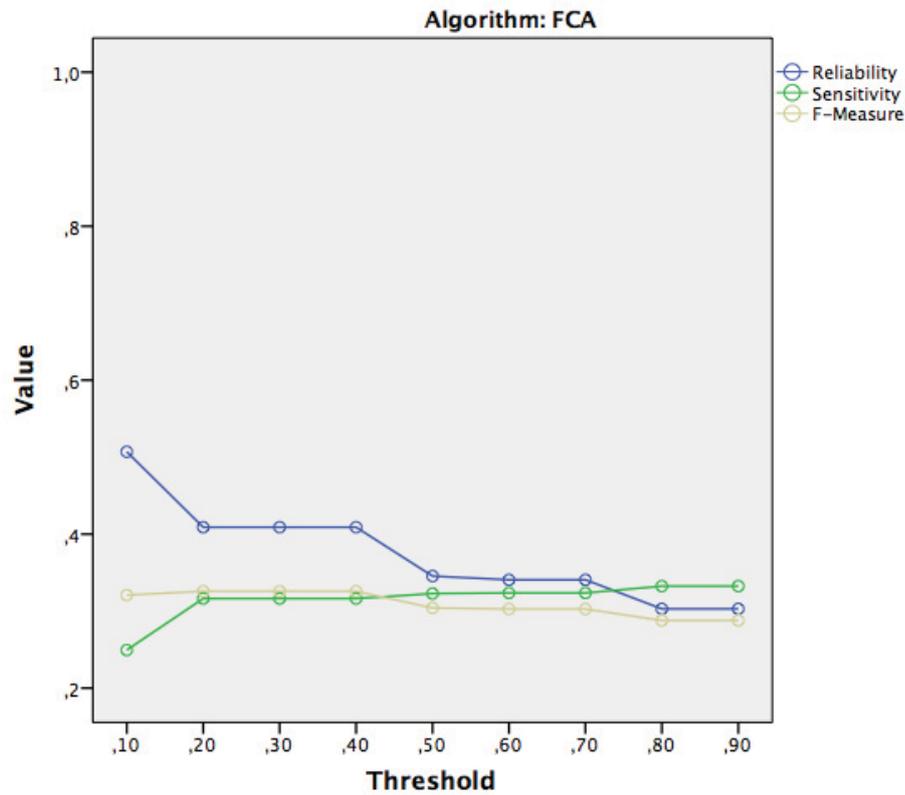


Figure 4.17 – FCA-based results for the different Stability Thresholds.

Number of Topics	Reliability	Sensitivity	F(R,S)
10	0.4470	<b>0.2108</b>	<b>0.1472</b>
30	0.4710	0.1424	0.1372
50	0.4860	0.1199	0.1307
70	0.5030	0.1131	0.1323
90	0.5160	0.1068	0.1307
100	0.5220	0.1053	0.1309
200	0.5630	0.0916	0.1275
400	0.6210	0.0834	0.1259
500	<b>0.6370</b>	0.0811	0.1236
<b>Best RepLab</b>	<b>0.4624</b>	<b>0.3246</b>	<b>0.3252</b>

Table 4.25 – LDA-based Results

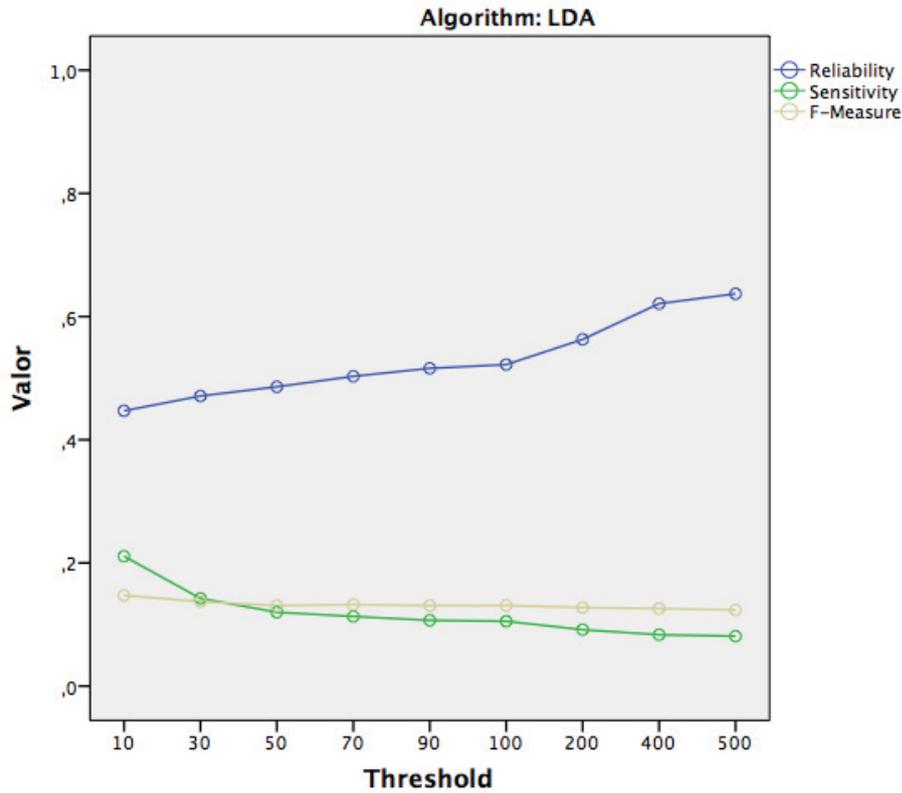


Figure 4.18 – LDA-based results for the different configurations.

#### 4.1.6. Discussion

This section presented the detection of thematically similar topics in streams of Twitter data. Instead of using clustering, classification or probabilistic techniques, mainly applied in the state of the art, we have proposed an approach based on Formal Concept Analysis (FCA).

FCA have demonstrated to be able to deal with the main problems related to the topic detection process: an unknown number of topics, the need to capture the implicit hierarchy of the topics, topic adaptation or feature selection. These limitations imposed by the task are shared by the recommendation task. In topic detection, the number of topics is not known a priori, neither the number of user-item groups (i.e., user preferences) in recommendation. In the same way than topics, user preferences have also an implicit hierarchical structure. In consequence, any recommendation model should be able to deal with it. User preferences have to be adapted in presence of new content, as topics should

also have. Therefore, the conclusions drawn from this experimentation are also valid in the recommendation context.

The approach has been tested within the scope of the RepLab 2013 Campaign in order to elaborate an evaluation framework based on real data and a real scenario. An extensive analysis of the results has been carried out. The objective of this analysis is to study the correlation between the main features related to the FCA computation; for instance, stability values, attribute reduction algorithm or the topic selection method. In this analysis, other aspects related to the task has been also addressed, such as the adaptation of the topic-detection approach to the appearance of new topics or the impact of the tweet-filtering process on the overall system performance.

This analysis draws the following conclusions. First of all, the direct correlation between information and the precision of the system; that is, the more information the system, the more precise. This correlation can be seen in the variation of the results based on the stability value or the threshold of the attribute reduction algorithm. Related to this, the following conclusion is that not only is it important to have more information but also “quality” information.

The system adaptability to the appearance of new topics can be also highlighted. In the experimental results, it has been demonstrate as our FCA-based approach, unlike clustering approaches, is able to integrate previous knowledge about the prior topics (contained in the training set) without losing the ability to detect unseen topics.

Other important remark is the high general performance of the proposed approach, improving the official best result of RepLab 2013. These results, at the top of the state of the art, confirm FCA as a suitable technique for Topic Detection, improving the performance of traditional classification, clustering or probabilistic approaches. Related to these results, it must be highlighted that they were obtained when the Reliability and Sensitivity values tend to be equal. It means that the approaches that result in a system extreme behaviour do not achieve good general results. This is due to a high Reliability (or Sensitivity) value causing an extremely low Sensitivity (or Reliability) value, damaging the overall performance.

Some further researches had proven that some configuration of HAC and LDA algorithms obtain very successful results for this task, ever better than the ones obtained to our approach. However, to our view, those results are only telling a side of the truth because, while it is clear that a higher-performing system is preferable, some aspects are not covered by traditional (external) evaluations. To have a clear picture about how to define

the quality of a topic detection approach, we also presented a further analysis, based on an internal evaluation applying clustering quality metrics.

In this sense, the main hypothesis was that the FCA performance for content organization might be successfully applied to a topic detection task. To test that claim, it has been presented the experimental results obtained by our FCA-based proposal in comparison to other state-of-the-art techniques (HAC and LDA) in terms of the quality of the generated topics and their performance in their application in a real scenario: the Topic Detection Task at RepLab 2013.

Table 4.26 shows a summary of the results by technique, as well as their comparison as regards different aspects. In particular, this table shows the results for the internal evaluation metrics (shown as [min\_value, average  $\pm$  std\_deviation, max\_value] for each metric), the external evaluation metrics (shown as [min\_value, average  $\pm$  std\_deviation, max\_value] for each metric), the complexity of the different techniques, and the need to predefine the number of topics a priori and the ability to handle new topics, unseen in the training process. Focusing on the internal evaluation, the results offered by the different indexes confirm our hypothesis that FCA provides a clustering representation with more *quality* and with more stable behaviour than HAC and LDA. If we consider the external evaluation, FCA also appears to be preferable to LDA and HAC, even though this latter improves the FCA performance for some configurations. FCA, in contrast to HAC, demonstrates a more homogeneous performance, obtaining high results no matter what the configuration might be, as the internal evaluation has already pointed out. All methodologies suffer from high complexity, although it is FCA which presents the worse-case scenario. Finally, only HAC and FCA are able to detect topics unseen in the training process.

To sum up, FCA appears as a more suitable approach for topic detection because: 1) FCA creates quality and cohesive topics, a desired aspect for the proposed scenario and, 2) the quality of the clusters results in an accurate topic detection proposal in terms of the Replab Evaluation.

In addition, unlike HAC and LDA, FCA does not need to settle a series of parameters in order to compute the final topic detection results, given its homogeneous behaviour along the different configurations. It is a desired characteristic in a real production environment where no annotations for configuring the topic detection systems are available.

	HAC		LDA		FCA	
<b>Internal Evaluation</b>						
Dunn Index	[0.036, 0.05 ± 0.01, 0.068]	[0.0002, 0.0004 ± 0.0001, 0.006]	[0.019, 0.026 ± 0.0001, 0.039]			
Davies-Bouldin	[7.8, 9.73 ± 2.27, 13.55]	[20.99, 26.07 ± 5.19, 38.36]	[2.25, 3.54 ± 0.8, 4.69]			
Silhouette Coefficient	[-0.005, -0.002 ± 0.002, -0.0001]	[-0.22, -0.076 ± 0.068, -0.019]	[-0.03, -0.017 ± 0.007, -0.01]			
Calinski-Harabasz	[0.01, 0.066 ± 0.049, 0.17]	[0.24, 0.4 ± 0.119, 0.64]	[0.15, 0.20 ± 0.03, 0.0389]			
<b>External Evaluation</b>						
Reliability	[0.093, 0.77 ± 0.3, 0.994]	[0.447, 0.53 ± 0.062, 0.64]	[0.3, 0.37 ± 0.062, 0.51]			
Sensitivity	[0.119, 0.315 ± 0.3, 0.9612]	[0.08, 0.117 ± 0.037, 0.21]	[0.25, 0.315 ± 0.02, 0.33]			
F-measure	[0.19, 0.26 ± 0.08, 0.407]	[0.12, 0.13 ± 0.0065, 0.147]	[0.29, 0.31 ± 0.01, 0.326]			
Complexity	$\mathcal{O}( N ^3)$	$\mathcal{O}( K  N  V )$	$\mathcal{O}( C  G  M ^2)$			
Predefine Topic Number	NO	YES	NO			
Handle new Topics	YES	NO	YES			

Table 4.26 – Summary of the comparison among HAC, FCA and LDA

---

## 4.2 Application Scenario 2: Image Diversification

---

Diversification refers to, given a user query, the creation of a result list that fulfils this query, which maximizes the diversity among the items. The rationale is that the users are not only interested in accurate results but also in results covering the different aspects related to them. This is especially interesting in the recommendation scenario. A recommendation list including diverse and relevant items is far more interesting than other list including quite related, but very similar items (e.g., a recommendation list including all the “Star Wars” movies). More on this regard can be found at the review of the diversity applied to recommendation systems in [Castells et al., 2014]. Other related issues about system summarizing results according to the different query aspects that are more likely to cover the users' information needs can be consulted in the work of [Ionescu et al., 2014b], which offers an overall view on the task. Finally, from the IR-based point of view, a good definition of the problem and a review of the state of the art can be found at [Agrawal et al., 2009].

Diversification problem is herein prompted by the context of the image retrieval systems. To tackle this task several approaches has been proposed in the state of the art [Rudinac et al., 2013; Taneva et al., 2010]. However, to our point of view none of them is able to grasp the gap between the ideal diverse result list (covering all the latent aspects in the knowledge domain related to this result list) and the actual result list. This gap is related to the impossibility of creating an accurate modelling of these latent aspects. This problem is also highly relevant in the recommendation task when addressed from a Content-based paradigm. Recommendation systems should capture the latent aspects in the content of the items to be recommended (e.g., topic in news reports, genres in movies) to describe the user preferences as a combination of this aspects (e.g., a user likes news reports about soccer and basketball). Finally, the recommendation list offered to the user should cover all the captured latent aspects by offering items related to them. In consequence, we think the application scenario presented in this section is useful to evaluate to what extent our FCA proposal is able to model these diverse aspects addressed by the contents (i.e., image captions in this case).

In more detail, to cope with the diversification problem, we apply FCA to create an image representation using the concept/s covered by the images. To that end, the textual

information related to the images caption is used to create a concept-based representation. FCA is applied to discover the latent topics addressed by the image set, trying to cover as much different topics as possible. We experimented with different kinds of data related to the images (e.g. raw-text description of the images, social tags, user information, date information...). We expect the proposed FCA-based representation will improve the diversification by making explicit the latent concepts in the images as well as the relationships between them.

Our proposal is tested in the Retrieving Diverse Social Images Task at the MediaEval international forum. This task provides an experimental test-bed to allow the comparison of diversification systems in a real scenario: the diversification of a list of Flickr images. This scenario is particularly interesting because it provides a social-based scenario (i.e., social annotations made by Flickr users) with similar requirements to that proposed for the recommendation scenario. In particular, we have made use of the experimental environment provided by the Retrieving Diverse Social Images Task at MediaEval 2014 [Ionescu et al., 2014a] and 2015 [Ionescu et al., 2015].

#### 4.2.1. Test bed: Retrieving Diverse Social Images Task at MediaEval

---

This task addresses the problem of result diversification in the context of a social photo retrieval system (Flickr<sup>11</sup>). To exemplify the problem addressed by this task, the organizers present the use case of a tourist trying to find more information about a place they are potentially visiting, but with only a vague idea about the location (e.g., name of the location). By means of the name of the location, this tourist aims to search for additional information, expecting to obtain a complete description of the place; that is, a list of images related to the location but covering the different aspects about the location.

More specifically, systems participating in the task are requested to; given a list of photos retrieved from Flickr ranked according to the Flickr's default "relevance" algorithm, diversify the results. To that end, the systems have to provide a set of 50 images that depict different views/aspects of the location at different times of the day/year and under

---

<sup>11</sup> <https://www.flickr.com/>

different weather conditions, creative views, etc. In what follows, the dataset offered by the task, as well as the provided evaluation environment is explained.

The novelty of the 2015 is the inclusion of multi-concept queries, which are related not only to locations but also to events associated with these locations, e.g., "Oktoberfest in Munich" or "Bucharest in winter".

#### 4.2.1.1. Dataset

Both datasets (2014 and 2015) are similarity organized. They consist of information on locations (e.g., Eiffel Tower, Palazzo delle Albere). For each location, it is provided a ranked list of photos (around 300 per location) retrieved from the Flickr search engine. Some general-purpose visual descriptors, text descriptors and user-annotated credibility descriptors are also provided for each image.

For the experimentation, the dataset is organized in a development dataset (devset) to be used for designing/tuning the methods and a testing dataset (testset) to be used for the final evaluation. In order to allow the evaluation of the proposed approaches, all the images in the dataset have been annotated in terms of relevance to the query and diversity. Expert annotators with advanced knowledge of the location characteristics have annotated the dataset. More information about the dataset, its annotation and the metadata associated to the images can be consulted in [\[Ionescu et al., 2014c\]](#).

The specific figures about both datasets can be seen in [Table 4.27](#). The locations in both dataset are different. As said before, the 2015 MediaEval dataset contains one-topic queries (i.e., the same kind of queries than those in the 2014) and multi-topic queries (i.e., new at MediaEval 2015).

	2014	2015
<b>Devset Queries</b>	30	153 (all one-concept)
<b>Testset Queries</b>	123	139 (69 one-concept — 70 multi-concept)
<b>Devset Images</b>	45,000	45,375
<b>Testset Images</b>	50,000	41,394

**Table 4.27** – Figures about MediaEval 2014 and 2015 Dataset

#### 4.2.1.2. Evaluation Setup

The system performance is assessed in terms of Cluster Recall at X (CR@X), which measures how many clusters from the ground truth are in the top X results, Precision at X (P@X), a measure for the number of relevant photos in the top X, and F1-measure at

X, defined as the harmonic mean of the previous measures. Several cut off points have been considered (from X=5 to X=50). However, in the official evaluation, the systems were ranked according to the F1-measure (F1@20). It tries to imitate the real context of a typical Web image search engine, where a user inspects only the first result page.

## 4.2.2. FCA-based Work Proposal

---

Our proposal is based on taking the information related to the images to be diversified (i.e., the original image result set coming from the Flickr retrieval system) to create a concept-based representation of them. This representation intends to discover latent concepts in the data and the relationships between them. To this purpose, we propose the application of Formal Concept Analysis to organize the image set to be diversified according to their latent concepts.

After the FCA application, a hierarchy organizing the images according to the latent concepts discovered (in the form of *formal concepts*) is obtained. It remains the diversification of the images according to these latent concepts. To that end, we apply a HAC algorithm to group together those formal concepts that could be considered as similar (belonging to the same aspect/topic). After this grouping, each HAC cluster may be considered as the image set related to a given topic/aspect and, therefore, the set of clusters may be considered as the different topics/aspects addressed by the images. The generation of the final diversified result list will be then based on selecting the best image in each cluster. In the following subsections, this process is explained in more detail.

### 4.2.2.1. Image Content Representation

Before the FCA application, an important previous step is related to the representation of the images by means of the set of features best describing them. It can be done by: 1) the visual features related to the images (colours, shapes, etc.) or, 2) the textual contents related to the images (description, title, tags, etc.). The approach herein proposed is only focused on the textual content of the images. In more detail, the content representation is based on identifying the most representative textual features of each image. To that end, Kullback Leibler Divergence (KLD) [Kullback and Leibler, 1951] is applied to identify the most representative features of each image. An example of this KLD-based representation for an image related to the Valencia's opera house is the Table 4.28, where each feature is weighted according to their KLD values. In order to select the most

interesting ones, for each image, the features only included in the first tercile (i.e., first 1/3) according to the KLD weights are selected to represent the image.

KLD Feature	KLD Value
house	0.5255
bezembinder	0.4408
opera	0.4269
turia	0.3116
valencia	0.2895
calatrava	0.2666
City-of-the-arts-and-sciences	0.1679
...	...

**Table 4.28** – KLD-based representation example

Different configurations to represent the images have been developed. These configurations are focused on testing the performance of different features related to the images that might be interesting to diversify the image set (e.g., textual features describing the images, date related features, geo-location features, or user-related information). More in detail, the different kinds of features proposed are:

- **Tags Info:** Tags related to the image. It can be seen as a description of the image in terms of keywords. (e.g., bridge, London, Tower Bridge, United Kingdom, River Thames, Suspension Bridge, City of London).
- **Description Info:** Detailed textual description about the image, provided by the user. It is presumably the field including more information. However, as it is written in natural language, the information is not as clearly expressed as it can be in the previous field.
- **All Text Info:** Both previous textual fields together and some other textual features like the title of the images. The idea is to combine all the information offered by the textual features.
- **Date Info:** Date in which the image was taken by the user. The date can be interesting in the diversification since final users can be interested in images of the location in different dates (e.g., day and night or winter and summer) or images of the location in a specific date (e.g., images of the Eiffel Tower in Christmas).
- **Geo Info:** Geolocation of the image; that is, where has the image been taken? Different geolocations might mean different views or perspectives of the location's images.

- **User Info:** User (username) who has taken the image. Different users are susceptible to offer different views of the same location. For instance, one could be more interested in specific details about the location.
- **All Info:** All the previous information about the image.

#### 4.2.2.2. FCA-based Modelling

The objective of this step is to create the concept lattice organizing the images according to their shared features in a hierarchical structure; being the images the objects and their related features the attributes of the formal context. The concept lattice organizes the images putting together, in *formal concepts*, the similar ones according to their features. Given that each group is based in the textual description of the images, each *formal concept* is susceptible to represent one topic or aspect related to the image set.

Figure 4.19 shows an example of the concept lattice for the images related to the Chrysler Building. An example of one of these *formal concepts* and the images it contains is shown in Figure 4.20.

The diversification might be conducted now by selecting an image from each of the obtained formal concepts. However, because of the nature of *formal concepts* and the way they are depicted in the lattice (i.e., two related formal concepts that are close in the lattice structure or a parent/child pair of formal concepts are susceptible to refer to similar aspects, see section 3.1 for more details) different formal concepts can be quite similar. This issue is reflected in the large number of formal concepts (as it can be appreciated in the Figure 4.19).

To address this issue, we propose the application of a HAC algorithm to cluster similar formal concepts, as described in the next section.

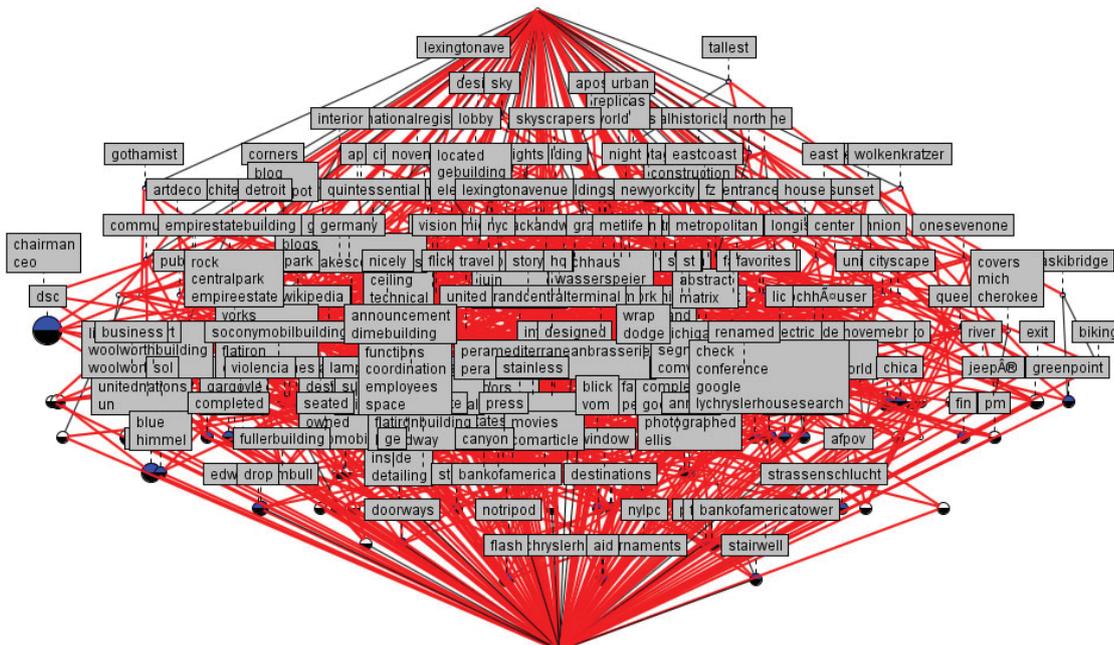


Figure 4.19 – Concept Lattice for the Chrysler Building images

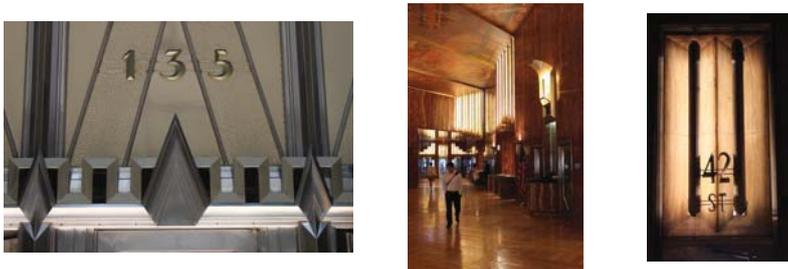


Figure 4.20 – Example of Formal Concept Images (with intension: ArtDeco, National Historical Landmark)

#### 4.2.2.3. HAC-based grouping

As a result of the FCA-based modelling, a hierarchy organizing the obtained set of formal concepts is obtained. In this kind of representation is easy to find similar formal concepts (those sharing a similar KLD feature set) but also the most different ones (those with a different KLD feature set). This aspect is easily visible in the lattice structure; the most similar formal concepts are close in the lattice while the most different ones are separated in the structure.

Nevertheless, it remains the creation of a set of diverse image groups based on this organization by applying a single linking HAC algorithm. This algorithm cluster similar image groups (i.e., *formal concepts*) together. In order to set the "similarity" of two *formal concepts*, the Zeros-Induces index has been applied [Alqadah and Bhatnagar, 2011]. The

Zeros-Induces index is based on the number of features that two *formal concepts* share. In particular, the zeros-induces index was defined in [Alqadah and Bhatnagar, 2011] as follows:

**Definition 6.** Given two *formal concepts*  $C_1 = (A_1, B_1)$ ,  $C_2 = (A_2, B_2)$  of a context  $\mathbb{K}$ , then the zeros induced by  $C_1$  and  $C_2$ , denoted by  $z(C_1, C_2)$ , is the number of zeros enclosed by the sub-matrix induced by rows  $(A_1 \cup A_2)$  and columns  $(B_1 \cup B_2)$  in  $mat(\mathbb{K})$

**Definition 7.** The zero-induces index  $S_z$  of the concept  $C_1$  and  $C_2$  is equal to

$$S_z = \frac{|A_1 \cup A_2| * |B_1 \cup B_2| - z(C_1, C_2)}{|A_1 \cup A_2| * |B_1 \cup B_2|}$$

The rationale is that the more similar two *formal concepts*, the less empty cells in their *formal context*. In case of two equal *formal concepts*, no empty cells will exist and the zero-induced value  $z(C_1, C_2)$  will be 0. In this case, the numerator and the denominator in the Definition 7 will be equal, being the index value equal to 1 (i.e., the maximum similarity value). On counterpart, two completely different *formal concepts* will have all the cells in their associated formal context empty. In this case, the  $z(C_1, C_2)$  value will be equal to the number of cells in the sub-matrix (i.e., equal to  $|A_1 \cup A_2| * |B_1 \cup B_2|$ ). Consequently, the numerator, and the index value, will be equal to 0 (i.e., the minimum similarity value). To exemplify how this index works, Figure 4.21 includes two example *formal concepts*  $C_1$  and  $C_2$  (in the left) and the *formal context* related to them (in the right).

Extent		Intent	
$C_1$	$\{Image_1, Image_4\}$	$\{Feature_1, Feature_5\}$	
$C_2$	$\{Image_2, Image_3\}$	$\{Feature_2, Feature_3\}$	

	Feature <sub>1</sub>	Feature <sub>2</sub>	Feature <sub>3</sub>	feature <sub>4</sub>	features
Image <sub>1</sub>	×				×
Image <sub>2</sub>		×	×		
Image <sub>3</sub>	×	×	×		
Image <sub>4</sub>	×				×

Figure 4.21 – Example of the Zeros-induced Index

The zeros-induces index of these two *formal concepts* would be equal to:

$$S_z = \frac{|A_1 \cup A_2| * |B_1 \cup B_2| - z(C_1, C_2)}{|A_1 \cup A_2| * |B_1 \cup B_2|} = \frac{4 * 5 - 11}{4 * 5} = 0.45$$

After this execution, the resultant clusters can be considered as the different topics/concepts addressed by the images, because otherwise they would have been grouped together in the clustering process.

#### 4.2.2.4. Result List Creation

Finally, after the identification of the different topics addressed by the images, by means of the HAC execution, for each cluster, the best image in the group (i.e., that with a higher ranking according to the original ranking provided by the task, taken from Flickr) will be taken and offered as result. The final result list will be a ranked list of images, ordered according to the provided ranking. An example of a diversified result list related to the Chrysler Building location can be seen in the [Figure 4.22](#).



Figure 4.22 – Example of a diversified result list for the Chrysler Building images

---

### 4.2.3. Experimentation

---

Table 4.29 shows the results obtained by our approaches for the official metrics and the result of the best textual and best performing approaches (BEST-Textual and BEST) for the 2014 and 2015 tasks.

The best 2014 textual approach is based on a pseudo-relevance feedback that feeds a Hierarchical Clustering scheme [Boteanu et al., 2014], while the best 2014 system applies a multimodal approach (i.e., including textual and visual information) also based on a Hierarchical Clustering [Dang-Nguyen et al., 2014].

Regarding the 2015, Supervised Maximal Marginal Relevance, trained using relevant and irrelevant examples from queries for the one-topic and overall results [Spyromitros-Xioufis et al., 2015], and the clustering of the textual and visual features for the multi-topic results [Dang-Nguyen et al., 2015] present the best results. Focusing on textual approaches, the best results (for one-topic, multi-topic and overall) are achieved by [Spyromitros-Xioufis et al., 2015].

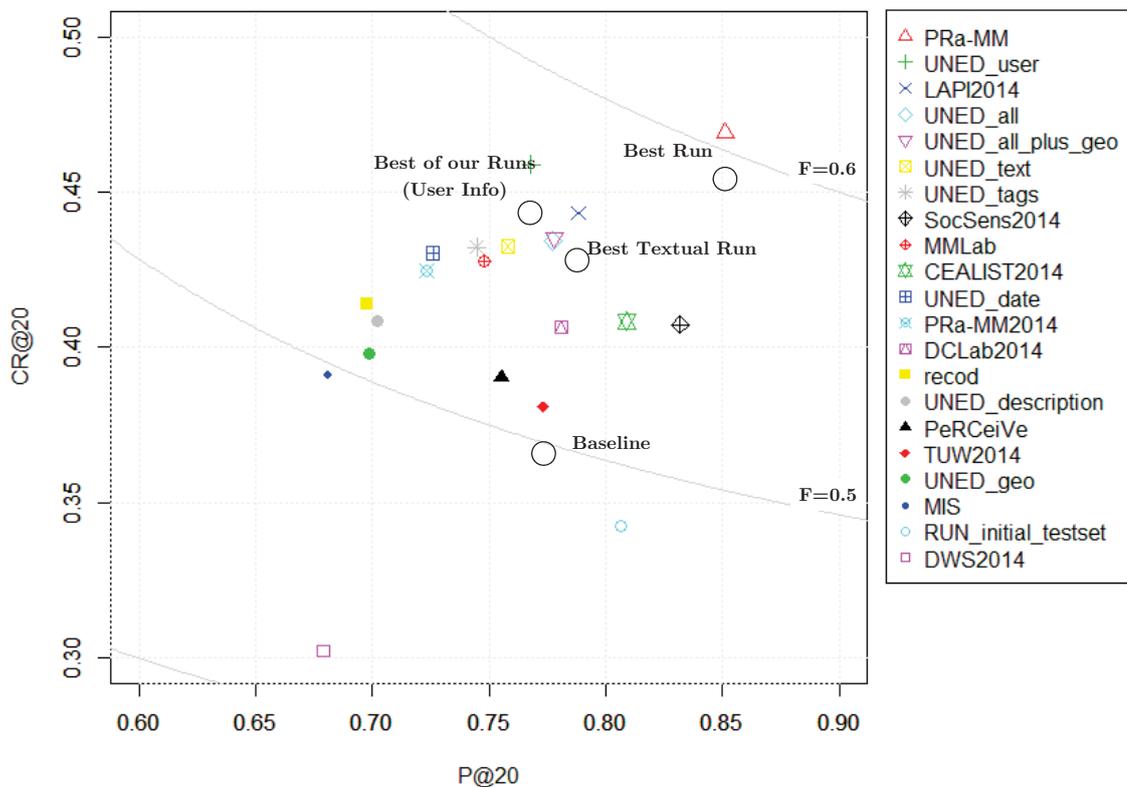
The approach herein proposed is only based on textual data; so, in order to compare us to other approaches we will focus from now on only-textual approaches. Yet, we include multimodal approaches in the analysis in order to frame our result in the overall task performance. For more details about all the approaches participating in the MediaEval campaign, please refer to the MediaEval Proceedings in [Larson et al., 2014] for the 2014 edition and to [Ionescu et al., 2015] for the 2015 proceedings.

As seen in Table 4.29, the general performance of our approach is at the same level as the best approaches in the state of the art for both 2014 and 2015, according to the official results of the task. Furthermore, when we compare our approach to the other textual approaches in the 2014 edition, it achieves the best results from among all the approaches in the task for one of the configurations (user-based information) and, for the rest of the configurations; results are also at the level of the top-performing systems. As regards the 2015 edition, our results are lower, but still in the same level than those of other systems.

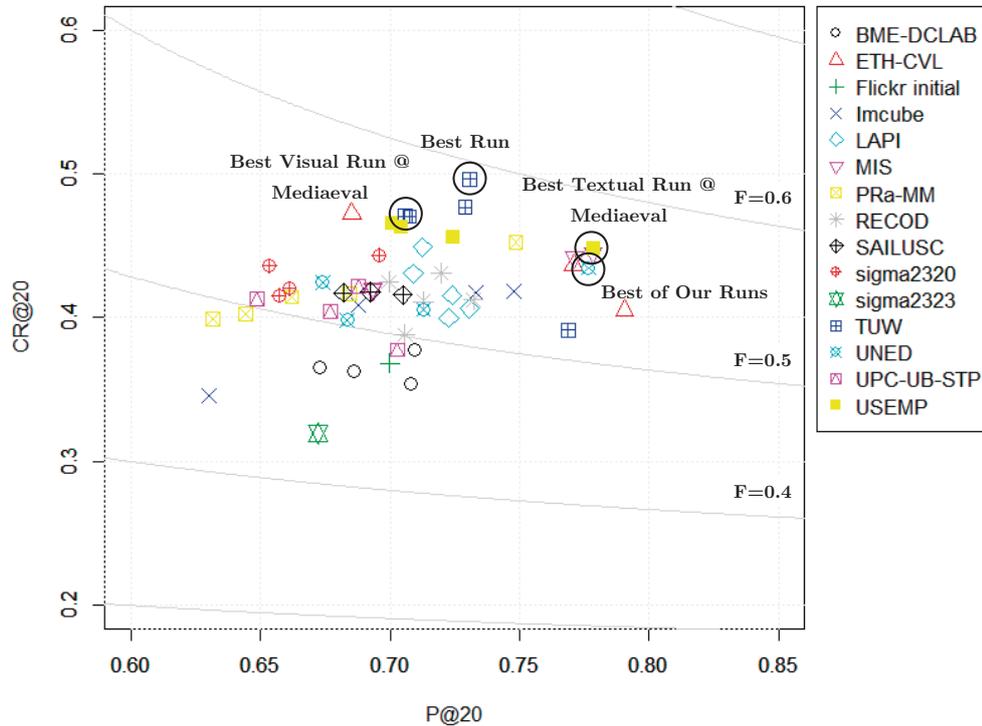
Approach	2014			2015								
				One-Topic			Multi-Topic			Overall		
	P@20	CR@20	F1@20	P@20	CR@20	F1@20	P@20	CR@20	F1@20	P@20	CR@20	F1@20
<b>BEST</b>	0.8512	0.4692	0.6049	0.8333	0.5044	0.6177	0.7607	0.4753	0.567	0.7309	0.4963	0.5727
<b>BEST-Textual</b>	0.7882	0.4431	0.5583	0.8239	0.4549	0.5692	0.7343	0.4417	0.5299	0.7788	0.4483	0.5494
<b>All</b>	0.7772	0.4343	0.5502	0.6933	0.4128	0.5087	0.6829	0.4282	0.5087	0.691	0.4206	0.5087
<b>All Text</b>	0.7581	0.4325	0.5429	0.6949	0.3937	0.496	0.6807	0.4282	0.5116	0.6877	0.4111	0.5039
<b>Date</b>	0.726	0.4304	0.5339	0.6768	0.4085	0.4975	0.6693	0.4279	0.5041	0.673	0.4183	0.5008
<b>Description</b>	0.7024	0.4084	0.5105	0.6085	0.3774	0.4529	0.5786	0.387	0.4479	0.5921	0.3822	0.4504
<b>Geo</b>	0.6988	0.3982	0.5007	0.6645	0.3783	0.4746	0.5114	0.345	0.3981	0.5874	0.3615	0.4361
<b>Tags</b>	0.7447	0.4322	0.54	0.6696	0.3961	0.4914	0.6529	0.4304	0.4996	0.6612	0.4134	0.4955
<b>User Info</b>	0.7679	0.4589	0.567	0.7087	0.4051	0.5015	0.5736	0.4109	0.4618	0.6407	0.408	0.4815

**Table 4.29** – Official Results of our approaches for Retrieving Diverse Social Images Task

In this regard, [Figure 4.23](#) shows a comparison of our approaches to the textual systems participating in MediaEval 2014 and [Figure 4.24](#) to those in MediaEval 2015. In these figures, the system performance is shown in terms of precision, Cluster Recall and F1 (height-lines in the figure) measures, being the best ones those that maximize both values (i.e., the ones closer to the upper right corner of the figure). As seen in these figures, our approaches are among the best performing ones in comparison to others in the task.



**Figure 4.23** – Official Results of All the Text-based systems for Retrieving Diverse Social Images Task 2014



**Figure 4.24** – Official Results of All the Text-based systems for Retrieving Diverse Social Images Task 2015

Focusing in the different applied features, a more detailed analysis of these configurations for Precision and Diversity values for the 5 to 50 first results is shown in [Table 4.30](#) for the 2014 results and in [Table 4.31](#) for the 2015 results. The table also includes the best MediaEval textual run. For the values denoted by \* the **user-based** approach is significantly better according to a Wilcoxon test with a p-value equals to 0.05 for the given ranking level. For the values denoted by †, the **all-based** approach is significantly better. Finally, for the values denoted by ‡, the **all-text-based** approach is significantly better. The best result for each level and for each measure is denoted in **bold** in the table and the best result of our approaches in *italic*. The values for the 20 first results (P@20, CR@20 and F1@20), which are the ones used in the official evaluation to set the system performance, are shown in red in the Table.

	Best Textual	User	All	All Text	Tags	Date	Description	Geo
P@5	<b>0,8211</b>	0,774	<i>0,7772</i>	0,7642	0,7545	0,7122 <sup>*†‡</sup>	0,6846 <sup>*†‡</sup>	0,735 <sup>†</sup>
P@10	<b>0,7984</b>	0,7724	<i>0,7748</i>	0,7561	0,7634	0,7195 <sup>*†‡</sup>	0,7008 <sup>*†‡</sup>	0,7252 <sup>*†</sup>
P@20	<b>0,7882</b>	<b>0,7679</b>	<b>0,7772</b>	<b>0,7581<sup>†</sup></b>	<b>0,7447<sup>*†</sup></b>	<b>0,726<sup>*†‡</sup></b>	<b>0,7024<sup>*†‡</sup></b>	<b>0,6988<sup>*†‡</sup></b>
P@30	<b>0,7718</b>	0,5352 <sup>†‡</sup>	<i>0,7607</i>	0,755	0,7122 <sup>†‡</sup>	0,7144 <sup>†‡</sup>	0,6732 <sup>†‡</sup>	0,6756 <sup>†‡</sup>
P@40	<b>0,7632</b>	0,4014 <sup>†‡</sup>	0,7033 <sup>‡</sup>	<i>0,739</i>	0,6618 <sup>†‡</sup>	0,6748 <sup>†‡</sup>	0,6242 <sup>†‡</sup>	0,6266 <sup>†‡</sup>
P@50	<b>0,76</b>	0,3211 <sup>†‡</sup>	0,6122 <sup>‡</sup>	<i>0,7065</i>	0,592 <sup>‡</sup>	0,6199 <sup>‡</sup>	0,5602 <sup>†‡</sup>	0,5732 <sup>†‡</sup>
CR@5	0,1553	<b>0,1559</b>	0,1538	0,1503	0,1489	0,141 <sup>*†‡</sup>	0,1379 <sup>*†‡</sup>	0,147
CR@10	0,2661	<b>0,2675</b>	0,2679	0,266	0,2619	0,2547 <sup>‡</sup>	0,2481 <sup>*†‡</sup>	0,2561
CR@20	<b>0,4431</b>	<b>0,4589</b>	<b>0,4343<sup>*</sup></b>	<b>0,4325<sup>*</sup></b>	<b>0,4322<sup>*</sup></b>	<b>0,4304<sup>*</sup></b>	<b>0,4084<sup>*†‡</sup></b>	<b>0,3982<sup>*†‡</sup></b>
CR@30	0,5479	0,4696 <sup>†‡</sup>	0,5439	<b>0,5482</b>	0,5423 <sup>†‡</sup>	0,5421	0,4952 <sup>†‡</sup>	0,4913 <sup>†‡</sup>
CR@40	0,6105	0,4696 <sup>†‡</sup>	0,5974 <sup>†</sup>	<b>0,6207</b>	0,5887 <sup>‡</sup>	0,6162	0,5536 <sup>†‡</sup>	0,5441 <sup>†‡</sup>
CR@50	0,6698	0,4696 <sup>†‡</sup>	0,6187 <sup>†</sup>	<b>0,6803</b>	0,6163 <sup>‡</sup>	0,6541 <sup>‡</sup>	0,58 <sup>†‡</sup>	0,5743 <sup>†‡</sup>
F1@5	<b>0,2581</b>	<i>0,2571</i>	0,2546	0,2493	0,2469	0,2334 <sup>*†‡</sup>	0,2279 <sup>*†‡</sup>	0,2421
F1@10	<b>0,3928</b>	0,3927	<i>0,3932</i>	0,3882	0,3852	0,3727 <sup>†‡</sup>	0,3631 <sup>*†‡</sup>	0,3737
F1@20	<b>0,5583</b>	<b>0,567</b>	<b>0,5502<sup>*</sup></b>	<b>0,5429<sup>*</sup></b>	<b>0,54<sup>*</sup></b>	<b>0,5339<sup>*</sup></b>	<b>0,5105<sup>*†‡</sup></b>	<b>0,5007<sup>*†‡</sup></b>
F1@30	<b>0,6312</b>	0,4935 <sup>†‡</sup>	0,6256	<i>0,6267</i>	0,5949 <sup>†‡</sup>	0,608 <sup>†‡</sup>	0,5632 <sup>†‡</sup>	0,5608 <sup>†‡</sup>
F1@40	<b>0,669</b>	0,4271 <sup>†‡</sup>	0,6359 <sup>‡</sup>	<i>0,6646</i>	0,6134 <sup>†‡</sup>	0,6344 <sup>‡</sup>	0,5774 <sup>†‡</sup>	0,5732 <sup>†‡</sup>
F1@50	<b>0,7016</b>	0,3766 <sup>†‡</sup>	0,6051 <sup>‡</sup>	<i>0,6814</i>	0,5931 <sup>‡</sup>	0,6217 <sup>‡</sup>	0,5592 <sup>†‡</sup>	0,5635 <sup>†‡</sup>

Table 4.30 – Official Results of our approaches for Retrieving Diverse Social Images  
Task 2014 — Extended Results

	Best Textual	User	All	All Text	Tags	Date	Description	Geo
P@5	<b>0.764</b>	<i>0.7165</i>	0.7108	0.7036	0.7151	0.682 <sup>*†</sup>	0.659 <sup>*†‡</sup>	0.6892 <sup>*†‡</sup>
P@10	<b>0.7583</b>	0.7093	<i>0.7137</i>	0.7086	0.6899 <sup>*†</sup>	0.672 <sup>*†‡</sup>	0.6338 <sup>*†‡</sup>	0.6547 <sup>*†‡</sup>
P@20	<b>0.7788</b>	<b>0.6407<sup>†</sup></b>	<b>0.691</b>	<b>0.6877</b>	<b>0.6612</b>	<b>0.673</b>	<b>0.5921<sup>*†‡</sup></b>	<b>0.5874<sup>*†‡</sup></b>
P@30	<b>0.7731</b>	0.5588 <sup>†‡</sup>	0.6415	<i>0.6449</i>	0.6247	0.6372	0.5405 <sup>*†‡</sup>	0.5444 <sup>†‡</sup>
P@40	<b>0.7714</b>	0.4856 <sup>†‡</sup>	0.5874	0.5916	0.5723	<i>0.6036</i>	0.4971 <sup>†‡</sup>	0.4951 <sup>†‡</sup>
P@50	<b>0.7662</b>	0.4305 <sup>†‡</sup>	0.5335	0.5381	0.5128 <sup>†</sup>	<i>0.5674</i>	0.4449 <sup>†‡</sup>	0.4557 <sup>†‡</sup>
CR@5	<b>0.1846</b>	0.1637	0.1588	0.1623	<i>0.1682</i>	0.1616	0.1516 <sup>*</sup>	0.1584
CR@10	<b>0.2913</b>	<i>0.2772</i>	0.2755	0.2664	0.2744	0.2665	0.2587 <sup>†</sup>	0.2545 <sup>†</sup>
CR@20	<b>0.4483</b>	<b>0.408</b>	<b>0.4206</b>	<b>0.4111</b>	<b>0.4134</b>	<b>0.4183</b>	<b>0.3822<sup>*†‡</sup></b>	<b>0.3615<sup>*†‡</sup></b>
CR@30	<b>0.5383</b>	0.4732 <sup>†‡</sup>	0.5032	0.4973	0.4997	<i>0.5141</i>	0.4497 <sup>*†‡</sup>	0.438 <sup>*†‡</sup>
CR@40	<b>0.6024</b>	0.5069 <sup>†‡</sup>	0.5506	0.5438	0.5509	<i>0.5729</i>	0.4992 <sup>†</sup>	0.4822 <sup>*†</sup>
CR@50	<b>0.6505</b>	0.526 <sup>†‡</sup>	0.5759	0.5762	0.5734	<i>0.6141</i>	0.5248 <sup>†</sup>	0.5036 <sup>*†</sup>
F1@5	<b>0.2906</b>	0.2616	0.2526	0.2581	<i>0.2661</i>	0.2548	0.2422 <sup>*</sup>	0.254
F1@10	<b>0.4089</b>	<i>0.3869</i>	0.3855	0.3779	0.3824	0.3723	0.3569 <sup>*†</sup>	0.359 <sup>*†</sup>
F1@20	<b>0.5494</b>	<b>0.4815</b>	<b>0.5087</b>	<b>0.5039</b>	<b>0.4955</b>	<b>0.5008</b>	<b>0.4504<sup>*†‡</sup></b>	<b>0.4361<sup>*†‡</sup></b>
F1@30	<b>0.6152</b>	0.4948 <sup>†‡</sup>	0.5467	0.547	0.5399	<i>0.5504</i>	0.4762 <sup>*†‡</sup>	0.4688 <sup>*†‡</sup>
F1@40	<b>0.6549</b>	0.4774 <sup>†‡</sup>	0.5513	0.5498	0.5452	<i>0.5682</i>	0.4825 <sup>†‡</sup>	0.4696 <sup>†‡</sup>
F1@50	<b>0.6802</b>	0.4541 <sup>†‡</sup>	0.5363	0.538	0.5241	<i>0.5682</i>	0.4665 <sup>†‡</sup>	0.4588 <sup>†‡</sup>

Table 4.31 – Official Results of our approaches for Retrieving Diverse Social Images  
Task 2015 — Extended Results

Two clear tendencies appears in the results. Firstly, in general, the more values of the ranking, the worse the Precision (i.e.,  $P@50 \ll P@5$ ). It is an expected behaviour; precision usually decrease when more results are taken into account, because more wrong items are expected to be retrieved in the lowest positions of the ranking. This behaviour is clearly shown in [Figure 4.25](#), wherein Precision-based results are detailed for both 2014 and 2015 datasets. It is especially marked for the **user-based** results (the best-performing approach), where the precision loss is sharper than in the other configurations. This may be explained by the fact that the diversification carried out by offering results from different users strongly affects the quality of the result set. It is reasonable given that the precision results are based on the relatedness of the images to the original location query (e.g., Eiffel Tower). Consequently, the feature related to the user taking the photo does not appear as a strong signal for this relatedness (see [Table 4.30](#) and [Table 4.31](#)).

Taking into account only these Precision-based values, the approach applying all the available information (**All**) about the images offers the best performance among our approaches, only outperformed by the best textual approach at Mediaeval. It makes sense that in this scenario the more information about the image, the better the accuracy: the image representation (based on all the available image features) is likely to be richer than when these features are considered by separate.

On the other hand, focusing on the diversity, the larger the result list, the better the diversity. Since the diversity values are based on the number of clusters in the gold standard covered in the final result list, more items lead to more clustering coverage and, therefore, to more diversity. However, from the 20<sup>th</sup> position, and especially from the 30<sup>th</sup>, it tends to stabilize. It appears that at some point the inclusion of more items does not lead to more cluster coverage. Regarding the feature type, in general, the inclusion of all the information (especially textual information) related to an image seems to improve the results slightly. Including therefore different kinds of data improves the diversity of the results. Furthermore, as said before and seen in [Table 4.30](#) and the subsequent figures, Precision-based results are also favoured by the inclusion of different features.

To sum up, the aggregation of all the different features related to the images (**All** and **All Text** approaches) appears as the most suitable approach for both the Precision and Diversity. Only **user-based** information, especially for the 2014 experimentation, outperforms the results of these aggregation-based approaches for some ranking levels (among them the @20 used to evaluate the systems in the MediaEval task). However, given the high variance of this **user-based** result, it appears that these results are due to an over-specialization for some ranking levels instead of a general improvement.

No significant differences are observed among the results of the other features when taken by separate (**tags**, **description**, **date** or **geo**). It is remarkable the low performance of the **description** approach. A priori, this approach could be considered as the more informative one; however, the obtained results have proven the opposite. In this regard, a keyword-based representation, like that based on **tags**, has been demonstrated to be more effective than a natural language based representation, like the **description** approach.

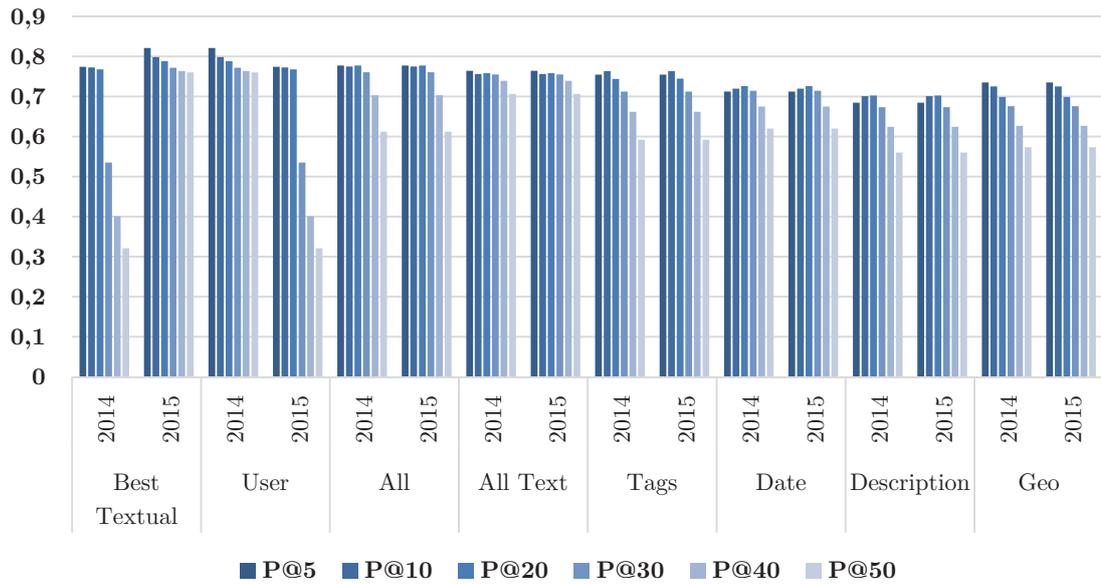


Figure 4.25 –Extended Results: Precision Based Results

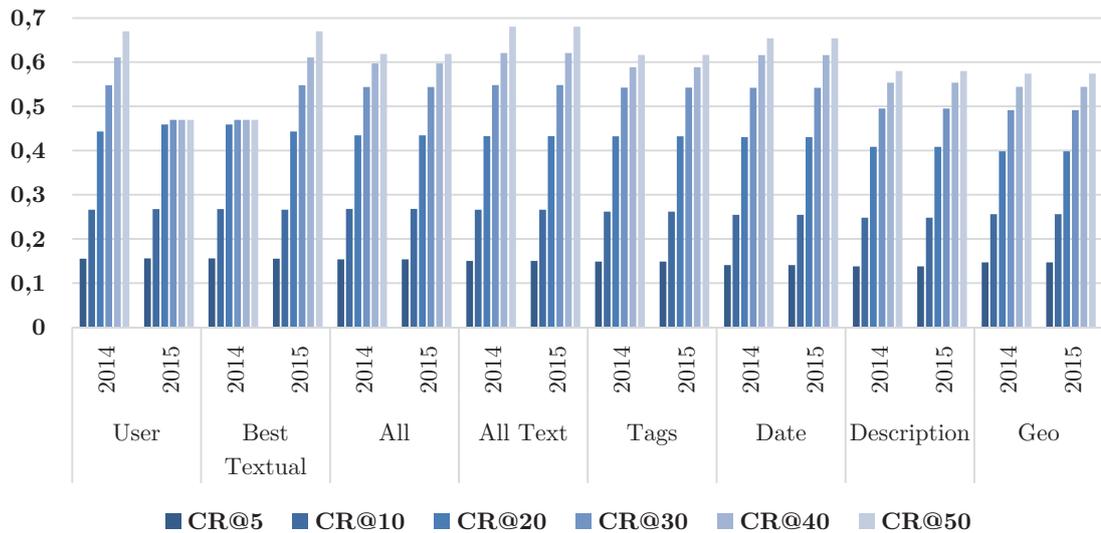


Figure 4.26 – Extended Results: Cluster Recall Based Results

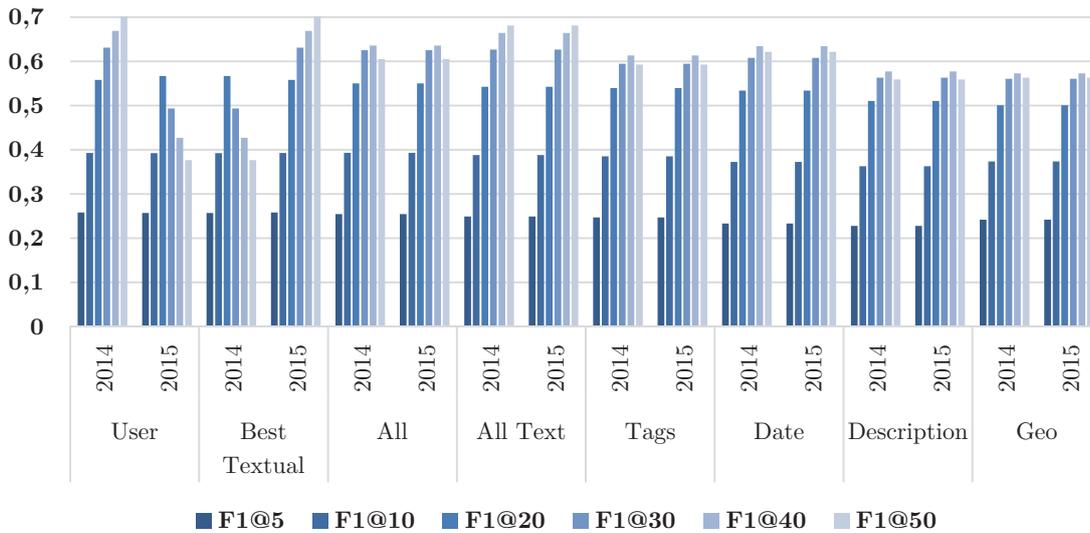


Figure 4.27 – Extended Results: Cluster Recall Based Results

#### 4.2.4. Discussion

This section detailed the second of the proposed application scenarios — image diversification — for the evaluation of our FCA-based representation proposal. FCA is herein applied to create an image representation based on the textual content related to the images.

By means of FCA, it was intended to infer the latent topics addressed by the images to then offer results covering all the detected topics, expecting to improve in this way the image diversity. Different kinds of textual features describing the images (textual features describing the images, date related features, geo-location features or user-related information) that might be useful for accurately represent the images and for covering the different aspects related to the images have been evaluated.

To that end, we have taken advantage of the experimental framework provided by the Retrieving Diverse Social Images Task at MediaEval. The obtained results prove our proposal as suitable to offer accurate and diverse results. Our FCA-based approach achieves results in the same level than the best textual ones in the Retrieving Diverse Social Images Task (both for 2014 and 2015 editions), outperforming them for some configurations. Moreover, some of these configurations are even in the same level than the top-performing approaches, based on multimodal methodologies.

Focusing on the different features applied to represent the images, the grouping of all the textual-based features together seems to be the best choice. Precision- and Diversity-based results offer better performance than when the features are used separately. Among the individual features, the *a-priori* most sensible choice is the use of the description of the images, written by the user that uploaded the image to Flickr, to represent them. Nevertheless, keyword-based representations (e.g., tags) lead to better representations in terms of both, precision and clustering recall.

Summarizing, results have confirmed (as in section [4.1](#)) that the proposed FCA-based representation proposal is able to detect the latent aspects that cover the different aspects addressed by the images. This is of special importance for its latter application in the recommendation context. Recommendation is highly dependent on the discovery of these latent aspects in order to identify the different user tastes and the items fulfilling them.

# 5

## LOD-based Representation Enhancement

---

*This chapter presents the experimentation conducted in regards to knowledge modelling by applying Formal Concept Analysis. It is considered a step forward of the work conducted in the previous chapter*

### Content

---

5.1	KNOWLEDGE-BASED REPRESENTATION PROPOSAL .....	158
5.1.1.	DBpedia Modelling .....	159
5.1.2.	EWN Modelling .....	164
5.2	APPLICATION SCENARIO: IMPROVING THE TOPIC DETECTION .....	167
5.3	DISCUSSION .....	177

---

In the previous chapter, it has been presented a methodology to model text-based representations by means of FCA. This methodology has proved to be able to infer the latent aspects that better describe the content, thus improving other state-of-the-art methodologies in the two proposed experimental scenarios. In this section, we intend to go a step further in the representation of textual content. In spite of the improvement enabled by the FCA modelling, the representation proposed in the previous section still relies on textual content, and therefore some of the problems related to textual representation still remains.

In this regard, this chapter proposes the use of a higher-level feature description (e.g., Linked Open Data features related to an item instead of its textual content). These higher-level features enable the unambiguous representation of data (e.g., EuroWordNet avoid ambiguous textual representations by grouping them in cognitive synonyms or *synsets*), as well as it allows more compact representations (i.e., instead of using the textual description of a movie, the entities related to it in DBpedia — director, actors, genre, etc. — may be used). Furthermore, these representations may be helpful in scenarios where the system only have limited information or even not information at all about the items. In such contexts, the features coming from sources like DBpedia or EuroWordNet may serve to enrich the item description. This is a common issue in social networks, like for instance Twitter, where users express themselves by means of short texts.

In order to validate the FCA-based representation relying in these higher-level features, we have applied the same evaluation scenario than in section 4.1: the detection of topics on a stream of Twitter data. In this way, it is going to be possible to compare the improvement derived from the application of knowledge-based features. Two knowledge-based resources have been applied for the experimentation: DBpedia to collect semantic information and EuroWordNet to collect linguistic information on the tweet content.

FCA is then applied to these resources and the resultant FCA models are used to extract the most interesting information to represent the tweets. Later on, a state-of-the-art algorithm is applied to the tweet representations to perform the topic detection. We also propose several baselines with which to compare our FCA-based approach, where everything except the modelling process is identical. Consequently, the observed variations in the topic detection results are only going to be based on the modelling approach performance. By means of these experimental results, it is going to be proved how the content representation provided by the FCA-based modelling of these knowledge

features achieves a significant performance improvement for topic detection in relation to the state of the art and to the results in the previous chapter.

## 5.1 Knowledge-based Representation Proposal

---

The main aim is to apply a FCA-based to create a content representation relying on higher-level knowledge features. To that end, FCA is applied in order to infer the best knowledge features in order to describe the content, taking advantage of the subjacent knowledge-based structure.

Two knowledge-based resources have been applied: DBpedia to collect semantic information and EuroWordNet to collect linguistic information. DBpedia is a Linked Open Data (LOD) resource including the information contained in the Wikipedia Infoboxes. Due to the large amount of data it contains and its semantic organization (allowing the easy process of these data), it constitutes the main repository for content enrichment. More in detail, we have used the 3.9 version of the DBpedia Dataset. This dataset includes about 4 millions of instances described by means of 470 million of attributes. The instances and their attributes are structured according to an ontology specifically developed for DBpedia.

On the other hand, EuroWordNet (EWN) is a multilingual database including different European languages: Dutch, Italian, Spanish, German, French, Czech, and Estonian [Vossen, 2004]. It is structured in the same way as the English WordNet: *synsets* (sets of synonymous words) with semantic relations between them. EWN offers a dataset of concepts, categorized in lexical groups (i.e., nouns, verbs, adjectives and adverbs) and organized in a hierarchical structure according to semantic relationships (e.g., hypernyms, hyponyms). For example, the word group (car, auto, automobile, machine, motorcar) constitutes a *synset*, which is described as: *4-wheeled; usually propelled by an internal combustion engine*.

Taking a closer look to the DBpedia ontology class information, it can be seen as a *formal context*: a set of DBpedia entries (objects of the formal context  $G$ ) a set of DBpedia features describing the entries (attributes of the formal context,  $M$ ) and a relationship ( $I$ ), indicating that an entry  $G$  has a feature  $M$ . Likewise, EWN information may be also represented as a *formal context*: the *synsets* are the objects of the relationships in the

EWN structure (e.g. hyperonymy, hyponymy, meronymy, synonymy, and antonymy) the attributes to describe the synsets.

After the FCA execution, a lattice structure, organizing the set of obtained *formal concepts*, will be obtained. The resultant lattice can be understood as a semantic- or linguistic-driven organization of the knowledge: the *formal concepts* will put together those “similar” entries, according to their shared features; while the lattice will order these *formal concepts* from the most generic to the most specific ones. This latter organization offers all the advantages derived from the FCA and lattice theory; that is, an easy- readable and treatable data representation and the discovering and exposure of the latent data relationships. The idea is now, given a content to be represented, use this FCA models to represent it. The hypothesis is that the application of the acquired knowledge (from the FCA lattices) should lead to a better content enrichment by selecting the most suitable information related to the content.

In what follows, the specific details on the modelling process for each of the resources (DBpedia and EuroWordNet) as well as its application for content representation is detailed.

### 5.1.1. DBpedia Modelling

DBpedia content is categorized according to the DBpedia Ontology. This ontology has a set of classes describing the different knowledge domains addressed in DBpedia (e.g., soccer, movies, automotive, etc.). In order to take into account the specific characteristics of each knowledge domain, the DBpedia data have been modelled by creating a FCA-based model for each one.

In order to generate the DBpedia models, an FCA-model has been created for each of the ontology classes corresponding to the knowledge domains addressed in the RepLab dataset (automotive, banking, university and music). In particular, the formal context  $\mathbb{K}_{DBpedia} := (G, M, I)$  associated to each model includes the set of DBpedia entries in the given domain as the objects of the formal context ( $G$ ) and the set of DBpedia features related to these entries as the attributes ( $M$ ). By DBpedia features we refer to the pairs **property**-value related to the DBpedia entry (e.g., for the entry **BMW**: **manufacturerOf**-BMW-M5, ..., **manufacturerOf**-BMW\_M3, ... , **subject**-German\_brands, ...).

The binary relationship  $I$  will therefore indicate that a DBpedia entry  $g \in G$  has a given feature  $m \in M$  (e.g., the entry **BMW\_M3** has the feature `German Car` or the entry **BMW\_M3** does not have the feature `French Car`). It might potentially result in a huge formal context. DBpedia entries per domain are in the number of thousands, as are their properties. In the worst-case scenario each property could take a different value for each entity (e.g., **BMW\_M3: locationCountry-Germany**, **Renault\_Megane: locationCountry-France**, **Ford\_Mustang: locationCountry-USA**), leading to millions of different features. However, most of these features have a small frequency (e.g., the value of the property **assets**) and, consequently, they are barely representative. On the other hand, there are frequent features such as **locationCountry-Germany** or **subject-Sedans**, which are more likely to describe the entities better.

Applying this assumption, we have applied the reduction in the *formal context* already applied in section 4.1.3.2 [Cigarrán et al., 2004] in order to select the features that best describe the DBpedia entries by means of the algorithm. This algorithm produces a smaller and denser feature representation, maintaining the representativeness. For this experimentation the parameter for the lower threshold is set to 1% and the upper threshold to 50% (refer to [Cigarrán et al., 2004] for details on these thresholds). In previous experimentations, these parameters have proven to provide the best experimental results [Castellanos et al., 2013; Cigarrán et al., 2016].

In spite of this reduction, the complexity of the FCA computation and the scalability issues are definitely something to take into account. As it has been extensively studied and pointed out in the FCA literature, the complexity of the FCA algorithms is one of its main drawbacks (see section 4.1.4.3). Scalability has to be taken into account especially in scenarios like that addressed in this work (Twitter), where large numbers of data streams are expected. At this point, it is important to highlight that the FCA computation is carried out off-line: the FCA models are firstly computed and then applied to enrich and represent Twitter data. Consequently, the FCA complexity has no limitation when dealing with Twitter data. In this regard, the complexity and computation times of this off-line process applied to Web of Data (WOD) resources are doable, as proved in [Kirchberg et al., 2012].

Some figures on the FCA computation, like the object and feature count and the formal concepts they create are detailed in Table 5.1. As shown in the Table, after applying FCA the initial number of DBpedia relationships is reduced by an order of magnitude in terms of the number of *formal concepts* generated. In other words, FCA has created a

more abstract representation of the DBpedia data, which groups together (in terms of formal concepts) similar objects according to their shared relationships.

	Object Count	Attributes Count	FCA Relationships Count	Formal Concept Count
<b>Automotive</b>	8,383	115	7,582	1,722
<b>Music</b>	42,419	101	90,136	6,021
<b>University</b>	15,87	55	16,69	613
<b>Banking</b>	20,818	36	8,985	558

**Table 5.1** – FCA DBpedia Knowledge Organization Statistics

Taking these considerations into account, the DBpedia data can therefore be (re)organized in a lattice structure by applying the FCA theory. [Figure 5.1](#) shows a toy-example including an excerpt of the knowledge domain representation related to the automotive domain. For the sake of simplicity, this example contains only a bunch of entities and features, but the concept lattice created for the experimentation contains the whole set of entries and features in the DBpedia data. It is going to be proved that the more abstract representation FCA provides, grouping similar objects together and organizing them in the lattice structure, may allow the easier identification of the most interesting information in the DBpedia data.

To exemplify this main point, the lattice in [Figure 5.1](#) organizes the features from the most generic (at the top of the lattice) to the most specific (at the bottom of the lattice). For example, the feature `Car` is quite a generic one, related to many entries (i.e., most of the entries in the automotive domain are cars). In contrast, the feature `M-Series` can be considered as a specific one: it is only related to a small set of entries (i.e., only a handful of cars belong to the BMW `M-Series`).

Therefore, given a tweet related to the automotive domain, instead of taking all the information in DBpedia related to the entities in the tweet, it is now possible to know which information is more suitable to create an accurate data representation: the features closely related in the lattice structure. For instance, taking a look at [Figure 5.1](#), if we want to represent the entity `M7`, the features close to its concept (the one labelled as `M7`) in the lattice structure (e.g., `Concept_Car`, `M-Series` and `High_Performance_Car`) are more suitable to represent it than other features that are not so close in the lattice structure, like for example `German_Car` or `Car`.

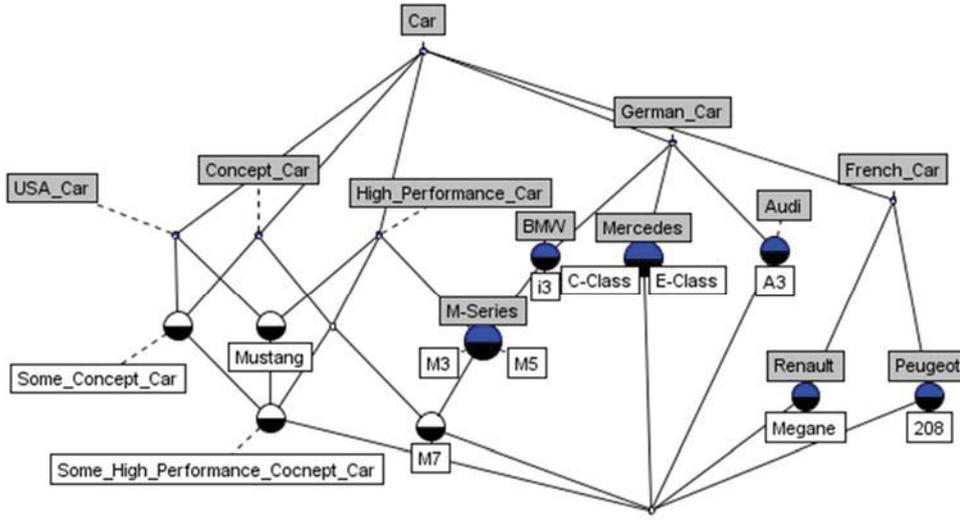


Figure 5.1 – FCA-based LOD Modelling

Generalizing this example, the vast amount of information available in DBpedia will allow learning a large number of informational patterns. It is expected that the application of this acquired knowledge will drive to a better representation of data belonging to this domain by making use of the FCA representation.

In more detail, in order to represent a given  $Content_i$ , composed by a set of named entities  $Content_i = (NE_1, \dots, NE_n)$ , which have been previously detected by applying the Textalytics Semantic API (see section 4.1.5.1), and the DBpedia model related to the knowledge domain of the tweet (automotive, banking, music or university)  $\mathfrak{B}_{DBpedia}(G, M, I)$ , it is defined:

- 1) a set  $OC = (OC_1, OC_2, \dots, OC_i)$  including all the *object concepts* related to each of the named entities  $NE$  in the  $Content_i$  which are included in  $\mathfrak{B}_{DBpedia}(G, M, I)$ , such as:

$$\forall OC_x \in OC, \exists NE_x : \gamma NE_x \text{ in } \mathfrak{B}_{DBpedia}(G, M, I) = OC_x$$

- 2) a set  $UN = (UN_1, UN_2, \dots, UN_j)$  that includes the *upper neighbours* of the *object concept* set  $OC$ , such as:

$$\forall UN_x \in UN, \exists OC_x \in OC : UN_x > OC_x \text{ in } \mathfrak{B}_{DBpedia}(G, M, I)$$

- 3) and, finally, a set of *candidate concepts* ( $CC$ ) including the *object concepts* and their *upper neighbours*, such as:

$$CC = (CC_1, CC_2, \dots, CC_k) = OC \cup UN$$

The final representation for the  $Content_i$  will be a label set  $LS = (m_1, m_2, \dots, m_l)$  that includes all the attributes in the *intension* of the *candidate concepts*, such as:

$$\forall m_x \in LS, \exists CC_x = (A_x, B_x) \in CC : m_x \in B_x$$

Table 5.2 shows an example applying this process (adapted to the case study) based on the FCA-based DBpedia Model in the Figure 5.1. One way of integrating the DBpedia data could be based on taking all the DBpedia features related to the named entities which belong to the tweet domain: **M3**, and **M5**. This is the approach in the second column of Table 5.2 (LOD Representation). Nevertheless, it is likely to include noisy or too generic information (e.g., an **M3** is a car vs. **M3** is a High Performance German Car). This process is susceptible of being refined by, for example, only selecting the potentially most interesting DBpedia features, such as `locationCountry`, `manufacturer`, `subject`.... However, this refinement is not straightforward: it is dependent on the data to be modelled, the task to be addressed or the DBpedia data.

Original Text	LOD Representation	LOD_FCA
<b>M3</b> & <b>M5</b> M Performance Editions - Web Exclusive: M enthusiasts in the UK will be rejoicing at the announc...	<b>M3</b> : [BMW, German_Car, M-Series, Car, High_Performance_Car...] <b>M5</b> : [BMW, German_Car, M-Series, Car, High_Performance_Car...] <b>UK</b> : [Country, ..., Europe]	<b>M3</b> : BMW, M-Series, High_Performance_Car <b>M5</b> : BMW, M-Series, High_Performance_Car <b>UK</b> : NOT USED
<b>M7</b> Rendering Released: There are many car enthusiasts and lovers who would want a beefed up 7 Series...	<b>M7</b> : [BMW, German_Car, M-Series, Car, High_Performance_Car, Concept_Car...]	<b>M7</b> : M-Series, High_Performance_Car, Concept_Car
The latest <b>C-Class</b> has only just gone on sale, but you can already get nearly £4k off. Now that's a deal!	<b>C-Class</b> : [Car, Mercedes, German_Car]	<b>C-Class</b> : Mercedes

**Table 5.2** – Example of Data Representation

In contrast, our FCA-based proposal allows this refinement to be performed automatically by taking advantage of the knowledge modelling provided by the concept lattice. In more detail, the first step to represent the tweet is to look up the named entities included in the tweet in the DBpedia model ( $\mathfrak{B}_{DBpedia}(G, M, I)$  in Figure 5.1) related to the knowledge domain of the tweet (i.e., automotive):  $Tweet_i = (M3, M5)$ . The aforementioned content representation is applied hereafter to represent the tweet

(the third column in the [Table 5.2](#)). By looking at the Table, it can be seen that the LOD-FCA representation includes more specific terminology than the LOD representation, which includes all the potentially related DBpedia features.

### 5.1.2. EWN Modelling

This modelling is similar to the one applied to the DBpedia data. The DBpedia data is useful to represent the tweets by means of the named entities mentioned in them. In environments such as Twitter, the named entities are a strong signal in order to model the contents. However, it is also reasonable to think that the text in the tweets may also offer some more information not covered by the named entities. As in the previous experimentation, the aim was to create an extra knowledge-layer on top of the EuroWordNet (EWN) information. More specifically, the single English (EWNen) and Spanish (EWNes) versions has been used (i.e., these are the languages covered by the RepLab collection).

Applying the FCA rationale to the formal context associated to the EWN data  $\mathfrak{B}_{EWN}(G, M, I)$ , we have the synsets to be taken as the objects of the formal context,  $G$  and the relationships in EWNen and EWNes (e.g. hyperonymy, hyponymy, meronymy and synonymy and antonymy) to be taken as the attributes  $M$  to describe the synsets. The relationship  $I$  indicates that a synset  $g \in G$  is related to a EWN relationship  $m \in M$  in the form of **relationship**-value. For example, the synset **auto** is associated to the EWN relationship **hyperonym-motor\_vehicle**.

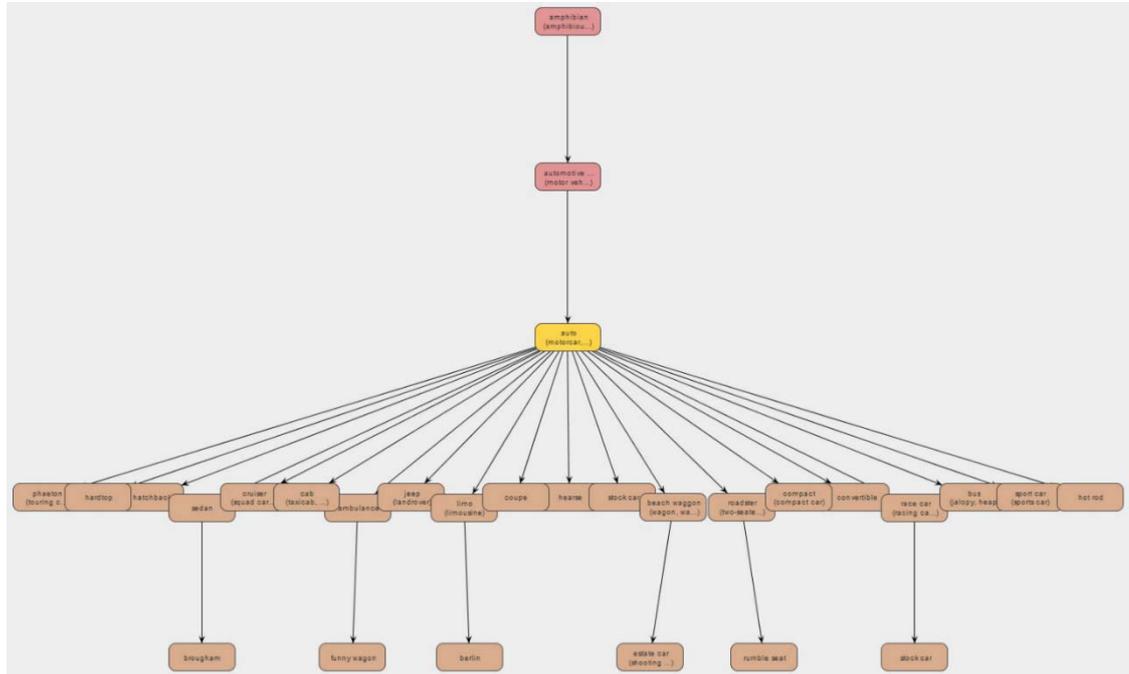
The final figures on the FCA computation are shown in [Table 5.3](#). As in the DBpedia modelling, the abstraction enabled by FCA reduces the number of EWN relationships. Therefore, the concept lattice  $\mathfrak{B}_{EWN}(G, M, I)$ , will put together those "similar" synsets, according to their shared relationships, and will organize them from the most generic to the most specific. To create these concept lattices, the formal context reduction used to create the DBpedia lattices has been applied.

	Object Count	Attributes Count	FCA Relationships Count	Formal Concepts Count
EWNen	91,555	1,192	165,098	4,972
EWNes	28,093	777	64,606	2,351

**Table 5.3** – FCA EWN Knowledge Organization Statistics

An example of this process is shown in [Figure 5.2](#) and [Figure 5.3](#). [Figure 5.2](#) includes an example of the EWNen hierarchy for the synset “auto” (it has been created by means of the LexiRes software [[De Luca and Nürnberger, 2006](#)]). Conversely, [Figure 5.3](#) includes an example of the FCA-based lattice structure is shown. From a closer look, it can be seen as [Figure 5.3](#) organizes the relationships according to its specificity. For instance, the **hyperonym**-wheeled\_vehicle relationship can be understood as a very general one (i.e. at the end all the synsets in EWN are entities). The original EWN structure presents a similar organization based on the hierarchical relationships (hyperonymy, hyponymy). However, in the lattice structure the non-hierarchical relationships are also taken into account to infer the inherent hierarchical structure. For example, the **synonym**-auto relationship is shown as a specific relationships, more related to the synset “machine” than, for instance, the relationship **hyperonym**-wheeled\_vehicle, being more appropriated to represent “machine” than some other more generic relationship.

Although the [Figure 5.3](#) is only an example, the final model will include all the information in EWNen and EWNes. As much more information is used, much more knowledge will be also inferred.



**Figure 5.2** – Example of EuroWordNet hierarchy

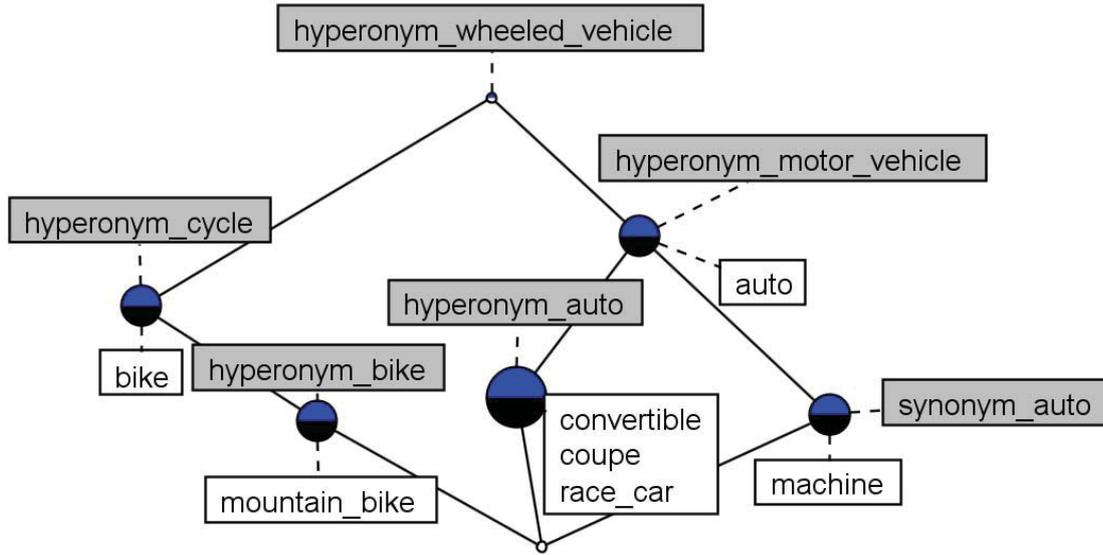


Figure 5.3 – FCA-based EuroWordNet Model

Taking advantage of this FCA-based representation, to represent a given  $Content_i$  to be represented, composed by a set of representative terms  $Content_i = \{TermSet\} = (term_1, \dots, term_n)$ , and the EWN model related to the language of the content (English or Spanish)  $\mathfrak{B}_{EWN}(G, M, I)$ , it is defined:

- 1) a set  $OC = (OC_1, OC_2, \dots, OC_i)$  including all the *object concepts* related to each of the terms  $TermSet$  in the  $Content_i$  which are included in  $\mathfrak{B}_{EWN}(G, M, I)$ , such as:

$$\forall OC_x \in OC, \exists term_x : \gamma term_x \text{ in } \mathfrak{B}_{EWN}(G, M, I) = OC_x$$

- 2) a set  $UN = (UN_1, UN_2, \dots, UN_j)$  that includes the *upper neighbours* of the *object concept* set  $OC$ , such as:

$$\forall UN_x \in UN, \exists OC_x \in OC : UN_x > OC_x \text{ in } \mathfrak{B}_{EWN}(G, M, I)$$

- 3) and, finally, a set of *candidate concepts* ( $CC$ ) including the *object concepts* and their *upper neighbours*, such as:

$$CC = (CC_1, CC_2, \dots, CC_k) = OC \cup UN$$

The final representation for the  $Content_i$  will be a label set  $LS = (m_1, m_2, \dots, m_l)$  that includes all the attributes in the *intension* of the *candidate concepts*, such as:

$$\forall m_x \in LS, \exists CC_x = (A_x, B_x) \in CC : m_x \in B_x$$

[Table 5.4](#) shows an example of the data representation by applying the aforementioned process. For instance, by applying the EWN concept lattice in [Figure 5.3](#), the most specific information related to the terms in the tweet (e.g., `hyperonym_wheeled_vehicle` vs. `synonym_auto`) is going to be used to represent the tweet (EWN-FCA in [Table 5.4](#)). In contrast, the basic methodology (EWN representation in [Table 5.4](#)) does not perform any kind of selection or refinement and includes all the information in EWN related to the content.

Original Text	WN Representation	WN-FCA
BMW M5, what a <b>machine!</b>	[ <code>hyperonym_entity</code> , <code>hyperonym_wheeled_vehicle</code> ... ]	[ <code>synonym_auto</code> , <code>hyperonym_motor_vehicle</code> ]

**Table 5.4** – Example of Data Representation using the EWN Representation

To summarize, based on the notion of *object concept*, a representation methodology has been proposed to infer the most specific information automatically from the previously generated FCA models, which in turns leads to a content representation including less but more specific information, more related to the tweets, thus improving the representativeness of the content description.

## 5.2 Application Scenario: Improving the Topic Detection

In order to test to what extent the knowledge-based representation proposed herein improves the content representation, it has been settled an experimental framework, based on the aforementioned RepLab 2013 (see section [4.1.1](#)). This section is going to prove that the Topic Detection process is going to be improved by the application of the (re)organized EWN and DBpedia models (see section [5.1](#)).

To frame the performance obtained by the FCA-based proposal, three different groups of experiments have been settled: textual, raw models (i.e., using the knowledge resources, EWN and DBpedia, without the FCA Modelling) and FCA-based (i.e., applying the FCA-based EWN and DBpedia models). All these approaches have been used to enrich and model the tweet contents and the same Topic Detection methodology has been applied.

As the topics to be detected are mainly thematically based (i.e. they try to divide the data according the different topics addressed), textual information appears as a straightforward approach to represent the data. On the other hand, in order to set "how much" of the possible performance improvement (with respect to the baseline representations) might be attributable to the FCA approach and "how much" to the knowledge-based information itself, different LOD and EWN representations are proposed. More in detail, the experimental configurations are explained below. In the same way, the proposed configurations applied to an example tweet ---**New BMW M5. More than just a car**--- can be seen in [Table 5.5](#):

- **Text:** Each tweet is represented only by its textual content (after removing stop-words, and stemming and taking into account the special Twitter signs like hashtags and references)
- **LOD:** Each tweet is represented with the DBpedia Information (without the FCA Modelling) related to the named entities appearing in the tweet.
- **EWN:** Each tweet is represented with the EWN Information (without the FCA Modelling) related to the terms in the tweet. In this case, stopwords have been also removed, but the terms have not been stemmed and the special signs have not been used.
- **Text + LOD:** Each tweet is represented with the textual content (processed in the same way than in the Text approach) plus the DBpedia information (in the same way than in the only LOD approach).
- **Text + EWN:** Each tweet is represented with the textual content plus the EWN information.
- **Text + EWN + LOD:** Each tweet is represented with the textual content plus the EWN and LOD information.
- **LOD FCA:** Each tweet is represented with the DBpedia FCA-based Information (the FCA-based models obtained from the DBpedia information) related to the entities appearing in the tweet.
- **EWN FCA:** Each tweet is represented with the EWN FCA-based Information (the EWNen and EWNes FCA-based models) related to the terms appearing in the tweet.
- **Text + LOD FCA:** Each tweet is represented with the textual content plus the LOD FCA based information.

- **Text + EWN FCA:** Each tweet is represented with the textual content plus the EWN FCA based information.
- **EWN + LOD:** Each tweet is represented with the raw EWN Information (without the FCA Modelling) related to the terms in the tweet plus the raw LOD Information (without the FCA modelling) related to the entities in the tweet.
- **EWN FCA + LOD FCA:** Each tweet is represented with the EWN FCA based Information (the EWNen and EWNes FCA-based models) related to the terms appearing in the tweet plus the LOD FCA based information related to the entities appearing in the tweet.
- **Text + EWN FCA + LOD FCA:** Each tweet is represented with the EWN FCA based Information related to the terms appearing in the tweet plus the LOD FCA based information related to the entities appearing.

Approach	Description	Representation
Text	Textual content of the tweet after removing stop-words and stemming	new, BMW M5, more, car
LOD	DBpedia information related to the tweet entities without the FCA Modelling	BMW, German Car M-Series, Car, High Performance Car. . .
EWN	EWN information related to the tweet terms without the FCA Modelling	hyperonym vehicle, hyperonym machine, hyperonym wheeled vehicle...
EWN+LOD	EWN plus LOD information	hyperonym vehicle, hyperonym machine, hyperonym wheeled vehicle..., BMW, German Car M-Series, Car, High Performance Car...
Text + LOD	Text plus LOD information	new, BMW M5, BMW, German Car M-Series, Car, High Performance Car..., more, car
Text + EWN	Text plus EWN information	new, BMW M5, more, car, hyperonym vehicle, hyperonym machine, hyperonym wheeled vehicle...
Text + EWN+LOD	Text plus EWN plus LOD information	new, BMW M5, BMW, German Car M-Series, Car, High Performance Car..., more, car, hyperonym vehicle, hyperonym machine, hyperonym wheeled vehicle...
LOD-FCA	DBpedia information in the FCA-based model related to the named entities in the tweets	BMW, M-Series, High Performance Car
EWN-FCA	EWN information in the FCA-based model related to the terms in the tweets	synonym auto, hyperonym wheeled vehicle, hyperonym machine
EWN-FCA + LOD-FCA	EWN-FCA the LOD-FCA information	BMW, M-Series, High Performance Car, synonym auto, hyperonym wheeled vehicle, hyperonym machine
Text + LOD-FCA	Text plus LOD-FCA	new, BMW M5, BMW, M-Series, High Performance Car, more, car
Text + EWN-FCA	Text plus the EWN-FCA	new, BMW M5, more, car, synonym auto, hyperonym wheeled vehicle, hyperonym machine
Text+EWN-FCA+LOD-FCA	Text plus the EWN-FCA plus the LOD-FCA information	new, BMW M5, BMW, M-Series, High Performance Car, more, car, synonym auto, hyperonym wheeled vehicle, hyperonym machine

Table 5.5 – Tweet Representation Example

The obtained results are detailed in the [Table 5.6](#). For the values denoted by †, there is a statistically significant improvement in the F-measure values compared to the Textual baseline approach, according to a Wilcoxon test with a p-value equals to 0.05. The same results are detailed in the [Figure 5.4](#), where all results are shown together, and in [Figure 5.5](#), where the results are separately presented by Reliability, Sensitivity and F-measure.

Approach	Reliability	Sensitivity	F-Measure
<b>Best-Replab</b>	0,4624	0,3246	0,3252
<a href="#">[Spina et al., 2014]</a>	[0.608 ± 0.33]	[0.4292 ± 0.3]	[0.3239 ± 0.12]†
<a href="#">[Berrocal et al., 2013]</a>	0.31	0.43	0.29
<b>Text</b>	[0.8535 ± 0.16]	[0.1923 ± 0.13]	[0.2829 ± 0.15]
<b>LOD</b>	[0.1437 ± 0.26]	[0.5698 ± 0.31]	[0.2003 ± 0.12]
<b>EWN</b>	[0.7334 ± 0.18]	[0.2184 ± 0.15]	[0.2945 ± 0.16]†
<b>EWN+LOD</b>	[0.2362 ± 0.29]	[0.7908 ± 0.34]	[0.1816 ± 0.15]
<b>Text + LOD</b>	[0.14 ± 0.26]	[0.5678 ± 0.32]	[0.2090 ± 0.12]
<b>Text + EWN</b>	[0.2424 ± 0.29]	[0.7798 ± 0.34]	[0.1895 ± 0.15]
<b>Text + EWN+LOD</b>	[0.2977 ± 0.35]	[0.7507 ± 0.35]	[0.1953 ± 0.16]
<b>LOD-FCA</b>	[0.3974 ± 0.24]	[0.5594 ± 0.26]	[0.3385 ± 0.13]†
<b>EWN-FCA</b>	[0.4031 ± 0.24]	[0.3946 ± 0.20]	[0.3035 ± 0.11]†
<b>EWN-FCA + LOD-FCA</b>	[0.4088 ± 0.24]	[0.5544 ± 0.27]	[0.3088 ± 0.13]†
<b>Text + LOD-FCA</b>	[0.4399 ± 0.23]	[0.5205 ± 0.26]	[0.3570 ± 0.14]†
<b>Text + EWN-FCA</b>	[0.2977 ± 0.35]	[0.7507 ± 0.36]	[0.1953 ± 0.16]
<b>Text+EWN-FCA+LOD-FCA</b>	[0.2577 ± 0.27]	[0.7414 ± 0.31]	[0.2206 ± 0.12]

**Table 5.6** – Topic Detection Results

In order to frame the performance of the different configurations that we propose, [Table 5.6](#) also includes the results for the best-performing approach in RepLab 2013 [\[Spina et al., 2013\]](#) and the best results reported for the task (i.e., those obtained by [\[Spina et al., 2014\]](#)). The former applies a "wikified" tweet clustering approach (i.e., Tweet content is enriched with Wikipedia entries semantically related to the tweet), using the Jaccard similarity as similarity threshold; the latter also proposes a HAC approach but applying a supervised process to learn a similarity function that drives the clustering implementation. An important remark is that, unlike the approaches proposed herein, the approach of [\[Spina et al., 2014\]](#) follows a supervised approach (i.e., it needs an annotated training set).

In addition, the comparison includes other approach that has also offered good results for the RepLab dataset: the graph-based approach in [\[Berrocal et al., 2013\]](#), which proposes a Community detection methodology applying a VOS clustering.

The results of our approach, and those of [\[Spina et al., 2014\]](#), depend on the different similarity thresholds applied to the HAC algorithm (from 0.0, increasing by 0.1, to 1.0). Consequently, in order to take into account the performance of the approach for the different thresholds, the results are displayed in terms of [average  $\pm$  standard deviation] for the different thresholds. The results are presented in this way to have insight into the actual performance of the algorithm, instead of focusing on the results for the optimal threshold configuration. This configuration may vary along the different approaches and it is not possible to extrapolate it to other experimental set-ups or datasets. Consequently, we considered that the kind of analysis, focused on the overall algorithm performance that we propose was preferable.

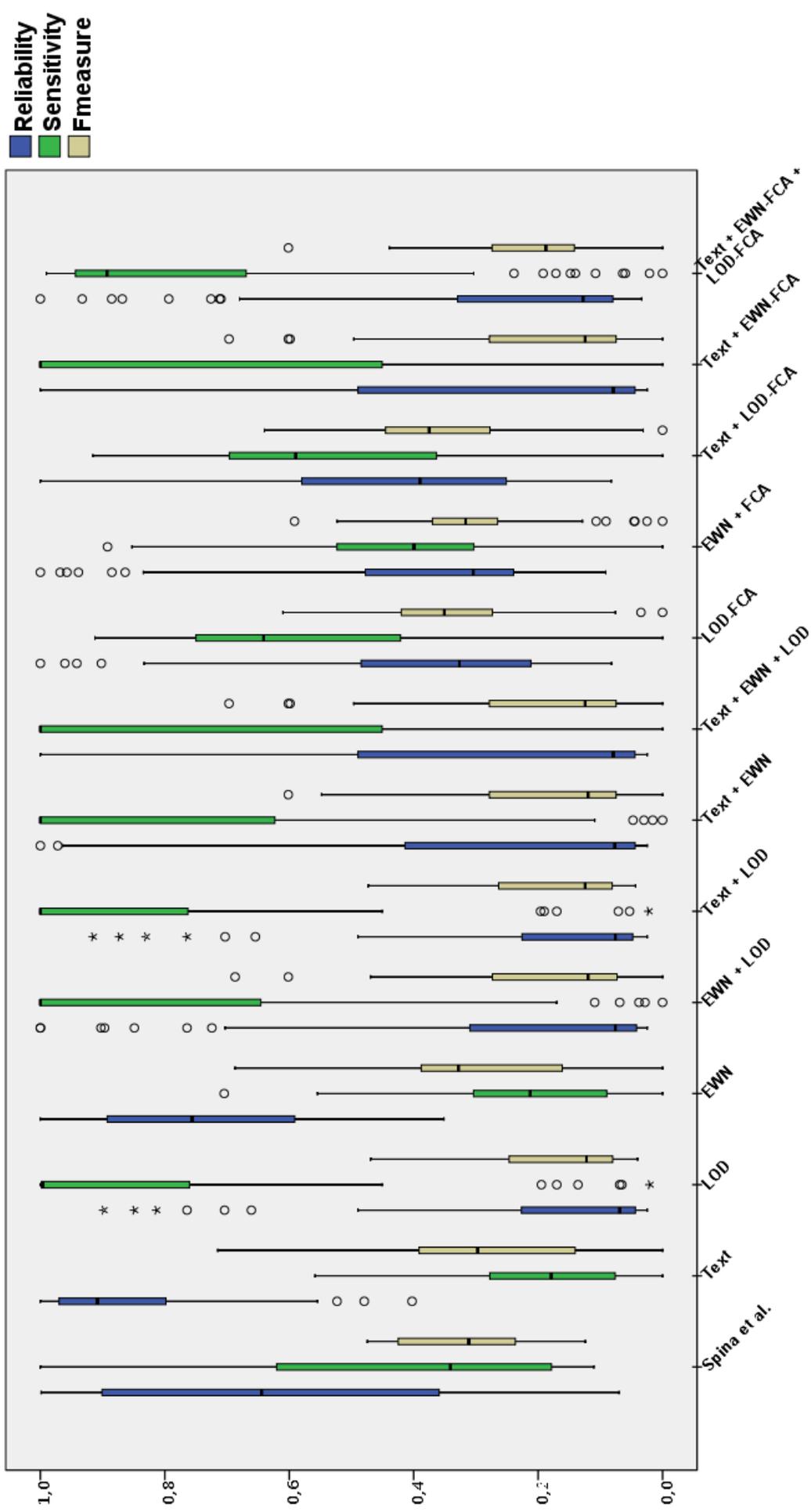
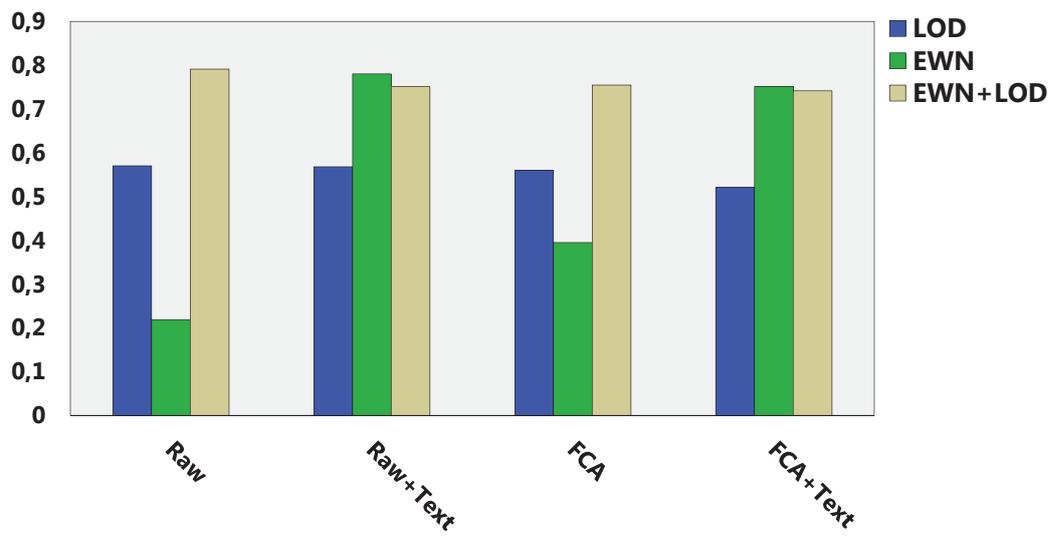
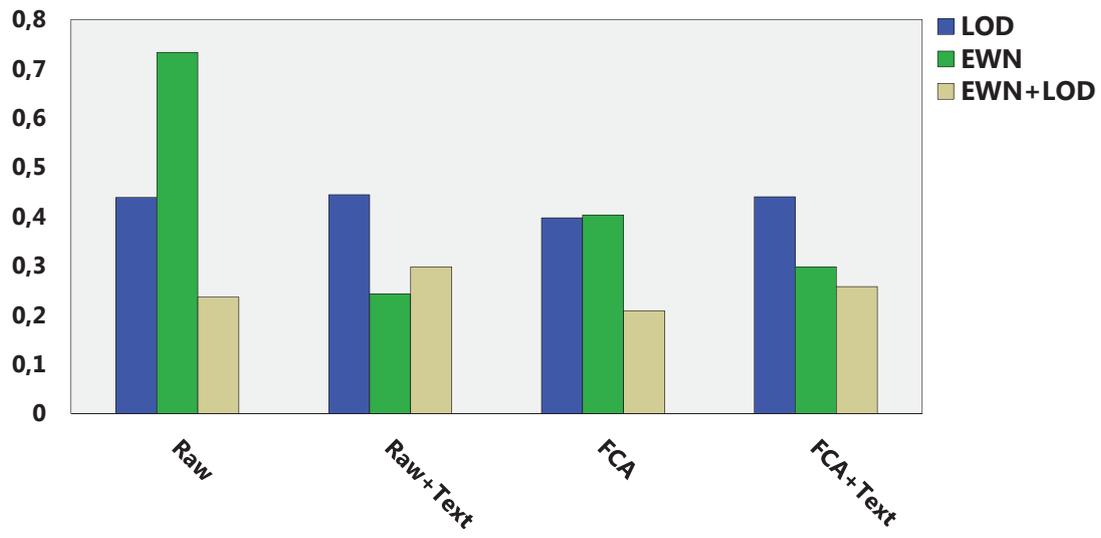
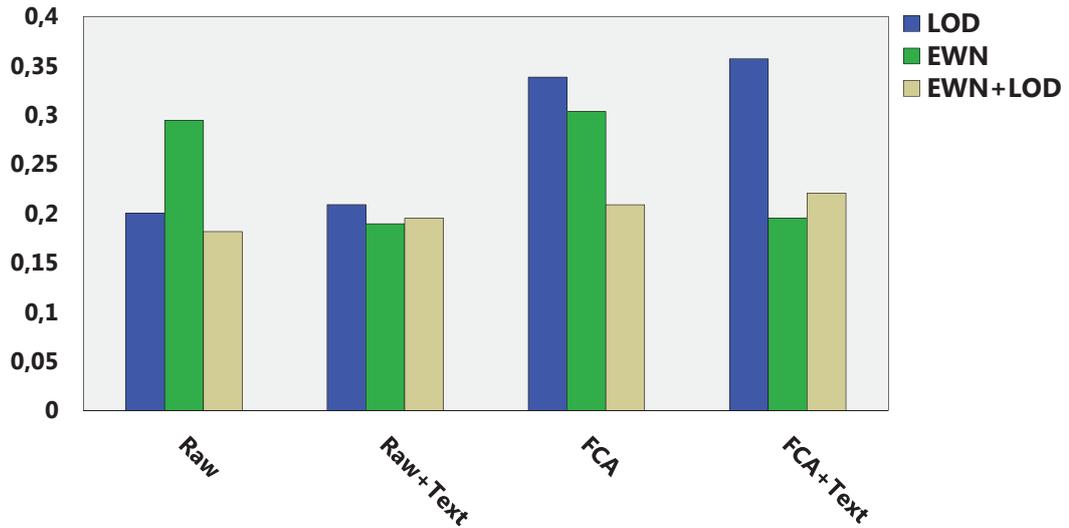


Figure 5.4 – Topic Detection Results





**Figure 5.5** – Topic Detection Results (Reliability, Sensitivity and F-measure from left to right)

The first point to highlight is the high-performance of the text-based representation: text is a strong signal to detect thematically based topics. However, when DBpedia information is added, the performance is lower than in the baseline, either individually (**LOD** in [Table 5.6](#)) or combined with the textual (**Text + LOD** in [Table 5.6](#)) or EWN information (**EWN + LOD** in [Table 5.6](#)). As we hypothesized before, the adding of DBpedia information might add valuable but also noisy information.

As regards EWN information, results are different. The use of raw (i.e. unmodelled) information (**EWN** in [Table 5.6](#)) reaches (and even outperform) the text-based results. It is reasonable being that, EWN data cover the same aspects that textual information, but represented in a more structured way.

Nevertheless, when **EWN** information is combined with textual information (**Text+EWN** and **Text+EWN+LOD** in [Table 5.6](#)), results are lower than those considering both separately. As EWN and textual information cover the same aspects, the application of both kinds of data together leads to a redundant representation. The issue is reflected in the Reliability (precision-based) and Sensitivity (recall-based) results. The redundancy in the representation makes more features (textual- or EWN-based) to be shared by the tweets, even though these new-shared features do not mean that new relationships have been discovered among the tweets. Therefore, the clustering process will create bigger clusters, less precise but with more coverage. This behaviour holds for the rest of the configurations applying text and EWN data together, even when the FCA

models are applied: **Text+EWN-FCA+LOD-FCA** and **Text+EWN-FCA** in [Table 5.6](#).

Focusing on the application of the proposed FCA models, the (re) organization of the LOD and EWN knowledge enabled by FCA improves the topic detection performance for all the configurations, using either textual (**Text+LOD-FCA**, **Text+EWN-FCA**, **Text+EWN-FCA+LOD-FCA**) or non-textual information (**LOD-FCA**, **EWN-FCA** and **EWN-FCA+LOD-FCA**), outperforming the un-modelled **EWN** and **LOD** approaches.

This improvement is much greater in the **LOD-FCA** than in the **EWN-FCA** (see [Table 5.6](#)). At this point it is important to highlight that originally EWN had a more formal and better defined structure than DBpedia. Consequently, the potential improvement in the EWN structure by reorganizing it by means of FCA is expected to be smaller than the improvement achieved by the DBpedia reorganization, whose original organization was not so well defined.

In fact, the performance of the approach using FCA to model DBpedia and textual information (**Text+LOD-FCA**) outperforms the best results obtained in the RepLab Campaign. Moreover, although for some HAC configurations the approach presented in [\[Spina et al., 2014\]](#) performs better than our proposal, taking into account the performance averaged for the different HAC thresholds, our proposal is able to improve that in [\[Spina et al., 2014\]](#), even though the latter proposal applies a supervised methodology.

To sum up, FCA knowledge modelling enables a better data representation, leading to a higher performance in a data organization task such as Topic Detection. Both, linguistic and LOD resources add valuable information not originally covered by the Twitter data. However, the inclusion of the un-modelled information is not enough to improve the data representation; in fact, it seems to add noisy but not valuable information. It is when our FCA-based approach is applied when the most interesting information is extracted, inferred from the organization of the knowledge in the proposed resources (EuroWordNet and DBpedia).

## 5.3 Discussion

---

The initial hypothesis of this section has been confirmed by the experimental results. Where baseline approaches, applying DBpedia or EWN data, were not able to add more valuable information to that already contained in the tweets, the approaches applying the FCA-based models were able to do so. In fact, the best performance is obtained by combining the FCA-modelled LOD information with the textual information. The approaches only applying the FCA-modelled information (either LOD or EWN) without the text were also able to achieve a similar performance. Special attention has to be paid to the combination of Textual and EWN data. Although EWN offers valuable information for the tweet representation, outperforming the text-based approach, the redundancy included when it is combined with textual information results in a less precise, and consequently a worse data representation than when only EWN data is taken into account. As EWN and textual information cover the same aspects, the application of both kinds of data together leads to a redundant representation. In this regard, an important remark is that when knowledge-based information is used to enrich or represent a set of contents, not only is important the information added but also the way in which it is added.

Summing up, the evaluation carried out in this chapter has demonstrated that representations based on higher-level knowledge-based features are preferable to those based on raw textual-based representation. In particular, representations enabled by these higher-level features are able to improve the results presented in the previous section by textual-based representations. Moreover, the results in this section outperform those obtained by state-of-the-art approaches in the task.



# Part III

## RECOMMENDATION PROPOSAL

*“We are leaving the age of information and entering the age of recommendation”*

Chris Anderson in “The Long Tail”

# 6

## Recommendation: From Formal Concepts to User Preferences

---

*This chapter presents the recommendation proposal applying the modelling approach presented in the first part of the thesis to the recommendation scenario. It covers the different experimentations in recommendation, from the first basic models to the final common representation space*

### Content

---

6.1	FCA FOR RECOMMENDATION: PRELIMINARY APPROACH .....	182
6.1.1.	Recommendation Approach.....	182
6.1.2.	Application Scenarios .....	183
6.2	TOWARDS A COMMON CONCEPT-BASED REPRESENTATION SPACE.....	193
6.2.1.	Recommendation Approach.....	196
6.2.2.	Application Scenarios .....	198

---

In the previous sections the performance of FCA for content modelling has been proven. In Section 4, FCA was proposed to create a concept-based representation for text modelling. This proposal was tested in two different experimental scenarios (the RepLab 2013 Evaluation Campaign and the Retrieving Diverse Social Images Task at MediaEval) in which the modelling of textual contents played a crucial role. The results demonstrated the suitability of the proposed FCA-based modelling, improving the state-of-the-art approaches for both tasks.

In Section 5, the same FCA-based approach was then applied for knowledge modelling. When applied to text, FCA was able to create a meaningful representation; thereupon, it was expected that applied to knowledge resources, it would be able to create meaningful knowledge-driven representations. This idea was applied in the context of the RepLab 2013 Evaluation Campaign to improve the tweet representations by means of the information inferred from the FCA-based knowledge-driven representation of the DBpedia data. It was experimentally proven that the FCA-based representation actually enabled a better tweet representation.

In this regard, this section studies the development of a recommendation proposal based on this FCA-based model. Firstly, the first experiments that were carried out in the context of this thesis are presented. This experimentation was based on a basic FCA model (i.e., only modelling items according to their content without considering the user dimension). In this first experimentation, some problems were identified, mainly related to the user-item dimension gap and to the shallow content representation, as it exposed in the review of the literature.

From the conclusions drawn in these first experimentations in section 6.1, the recommendation approach presented in this thesis (at section 6.2) was developed. In particular, this approach delves into the idea of the development of a common representation space for recommendation. As discussed in the review of the recommendation field literature, the existence of a common representation to represent users and items might reduce the representation gap between user preferences and item descriptions. If we are able to join both spaces in a common layer, the recommendation will only need to look for the items closer to a user profile (i.e., in terms of distance in the common space).

Along the different experiments presented at this section, it is going to be proved how this model is able to enhance the recommendation process, outperforming state-of-the-art recommendation systems.

## 6.1 FCA for Recommendation: Preliminary Approach

---

Formal Concept Analysis has previously demonstrated (in the experimentation proposed in chapters 4 and 5 of this work and in others proposed in the literature) to be a suitable content modelling technique. In particular, it has proved to create an abstract conceptual representation of objects according to the relationships among them, described by means of the attributes they share. This representation has been also proved able to discover the latent aspects in the contents (e.g., topics).

In a Content-based recommender system, items and users are commonly represented by textual features (from the basic bag-of-words approaches to the most complicated ones based on statistically-inferred latent representations). In this context, recommendation is understood as the finding of items similar to those already consumed by the users.

Therefore, the first working hypothesis in this context was to test the actual performance of FCA in the modelling of items to be recommended. The rationale was that the conceptual organization of such items might facilitate the identification of interesting recommendations. In this sense, these first experimentations intended to validate this hypothesis. In particular, in the following subsections the specific FCA-based recommendation methodology and the experimentation conducted in this regard are presented.

### 6.1.1. Recommendation Approach

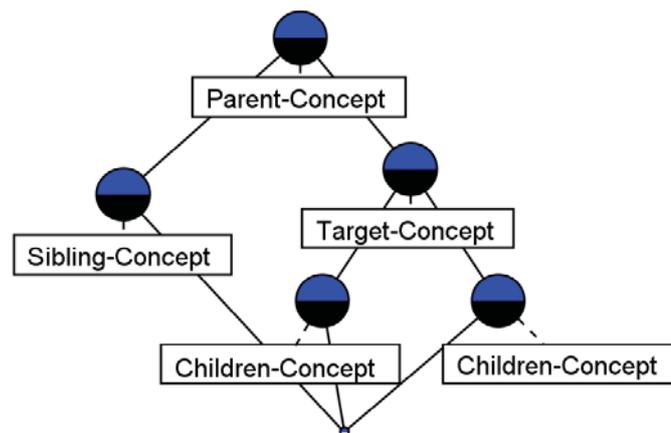
---

The idea followed by the recommendation approach is along the line of that proposed in [du Boucher-Ryan and Bridge, 2006]. The rationale is to take advantage of the relationships represented in the structure of the *concept lattice* to find suitable recommendations.

The algorithm bases its operation on the navigation across the lattice. In more detail, given a target item, the algorithm looks for the most similar items to be offered as recommendations. The navigation process starts at the *object concept* of the target item in the lattice. The *object concept* is selected as the starting point because it is the most specific *formal concept* (that with a larger number of attributes in the *intent* and,

therefore, with more information) in which the target item is included. The rationale is that the more information and more specific about an item, the more accurate the recommendation based on similar items. For example, the recommendation of a movie because it stars the user's favourite actor and directed by the user's favourite director appears to be more interesting than a recommendation because the movie is a comedy and the user likes comedy movies.

Starting at the object concept, the algorithm navigates across the lattice taking those *formal concepts* included in the navigation path. The navigation is carried out by taking the closest *formal concepts* in the lattice structure, which are expected to be those user preferences most closely related to the object concept. In particular, the algorithm uses the *children concepts* (i.e. those linked immediately below in the lattice) and the *sibling concepts* of the target concept (i.e. the *children concepts* of the *parent concepts*, except the *target concept* itself). For an example of these type of concepts, see [Figure 6.1](#).



**Figure 6.1** – Example of Children, Parent and Sibling Concepts of a Target Concept

The navigation is an iterative process, fixed by an  $N$  value that sets the number of levels that the algorithms should visit (up and down). The recommended items are those included in these formal concepts, not still consumed by the user.

### 6.1.2. Application Scenarios

In what follows two different applications scenarios are detailed. In both, the recommendation is based on finding items whose content is similar to the content of the

items already consumed by a given user. Therefore, the aforementioned FCA-based recommendation proposal was applied to find such items.

### 6.1.2.1. Open Recommendation Platform (ORP): News Recommendation

The task is based on the news recommendation in real-time. This scenario is especially challenging. Beyond the specific requirements of the recommendation systems, it includes other challenges: a high response speed, scalability to be able to manage the real-time data stream, the ability to compute recommendation models in real-time to adapt them to the continuous information streams and to integrate them into the recommendation approach.

The particular scenario proposed for the task is the Open Recommendation Platform (ORP), operated by *plista*<sup>12</sup>. ORP provides a framework in which systems can be deployed into a real recommendation environment. This ORP provides an evaluation framework based on a real user study that evaluates the recommender systems by means of the explicit feedback (i.e., whether they click in the recommendations) of the *plista* end users. The evaluation will focus on click events: the absolute number of clicks and the relative number of clicks to recommendation requests.

In what follows it is detailed the experimental test-bed based on the ORP, the results by the proposed recommendation proposal and the conclusions extracted from the achieved results.

## Experimentation

The experimentation is based on the context of the News Recommendation Evaluation Lab (NEWSREEL 2014) [Hopfgartner et al., 2014], in particular in the Task 2: Recommend news articles in real time [Kille et al., 2014]. This task requested the participants to connect their systems to the ORP and to provide recommendations in real time of news reports for users of news portals. These recommendations should be based on the past user profile that the system may have generated and the content of the news report that the user is currently visiting. To that end, after registering with the ORP, the systems receive updates of the new contents available to recommend and the activity of the users with these contents. This information is valid to train or adapt the operation of the recommender systems as well as to know the items that are susceptible to be recommended. In addition, while the ORP offers this real-time information to the

---

<sup>12</sup> <http://orp.plista.com/index>

registered systems, it also requests for recommendations triggered by the interaction of a user with a given news report.

In this regard, our experimentation focuses on testing the performance of the proposed FCA-based approach in this real Content-based recommendation scenario. Besides the aforementioned problems related to the real-time aspect, there is another problem related to this scenario: the system has no previous information on the items or users at the beginning. The system should record all the coming information (new users, new items, new item content and new interactions) in order to compile background information to offer recommendations. To facilitate this point, the organizers also provided an offline dataset, containing user-item interaction for fifteen time slots before the release of the dataset.

Nevertheless, there remains the problem of how much information should the system store. It might be thought that the more information stored, the better the performance of the system. However, this approach has some disadvantages. Firstly, there is the problem of how to manage such a large amount of data. Given that recommendations should be made in real time, it is not possible to compute such an amount of data online: as can be seen in [Figure 6.2](#) to [Figure 6.5](#), the systems have to process between 12 and 35 thousand requests per day. Furthermore, the systems should also store and process the information related to these interactions. Therefore, only offline computation is suitable in order to create a recommendation model. However, even offline computation is too expensive in terms of complexity. In this sense, instead of considering all the data, we propose two approaches: 1) consider the most novel 1,000 items, or 2) consider the 1,000 top-scored items.

Another disadvantage is related to the time dimension. Not all the data have the same importance: it is reasonable to think that the most novel (i.e. the latest data to arrive at the system) or the top scored (i.e. the data most consumed by the users) is more interesting for the users. Taking into account both the complexity problem and the time dimension, we decided take into account only the data belonging to the last previous 24 hours for the recommendation computation (i.e. every 24 hours the previous information is removed and a new FCA model of the new data is generated). Nevertheless, considering that some information could be considered interesting for the users from day to day, the system keeps the most consumed items (the top 100) for the next day.

Once has been agreed in the data to be used, we have applied the recommendation proposal in section [6.1.1](#) for the computation of the FCA-models for recommendation. In order to measure the suitability of our system, we also developed two baseline systems

to show the improvement in the results due to the application of FCA. It is important to note that, for the baselines and for the FCA based approach the input data are the same (the data of the last 24 hours plus the top-scored items).

- **Baseline 1 - Most Novel Items:** Given a recommendation request, this system uses the set of most novel items. That is, the last items that have arrived at the systems are offered as recommendations.
- **Baseline 2 - Top Scored Items:** Given a recommendation request, this system offers the set of top-scored items. The score of an item is set by the number of times that it has been accessed or it has been clicked.
- **Approach 1 - FCA-Based Recommendation – Most Novel:** To fulfil the recommendation requests, the item contained in the request is used to look for similar items in the lattice. In more detail, the algorithm looks for the object concept of the item included in the request. Thereafter, it recommends the closest items in the lattice structure according to the aforementioned proposal (see section [6.1.1](#)). The information to create the concept lattice is updated daily based on the most novel items.
- **Approach 2 - FCA-Based Recommendation – Top Scored:** This approach applies the same methodology than the previous one. The difference relies in the information used to create the concept lattice. The information, daily updated is herein based on the top-scored items.

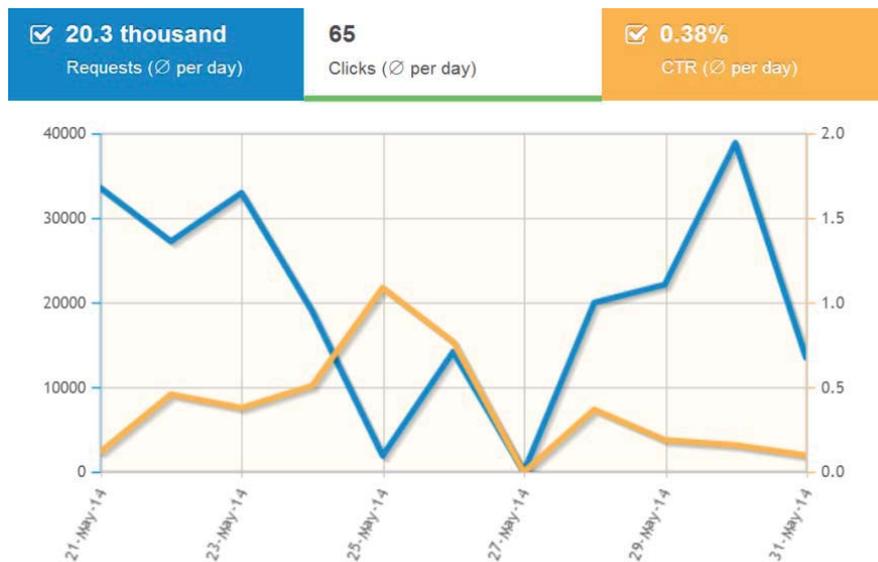
## Results

The proposed recommendation approaches and baselines were deployed for several weeks into the ORP. The results herein presented correspond to the last weeks, when the systems were really in the production phase and the official NEWSREEL evaluation was conducted. These results are provided by the ORP based on the Click Through Rate (CTR) per day; i.e., the percentage of the recommendations offered by the systems in which the users have clicked.

[Figure 6.2](#) and [Figure 6.3](#) show the results of both baselines. Their behaviour is similar, however the results of that based on most novel items outperform those of the approach based on the top scored. It points out that users prefer a novel although not-so-accurate recommendation, at least in the scenario proposed by this task: news recommendation. It does make sense that users are not interested in one-week-old news reports, no matter their content could be. This aspect has been also pointed out by some works in the literature (see section [2.3.2.3](#)).



**Figure 6.2** – Results for the Baseline 1 – Most Novel Items

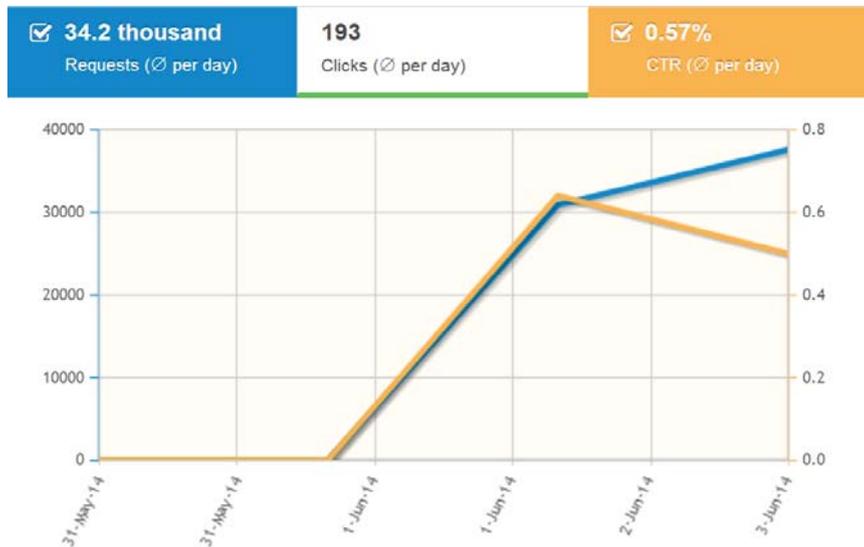


**Figure 6.3** – Results for the Baseline 2 - Top Scored Items

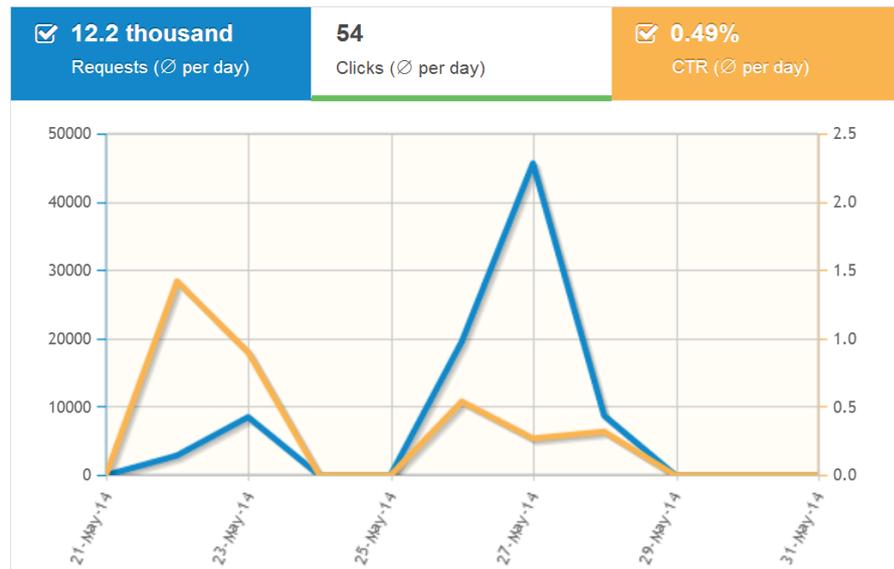
The performance of the FCA-based approach is presented in [Figure 6.4](#) and [Figure 6.5](#). Note that [Figure 6.4](#) has a time deviation with respect to other figures, due to problems with the computation of this approach. However, the results can still be compared to the previous ones given that: 1) the system, 2) the environment and 3) the amount of data processed is the same for all the approaches.

The first point to highlight is that both approaches outperform their respective baselines (compare [Figure 6.4](#) and [Figure 6.2](#) for the most novel-based approaches and compare [Figure 6.5](#) and [Figure 6.3](#) for the top score-based approaches). It is the expected result. It sounds reasonable that the modelling of the content provided by FCA may improve the performance of the non-personalized baseline approaches. Nevertheless, the important point to prove here was whether this improvement justifies the increase in complexity that the FCA-based approach entails. As can be seen in the figures, results are improved by a 20% where the FCA modelling is applied, which seems to justify its application.

Regarding the different information used to compute the model, as was highlighted in the previous results, the most novel-based approach again improves the results of the top scored.



**Figure 6.4** – Result of the Approach 1: FCA Based Recommendation – Most Novel



**Figure 6.5** – Results of the Approach 2: FCA-Based Recommendation – Top Scored

## Discussion

This section presented the first of the preliminary works developed in regards the application of FCA for the recommendation process. This approach intended to bring the proven performance of FCA for content modelling (see section 4 and 5) to the recommendation scenario. The idea behind this application was that if FCA were able to accurately model the content of the items to be recommended, it would facilitate the process of finding the most similar items to a target one. In this way, given a user that is reading a given news report, system should be able to recommend similar ones.

To test this intuition, we made use of the ORP and specifically the recommendation scenario provided by the NEWSREEL Challenge. We experimented with a recommendation approach based on Formal Concept Analysis. To take the special requirements of this real-time environment into account, being able to compute recommendations in the required time, our framework proposes a daily FCA-based item modelling. This item modelling is then used to look for similar items related to the recommendation request.

For the experimentation, two baselines were computed covering two basic proposals: recommending most novel and top-scored items. Based on these two baselines, two FCA-based proposals were presented. The FCA-based results seem to be promising when compared to the baselines. Nevertheless, if we consider the results of the other participants, the FCA-based system does not reach the overall performance [Kille et al., 2014]. In this sense, one main aspect has affected the performance of the FCA proposal: 1) the need to better describe the items to create our FCA-based representation.

To sum up, the viability of applying FCA to such a challenging scenario was demonstrated, even if it is still far from the performance of the state-of-the-art systems in this field. FCA has demonstrated to enable item representations that improve the recommendation process, achieving better results than those obtained by the baseline (unmodelled) approaches. In this regard, the main limitation of our approach is the lack of a proper item description. In this sense, in the following section we have applied our proposal in a scenario where higher-level semantic-based representations are available. In this way, it is expected to address the problem pointed out in this section (i.e., the lack of a proper item description), thus improving the final recommendation results.

#### 6.1.2.2. The ESWC 2014 Recommendation Challenge

This experimentation is focused on the *ESWC 2014 Recommendation Challenge*. This challenge pursues the experimentation through the application of Linked Open Data (LOD) to the recommendation task. In more detail, LOD is applied to generate item descriptions based on semantic features related to the item. For instance, a movie is represented by means of the properties related to the DBpedia entry of the movie (genre, author, director, etc.).

As stated in the previous section, the lack of proper item description strongly affects the recommendation process. Even though, the FCA-based proposal outperformed baseline approaches where no modelling at all was carried out, the results were lower than those of the best-performing systems in the task. In this sense, this experimentation apply the FCA-based proposal to a scenario where a higher-level semantic item description is available.

Focusing on the *ESWC 2014 Recommendation Challenge*, the organizers provided the participants with an experimental dataset and with the definition of three experimental tasks [\[Di Noia et al., 2014\]](#):

- **Task 1:** Predict Missing Ratings
- **Task 2:** Order an item set according to a predicted recommendation score
- **Task 3:** Generate a Top-20 Recommendation List

In more detail, the DBbook dataset is made up of (a) more than 70,000 interactions between users, (b) an item set made up of more than 8,000 books and (c) the DBpedia endpoint of each item in the collection. The DBbook dataset did not include any kind of content data of the items, but the DBpedia endpoints are useful to gather the information in the DBpedia page to enrich the annotation of the items.

Although DBpedia provides a large repository of semantic information about the entities, it also includes a lot of noisy or uninteresting information. This is related to the fact that DBpedia has been created by automatically gathering the information in Wikipedia. For instance, the DBpedia page of a book in the dataset **Peter Pan in Scarlet** ([http://dbpedia.org/page/Peter\\_Pan\\_in\\_Scarlet](http://dbpedia.org/page/Peter_Pan_in_Scarlet)) includes information such as external links, `wikiPageID`, `wikiPageRevisionID` or `coverArtist` among others that do not appear to be useful to describe the item. Therefore, considering the type of item in the dataset (books), only those data considered more related to the interest of a given user to a given item were selected: `abstract`, `literaryGenre`, `country`, `language`, `name` and `subject`.

By taking advantage of this scenario, we have proposed different experimentations that we detail in more detail in the following sections.

### Experimentation

The experimentation carried out focuses on the Task 2. In this task, given an item set to be recommended, the systems have to offer a top-N recommendation list based on this item set. To that end, the recommendation approach detailed in section 6.1.1 has been applied to the items described according to the aforementioned DBpedia data. The official results released by the *ESWC 2014 Challenge* organizers are shown in the Figure 6.6 (ours identified by UNED). This figure shows the results obtained by all the participant groups (in the horizontal axis) after the evaluation period had finished, sorted by the F-Measure value (in the vertical axis).

As can be seen in the Figure 6.6, the results achieved by the FCA-based proposal are within the average performance of the systems participating in the task (note the figure is skewed, there is a bias in the values of the vertical axis which only range from 0.48 to 0.58).

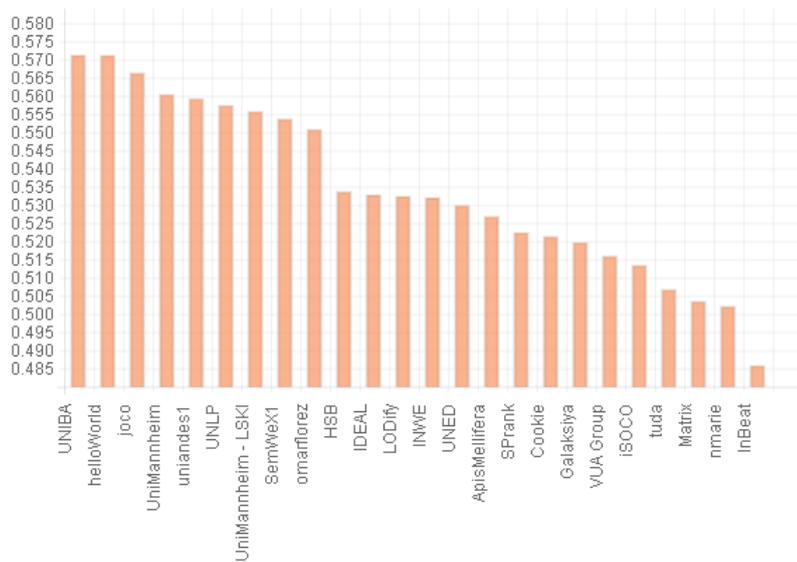


Figure 6.6 – ESWC2014 Interest Prediction Results

## Discussion

From the analysis of the obtained results, it can be concluded that FCA is a suitable technique to be applied in this kind of recommendation scenario. In the previous task, where no accurate item descriptions were available its overall performance was lower than other systems in the literature. In contrast, when such representations are available (as in this section) the results are similar to others in the state of the art.

Nevertheless, the advantages of the inclusion of LOD in the proposed modelling is still not clear. It seems that an accurate conceptual modelling based on LOD information could lead to better recommendations. However, the use of LOD information also entails the inclusion of noisy or unrelated information that can hurt the accuracy of the item representation. In this section, this point was addressed by manually select those properties expected to contain the most interesting information. However, the identification of such information is not always so easy. In any case, it would be desirable to avoid the manual process of selecting the most interesting properties.

To cope with the aforementioned process, the following section details the process related to the modelling of the DBpedia data for the recommendation process, based on the knowledge modelling presented at section 5.

In addition, the experimentation proposed in this section addressed the recommendation task by only modelling the item descriptions to then recommend similar items. In the following section, the proposal of a common representation space for recommendation is going to be introduced. It is therefore expected to improve the recommendation process,

coping with the user-item gap, by taking advantage of the enhanced item representation to generate a common space in which model user and items.

## 6.2 Towards a Common Concept-based Representation Space

---

After the preliminary experimentation, presented in the previous section, a series of problems were identified. As previously hypothesized, they seemed to be related to: 1) the gap between the user and item spaces; and, 2) the difficulties on discovering the most interesting LOD-based item representation.

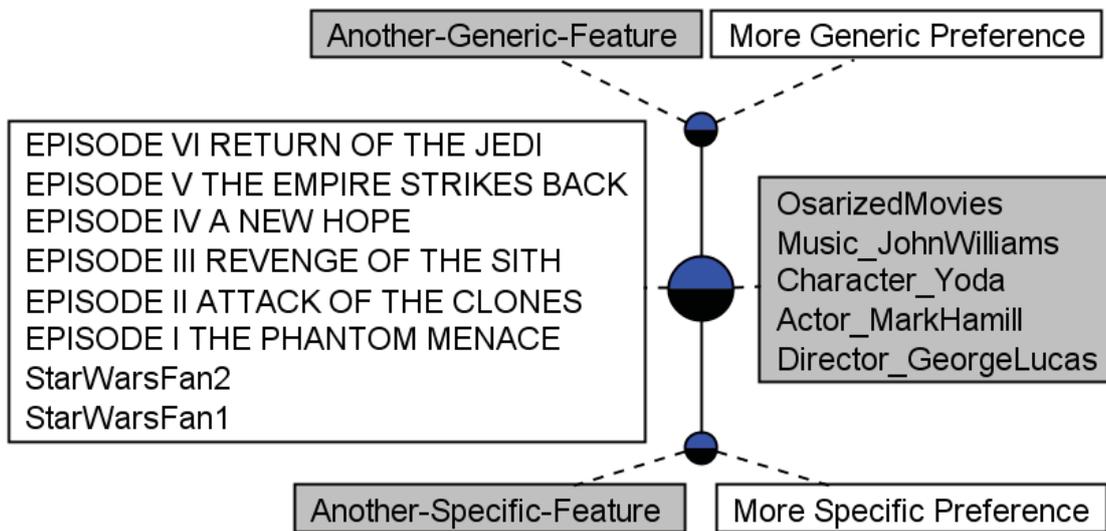
In this regard, this section presents the proposal to create a common representation space that intends to go a step further by proposing a concept-based user and item representation. Applying this approach, items and users will not be described by textual features but by concepts automatically inferred from the content of the item descriptions and the user profiles. In this way, each item or user profile can be understood as a distribution over the set of concepts addressed in their content. It will create a more abstract representation, facilitating the better identification of user-item relationships, which can be understood as user preferences. In addition, the inferred concepts will put together items and user profiles related to them.

In more detail, in the Content-based recommendation scenario, you have a set of users described by a set of features; i.e., those features related to the items already consumed by the user. It is reasonable to think that if they have liked some type of items in the past they will like similar ones in the future. For instance, if a user has consumed two items  $\mathbf{Item}_1 = \{\text{Feature}_a, \text{Feature}_b\}$  and  $\mathbf{Item}_2 = \{\text{Feature}_b, \text{Feature}_c\}$ , they will be represented as  $\mathbf{UserProfile} = \{\text{Feature}_a, \text{Feature}_b, \text{Feature}_c\}$ . In other words, a user profile representation can be seen as an aggregation of item representations. Therefore, given that users and items are represented by means of the same feature set, it seems also reasonable to model them together in the same representation space.

In this scenario, the recommendation context can be interpreted as a bipartite graph, partitioned into objects  $O$  (users and items) and features  $F$  (the features representing the items). Following the FCA theory and the conceptual modelling proposed (see section 3), the bipartite graph can be interpreted as a *formal context*  $\mathbb{K} := (O, F, I)$ , where  $I$  is a

binary relationship that sets that a given object (user or item) can be described by a given feature. From this *formal context*, a set of *formal concepts*  $(A, B)$  can be inferred, where  $A$  is the set of users and/or items sharing the feature set  $B$  (i.e. users/items in  $A$  are described by the features in  $B$ ). This set of formal concepts can be therefore understood as the set of user preferences inferred from the user profiles and the items associated to these preferences.

To exemplify this point, let us consider the following example. The concept lattice in [Figure 6.7](#) has been generated throughout the application of the this proposal. In this lattice you have a *formal concept* (the one in the centre) including into its *intent* (white filled squares) a set of users — StarWarsFan1 and StarWarsFan2 (i.e., the names are only for convenience, you do not know a priori whether they are actually Star Wars fans) — and a set of items (the Star Wars Movies). The *extent* of the *formal concepts* (grey filled squares) includes the features describing these objects (items and users), such as Oscarized Movies or Actor is Mark Hamil. This *formal concept* can be seen as the user preference I like Star Wars, including the set of users that share this preference — the Star Wars fans —; the set of descriptors of this preference — the features in the *extent* — and; finally, the set of items fulfilling this preference — the Star Wars movies.



**Figure 6.7** – Example of a user preference in the form of a formal concept.

Besides of discovering the *formal concepts*, by means of Formal Concept Analysis a lattice representation to organize in a hierarchical structure all of the inferred user-item groups

is generated. It enables the representation of coarse and fine-grained user preferences (i.e., the most generic user preferences are in the top of the lattice structure whereas the most specific ones are on the bottom).

Continuing with the example in [Figure 6.7](#), on top of the I like Star Wars *formal concept* there is another more generic formal concept, related to another more generic user preference. This formal concept is described by a more generic set of features and it only includes those items fulfilling this more generic user preference. This user preference might be, for instance, I like the movies in which Mark Hamil appears. Conversely, below the I like Star Wars *formal concept*, there is a more specific formal concept related to a more specific user preference, described by a more specific feature set and only containing the set of items fulfilling it. For instance, this preference could be I like the Star Wars Movies in which Darth Vader appears.

By generalizing this example to the whole dataset, it is possible to detect the preferences and the related items of the entire user set and modelling them in the concept lattice. The generated concept lattice offers a series of advantages for the recommendation task. Let suppose the modelling shown in the [Figure 6.8](#). Driven by our approach, we are able to:

- **Identify similar user:** `User_K` and `User_J` are related inasmuch they share a common interest of International Politics. In addition, the `Item_7` might be interesting for them since it address this topic.
- **Identify similar items:** `Item_9` and `Item_10` cover the same topics: Videos about the FC Barcelona in the Champions League.
- **Organize items according its specificity:** `Items_8`, related to news reports about Spain, is more generic than `Item_6`, related to Spanish Politics.
- **Identify items related to user profiles:** `Item_7` can be recommended to `User_K` and `User_J`, they are interested in news about International Politics and `Item_7` covers these topics.

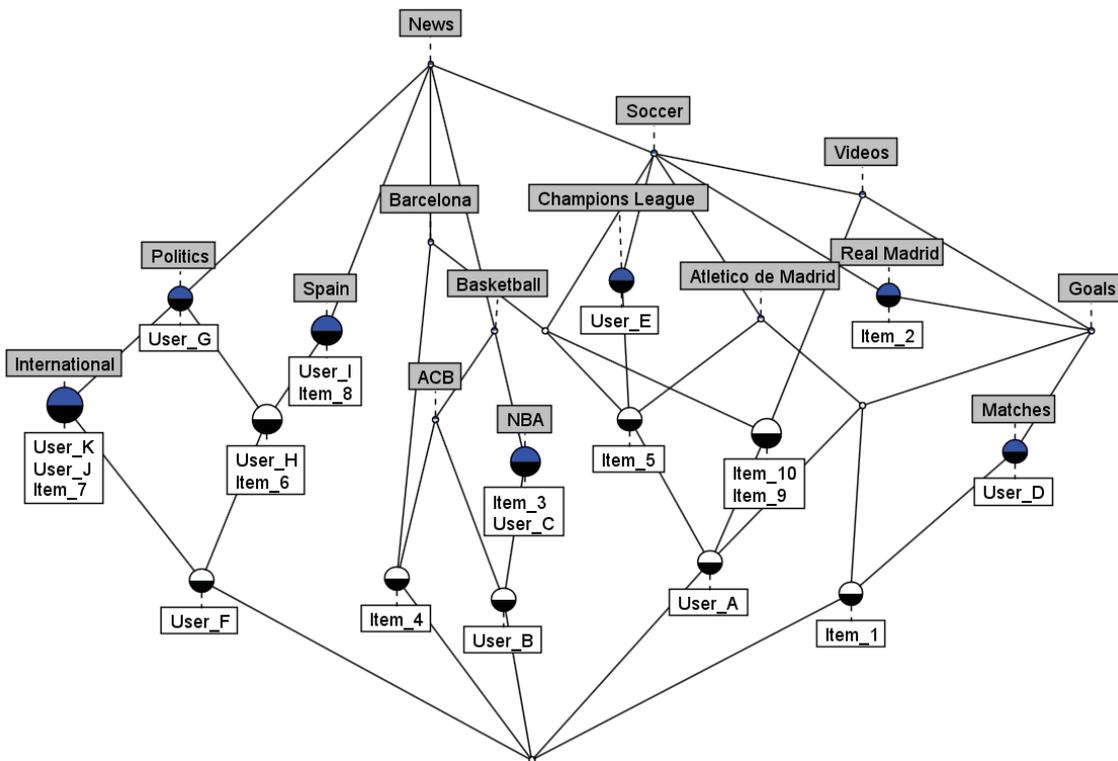


Figure 6.8 – Example of the concept-based common representation

### 6.2.1. Recommendation Approach

This approach is an extension of that presented at [6.1.1](#) in order to deal with the *concept lattice* in the form of a common representation space; i.e., including users and items as the objects of the formal concepts and item features as the attributes that describe them. In this context, as explained in [Figure 6.7](#), the *formal concepts* can be considered as user preferences. Consequently, the recommendation process will be based on taking advantage of these inferred preferences to recommend items fulfilling them.

In particular, given a target user, the algorithm will look for the most specific *formal concept* (i.e., user preference) in which the user appears; that is, the *object concept* ( $\gamma u$ ). As stated in section [6.1.1](#), the most specific formal concept is expected to lead to recommendations more accurate than if general preferences are considered (e.g., I like action movies vs. I like WWII movies). Once identified the *object concept*, the algorithm will take the *formal concepts* in the neighbourhood that are closely related to the *object concept*. To select these neighbour concepts, the navigational process presented in section [6.1.1](#) is applied: select the *children* and *sibling concepts* of the target concept.

Finally, the items in these formal concepts will be offered as recommendations. Each one of these neighbour concepts can be seen as the user preference more related to that represented by the *object concept*. Consequently, the items that these neighbours contain can be also seen as those items fulfilling the user preference.

Based on this rationale, [Figure 6.9](#) includes the pseudo-code of the recommendation algorithm applied for this experimentation.

---

### Algorithm Recommendation Algorithm

---

```

Require: UserSet:  $U = \{User_1, \dots, User_n\}$  not empty. The set of users to which recom-
mend items
for  $User_i \in U$  do
  Get object concept of  $User_i$ : ( $\gamma User_i$ )

   $targetConceptListForUser_i \leftarrow \gamma User_i$ .
  for level = 0 to N do
    for  $targetConcept \in targetConceptListForUser_i$  do
      Get  $children\_concepts([C_1, \dots, C_n])$  of  $targetConcept$ 
      for  $C_i \in [C_1, \dots, C_n]$  do
        if  $User_i \in C_i == \text{false}$  then
          Add  $C_i$  to the  $recoCandidateFCListForUser_i$ 
          Add  $C_i$  to the  $targetConceptListForUser_i$ 
        end if
      end for
    end for
  end for

   $targetConceptListForUser_i \leftarrow \gamma User_i$ 
  for level = 0 to N do
    for  $targetConcept \in targetConceptListForUser_i$  do
      Get  $sibling\_concepts([S_1, \dots, S_n])$  of  $targetConcept$ 
      for  $S_i \in [S_1, \dots, S_n]$  do
        if  $User_i \in S_i == \text{false}$  then
          Add  $S_i$  to the  $recoCandidateFCListForUser_i$ 
          Add  $C_i$  to the  $targetConceptListForUser_i$ 
        end if
      end for
    end for
  end for

  for Formal Concept  $FC = (A, B) \in recoCandidateFCListForUser_i$  do
    for  $item_i \in A$  do
      Add  $item_i$  to the  $RecommendationListForUser_i$ 
    end for
  end for

  return  $RecommendationListForUser_i\{item_1, \dots, item_n\}$ 
end for

```

---

Figure 6.9 – Recommendation Algorithm

Using the concept lattice in [Figure 6.8](#) as example, let us suppose that we want to provide User\_E with recommendations by applying the algorithm in [Figure 6.9](#). Firstly, the algorithm selects the object concept (i.e. that labelled with User\_E in the lattice) as the

starting point for the navigation process. Secondly, supposing a threshold  $N$  (number of levels to go over in the lattice) equals 1, the recommendation will include the items included in the *children concepts* (Item\_5) and the items in the *sibling concepts* (Item\_2). If the threshold  $N$  is greater, taking these objects as the starting point (*sibling* and *children concepts*), the process would be repeated  $N$  times.

### 6.2.2. Application Scenarios

---

This section details the different experimentations conducted in social networks in order to test the common representation space proposal for a recommendation task presented in this chapter. As pointed out in the motivation of this work, this context is especially challenging for the recommendation task, making unfeasible the application of traditional recommendation methodologies. In this regard, this experimentation intends to demonstrate the suitability of the proposal presented in this work in the specific environment of social networks.

In more detail, section [6.2.2.1](#) includes an experimentation focused on recommending news reports to Twitter users. To that end, the user profiles (containing the tweets posted by the users and the news reports that have been read by the users) and the items to be recommended — the news reports — are modelled in the proposed FCA-based common representation space. It is going to be proved that our proposal is able to outperform other 15 state-of-the-art approaches implemented in the developed evaluation platform. In addition, different features (textual and conceptual features) to represent the users have been tested in this scenario. From this comparison between features, it is going to be demonstrated that higher-level features (when available) are preferable to represent contents for the recommendation task (these results go in the same direction than those in section [5](#) for content representation).

With this goal, section [6.2.2.2](#) details our participation in the ESWC 2015 Recommendation Challenge. In this experimentation, the conclusion drawn in previous section [6.2.2.1](#) about the use of higher-level features for the representation of contents is confirmed, and it is also confirmed that our proposal as the best-performing one for the recommendation task.

### 6.2.2.1. Twitter-based News Recommendation

This section intends to evaluate our proposal in a real social environment: Twitter. In particular, the experimental configuration is based on that presented in the paper of [Abel et al., 2011] to evaluate the FCA common representation space for recommendation. To that end, an evaluation platform has been implemented, based on a top-N recommendation scenario [Aggarwal, 2016], including 15 state-of-the-art approaches to frame the performance of the FCA-based proposal. An extensive analysis, focusing on the impact the different features (text, entities and concepts), items (news and tweets) and parameters applied for the experimentation is carried out. This analysis confirms our initial hypothesis: FCA-based approach consistently outperforms the rest of the approaches along the different experimental configurations.

The following subsections detail the dataset applied for experimentation, the task definition, the experimental configuration, the achieved results and their analysis.

#### Dataset

For the experimentation, we have made use of the dataset presented by [Abel et al., 2011] in the context of the 19<sup>th</sup> International Conference on User Modeling, Adaptation Personalization (UMAP 2011) [Konstan et al., 2011]. This dataset (hereinafter called the UMAP Dataset) is publicly available at <http://www.wis.ewi.tudelft.nl/umap2011/>.

In order to generate this dataset, Twitter information streams of more than 20,000 users, who together published more than 10 million tweets, have been crawled. The gathered tweets have been linked to the news articles appearing in them. To that end, more than 60 RSS feeds of prominent news media (e.g., BBC, CNN or New York Times among others) have been monitored and more than 77,000 news articles have been aggregated. 458,566 Twitter messages were linked to these news articles, of which 98,189 relations were explicitly given in the tweets by URLs that pointed to the corresponding news article. The remaining 360,377 relations were obtained by comparing the entities that were mentioned in both news articles and tweets as well as by comparing the timestamps. Summing up, the dataset is made up of a set of users, the tweets posted by these users and the set of news reports appearing in these tweets.

In order to enhance the item representation, thus expecting to improve the Content-based recommendation process, each tweet and news report was semantically enriched to identify topics and entities mentioned in them by means of the Open Calais service<sup>13</sup>.

---

<sup>13</sup> <http://www.opencalais.com/>

Therefore, the items are not only described by their textual content, but also by means of these semantic features (e.g., topics and entities appearing in the tweets).

As noted by [Abel et al., 2011], the Twitter messages per user follow a power-law distribution: the majority of users published less than 100 messages and only a small fraction of users wrote more than 1,000 tweets. Consequently, [Abel et al., 2011] generated a sample of the dataset including the users who posted at least 20 tweets (1,619 users in total). In total, this sample contains more than 2.3 million tweets. In order to replicate the same environment as much as possible, we have used the same subset. In addition, although the dataset has been crawled between November, 2009 and January 2011, in a further analysis of the dataset we discovered that most of the tweets have been published in the time span of November to December 2010. Thus, we are going to limit our experimentation to the tweets in these two months.

### Task Definition

The task is proposed as a TOP-N Content-based recommendation scenario. That is, systems have to predict a ranking of N items for each user. To that end, the systems gather the content of the items previously consumed by the user and try to recommend items whose content are to some extent similar to those already consumed. Although this is the general definition of a top-N recommendation, the specific experimentation scenario addressed in this thesis is slightly different. In particular, in this scenario the following content is available:

- A set of user profiles:

**Definition 8. User Profile:** *The profile of a user  $u \in U$  is made up of the set of items  $i \in I$  that have been consumed by the user, where an item could be a news report  $n \in N$  or a tweet  $t \in T$  (i.e.,  $I = \{N \cup T\}$ ):*

$$P(u) = \{i \in I : r(u, i)\}$$

where  $r(u, i)$  is a binary relationship between the user  $u$  and the item  $i$  that is equal to 1 if the user has consumed the item  $i$  and 0 otherwise.

In particular, three different kind of user profiles are considered according to the item type:

- 1) set of tweets  $t \in T$  posted or retweeted by the users:

$$P_t(u) = \{t \in T : r(u, t)\}$$

- 2) set of news reports  $n \in N$  contained in the tweets posted or retweeted by the users:

$$P_n(u) = \{n \in N \mid \exists t \in P_t(u) : i(t, n)\}$$

where  $i(t, n)$  is a binary relationship that is equal to 1 if the tweet  $t$  includes the news report  $n$  and 0 otherwise.

- 3) the combination of both profiles:

$$P_{nt}(u) = P_t(u) \cup P_n(u)$$

- A set of item models, defined as:

**Definition 9. Item Model:** *The model of a item  $i \in I$  is made up of the set of features related to the item:*

$$M(i) = \{f \in F \mid \exists i \in I = \{N \cup T\} : c(i, f)\}$$

where  $c(i, f)$  is a binary relationships between the item  $i$  and the feature  $f$  that is true if the item contains the feature.

In particular, four different kind of features are proposed:

- 1) **Text:** The text in the items (tweets/news) after stop-words removing.
- 2) **Topics:** The topics detected in the items in the dataset.
- 3) **Entities:** Named entities detected in the items in the dataset.
- 4) **All:** All the previous features together.

- A set of user models, such as:

**Definition 10. (User Model).** *The model of a user  $u \in U$  is a set of features  $F$ , such as for each  $f \in F$  included in the user model, there is an interaction between the user  $u$  and an item  $i \in I$  in the user profile that contains  $f$  in its model. Formally defined, a user model is denoted by:*

$$M(u) = \{f \in F \mid \exists i \in I = \{N \cup T\} : i \in P(u) \wedge f \in M(i)\}$$

In particular, given the different user profiles and features considered, 12 different user models are generated: 3 different item types (tweets, news and tweet+news)  $\times$  4 different features (text, topics, entities and all together).

In this context, the recommendation task as defined by [\[Abel et al., 2011\]](#) (as well as in this work) is defined as:

**Definition 11. (Recommendation Task)** *Given a target user  $u \in U$  to offer recommendation to and taking its User Model  $M(u)$  as input, the recommender systems should then offer a recommendation list  $R(u)$  including the set of news reports  $n \in N$  most interesting for the corresponding user, ordered by relevance  $R(u) = \{n_1 \dots n_k\}$ .*

### Evaluation Set-up

As stated in the Motivation of this work in section 1.1, the focus of this research is the top-N recommendation task. This task aims to provide the users with a ranking of relevant items. In consequence, an evaluation set-up to test the recommendation performance in terms of top-N recommendation is proposed.

To that end, the UMAP dataset (tweets — and the mentioned news — of each user over a span of time) has been split into training and test. The training set is applied to generate the recommendation models and learn the user preferences. The test set is applied to evaluate the recommended news reports. It is considered that a news report has been correctly recommended if there is a user interaction in the test set (i.e., the user has tweeted or retweeted a tweet containing the recommended news report). In order to split the dataset and evaluate it, we have applied  $k$ -fold cross-validation (with  $k=5$ ) i.e., the dataset  $\mathcal{D}$  is randomly split into five mutually exclusive subsets  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_5$ . Thereafter, the recommender system is trained and tested  $k$ -times; each time  $t \in \{1, 2, \dots, 5\}$ , it is trained on  $\mathcal{D} \setminus \mathcal{D}_t$  and tested on  $\mathcal{D}_t$ . Finally, the results are averaged over the five different folds.

By following this set-up, the final performance of the different systems is evaluated according the following two kind of metrics.

### *Precision-based Metrics*

These metrics aims to measure the accuracy of the recommendation results by comparing how many of the recommended items are actually of interest for the user. In particular, these metrics include Precision, Recall and the F-measure of both.

#### *Precision*

Precision (P) is a common metric in many fields related to information retrieval. It measures the probability of a recommended item fulfilling the user preferences. In more detail, precision is the ratio of items correctly recommended in relation to the total number of recommended items:

$$\text{Precision} = \frac{|\{\text{relevant items}\} \cap \{\text{recommended items}\}|}{|\{\text{recommended items}\}|}$$

In our context, we are interested in the precision at different levels of the ranking (i.e., the precision of the recommended items on the first 10 positions of the ranking). Formally, the precision at a given ranking level is equal to:

$$P@K = \frac{|\{\text{relevant items}\} \cap \{\text{first } K \text{ positions}\}|}{|\{\text{first } K \text{ positions}\}|}$$

### *Recall*

This metric is inversely related to the precision and both are commonly reported together in order to understand the overall performance of an information retrieval system. Recall ( $R$ ) measures the probability that of a relevant item to be recommended by calculating the ratio of recommended items which results to be relevant to the total number of relevant items:

$$\text{Recall} = \frac{|\{\text{relevant items}\} \cap \{\text{recommended items}\}|}{|\{\text{relevant items}\}|}$$

As with Precision, we are interested in the Recall value at different positions of the ranking, defined as:

$$R@K = \frac{|\{\text{relevant items}\} \cap \{\text{first } K \text{ positions}\}|}{|\{\text{relevant items}\}|}$$

### *F1-Measure*

This metric is useful to measure systems according to both precision and recall (i.e., it is high only when both precision and recall are high). The F1-measure ( $F$ ) of a system is defined as the harmonic mean of its Precision and Recall. In particular, as in the previous cases, we are interested in the F1 value at different K-ranking positions, defined as:

$$F@K = \frac{2 * P@K * R@K}{P@K + R@K}$$

### ***Ranking-based Metrics***

The recommendation list is offered as a ranking of items, ordered according their relevance. In this sense, the metrics included in this section measure the quality of a recommendation list by evaluating not only the items included in the ranking but also their position in the ranking.

*Success at rank K (S@K)*

This metrics stands for the mean probability that a relevant items appearing within the top-K position in the ranking.

$$S@K = \frac{|\{relevant\ items\}|}{K}$$

It is similar to precision but restricting its coverage to the first K positions of the ranking. In this way, it tries to replicate a real recommendation scenario, where a target user only checks these first results instead of the whole result list.

For more details, the behaviour of this metric, as well as the aforementioned for different results can be consulted at [\[Schröder et al., 2011\]](#), Figure 1.

*Area under the ROC Curve (AUC)*

The ROC curve is a graphical plot that relates the true-positive rate or recall (the relevant items identified as such) to the false-positive rate or fall-out (the un-relevant items identified as such) [\[Fawcett, 2006\]](#). If this curve is plotted in a 2-dimension figure (i.e., recall in the x-axis and fall-out in the y-axis), the area under this curve (AUC) is a measure of how good is a system: the closer to the upper left coordinate (.1) of the figure, the better the system (AUC equal to 1). In contrast, a completely random guess would result in a diagonal line in the figure and (AUC equal to 0.5). The quality of a system in terms of AUC value is roughly as follows:

- .9-1 = excellent (A)
- .8-.9 = good (B)
- .7-.8 = fair (C)
- .6-.7 = poor (D)
- .5-.6 = fail (F)

*Mean Reciprocal Rank (MRR)*

All of the previous measures focus on measuring the number of successfully recommended items in a given recommendation list without considering its order. On the other hand, this and the following measures do take into account this item ranking, assuming that if an item has a higher ranking, it means that it is more relevant for the user. Therefore, recommendation lists with successful recommendation at higher-ranking positions are more desirable. In particular, MRR, as defined by [\[Chakrabarti et al., 2008\]](#) refers to the inverse position of the first relevant item in the ranking:

$$MRR = \frac{1}{rank_{first\ relevant\ item}}$$

*Mean Average Precision (MAP)*

MAP provides a single-figure measure of quality across recall levels [Beitzel et al., 2009]. Formally defined is the mean of the Average Precision (AP) for each recommendation list (i.e., for the recommendation list of each of the  $q$  users to which recommend items).

$$MAP = \frac{\sum_{q=1}^Q AP(q)}{Q}$$

where the  $AP$  is equal to the average for the precision at each “seen” relevant item in the recommendation list:

$$AP = \frac{\sum_{i=1}^R \frac{rel_i}{rank_i}}{R}$$

where  $R$  is the number of relevant documents for the user and  $rel_i$  is equal to 1 if the item is relevant for the user and 0 otherwise.

*Normalized Discount Cumulative Gain (NDCG)*

NDCG evaluates a recommendation list based on the grade of the recommended items [Järvelin and Kekäläinen, 2002]. It is based on the concept of Discounted Cumulative Gain (DCG) that measures how much the overall quality of a given ranking improves by the appearance of a document with a given relevance (that offered by the recommendation algorithm) in a given ranking position. Formally defined, DCG is equal to:

$$DCG_k = rel_1 + \sum_{i=2}^k \frac{rel_i}{\log_2 i}$$

where  $k$  is the number of recommended items,  $rel_1$  is the relevance of the first item in the ranking and  $rel_i$  is the relevance of the item at the  $i^{th}$  ranking position if the item is relevant for the user and 0 otherwise. However, this measure entails the following problem: if any recommendation algorithm consistently offers a high relevance value to all the recommendation items, it will result in better DCG values no matter what its quality may be. To deal with this issue, NDCG applies a normalization of the DCG value thus:

$$nDCG_k = \frac{DCG_k}{IDCG_k}$$

where  $IDCG_k$  is the optimal  $DCG$  value until the  $k^{th}$  ranking position, i.e., all the recommended items are relevant.

## Performed Experimentation

Based on the previously presented test bed, different experimental configurations are proposed with which compare our recommendation proposal. In what follows, these configurations, as well as the results are presented.

### *Configurations for the experiments*

This section presents the different experiments proposed to validate our proposal, applying the UMAP dataset and the aforementioned task definition. In order to frame the performance of our FCA-based proposal, different recommendation baselines and state-of-the-art algorithms, detailed in the following subsections, have been implemented.

In more detail, section [Content-based Approaches](#) includes 4 different approaches that make use of the Content-based features contained in the UMAP Dataset to carry out the recommendation process. Section [State-of-the-art Recommendation Algorithms](#) details 11 different recommendation proposals, which go from basic non-personalized recommendation approaches to some of the most sophisticated techniques in the state-of-the-art. Finally, section [FCA-based Recommendation](#) presents the results achieved by the FCA-based common space recommendation approach proposed in this research.

### *Content-based Approaches*

This section details the different Content-based baselines that make use of the experimental environment defined in the [Task Definition](#) section. This section details the different Content-based baselines that make use of the experimental environment defined in the Task Definition.

In more detail, the algorithm originally proposed by the authors of the UMAP Dataset, called **Abel et al. Proposal**, have been implemented. Three baseline proposals of Content-based recommenders have been also implemented. These three approaches are based on a methodology applying the K-Nearest Neighbour paradigm over the item model features; i.e., the recommendation is made by finding the K-most similar items to a given one. The difference between these three approaches relies on the similarity

measure applied to find these  $K$ -most similar items. The four approaches — Abel et al. and the three baselines — are detailed in the following:

- **Abel et al.:** This approach applies the Content-based recommendation proposed in the work of [Abel et al., 2011] in which the dataset used for experimentation was presented and analysed. They also proposed different user profiles and item models to create different types of user model. The recommendation is then made by applying a lightweight Content-based algorithm that recommends the news reports more similar to the user models. This process can be seen as a search-ranking problem, the user model being interpreted as query. In particular, the algorithm is formalized as follows:

**Definition 12. Abel et al. Recommendation Algorithm:** *Given a User Model  $M(u)$ , including a set of features  $f$ , in vector representation and the set of models  $M(n)$ , also including a set of features  $f$ , related to the candidate news reports to be recommended  $N = \{M(n_1), \dots, M(n_n)\}$  represented using the same vector representation, this algorithm ranks the candidate news reports according to their cosine similarity to  $M(u)$  defined as:*

$$sim_{cosine}(u, n) = \frac{|M(u) \cap M(n)|}{\sqrt{|M(u)|} \cdot \sqrt{|M(n)|}}$$

- **baseline-CB-cosine:** This algorithm is based on the  $k$ -nearest neighbour algorithm, which is applied, given an item, to find the most similar ones according to their feature representations. The recommendation is thus made, given a user profile containing the set of items consumed by the user, by offering those news reports more related to these consumed items. In particular, the recommendation process is formalized as follows:

**Definition 13. Cosine Recommendation Algorithm:** *Given the set of news report representations ( $N = \{M(n_1), \dots, M(n_n)\}$ ), a matrix of the  $k$ -nearest neighbours  $K(n \times k)$  is generated, which includes the set of news reports  $n \in N$  in the rows and the  $k$ -most similar news reports to each  $n$  in the columns (i.e., given a row  $i$  related to a given news report  $n_i \in N$ , then the elements in this row  $k_{i,*}$  will be the most similar news reports to  $n_i$ ). This similarity is based on the similarity between the features  $f$  in the news report models  $M(n)$ , calculated by means of the cosine similarity, defined as:*

$$sim_{cosine}(M(n_j), M(n_i)) = \frac{|M(n_j) \cap M(n_i)|}{\sqrt{|M(n_j)|} \cdot \sqrt{|M(n_i)|}}$$

Once the matrix of  $k$ -nearest neighbours  $K(n \times k)$  is calculated, the recommendation for a target user  $u \in U$  will be the rank of most similar news reports, the similarity given a news report  $n \in N$  to the user  $u$  being the normalized number of times that the features included into the models of the nearest neighbours of  $n$  ( $NN = \{nn_1, \dots, nn_k\}$ ) appears in the user model  $M(u)$ , such as:

$$sim(u, n) = \frac{|f \in M(nn) : nn \in k_{n,*} \wedge f \in M(u)|}{k * |M(u)|}$$

- **baseline-CB-jaccard:** It applies the same rationale that the previous approach, but using instead the Jaccard similarity to create the  $k$ -nearest neighbour matrix. This algorithm is formally defined as:

**Definition 14. Jaccard Recommendation Algorithm:** Given the set of news report representations ( $N = \{M(n_1), \dots, M(n_n)\}$ ), a matrix of the  $k$ -nearest neighbours  $K(n \times k)$  is generated, which includes the set of news reports  $n \in N$  in the rows and the  $k$ -most similar news reports to each  $n$  in the columns (i.e., given a row  $i$  related to a given news report  $n_i \in N$ , then the elements in this row  $k_{i,*}$  will be the most similar news reports to  $n_i$ ). This similarity is based on the similarity between the features  $f$  in the news report models  $M(n)$ , calculated by means of the jaccard similarity, defined as:

$$sim_{jaccard}(M(u), M(n_i)) = \frac{|M(u) \cap M(n_i)|}{|M(u) \cup M(n_i)|}$$

Once the matrix of  $k$ -nearest neighbours  $K(n \times k)$  is calculated, the recommendation for a target user  $u \in U$  will be the rank of most similar news reports, the similarity given a news report  $n \in N$  to the user  $u$  being the normalized number of times that the features included into the models of the nearest neighbours of  $n$  ( $NN = \{nn_1, \dots, nn_k\}$ ) appears in the user model  $M(u)$ , such as:

$$sim(u, n) = \frac{|f \in M(nn) : nn \in k_{n,*} \wedge f \in M(u)|}{k * |M(u)|}$$

- **baseline-CB-pearson:** It applies the same rationale as the two previous approaches, but using the Pearson similarity instead to create the  $k$ -nearest neighbour matrix:

**Definition 15. Pearson Recommendation Algorithm:** *Given the set of news report representations ( $N = \{M(n_1), \dots, M(n_n)\}$ ), a matrix of the  $k$ -nearest neighbours  $K(n \times k)$  is generated, which includes the set of news reports  $n \in N$  in the rows and the  $k$ -most similar news reports to each  $n$  in the columns (i.e., given a row  $i$  related to a given news report  $n_i \in N$ , then the elements in this row  $k_{i,*}$  will be the most similar news reports to  $n_i$ ). This similarity is based on the similarity between the features  $f$  in the news report models  $M(n)$ , calculated by means of the pearson similarity, defined as:*

$$\rho_{M(u),M(n_i)} = \text{sim}_{\text{pearson}}(M(u),M(n_i)) = \frac{\text{cov}(M(u),M(n_i))}{\sigma_{M(u)}\sigma_{M(n_i)}}$$

*Once the matrix of  $k$ -nearest neighbours  $K(n \times k)$  is calculated, the recommendation for a target user  $u \in U$  will be the rank of most similar news reports, the similarity given a news report  $n \in N$  to the user  $u$  being the normalized number of times that the features included into the models of the nearest neighbours of  $n$  ( $NN = \{nn_1, \dots, nn_k\}$ ) appears in the user model  $M(u)$ , such as:*

$$\text{sim}(u,n) = \frac{|f \in M(nn) : nn \in k_{n,*} \wedge f \in M(u)|}{k * |M(u)|}$$

These three baselines have been applied to the 12 different user models considering the three different types of user profiles (tweets, news and both) and the different kinds of item models (text, concepts, entities, and all together) presented at section [Task Definition](#). In addition, different neighbourhood sizes, ranging from  $K=10$  to  $K=100$ , have been tested.

#### *State-of-the-art Recommendation Algorithms*

In order to put into context the performance of the FCA-based proposal with respect to not only that of the proposal of [Abel et al., 2011](#) but also to the state-of-the-art of the recommendation task, several algorithms reported in the literature have been implemented in the aforementioned experimental platform. To that end we have make use of the MyMediaLite Recommender System Library proposed by [Gantner et al., 2011](#). These algorithms are presented in detail below:

- **Random:** This baseline is based on a random guess. Given a target user, the system randomly generates a score for each item in the collection from a uniform distribution, offering the top- $N$  ranked items as recommendations. This approach does not apply any logic or information. Consequently, it can be considered as

the most basic baseline. Therefore, it sets the minimum performance that systems should overcome in order to consider that they are actually learning user preferences.

**Definition 16. Random Recommendation Algorithm:** *Given a target user  $u$  and a set of candidate news reports to be recommended  $N = \{n_1, \dots, n_n\}$ , this algorithm ranks the candidate news reports according to their similarity to user  $u$ , which is randomly generated from a normal distribution:*

$$sim_{random} = x : x \in X \sim \mathcal{N}(\mu, \sigma^2)$$

- **Most Popular:** This baseline does not consider user preferences. It only recommends the most popular news reports. Even though it is also a basic approach, it might offer good results in the proposed dataset, i.e., it is expected that a reduced number of “popular” news reports concentrate numerous likes. Consequently, the recommendation of these most popular news reports might result in good performance. In any case, any reasonable recommendation algorithm considering the user preferences should outperform this baseline.

**Definition 17. Most Popular Recommendation Algorithm:** *Given a target user  $u$  and a set of candidate news reports to be recommended  $N = \{n_1, \dots, n_n\}$ , this algorithm ranks the candidate news reports according to their popularity (i.e., number of times the news report has been consumed). Specifically, if  $\mathcal{M}$  denotes the user-news interaction matrix, the popularity of a news report is defined as:*

$$w_{n_i} = \sum_u \mathcal{M}_{u,n_i}$$

- **Bayesian Personalized Ranking:** This approach is an implementation of the Bayesian Personalized Ranking (BPR) proposed by [\[Rendle et al., 2009\]](#). This methodology focuses on the implicit feedback scenario and it is based on a general optimization criterion of the area Under the ROC curve (AUC), derived from the maximum posterior estimator for an optimal ranking.
- **Weighted Bayesian Personalized Ranking:** This approach proposes an extension of BPR with frequency-adjusted sampling.
- **Multicore Bayesian Personalized Ranking:** This is an extension of the BPR approach that is applied on multiple cores.
- **Soft Margin Ranking Matrix Factorization:** This approach applies a Matrix factorization model optimized for a soft margin ranking loss using stochastic

gradient descent. For more information on this process, please refer to the work of [Weimer et al., 2008].

- **Hybrid:** This proposal is based on a graph structure represented as a binary adjacency matrix of users and objects [Zhou et al., 2010]. This structure is used to create the recommendation list by means of a hybrid methodology that combines accuracy- and diversity-related metrics by applying a heat spreading (HeatS) algorithm.
- **Latent Dirichlet Allocation:** This approach is based on the application of the well-known Latent Dirichlet Allocation (LDA) to the recommendation process; more specifically the proposal presented by [Griffiths, 2002], based on the application of Gibbs Sampling. In particular, LDA is applied to cluster items together. Thereafter the recommendation process will be based on offering items in the same cluster to those already consumed by the users.
- **Learning to Rank Matrix Factorization (LRMF):** This approach applies the proposal of [Shi et al., 2010] that combines a list-wise learning-to-rank algorithm with matrix factorization (LRMF).
- **Rank ALS:** This approach applies the Alternating Least Squares (ALS) for Personalized Ranking proposed by [Takács and Tikk, 2012]. This methodology defines a ranking objective function without sampling to then use ALS for optimizing
- **Rank SGD:** This approach presented by [Jahrer and Töscher, 2012] proposes a ranking based method that tries to model the users' choice between item pairs implemented as the minimization of an objective function.

#### *FCA-based Recommendation*

This section details the FCA-based proposal based on the Common Space representation created through the application of Formal Concept Analysis proposed in this thesis. In particular, to build the common representation space, we have the formal context  $\mathbb{K} := (G, M, I)$  related to the recommendation scenario, where  $G$  is the set of users and items,  $M$  includes the set of features related to both users and items and  $I = \{G \times M\}$  is a binary relationship that is true if the user or item  $g \in G$  is related to the feature  $m \in M$ . In order to test the performance of our proposal for the different configurations, the 12 different user models in section [Task Definition] have been applied and, consequently, 12 different concept lattices (a.k.a. common spaces for recommendation) have been generated and evaluated.

### *Results at a glance*

This section includes the results of the different experimental approaches proposed in the aforementioned section. Given the large amount of results, in order to facilitate the reading, these results are included as supplementary materials in the [Annex I](#). In what follows, the results of the experimental configurations, divided into the three categories proposed in section [Performed Experimentation](#), are detailed.

#### *UMAP 2011 Approaches*

This section details the results of the approaches presented at section [Content-based Approaches](#). In particular, [Table Annex 1. 1](#) and [Table Annex 1. 2](#) detail the results for the baseline applying Cosine as similarity measure. [Table Annex 1. 1](#) includes the results for the precision-based metrics (Precision, Recall and F-measure) at different cut-off points (from 5 to 100) and [Table Annex 1. 2](#) the results of the ranking-based metrics (Success@K, AUC, MAP, NDCG and MRR). Similarly, [Table Annex 1. 3](#) and [Table Annex 1. 4](#) include the precision-based and ranking-based metrics for the approaches using Jaccard as similarity. Finally, [Table Annex 1. 5](#) and [Table Annex 1. 6](#) include the results for the Pearson Baseline.

The format of the tables is the same for the three different approaches. The results are divided into 12 different groups, one per user model, according to the different features and items applied to generate the models. Each one is named according the type of item from which the content of the model comes (News, Tweet and News and Tweets) and to the feature used to describe these items (Text, Topics, Entities and All). For instance, the model called News And Tweet All is generated by gathering all the information (text, topics and entities) appearing in the news and tweets related to a user. Conversely, the Tweet Text approach is created by using only the textual information in the tweets related to a user. The tables also show the values achieved by these different models for the proposed metrics in relation to the value of the neighbourhood size  $K$  from 10 to 100.

#### *Collaborative-Filtering Recommendation Algorithms*

In this section, the results of the different algorithms presented in section [State-of-the-art Recommendation Algorithms](#) are included. In particular, [Table Annex 1. 9](#) details the results of the state-of-the-art approaches for the Precision, Recall and F-measure and [Table Annex 1. 10](#) the results for the ranking-based measures. Given that these approaches do not make use of the Content-based information, there are no different

results for the different items and features as in the previous section. The neighbourhood size does not apply either to these results.

#### *FCA-based approach*

This section details the results of the FCA-based recommendation proposal. [Table Annex 1. 11](#) include the precision-based results and [Table Annex 1. 12](#) the ranking-based results. As described in section [FCA-based Recommendation](#), the results are presented according to the different models applied (based on the type of item and type of feature) and the different threshold values applied to carry out the FCA computation from (0% to 50%). The threshold parameterized the degree of reduction applied to the *formal context*; i.e., a threshold equal to 10% means that only those attributes appearing in more than 10% of the items would be considered to describe the items.

### **Result Analysis**

The different proposals, configurations, parameters and features have given rise to an overwhelming number of results (as can be seen in the Tables in the previous section contained in the [Annex I](#)). Therefore, in order to digest these raw results and facilitate their analysis and understanding, this section provides a more detailed and specific analysis of the results, according to different aspects.

In more detail, section [Overall performance](#) presents the comparison of the different Content-based, Collaborative Filtering algorithms and our FCA-based proposal in terms of their overall performance. This section aims to analyse, in general terms, which are the best methodologies and the best configurations of these methodologies to address this particular recommendation task.

Section [Analysis on the impact of the different features](#) focuses on the analysis of the Content-based approaches. In particular, this section analyses the impact on the performance of these algorithms of the features used to represent the user models and the items (text, entities, topics). As regards these items, section [Analysis on the impact of the different type of item](#) analyses the impact on the recommendation performance of the use of the different types of item applied to create the user models (tweets, news reports and both together).

Since some of the algorithms rely on some parameters to adapt their computation, section [Analysis on the impact of the different Content-based parameters \(K and Threshold Values\)](#) analyses the impact of these parameters on the final performance of those algorithms.

Finally, section [Take-Home Points](#) summarizes the most important points extracted from the analysis of the experiments conducted.

### *Overall performance*

[Figure 6.10](#) to [Figure 6.14](#) show the overall results in terms of Precision, Recall and the F-measure of both. Results are presented in three-dimensional figures with the different approaches in the X-axis, the different precision cut-offs in the Z-axis and their values in the Y-axis. In addition, there are 13 different bars for each approach and precision cut-offs related to the different features applied to creating the user models. Among these Content-based results, the different baselines (Cosine, Jaccard and Pearson) are averaged over the different k-values and our FCA-based approach is averaged over the different threshold values. In the case of the approaches based on Collaborative Filtering methodologies, which does not make use of the different Content-based features, there is only one bar (i.e., All).

In brief, [Figure 6.10](#) includes the Precision results at different cut-off points (from 5 to 50), [Figure 6.11](#) the Recall results and [Figure 6.12](#) the F-measure results at these cut-off points. Some important remarks have to be noted:

- Approaches making use of the content of the items achieves, in general, better results (in terms of the three measures). In this regard, it is only the **Hybrid** approach that achieves similar results to those of the Content-based approaches.
- Among these Content-based approaches, it is the **FCA-based** proposal that offers the best results for the different measures and cut-off points. The approach proposed by [\[Abel et al., 2011\]](#) achieves the second best results.
- Taking into account the different baselines, no big differences are observed.
- No different or unexpected behaviours as regards the cut-off points and measures are observed among the different features or approaches. In general, the best precision is achieved by the Precision at 5 and at 10. The larger the number of items taken into account, the worse the results. Recall values increase throughout the number of items in the ranking, as expected (i.e., it is more likely to include more relevant recommendations, in absolute values if more recommendations are included in the ranking). Finally, the most interesting point is to analyse the F-measure values, which balance the precision and recall values, to get a clearer insight into the algorithm performance. In this sense, the best values are obtained

when F1@30 to F1@50 values, in particular F1@40, are considered. Although precision gets worse, the larger increase in Recall for these cut-off values explains these results.

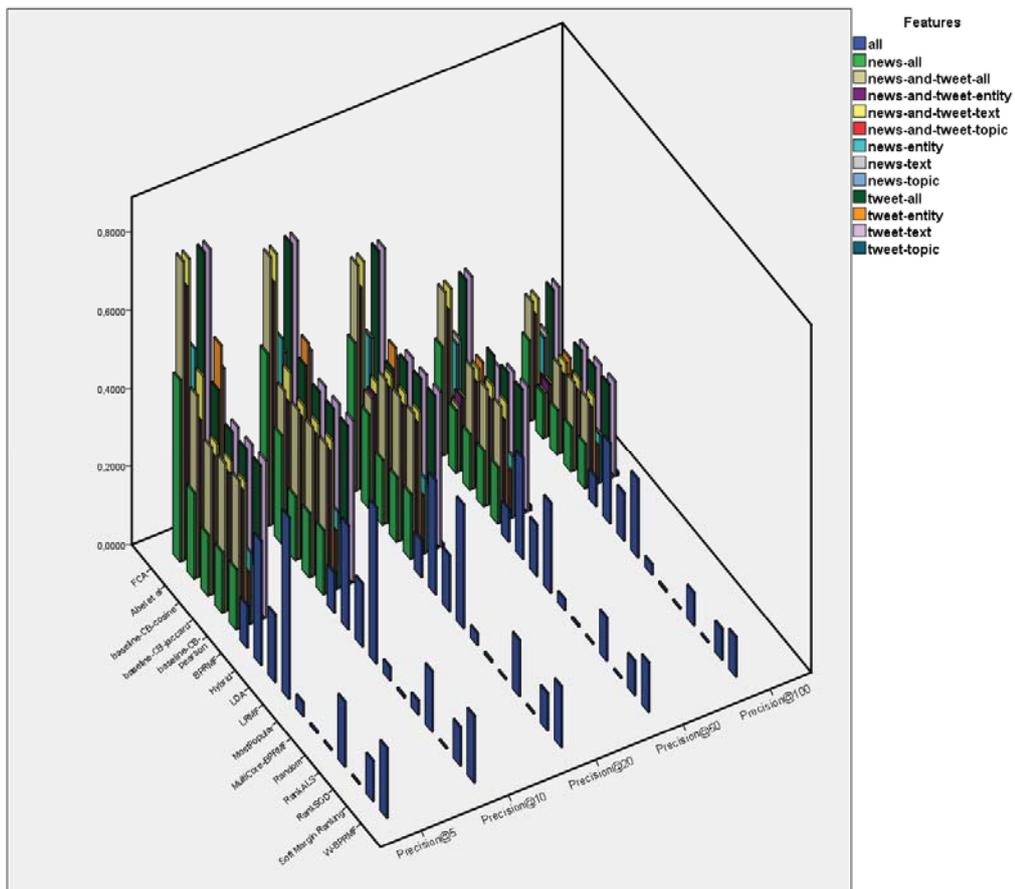


Figure 6.10 – Precision-based Overall Results

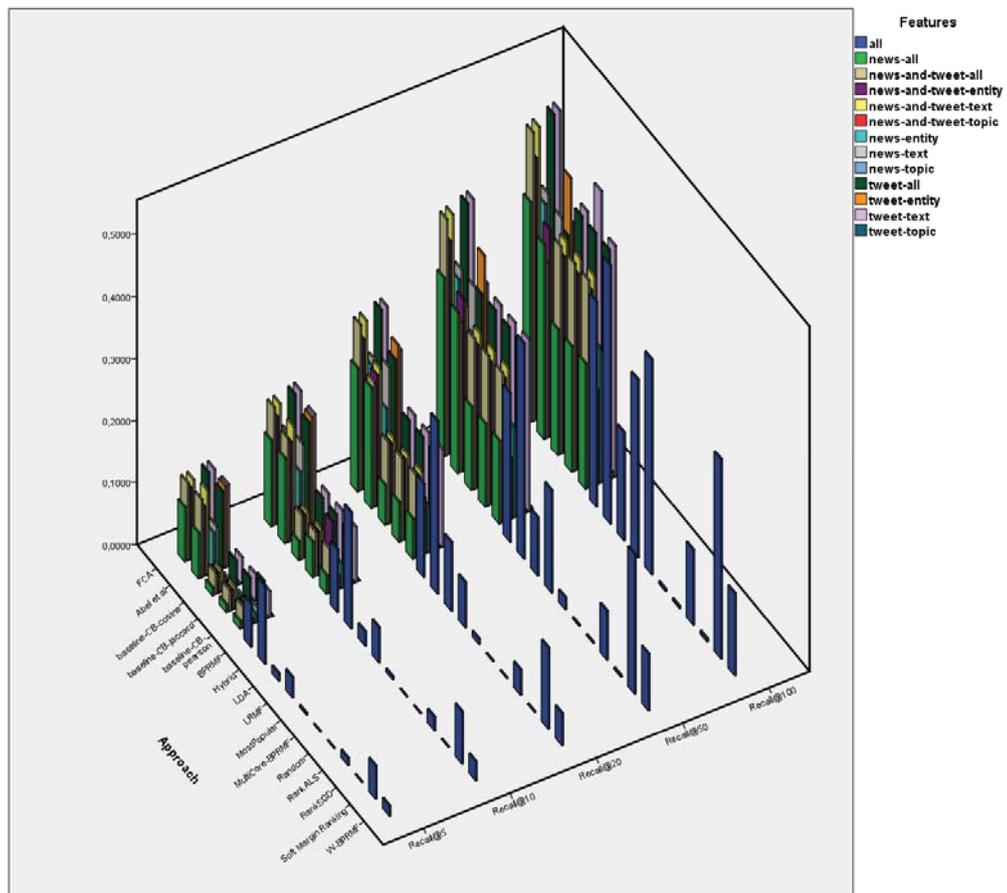


Figure 6.11 – Recall-based Overall Results

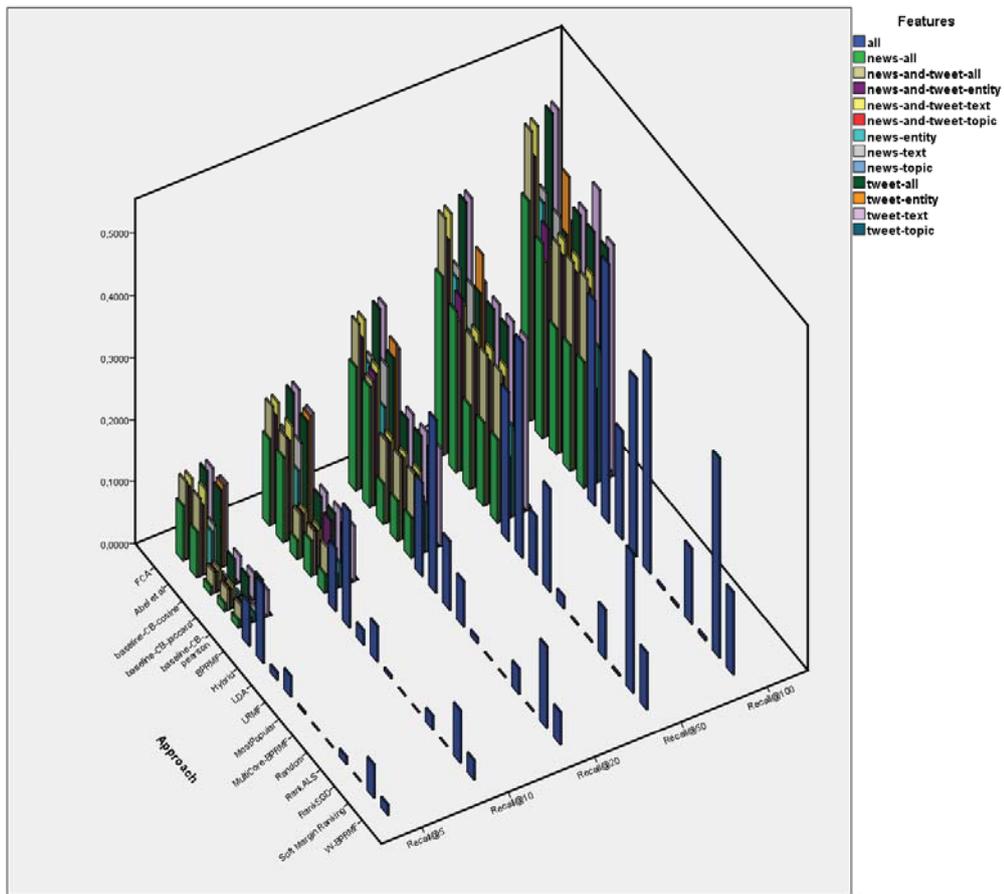


Figure 6.12 – F-measure-based Overall Results

Taking into account the ranking-based measures, and the Success at K measure in particular (in [Figure 6.13](#)), the results become more blurry. The differences between Content-based and the other approaches do not hold, at least not for all of them. For example, **Hybrid**, **BPRMF**, **LDA** and **LRMF** achieve similar results to those of Content-based approaches.

Although, the best results in general are still those of the **FCA-based** approach, the approach of [\[Abel et al., 2011\]](#) offers similar results, even better for some configurations and consequently no significant improvement can be seen. One interesting aspect is that, even at the cut-off point equal to 5, many approaches generates very accurate results, which in terms of this measure means that at least one relevant item has been offered as a recommendation in the first 5 results. Beyond that, the behaviour of the different approaches regards this measure is as expected: the larger the ranking, the better the results.

Finally, [Figure 6.14](#) shows the results in terms of the rest of the measures (i.e., Area Under the ROC Curve, Normalized Discount Cumulative Gain, Mean Average Precision and Mean Reciprocal Rank). The behaviour of the different approaches is comparable across the different metrics and is similar to that of the Success at Rank: Content-based approaches offer, in general, better results, some approaches like **Hybrid**, **BPRMF**, **LDA** and **LRMF** offer similar results and our FCA-based measure offers the best results for most of the configurations.

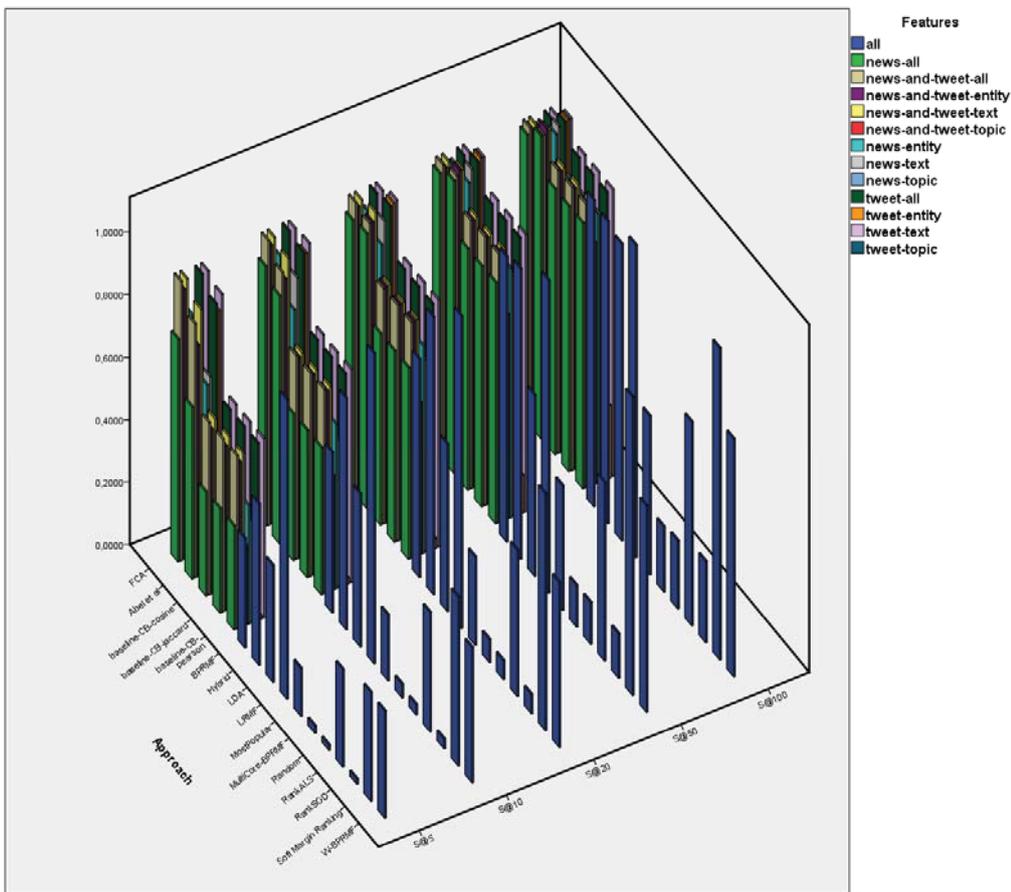


Figure 6.13 – Success@K-based Overall Results

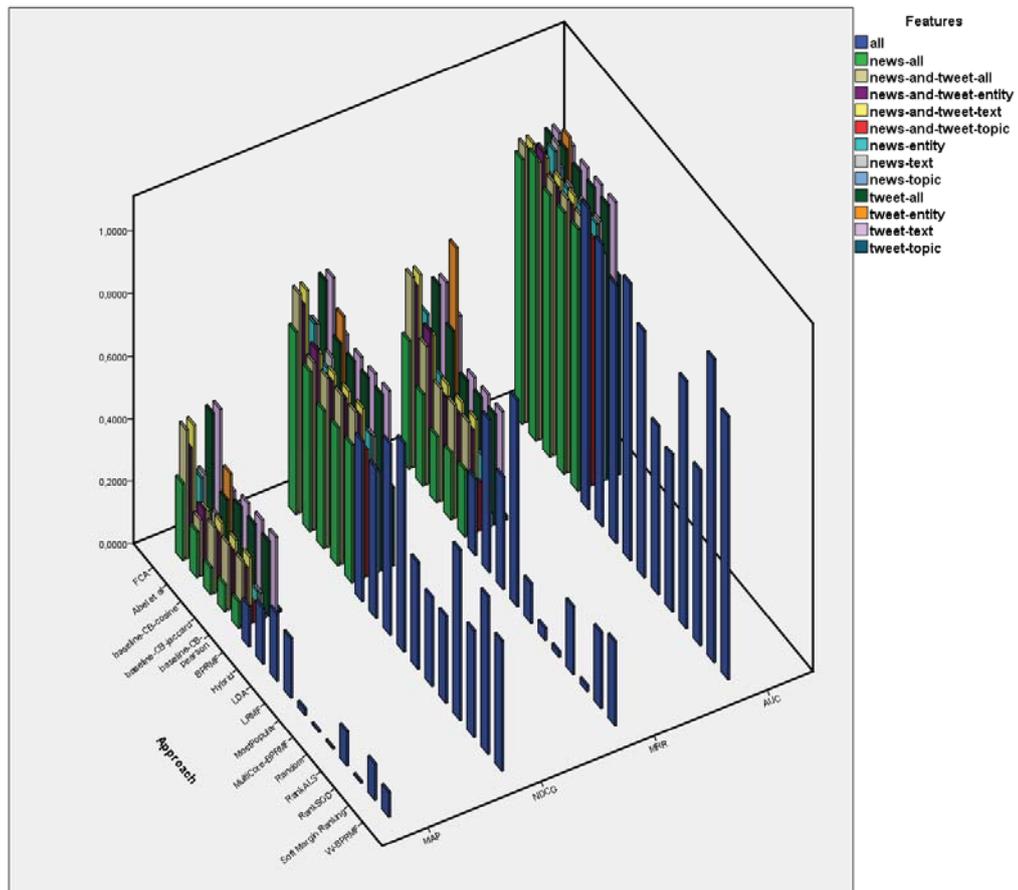
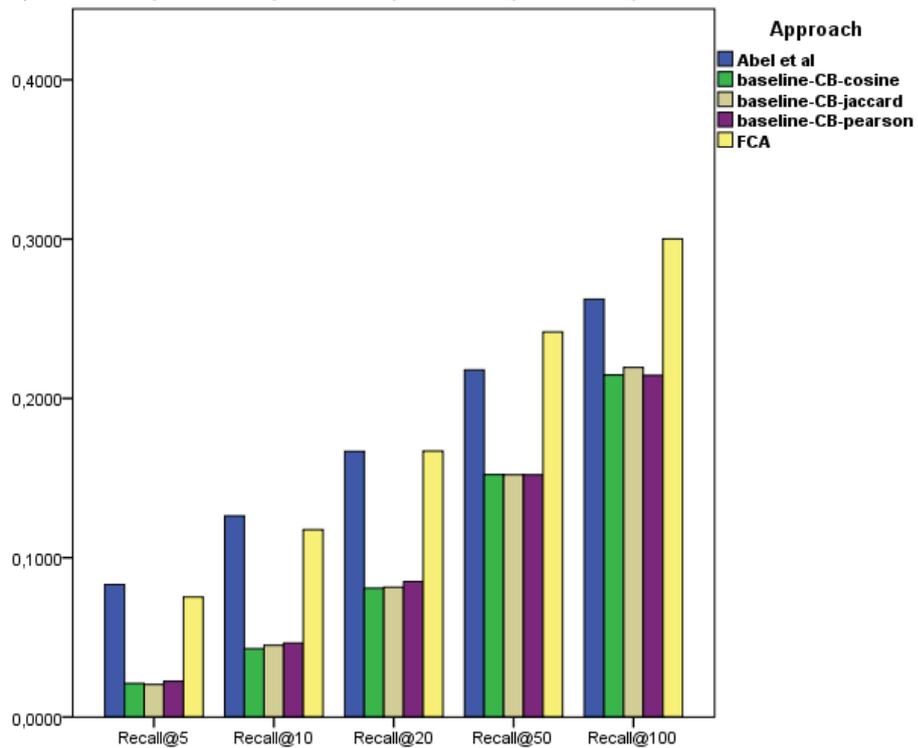
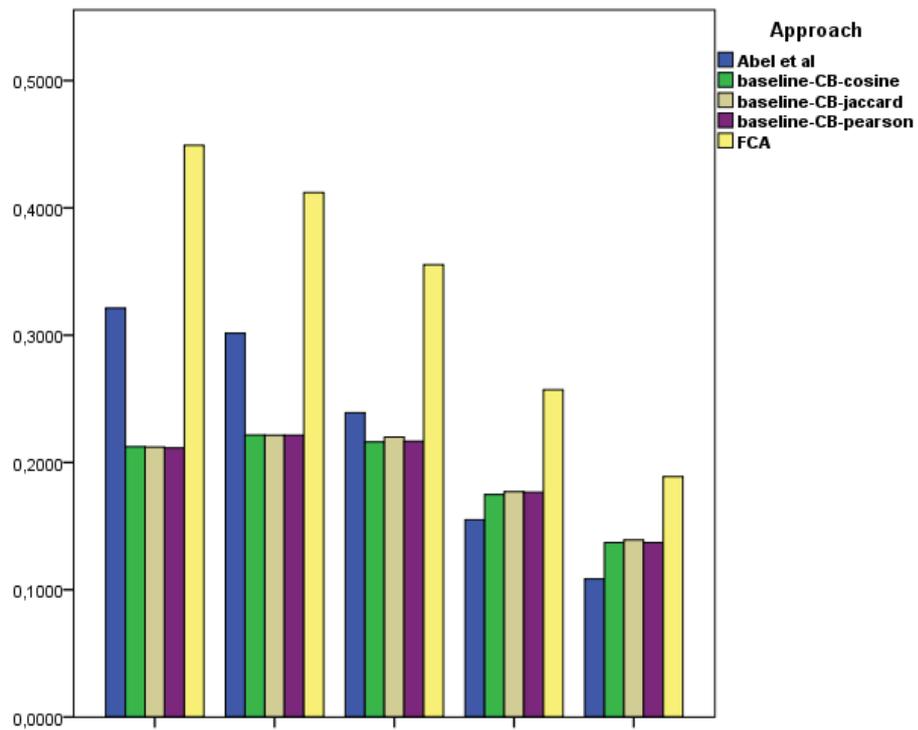
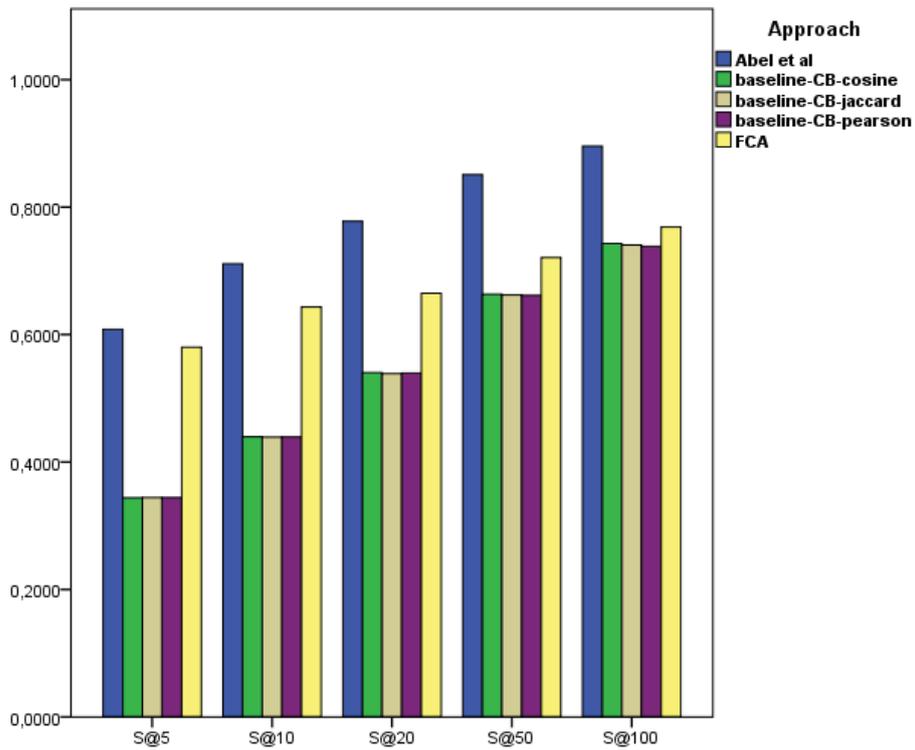
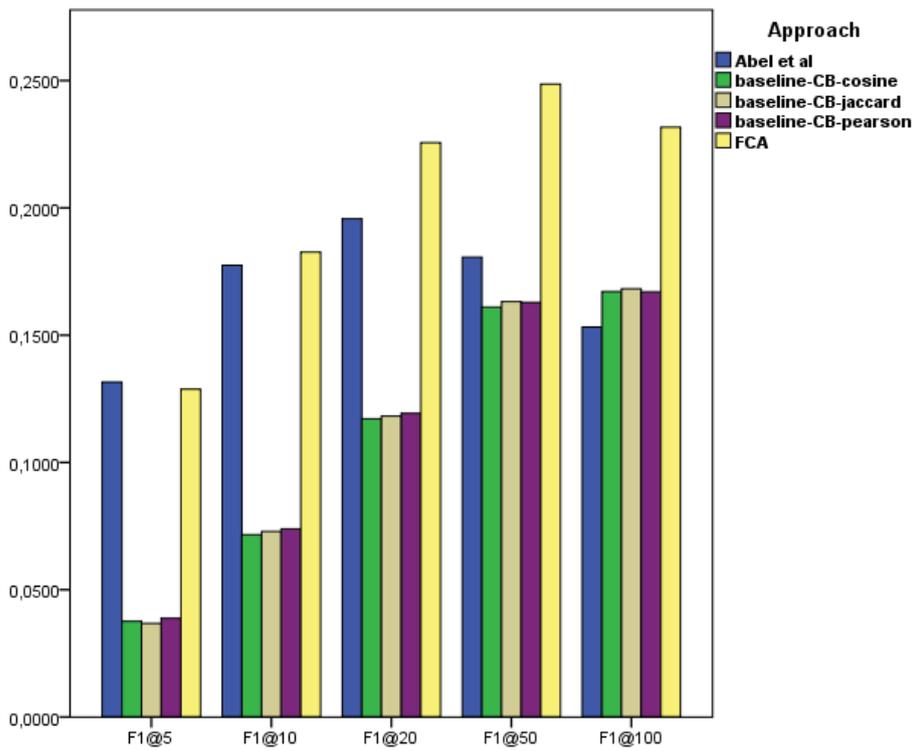


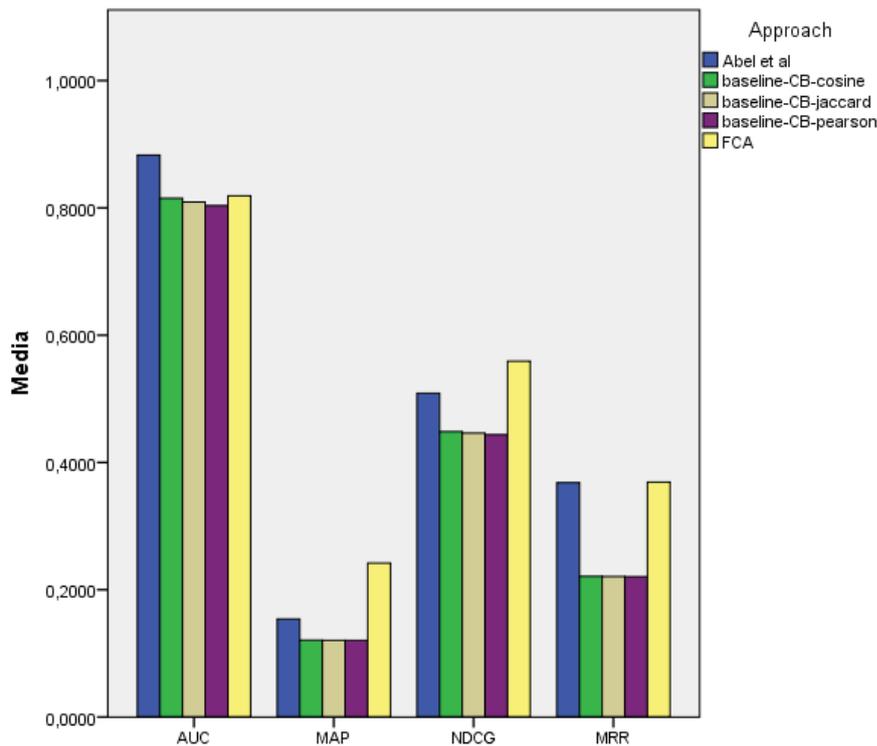
Figure 6.14 – Other Ranking-based Overall Results

*Content-based overall results*

Summing up from the previous section, Content-based recommendation approaches give rise to the best results across the different measures. In particular, our **FCA**-based approach does outperform other Content-based approaches for most of the configurations (i.e., for the Success at K metric, the approach of **Abel et al** achieves better results than ours). In this regard, [Figure 6.15](#) summarizes the Content-based results by showing the performance of the four different Content-based approaches according to the different measures. The results are averaged over the different feature types and k-values. This figure confirms our FCA proposal (yellow bars in the figure) as the best-performing approach. FCA offers the best results for most of the configurations in terms of the precision-based results. The Success@K metric produces different results; i.e., the results of the **Abel et al.** approach achieves the best results for these metrics. This means that the **Abel et al.** approach is more accurate in finding the first relevant recommendation. Nevertheless, the entire ranking is taken into account and evaluated according to the rest of the ranking-based measures. FCA is again the top-performing approach. As regards the baseline approaches, their performance is far from that achieved by the **FCA** and the **Abel et al.** approach. In addition, no differences are observed between the different similarity measures.







**Figure 6.15** – Content-based Overall Results (averaged over the different features)

#### *Collaborative-Filtering Results*

In order to give a clearer insight into the overall performance of the Collaborative Filtering, [Figure 6.16](#) shows the performance of all of them according to the different measures.

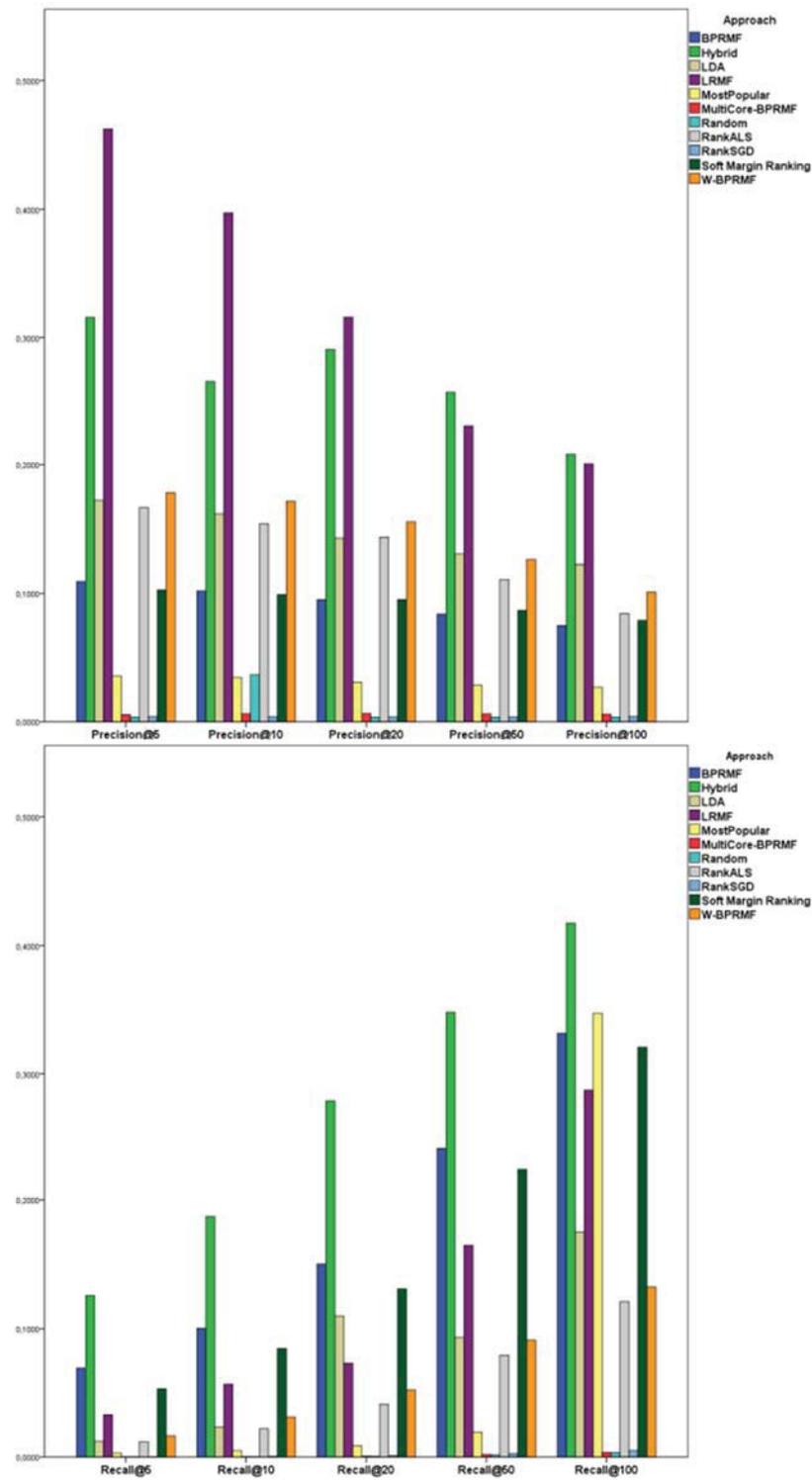
As regards the more basic proposals, the **Random**-based recommendation, as expected, does not achieve meaningful results. Although this aspect seems obvious, it is important to highlight the difference between the different approaches and the random recommendation. Any approach should improve this random baseline in order to prove that recommendation proposals are able to capture aspects related to user preferences beyond random guesses. Another interesting baseline proposal is the **Most Popular** recommendation. The kind of data in the experimental dataset (Tweets and the related News reports), offering the most popular items appear as a sensible first try in addressing the recommendation task. Both proposals offer quite low results, as expected, but they mark a bottom-line for the other approaches.

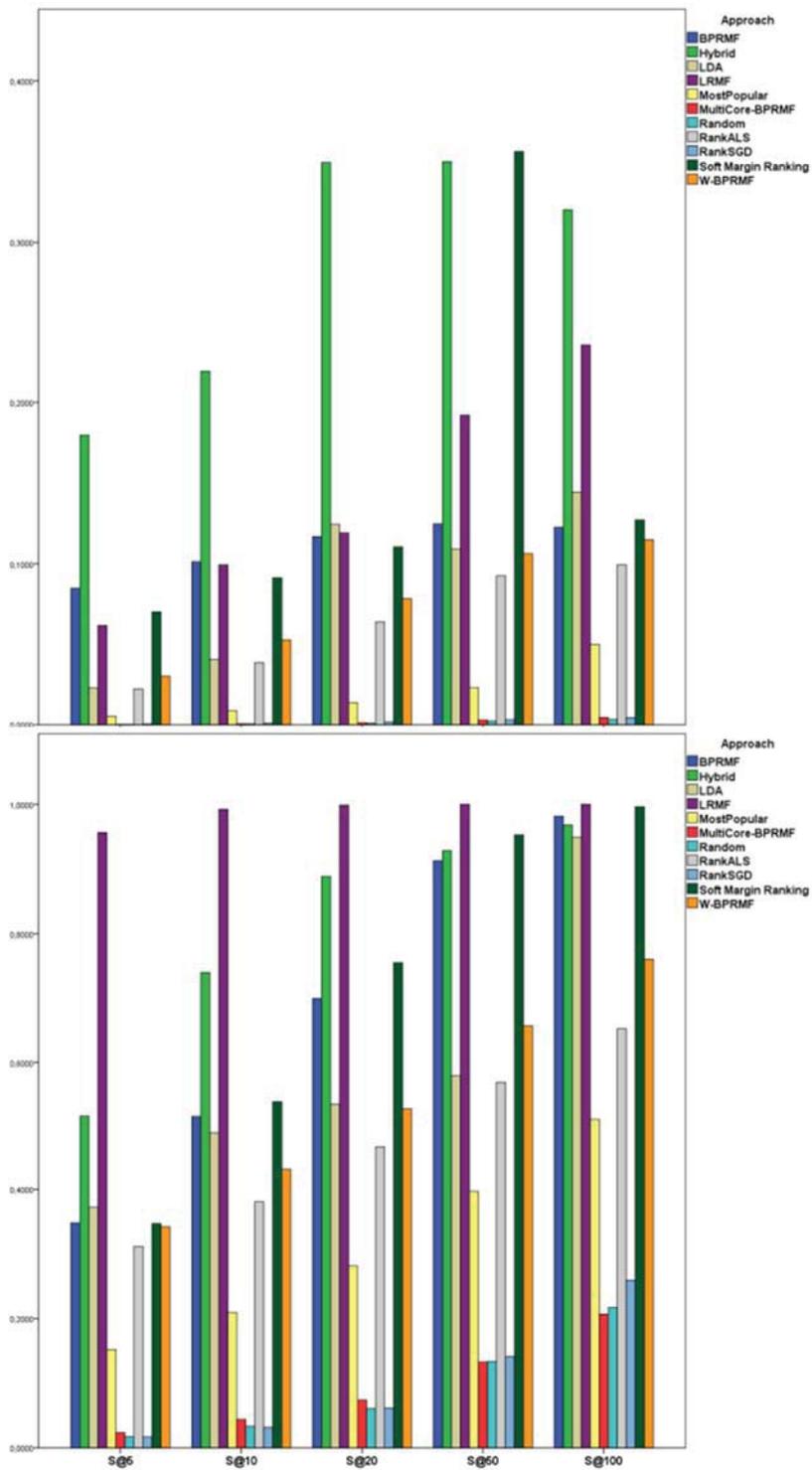
The best performing of this approach, not making use of Content-based features, is the **Hybrid** approach. This approach is somewhat similar to our proposal: it also tries to take advantage of the user-item incidence matrix in order to infer relationships that may

result in accurate recommendations. Another well-performing approach, especially in that referring to Recall-based results, is **LRMF**, which is based on the factorization of the user-item matrix. One interesting aspect is that this approach maintains the good performance across the different cut-off points. For instance, although as expected, precision values decrease when more items are considered, they do not do it to the same degree as other approaches. In this sense, another factorization-based recommender algorithm like **Soft Margin Ranking** also achieves a similar recommendation performance.

Other methodologies that obtain satisfactory results are: Bayesian Personalized Ranking (**BPRMF**) and its weighted implementation (**W-BPRMF**) that suppose an implicit feedback scenario (that applied in this work) to carry out the recommendation process, which seems to affect the recommendation process positively, and Latent Dirichlet Allocation (**LDA**). Related to the former approaches, BPR, the **Multicore** implementation, offers surprisingly low results. It seems that the **Multicore** implementation (i.e., split the implementation, and consequently the data, in several cores) hinders the actual performance of this algorithm. This kind of implementation might still make sense where huge amounts of data are computed.

Focusing on the approaches based on Ranking Optimization, **RankALS** performs well but **RankSGD** does not. In fact, this latter approach offers lower results than basic baselines such as **Most Popular** and **Random** recommendation. Both approaches apply a similar rationale to address the recommendation task. The main difference is that the former uses Alternating Least Squares for the ranking optimization and the latter applies Stochastic Gradient Descent. In previous experimentations, [Rendle et al., 2011] proved that SGD optimization depends largely on the learning rates and the number of iterations (i.e., it is much more sensitive to overfitting). In this way, [Rosenthal, 2016] also stands out in this direction, although no deeper analysis is carried out. In the specific context of Twitter data, where the user-item matrix is quite sparse, this tendency to overfit and suffer from popularity bias may explain the low **RankSGD** results.





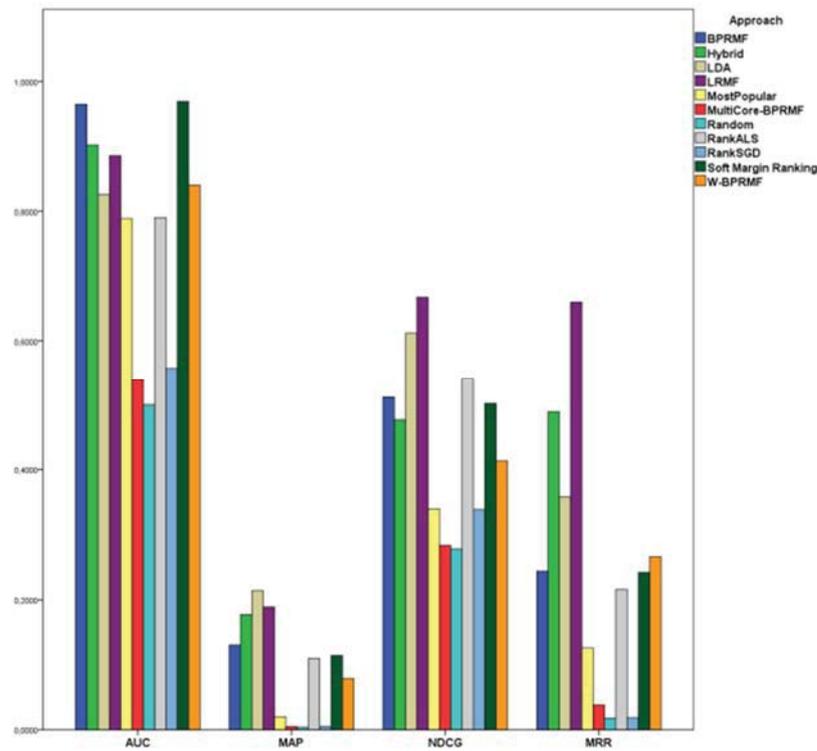


Figure 6.16 – Overall Collaborative Filtering Results

### *Analysis on the impact of the different features*

As we have seen in previous section, Content-based features do achieve the best results. However, the performance of the recommendation approaches using different kinds of features and items came up very different results. In more detail, [Figure 6.17](#) shows this comparison according to the different metrics averaging the results across the different approaches (to focus only on the feature performance) and the different thresholds. If we consider the different metrics, no significant differences are observed, i.e., the behaviour of the different features is similar to all of them. As regards the cut-off points, this behaviour is also the expected; that is, the precision decreases if more items are considered and vice versa for the recall values. F-measure achieves its highest performance for F1@40 and the Success at K is better while larger the number of items is considered. The other rank-based measures do not present variations either.

Differences can actually be observed between the different types of feature, all the recommendation approaches using textual features seem to work better than those applying higher-level features (entities or topics). In addition, when all features are taken together, results achieve similar results, although textual-based approaches are still better.

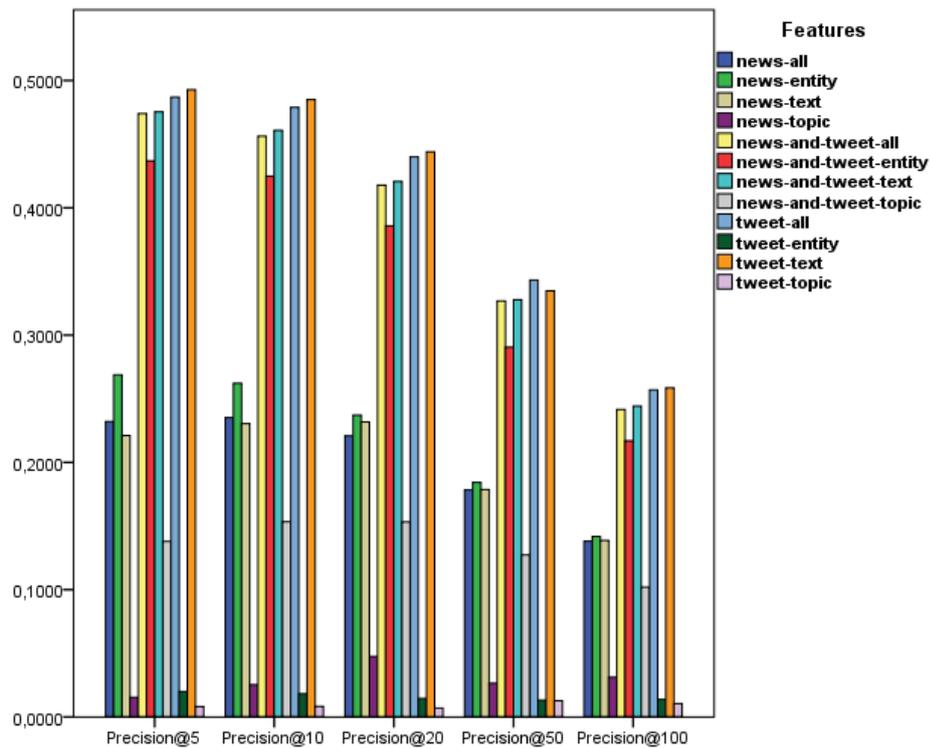
It may seem counterintuitive that higher-level features, such as entities and topics, are expected to better represent content, thus leading to more accurate recommendations. The explanation of this aspect is related to the cardinality of the relationships between the items and these features. In this regard, Table 6.1 shows the feature distribution of the different approaches (i.e., the average number of features included in each representation). As shown in the Table, the topic-based representations have no more than one feature per item, which lead to a barely informative representation, as seen in the results. Focusing on the results of the entity-based approaches, the `news-entity` approach outperforms the approach using textual features (`news-text`) or even the approach using all the features together (`news-all`). In the same way, the `news-and-tweet-entity` offers a similar performance to both the `news-and-tweet-all` and `news-and-tweet-text` approaches. In contrast, the `tweet-entity` approach does not offer similar results. It can be explained by looking at features per items of each approach in Table 6.1. While, `news-entity` and `news-and-tweet-entity` have almost 20 features per item, `tweet-entity` has only 1.6 features per item (similar to topic-based approaches), giving rise again to a barely representative approach.

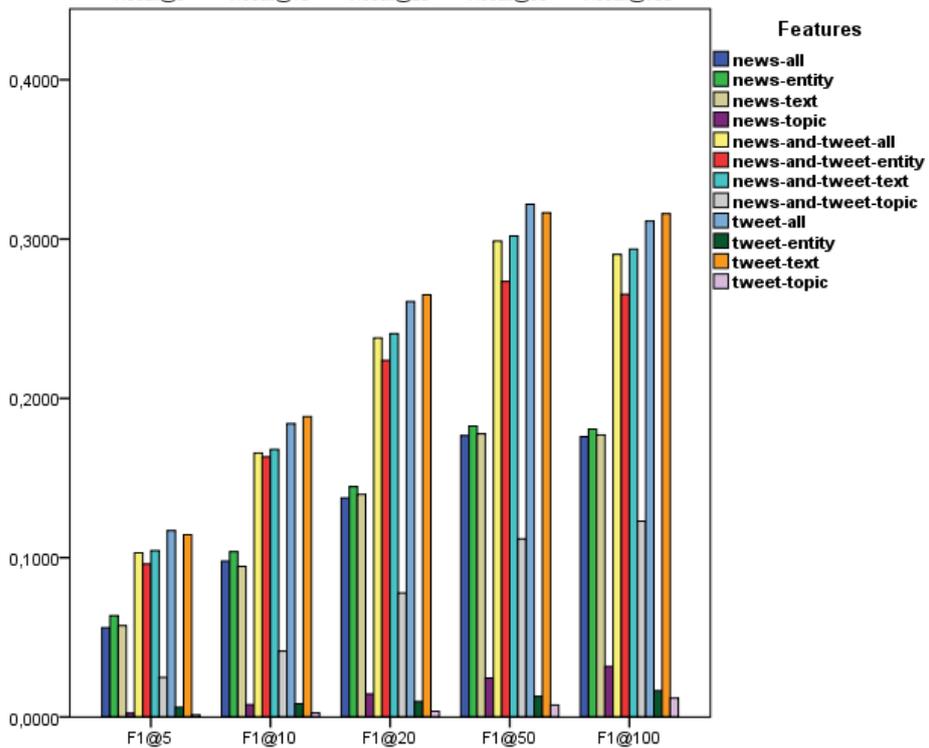
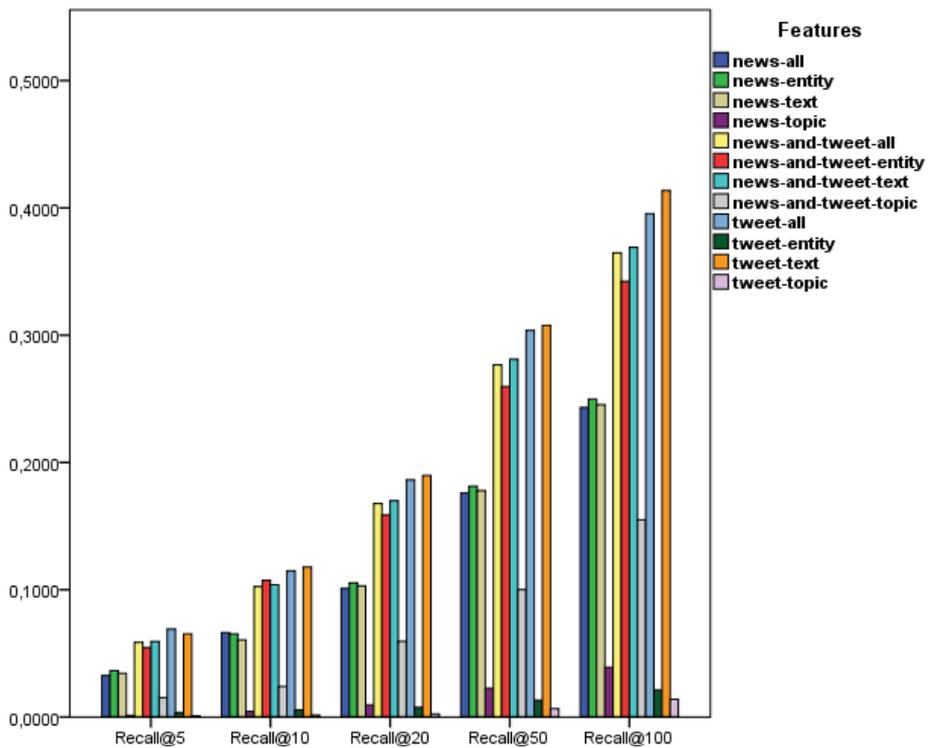
Approach	Features per Item
<b>News-all</b>	148.43
<b>News-Entity</b>	18.56
<b>News-Text</b>	128.59
<b>News-Topic</b>	1.28
<b>Tweet-All</b>	16.79
<b>Tweet-Entity</b>	1.60
<b>Tweet-Text</b>	14.66
<b>Tweet-Topic</b>	0.53
<b>News-and-Tweet-All</b>	148.43
<b>News-and-Tweet-Entity</b>	18.57
<b>News-and-Tweet-Text</b>	128.59
<b>News-and-Tweet-Topic</b>	1.28

**Table 6.1** –Feature Distribution Analysis

Therefore, to sum up, the better performance produced by the textual features in comparison to the higher-level features observed in the results does not mean that the textual features may represent the items in a more accurate way. The low performance of higher-level features is explained by the fact that the items in the dataset, in general, contain very few higher-level features (e.g., items have 1 topic per item in average). It is

confirmed by the representation using the entities. In this case, when there are enough features per item (`news-entity` and `news-and-tweet-entity`, which have an average of 18 features per item), the generated representations are informative enough. As a result, these approaches (`news-entity` and `news-and-tweet-entity`) achieve a better performance than those applying textual features. Nevertheless, the `tweet-entity` approach, where there are only 1.6 features per item, again offers a low performance. The intuition that higher-level features are preferable to represent the items is then confirmed; but it has to be enough features per item.





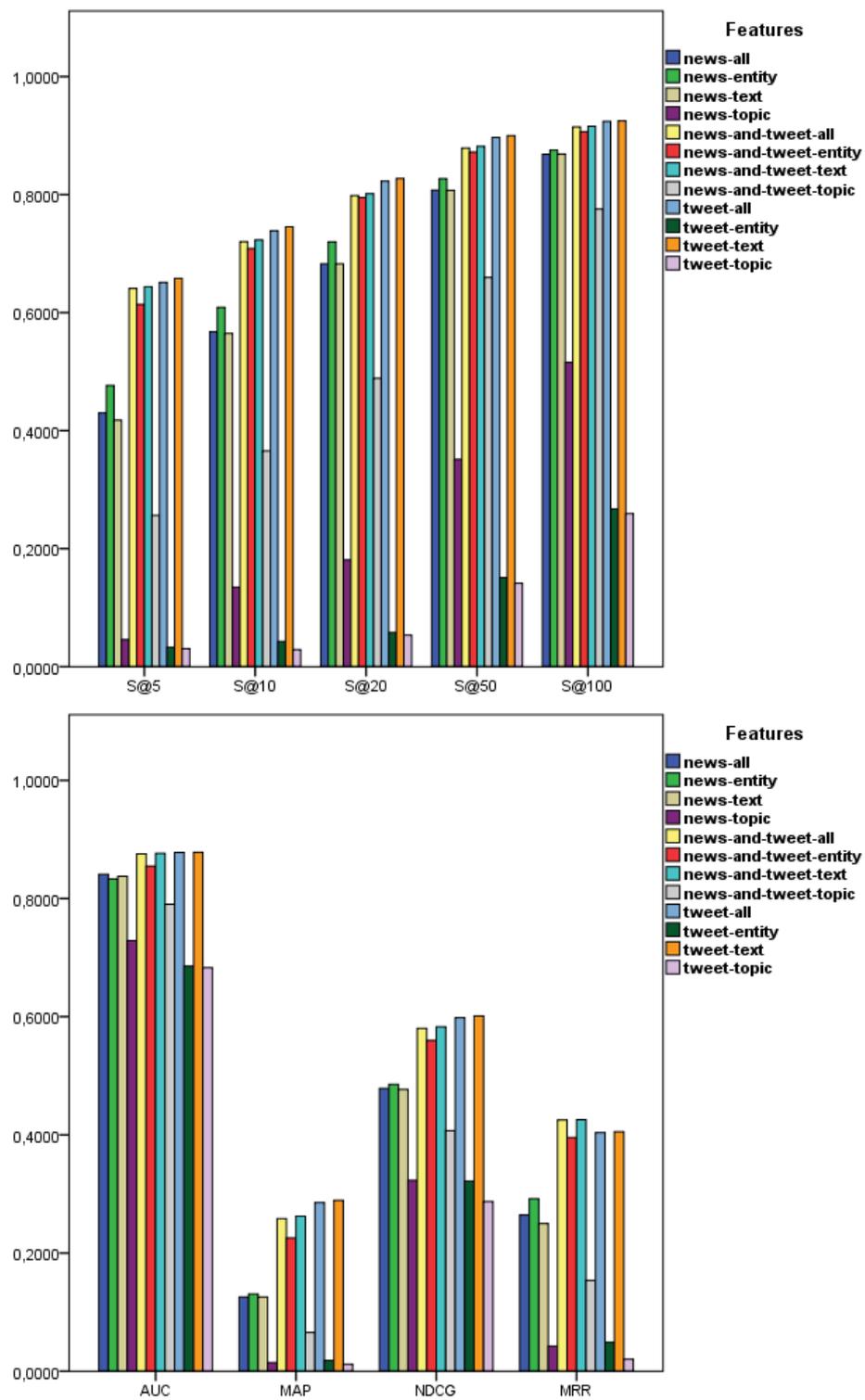


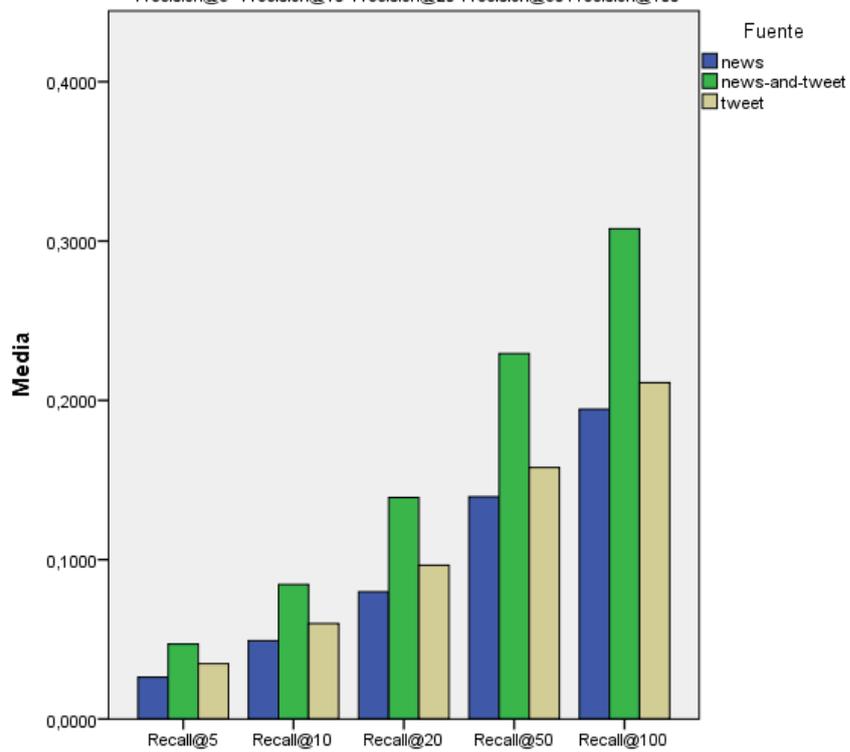
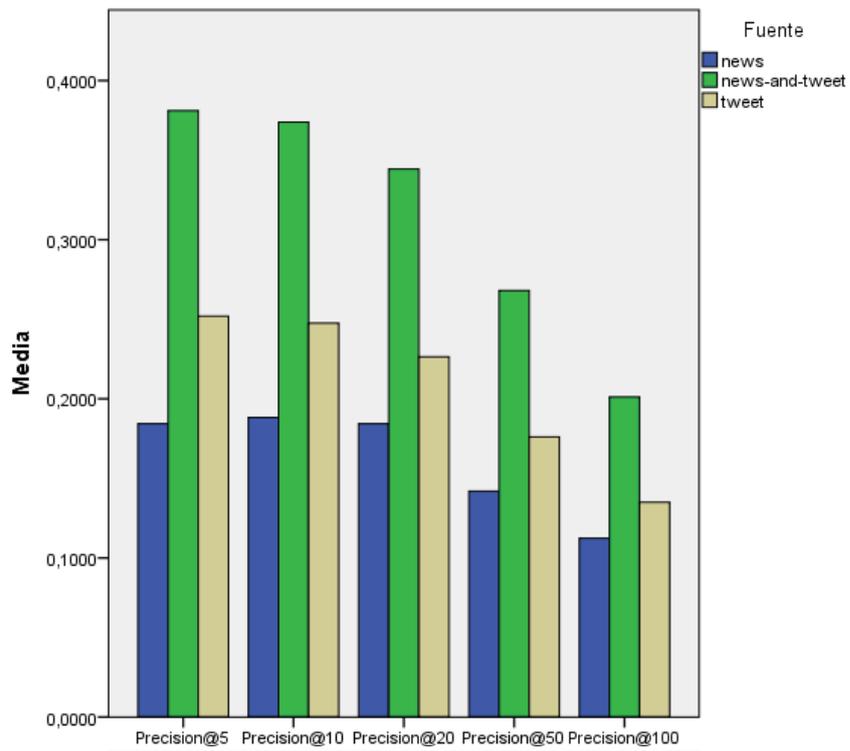
Figure 6.17 –Results according to the different features

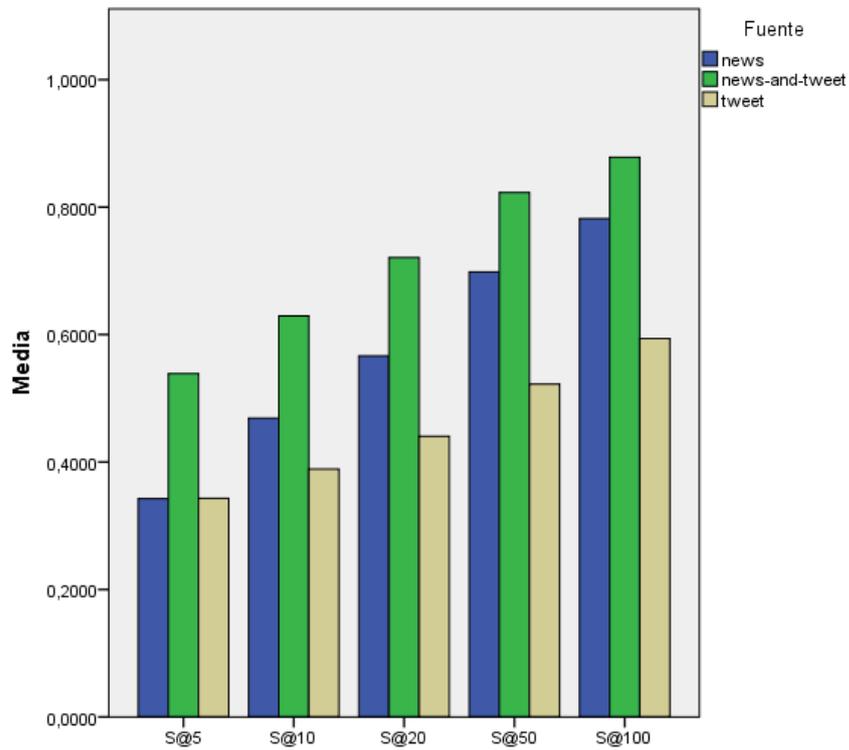
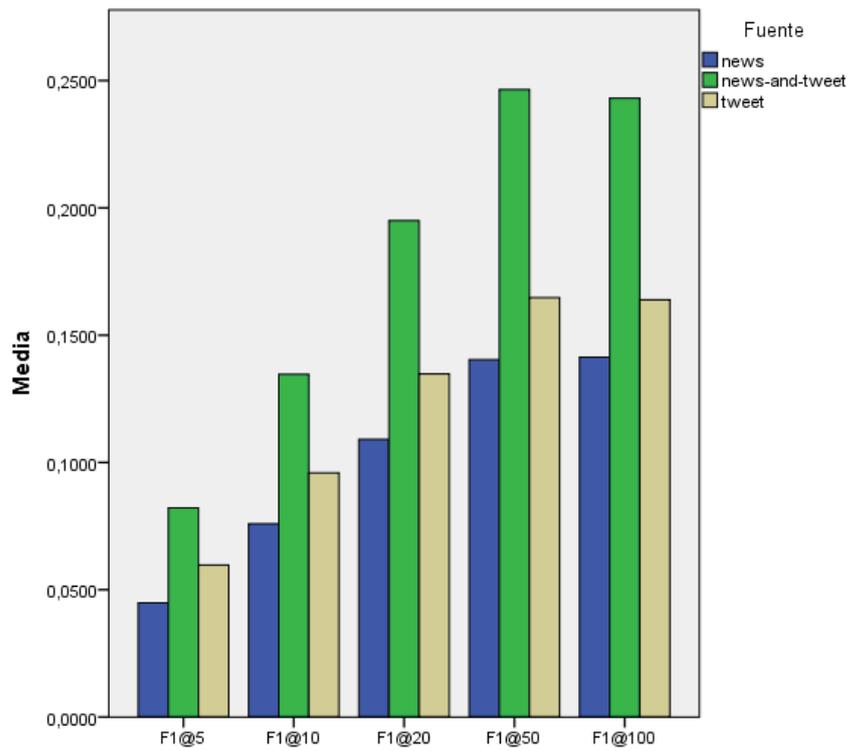
*Analysis on the impact of the different type of item*

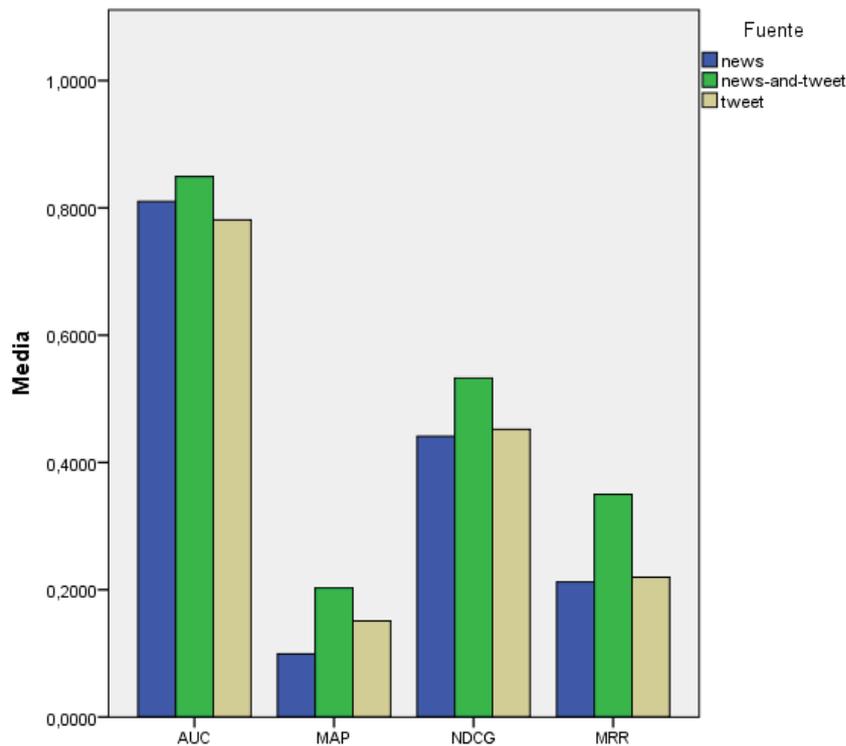
In the previous section, how the recommendation behaves when different features are applied has been analysed according to the different metrics. On the other hand, this section seek to analyse the performance of the different items used to create the user profiles, namely news, tweets and both together. For more detail on the different user profiles and their implications, refer to section [Task Definition](#).

[Figure 6.18](#) shows the results according to different metrics by item type. In order to focus only on these types, the results are averaged over the different approaches and features. As shown in the figure, the approaches using news and tweet representations together (`news-and-tweet`) outperforms the approaches using them separately (`news` or `tweet`). That is, the more information included in the representations, the better the performance of the recommendation algorithms. These results seem intuitive, but the important aspect is the difference between the results obtained. The improvement in performance should be enough to justify the increase in complexity entailed by the inclusion of more information in the representations. For instance, if using just the text in the tweets might achieve similar performance, there would be no reason to apply larger representations. Nevertheless, this is not the case in this experimentation. Approaches applying news and tweets together, not only improve the results, but they do it by 10% to 40% for the different measures. Consequently, the use of all the information available seems to be the most suitable way of generating the user representations

If news reports and tweets are compared, the results do not show any clear tendency. For some metrics, Precision and Recall (and consequently F-measure), it seems that the tweet-based approaches are preferred, while for the Success at K metric the news-based approaches work better.







**Figure 6.18** –Results according to the different items (averaged over the different approaches and features)

### *Analysis on the impact of the different Content-based parameters ( $K$ and Threshold Values)*

In the previous sections, in order to focus on the overall results and the studied parameters (features and items), the Content-based baseline results (baseline-CB-cosine, baseline-CB-jaccard, baseline-CB-pearson) were averaged across the  $K$ -values (i.e., the size of the neighbourhood applied to compute the recommendations). In the same way, the results of the FCA-based recommendation approach were averaged across the threshold applied to reduce the formal context.

In contrast, this section focuses on the impact of these thresholds in the final recommendation performance. [Figure 6.19](#) shows the variations in the different measures according to the values of the neighbourhood and [Figure 6.20](#) according to the threshold values for the FCA computation. These figures show the different values of the  $K$ -values and FCA threshold in the x-axis and the values for the different measures depicted in the figures in the y-axis (each line represents the value of the different  $K$  or threshold values for each particular metric). In particular, each graph in the figure shows the results

for the different measures (Precision, Recall, F-measure, Success@K and Ranking-based measures). Each line in the graph is related to the value of the metric for each cut-off point (from 5 to 50) in the case of Precision, Recall, F-measure and Success@K or to the value for each of the Ranking-based measures (AUC, MAP, MRR and NDCG in the bottom-right graph). Since the three different CB-baselines offer quite similar values, in order to focus only on the impact of the K-values, the results in [Figure 6.19](#) do not distinguish among them and results are averaged across the three baselines.

The results in [Figure 6.19](#) shows that the best results are achieved when  $K=10$  and that they decrease along with the K-value: the larger the K-value, the worse the results. It means that the inclusion of more news reports in the neighbourhood leads to worse recommendations. This issue is related to the neighbourhood-based recommendations; i.e., the methodology applied in these baselines. Increasing the neighbourhood size entails the inclusion of more information that might enhance the recommendation process, but also the inclusion of less related information. In this sense, this kind of approaches should find a compromise between these two issues. What happens in the experimentation is that a small neighbourhood size is enough to capture the user preferences. In this regard, the inclusion of more news reports leads to the inclusion of noise in the user models. At a closer analysis, it is seen that there is a large loss in performance from  $K=10$  to  $K=50$  and then the results stabilize.

In the same way, the results in [Figure 6.20](#) show the same behaviour for the FCA threshold: the larger the threshold, the worse the results. Nevertheless, in this case the explanation is different. Increasing the FCA threshold means that the formal context is smaller. Consequently, less data are considered for the computation of the concept lattice. As proven in previous experimentations, reducing the formal context inherently entails the generation of less informative concept lattices [Cigarrán et al., 2016](#). The interesting factor is to what degree is this information loss acceptable in contrast to the reduction in complexity that the formal context reduction allows.

In this regard, if the impact of this threshold in the final performance is compared to the impact caused by size of the neighbourhood (in [Figure 6.19](#)), it can be seen as the FCA threshold has much less impact. The decrease in performance is less steep. Even when the threshold is equal to 50%, the results are comparable to those achieved when no reduction at all is made (threshold equal to 0%). The reduction in the formal context is therefore justified. Even for large threshold values, the performance is barely affected and the complexity is greatly reduced. This aspect is related to the nature of the dataset. In social contexts like Twitter, only a small amount of frequent information (e.g.,

frequent terms, hashtags or, as in this experimentation higher-level semantic features) is enough to represent the content.

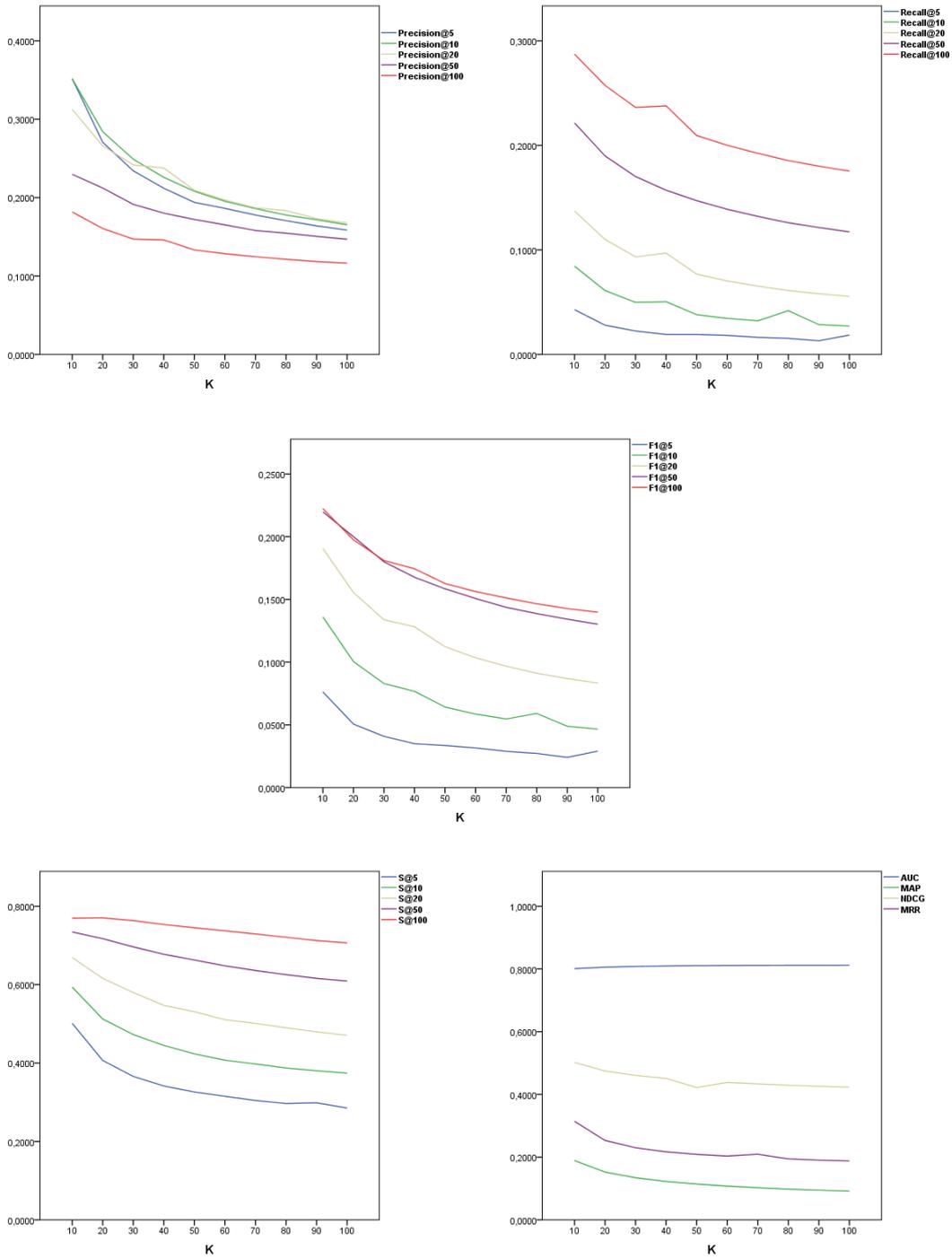


Figure 6.19 –Results according to the different K-values

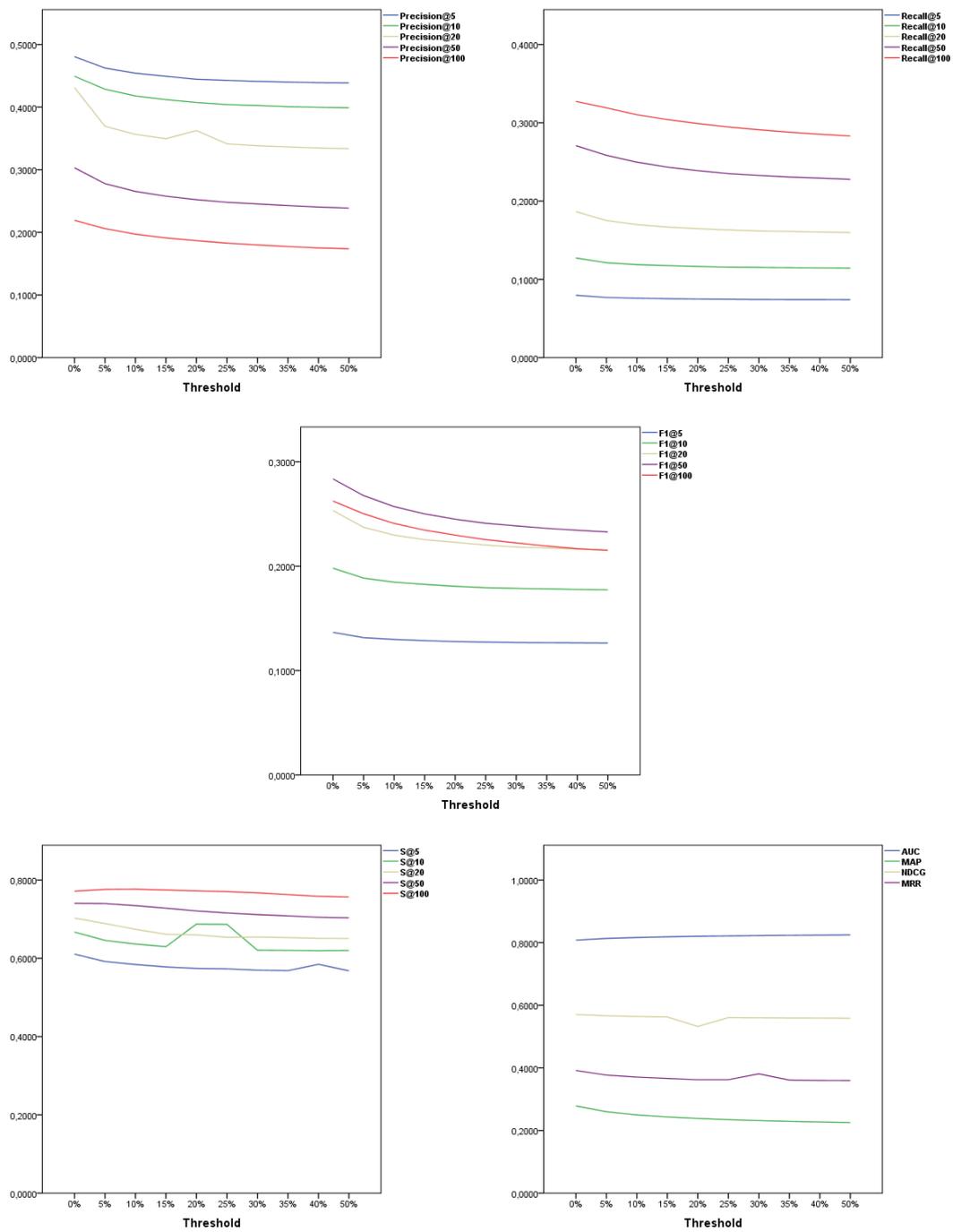


Figure 6.20 –Results according to the formal context reduction threshold

### *Take-Home Points*

The first and most important point to highlight is that, as seen in section `Overall performance`, the FCA-based proposal outperforms the other approaches, even those most sophisticated proposals in the state of the art.

In a more detailed analysis of the overall results, it can be seen that the Content-based approaches work better than the rest of the approaches. As regards these Content-based features, although the best results are achieved by text-based approaches or that approaches taking all features together into account, higher-level features seems to be desirable. The low performance of topic-based and some of the entity-based approaches is explained by the fact that these representations have very few features related to each item (see `Table 6.1`). Therefore, it leads to barely representative representations. In contrast, if there are a sufficient number of features per items (e.g., `news-entity` and `news-and-tweet-entity`), the results are at the same level (even better in the case of `news-entity`) than the best-performing approaches. In other words, these higher-level features are able to generate informative representations. In fact, this “informativeness” is achieved with a fewer number of features (around 148 features per items for the `news-all` features and 128 for the `news-text` vs. 18.56 for the `news-entity`).

As regards the type of items used to create the user profiles, the only clear conclusions to be drawn is that the combination of both, news and tweets, is the most suitable approach.

Focusing on the thresholds, the  $K$ -value of the neighbourhood in the CB-baselines has a significant impact on the performance of the final algorithm. Specifically, the best results are achieved when  $K=10$ , that is, when fewer news reports are included in the neighbourhood. The recommendation performance decreases as this value increases. As regards the threshold for the reduction of the formal contexts applied for the FCA approach, the algorithm performance is better when no reduction is applied (threshold equal to 0%) and it decreases as this threshold increases. However, this decrease is less steep than in the previous case; that is, it affects the final performance to a lesser degree.

If we focus on the rest of the state-of-the-art approaches, not making use of Content-based features, some remarks can be made. The best performing (`Hybrid`) uses the user-item incidence matrix to infer relationships to drive the recommendations (as FCA does), Matrix Factorization (`BPRMF`, `Soft Margin Ranking` or `LRMF`) methodologies performs well for all the metrics and, therefore, they seem a sensible choice. On the other hand, ranking optimization methodologies also offer satisfactory results. Finally, the

extremely low performance of some approaches like `Multicore-BPRMF` and `RankSGD`, which offer even worse results than baselines that do not apply any personalization like `Random` or `Most Popular`, is remarkable.

### Discussion

This chapter presented our proposal for a Top-N recommendation based on applying Formal Concept Analysis to create a common representation space. It aims to overcome the problems presented by traditional models related to the gap between user and item representations. In this sense, FCA is applied to generate a conceptual representation, in the form of a concept lattice, to model user profiles and item representation together.

This approach is experimented in the challenging scenario of News Recommendation in Twitter; in particular, the experimental dataset proposed by [\[Abel et al., 2011\]](#) in the context of the UMAP 2011 Challenge. In order to frame the performance of the FCA-based proposal, an evaluation platform, implementing different Content-based and state-of-the-art recommendation approaches and several quality measures has been implemented.

Results confirm our hypothesis that our proposal is better able to model user and items through the generation of a common representation space, based on the application of FCA. In more detail, our proposal is able to outperform all other state-of-the-art approaches. In addition, FCA also proved that is better able to model users and items than other Content-based approaches.

As regards the approaches that do not apply content-bases features, they are not able to reach the performance of Content-based approaches. As expected, the inclusion of such features to describe the content improves the recommendation process. Analysing their results in detail, the best performance is offered by the Hybrid approach. Among the rest of the approaches, those applying matrix factorization methodologies achieve satisfactory results, improving those approaches based on ranking optimization methodologies. Techniques based on Bayesian Personalized Ranking also achieve results at the same level.

Focusing on Content-based approaches, none of the items used to create these representations — news reports or tweets — achieve better results. Only the combination of both item representations together is able to improve the individual results.

As regards the parameter for adapting the computation of Content-based baselines, the smaller, the better. It seems that the inclusions of more items in the neighbourhood, leads to the inclusion of noisy information and therefore to the decrease in the

recommendation performance. The computation of the FCA-based approach is also reliant on the threshold used to reduce the formal context. As expected, the more restrictive this threshold, the worse the results, because less information is included in the formal context and consequently in the recommendation model.

In a more detailed analysis of the performance of these Content-based features, the first note is that higher-level features, when available, lead to more informative representations. In particular, approaches making use of entities in the items achieve similar results to those based on textual representations or even those including all the features together. Moreover, the resulting representations based on these higher-level features enables a lighter and less sparse model (i.e., they need fewer features per items in order to describe them). Nevertheless, other higher-level-based representations (e.g., topic-based) lead to poor results because there is no more than one feature per item and, consequently, the representation is barely informative.

From this latter remark it can be inferred that “rich-enough” higher-level representations based on conceptual features are desirable, not only because of the improvement of the recommendation accuracy, but also because of the advantages that these representations entails (i.e., less sparsity, less dimensionality). By “rich-enough” we mean representations where enough features per item are available. The availability of such representation should enable the enhancement of the recommendation process. In this sense, the following section is focused on experimentally prove this aspect.

#### **6.2.2.2. The ESWC 2015 Recommendation Challenge**

The previous section proved the suitability of the proposed FCA-based approach, achieving state of the art results. An important aspect derived from this experimentation is that higher-level features enable better representations than others based on raw text. These results confirm those in section [6.1.2.2](#), where semantic-based features outperformed the performance of raw textual representations for content modelling.

In this regard, this section aims to go a step further in this direction to demonstrate the improvement in the recommendation performance when FCA is applied to rich semantic data representations. To experimentally prove this hypothesis, the experimental environment proposed by the 2015 *Linked Open Data-enabled Recommender Systems* (hereinafter LOD-RecSys) challenge is proposed. The LOD-RecSys challenge focuses on the RS experimentation taking advantage of LOD with special attention to the diversity in the recommendations. This scenario is a follow-up of the 2014 edition (see section [6.1.2.2](#)). As in the 2014 edition, the experimental environment of the 2015 edition of the challenge offers a recommendation dataset annotated with semantic information related

to the items contained in DBpedia. By means of this annotation, it is possible to create the rich semantic representations that we hypothesized should enhance the recommendation process.

In particular, based on the knowledge modelling proposed at section 5 and the common representation space at section 6.2, a knowledge-based common representation space based on two steps is proposed:

- **A concept-based organization of the DBpedia data:** The rationale is that this organization may improve the DBpedia structure. As proven in section 5, it facilitates the identification of the most valuable information to be used to represent the items. Therefore, FCA is applied to model the DBpedia data and then connect the items to this modelling.
- **A common representation space for users and items:** In this space, items will be organized according to their DBpedia-features. Given that user profiles are the set of the DBpedia-features related to the items already consumed by the users, they are also modelled in the FCA-based DBpedia model. User profiles will be therefore related to the item/s containing the same DBpedia-features and they (user and items) will be grouped together and organized according to these features. Hence, each of these user/item groups may be seen as each of the user preferences and the item/s fulfilling these preferences.

In the following it is explained the work proposal, the conducted experiments and, finally, the obtained results and their analysis.

### LOD-RecSys Common Representation Space

The idea of a common representation space has been introduced in section 6.2 and in section 6.2.2.1 applied to the specific scenario of Twitter-based recommendation. In this section, we detail the specific aspects related to its application to the LOD-RecSys environment.

The common representation space for this experimental scenario has been created by following the Knowledge Organization proposal presented in section 5.1. In particular, there is a *formal context* containing a set of objects ( $G$ ) — the DBpedia-entries — a set of attributes ( $M$ ) — the DBpedia-features describing the entries — and a relationship ( $I$ ), indicating that an entry has a feature.

In more detail, to the later experimentation only the classes related to the item types in the dataset have been modelled (see section Dataset for more details): *movie*, *movie\_actor*, *movie\_character*, *movie\_director*, *movie\_genre*, *book*, *book\_writer*,

*book\_character*, *music\_album*, *music\_artist*, *music\_band*, *music\_composition* and *music\_genre*. With the information in these classes, two different models have been created, one for the movie-related information and another for music-related, including all the entries in the domain-related classes. The obtained FCA lattices will be then used to represent data belonging to these domains.

Some figures in the object and feature count, as well as the relationships between them and the *formal concepts* they create for the modelled DBpedia domains are detailed in Table 6.2. As shown in the table, FCA reduces by an order of magnitude the initial number of DBpedia relationships (i.e., Relationships vs. Formal Concepts), thus creating a more abstract representation of the DBpedia data. This observation confirms the results in section 5.1 where FCA also created a more abstract representation based on *formal concepts* of the DBpedia and EuroWordNet data.

	Objects	Features	Relationships	Formal Concepts
music	58,445	260	105,803	17,146
movie	37,547	72	85,893	29,599

Table 6.2 – DBpedia Model Statistics

An example of the movie domain representation is shown in Figure 6.21. This structure allows the inference of data relationships not explicitly defined in the original data. For instance, two instances that share the feature **basedOn-Tolkien\_Books** (e.g. **LOTR: The Two Towers** and **LOTR: The Return of the King**) are shown to be more closely related than two other instances sharing the feature **High\_Fantasy\_Films** (e.g., **LOTR: The Return of the King** and **The Hitchhiker's Guide to the Galaxy**). This inference is possible because the lattice structure makes explicit that a film based on a Tolkien book is by default a high fantasy film (i.e., it is a subconcept and, consequently, it is placed below in the lattice structure).

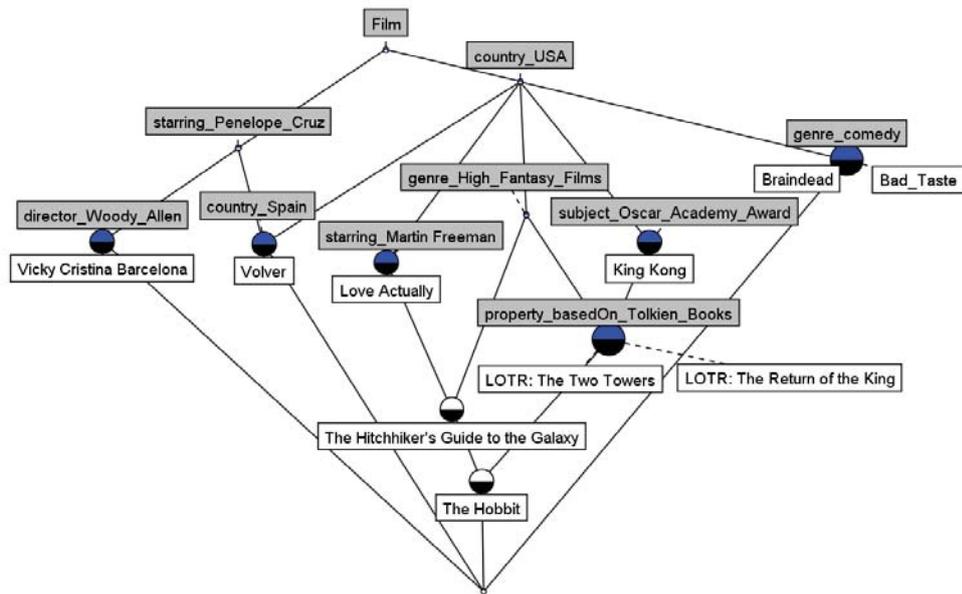


Figure 6.21 – Example of the DBpedia movie domain

By applying the aforementioned rationale, users may be therefore described by the aggregation of such features and, consequently, they can be modelled in that lattice structure. Following this example, the common knowledge-based representation space including user profiles will look like the example in the [Figure 6.22](#).

In a detailed look, some interesting relationships are discovered. For instance, users liking comedy movies (**Comedy\_Fan\_User** in the example) are likely to be interested in movies such as **Bad Taste** or **Braindead**. Another more elaborated example: users liking **Lord of the Rings** movies (**LOTR\_Fan\_User** in the example) might be also interested in a high fantasy film also starring Martin Freeman such as *The Hitchhiker's Guide to the Galaxy*.

The process is formally explained by the algorithm in [Figure 6.9](#) at section [6.2.1](#). In brief, given a user to offer recommendations, the algorithm looks for its object concept ( $\gamma g$ ), which is taken as the starting point. The *object concept* is the most specific concept in which the user appears. The most specific concept will lead to specific recommendations that are more likely to be the most interesting. Starting at the object concept, the algorithm navigates across the lattice taking those *formal concepts* included in the navigation path as already explained in the algorithm in [Figure 6.9](#).

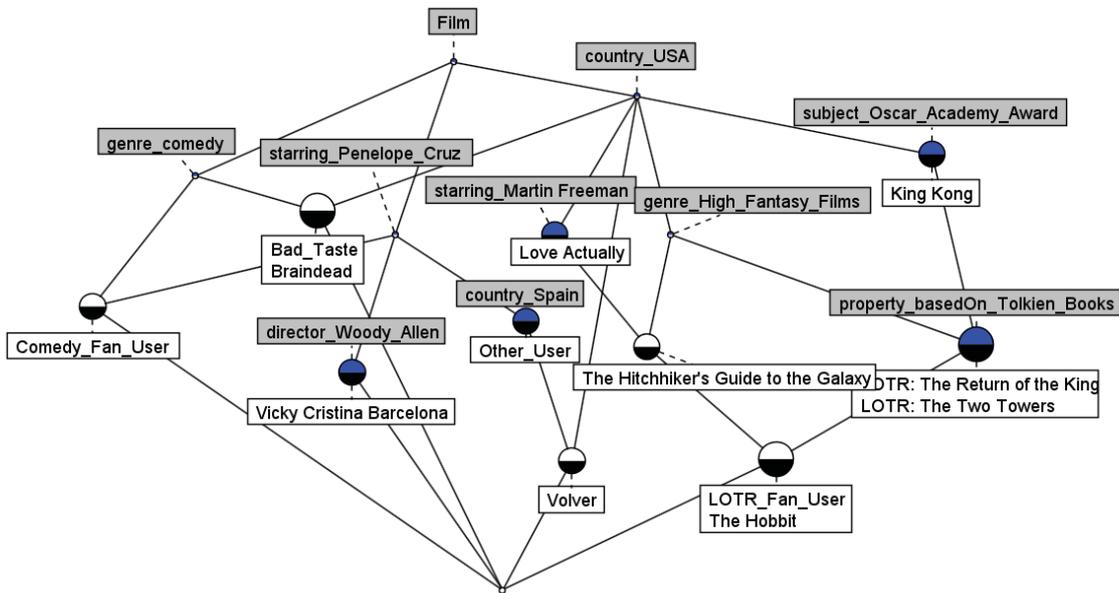


Figure 6.22 – Example of the common representation space

## Experiments: Results, Analysis and Conclusions

In what follows, the specific details of the experimentation is explained, focusing on the provided dataset, the evaluation set-up, the two tasks addressed and the analysis of the experimental results.

### *Dataset*

The LOD-RecSys challenge provides a dataset for experimentation that has been collected from Facebook profiles by gathering the 'likes' for items in three domains: movies, books and music (one for each different task in the challenge, for more details refer to <http://sisinflab.poliba.it/events/lo-d-recsys-challenge-2015/dataset/>). The items in the datasets are mapped to their corresponding DBpedia URIs. These mappings are useful to extract semantic information from DBpedia to be used by the recommendation approaches.

In particular, the experimentation focus only on the *music* and *movie* databases that contains 6,372 items and 52,072 users for the music domain, and 5,389 items and 32,159 users for the movie domain.

For its evaluation, the dataset is split into the training and test (or evaluation) set. The LOD-RecSys organizers only provided the participants with the training set to tune the recommender systems. This training set contains 854,016 ratings for the *music* domains and 638,268 ratings for the *movie* domains. An interesting aspect about these ratings is that they are unary ratings (i.e., a user rating only gives information about User<sub>1</sub> likes

Item\_1 but not about User\_1 dislikes Item\_1), implicitly collected from the user activity (i.e., without asking for explicit user feedback). Although early recommender systems were built on explicit feedback data, the "implicit scenario" proposed by this experimentation represents a more realistic scenario. Explicit ratings are normally hard to gather from the users, while, on the other hand, it is easy to obtain implicit user feedback. For instance, the simple act of a user buying or browsing an item may be viewed as an endorsement for that item.

### *Experimental Approaches*

This section details the different state-of-the-art recommendation algorithms compared to our system (**FCA+KO**) that applies the whole recommendation pipeline as explained before. We have also implemented different state-of-the-art recommendation approaches. The LibRec Java library for Recommender systems [Guo et al., 2015] has been used to implement these algorithms. More specifically, we propose the following algorithms<sup>14</sup>:

- **Random:** This baseline based the recommendation on a random guess. In particular, given a target user, the system randomly generates from a uniform distribution a score for each item in the collection. The top-N ranked items will be offered as recommendations.
- **MostPop:** This baseline y recommends the most popular items.
- **CF-baseline:** This baseline applies a user-based Collaborative Filtering recommendation implemented in Apache Mahout<sup>15</sup>. In more detail, the recommendation is reached using the *GenericBooleanPrefUserBasedRecommender* using Pearson Correlation as a similarity measure. We tested different neighbourhood sizes for the parametrization of the algorithm (from 10 to 100). The reported results are those generated with the best configuration (*neighbourhoodSize=80*); however, this parameter had a low impact in the final performance: the F-measure averaged over the 10 different configurations (from 10 to 100) is equal to 0.097 with a standard deviation equal to 0.005
- **CB-baseline:** This baseline applies a Content-based recommender. That is, it recommends items whose content is similar to the content of those already

---

<sup>14</sup> More details on the code, the configuration files for the execution of the different experiments and the proposed system outputs are publicly available at <https://github.com/AngelCastellanos/common-space-recommendation>

<sup>15</sup> <https://mahout.apache.org/>

consumed by the users. In more detail, the DBpedia features are considered as the content of the items. Thereafter, in order to set the item similarity this approach applies the Jaccard Similarity to the DBpedia features of each item. The recommendation is thus reached using the *GenericBooleanPrefItemBasedRecommender*, implemented in Mahout, which given a target user, it recommends the most similar items (according to the aforementioned criteria) to those in the target user profile.

- **BPR:** This approach is an implementation of the Bayesian Personalized Ranking proposed by [\[Rendle et al., 2009\]](#).
- **FISM:** This approach applies the Factorized Item Similarity Model proposed by [\[Kabbur et al., 2013\]](#). This proposal addresses the top-N Recommendation problem by learning an item-item similarity matrix as the product of two low dimensional latent factor matrices. Once calculated, the item-item similarity matrix is applied to generate recommendations by predicting the user ratings on unrated items. In particular, the variant that consider a loss function based on the optimization of the Area Under the ROC Curve (FISMAuc) has been applied.
- **LDA:** This approach is based on the application Latent Dirichlet Allocation (LDA) presented by [\[Griffiths, 2002\]](#), based on the application of Gibbs Sampling.
- **PRankD:** This approach applies the proposal of [\[Hurley, 2013\]](#) that presents a diversification criterion that can be incorporated into a ranking-based objective. The methodology is based on the use of a matrix factorization model to learn user- and item-feature vectors by minimising the ranking-based objective. It is expected to result in recommendation sets that are highly diverse, while remaining highly relevant.
- **RankSGD:** This approach propose a ranking based method which tries to model users' choice between item pairs executed as the minimization of an objective function [\[Jahrer and Töscher, 2012\]](#).
- **HYBRID:** This methodology applies the approach presented by [\[Zhou et al., 2010\]](#). This proposal is based on a graph structure that is represented as a binary adjacency matrix of users and objects (where a cell in the matrix will be equal to 1 if the related object has been consumed by the related user and 0 otherwise). This structure is used to create the recommendation list by means of a hybrid methodology that combines accuracy- and diversity-related metrics by applying a heat spreading (HeatS) algorithm.
- **AR:** This approach is based on the algorithm presented by [\[Kim and Kim, 2003\]](#). They propose a model to predict preferences for items by using Association Rule

mining. More in detail, recommendation is based on the discovering of rules that reflect relationships between items. In particular, these rules try to find items that frequently appear together. Thereafter, to predict the user preference to a given item, the system aggregates the confidence of the rules including this item in the result part and the items previously consumed by the user in the condition part. Although the original paper includes a refined version including multi-level rules, this implementation only considers regular association rules.

- **LRMF:** This approach applies the recommendation proposal of [Shi et al., 2010]. This proposal combines a list-wise learning-to-rank algorithm with matrix factorization.
- **FCA-Baseline:** In order to test the advantages offered by the Knowledge organization presented in section LOD-RecSys Common Representation Space, two different versions of our FCA-based proposal has been implemented: one with and one without this step. In particular, this approach applies the recommendation pipeline without this Knowledge Organization step: items are represented by means of all their related DBpedia features, as described at 6.1.2.2.
- **FCA + Knowledge Organization:** This approach applies the whole recommendation pipeline explained in Section LOD-RecSys Common Representation Space.

### *Evaluation Setup*

The evaluation makes use the environment provided by challenge<sup>16</sup>. In this framework, the experimental results have to be uploaded to the Challenge Evaluation Interface where they are evaluated in terms of Precision, Recall and F-Measure. This process intends to isolate the evaluation process, ensuring its reproducibility as well as the fair comparison of the systems in the challenge. In order to be able to compare the conducted experiments to the other systems participating in the LOD-RecSys challenge, we have made use of this evaluation setup.

The metric proposed in the challenge is the F-measure of the Top-10 item recommendation list (*F-measure@10*). For more details on this metric, please refer to its definition in section 6.2.2.1. Moreover, more details on this metric and its application for recommendation can be consulted at [Ricci and Shapira, 2011].

---

<sup>16</sup> <http://dec020.poliba.it:8181/eswc2015lodrecsys/>

### *Results*

In this task, the organizers proposed a classic recommendation scenario. Systems have to predict an item ranking related to each user: the higher the ranking, the more relevant the item.

To address this task, first the DBpedia model was created and then the user-item matrix was modelled by applying FCA (as described in section LOD-RecSys Common Representation Space) and, finally, the recommendation algorithm in section 6.2.1 has been applied.

Table 6.3 shows the results obtained by the FCA-based proposal, the different baselines and the proposed state-of-the-art proposals for the two different datasets (music and movie), according to the official evaluation. The evaluation framework of the LOD-RecSys Challenge only provides the final measures; consequently, no statistical significance analysis can be carried out. Nevertheless, the difference in the results are big enough to extract meaningful conclusions: the **FCA-based** proposal is at least 20 % than other approaches (except that denoted as **HYDRID**), even offering one or two orders of magnitude improvement for some approaches.

The first remark is that our **FCA-based** recommendation proposal presents the best performance, significantly outperforming the baseline approaches and achieving better results than all the state-of-the-art proposals, just slightly improved by the **HYBRID** approach for the music dataset.

Regarding the baseline proposals, the **Random**-based recommendation, as expected, does not achieve meaningful results. Although this aspect seems obvious, it is important to highlight the difference between the different approaches and the random recommendation. Given the nature of the task, the results are quite low (at the level of 0.1 for F-measure). Therefore, it may seem that recommendation approaches are not much better than random guesses. Nevertheless, the results in Table 6.3 prove that recommendation proposals are able to capture aspects related to user preferences beyond random guesses.

Approach	Movie Dataset			Music Dataset		
	Precision	Recall	F(P,R)	Precision	Recall	F(P,R)
<b>Random</b>	0.0142	0.0204	0.0168	0.0275	0.0754	0.0403
<b>MostPop</b>	0.0679	0.0977	0.0801	0.0518	0.1411	0.0758
<b>CF-baseline</b>	0.0849	0.1230	0.1005	0.0802	0.2209	0.1177
<b>CB-baseline</b>	0.0659	0.0948	0.0778	0.0605	0.1665	0.0887
<b>BPR</b>	0.0703	0.1011	0.0829	0.0653	0.1787	0.0955
<b>FISM</b>	0.0612	0.0904	0.073	0.0696	0.1946	0.1026
<b>LDA</b>	0.0842	0.1216	0.0995	0.0816	0.2227	0.1194
<b>PRankD</b>	0.001	0.0014	0.0011	0.009	0.0024	0.0013
<b>RankSGD</b>	0.0192	0.0277	0.0227	0.0178	0.0489	0.0261
<b>HYBRID</b>	0.1033	0.1502	0.1224	0.0987	0.2723	0.1449
<b>LRMF</b>	0.0016	0.0022	0.0019	0.0011	0.0028	0.0015
<b>AR</b>	0.0888	0.1283	0.1049	0.0825	0.2268	0.1208
<b>FCA + KO</b>	0.1005	0.1532	0.1249	0.0974	0.268	0.1429

**Table 6.3** – Official results for Top-N Recommendation Task compared to the baselines

Another interesting baseline proposal is the **Most Popular** recommendation. Given the kind of data in the experimental dataset (Facebook likes), offering the most popular items appears as a sensible approach. It is reflected in the good performance of the **MostPop** approach, improving more sophisticated baselines (e.g., Content-based for the movie dataset) and even improving some of the state-of-the-art recommendation proposals.

None of the other two baselines, that applying Collaborative Filtering outperforms the Content-based approach. It does make sense given that in Social Networks, such as Facebook, the user dimension is usually more interesting than the item content. In other words, items are more likely to be consumed, liked or shared by the users if some of their friends did so. In fact, this CF-baseline has proven to be a strong baseline, achieving results that are almost at the same level as the top-performing approaches.

Among all the state-of-the-art approaches, the **HYBRID** approach based on a graph-based representation offers the best results. This approach is somewhat similar to our proposal: it also tries to take advantage of the user-item incidence matrix in order to infer relationships that may result in accurate recommendations.

Another methodology that, according to [Table 6.3](#), seems to be suitable for this recommendation task is that based on Association Rule Mining (**AR**). Given a target user, it recommends items that frequently appear together with the items consumed by

the user. It implicitly offers items that are included in the union of the most popular item set and the set of items closely related to the target user. Consequently, it is expected that a more refined most popular approach might improve the original one.

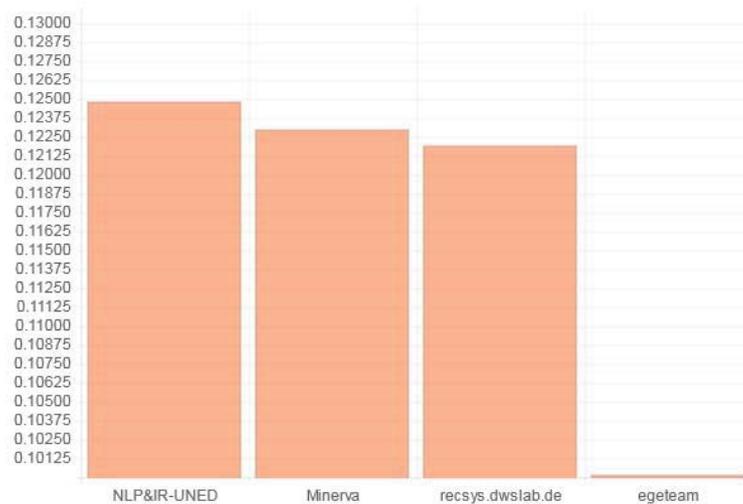
Other methodologies that obtain satisfactory results are: **Bayesian Personalized Ranking** (BPR) and **Latent Dirichlet Allocation** (LDA). The former supposes an implicit feedback scenario (that applied in this work) to carry out the recommendation process, which seems to affect the recommendation process positively. The latter applies a Content-based approach based on the well-known LDA methodology to group similar items together. In this regard, it is remarkable that in spite of LDA applying a much more sophisticated item modelling than the CB-baseline; it barely affects the final recommendation result. It is explained by the fact that items are described by their DBpedia features. In this kind of representation, considering features as binary is enough to define the content of the items and a simple co-occurrence based metric, such as the Jaccard Distance, is enough to capture the item similarities.

The rest of the recommendation methodologies do not achieve interesting results: their performance is lower than the **Most Popular** baseline. Neither Ranking Optimization-based approaches (**PRankD**, **RankSGD**), nor Matrix Factorization methodologies (**FISM**, **LRMF**) achieve valuable results for the recommendation task. **FISM**, based on matrix factorization, does improve most of the baseline algorithms (except the **CF-baseline**) for the music dataset, but does not for the movie dataset.

Especially notable is the extremely low performance of **LRMF** and **PRankD**. **LRMF** has been proven in the context of the MovieLens dataset [Shi et al., 2010]. The user-item incidence matrix in the MovieLens dataset has a high density (e.g., an average of 106 relationships between users and items and a user-item matrix density equal to 6.3 %). In contrast, the dataset used in this experimentation comes from a social network environment, where the user-item relationship density is much lower (e.g., in the music domain there is an average of 16 relationships between users and items and the density of the user-item matrix is equal to 0.25 %). Since **LRMF** relies greatly on the factorization of this user-item matrix, the low density affects the final recommendation performance significantly. On the other hand, **PRankD** tries to achieve a compromise between recommendation accuracy and diversity. Nevertheless, applied to this experimental configuration it seems to generate a barely accurate recommendation list.

To sum up, these results demonstrate that the proposed FCA-based approach improves the recommendation process through the inclusion of the refined DBpedia data and the use of the FCA-based common representation space. To confirm this point, [Figure 6.23](#)

shows the performance of our FCA-based approach in comparison to the other systems in the LOD-RecSys challenge (ours is denoted as NLP&IR-UNED)<sup>17</sup>. In particular, the figure indicates the results for Task 1 (see <http://sisinflab.poliba.it/events/lod-recsys-challenge-2015/tasks/> for more details), which makes use of the movie domain. Results for Task 2, applying the music domain are shown in [Figure 6.24](#). As shown in the figures, our system outperforms the others recommender systems in the task for both, the movie and music dataset.



**Figure 6.23** – Official results for Task 1

---

<sup>17</sup> Official results can be consulted at: <http://dee020.poliba.it:8181/eswc2015lodrecsys/leaderboard.html>

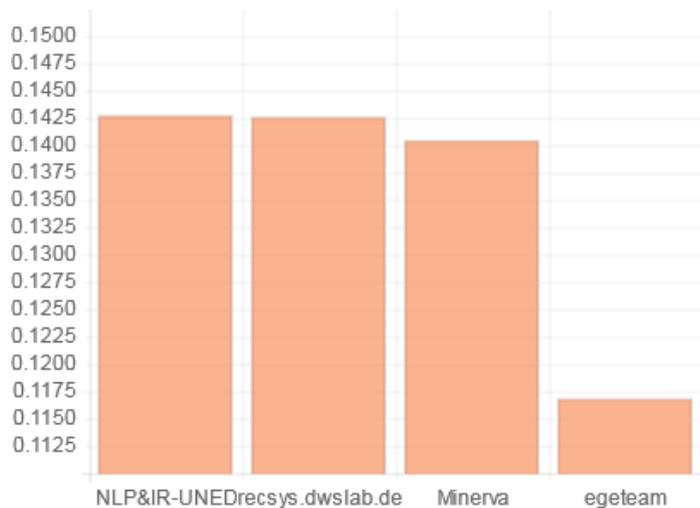


Figure 6.24 – Official results for Task 2

### General Remarks and Discussion

Section 6.2.2.1 proved that the representation of both, users and items, in the same representation space reduced the gap between both dimensions, thus improving the identification of user preferences and the linking of such preferences to the items fulfilling them. The experimentation in section 6.2.2.1 also pointed out that when higher-level representations were available, the recommendation process were improved by mitigating the problems related to more shallow representations, based on textual information or basic item features.

In this regard, the experimental scenario provided by the 2<sup>nd</sup> Linked Open Data-enabled Recommender Systems Challenge has been applied to experimentally confirm this latter aspect, as well as to apply the FCA-based proposal to an experimental task where the comparison to other state-of-the-art proposals is possible.

The first aspect to highlight is the overall performance of our proposal. In particular, results in Table 6.3 confirm that the proposed FCA common representation space, based on DBpedia data, is able to accurately represent items. This in turn leads to a more accurate recommendation process. In more detail, the approach applying the proposed FCA-based modelling outperforms the baseline proposals and the other state-of-the-art recommendation proposals for both tasks (i.e., for the second task, HYBRID approach achieves similar results).

Summing up, the proposed FCA-based common representation approach has again proved to enable the more accurate representation of recommendation data, providing an accurate recommendation (for the proposed tasks), being able to outperform other

state of the art methodologies. Furthermore, the results in this section prove that when this representation is based on higher-level semantic features, the recommendation process is improved, driven by the more accurate data representation.

# 7

## Conclusions

---

*This section details the conclusions of this thesis in regards to the general objectives, hypothesis and research questions.*

### Content

---

7.1	MAIN OUTCOMES .....	261
7.2	FUTURE DIRECTIONS .....	261

---

This thesis addressed the problem of content recommendation in social scenarios. Although the research in recommendation has a large tradition since its birth in the early 90s, with the advent of the social web, the challenges that recommender systems should face are larger and more complicated. In this context, the so-called top-N recommender systems have gained momentum, mainly fuelled by the problems of traditional rating-based recommender systems when applied to these social data. When focusing on this top-N scenario, there is still a large room for improvement for the recommender systems. In particular, we formulated the following problem:

**Recommendation task is usually addressed in the literature by modelling users and items by separate, resulting in a gap between both representation spaces.**

To cope with this problem, this thesis proposed a novel representation for user and items in recommender systems. This representation was based on generating a common representation space for both, users and items. It was expected to reduce in this way the user-item gap, thus improving the recommendation process. With this idea in mind, we proposed a methodology to create this common space based on a hierarchical representation of semantically based concepts, automatically inferred from the data. This methodology goes a step further in relation to those proposed in the state of the art based on Formal Concept Analysis.

Formal Concept Analysis has been proven in the literature as a powerful data organization technique. In consequence, we expected to take advantage of this high performance to model users and items in the scope of a recommender system, thus enabling the creation of the common representation space. In this regard, Section 4 has experimentally proven the suitability of FCA to model contents coming from social environments, namely Twitter. In more detail, the FCA was successfully applied to identify thematically similar contents and to hierarchically organize them according to its specificity. Furthermore, FCA demonstrated its performance in comparison to other state-of-the-art data representation approaches by achieving the best results for the scenario proposed for experimentation.

Section 5 goes deeper in the content representation. While the representation in section 4 was based on shallow textual representations, this section proposed the use of higher-level semantic features to describe the contents. It was demonstrated that these features were preferable. The high-performing results in the previous experimentation were improved when semantic features coming from DBpedia and EuroWordNet were applied to describe the contents.

Both sections 4 and 5 demonstrated the suitability of our FCA-based approach to accurately model contents in the scope of the latter recommendation task: social-based scenarios like Twitter or Facebook. These experimentations aimed to isolate the representation step from its actual application to the content recommendation. In this way, the FCA performance for content representation was proved independently of the latter application of this representation. Thereafter, section 6 presents the application of the FCA-based proposal for the generation of the common representation space. In order to prove whether the high performance of our FCA-based proposal for content representation enabled the improvement of the recommendation process, we proposed two different experimental configurations.

The first one is detailed in section 6.2.2.1 in the scenario of News Recommendation in Twitter. In this section, a large experimentation to evaluate the FCA-based proposal for several configurations and applying different input data as item representations has been developed. In addition, to frame the performance of this proposal, it is compared to several state-of-the-art recommendation proposals. This experimentation confirmed the initial hypothesis that a common representation space (i.e., that enabled by our FCA-based representation) should enhance the recommendation process. In addition, this experimentation proved, as it did in section 5 for data representation, that, when available, the use of higher-level semantic features improved the recommendation process.

Finally, section 6.2.2.2 presents the second recommendation scenario to experiment with the FCA-based proposal. It aims to compile all the previous experimentations in order to confirm the initial hypothesis. To that end, it was applied the final and more refined version of the proposal: a common representation space based on semantic features for the content recommendation in social scenarios. All the conclusions already drawn in previous sections are again confirmed: FCA performance for data representation, higher performance of semantic-driven representations, the suitability of the common representation space and state-of-the-art results for recommendation. In addition, when applied to the experimental scenario proposed in this section (based on recommending movies and music to Facebook users), not only the FCA system outperformed other state-of-the-art proposal, but also the rest of the approaches in the task.

To sum up, the initial hypothesis that gave risen to this thesis has been confirmed. FCA is able to generate a semantic-driven common representation for recommendation, which achieves state-of-the-art results for the different experimental scenarios.

## 7.1 Main outcomes

---

From the latter conclusions, it can be extracted a set of take-home points, which may be useful in the process of building a top-N based recommender system.

The first remark is that **the representation step is crucial in the recommendation task**. When accurate data representations are available, even a simple recommendation process, as that proposed in this thesis, are able to offer relevant recommendations. In this regard, the top-N recommendation task can be seen as a modelling task, in contrast to the “pattern discovery” view applied to rating-based recommender systems.

In relation to this representation step, **the use of higher-level features to model contents significantly improves the recommendation task**. Not only these features enable less sparse and lighter representations, but also they are able to better capture the semantics of the item content.

**Formal Concept Analysis presents itself as a powerful technique for recommendation**. It has extensively demonstrated in the experimentation that: 1) it is able to accurately model a set of contents by capturing their latent conceptual structure; and, 2) it enables the accurate model of recommendation data leading to relevant recommendations.

**The representation of both dimensions in the recommendation task, users and items, in a common space makes sense from the theoretical point of view, resulting in state-of-the-art results.**

## 7.2 Future Directions

---

The data representation by means of FCA has extensively demonstrated its suitability in terms of the evaluation of the quality of the generated representations. Nonetheless, there is still an open question related to the temporal dimension. The entire experimental setup has been based on the evaluation of static models. These models are snapshots at a specific time of the evolution of content description or user preferences along the time. Although our proposal has proven in the experimentation presented at section 6 that is able to integrate previous data and to update the representation when new data appear, it remains the formal evaluation of the implications of this aspect to model temporal

sequences of data. As a future work, the adaptation for these temporal sequences, especially in the context of recommendation, should be addressed.

In this thesis, FCA has been applied to offline scenarios, where a model is created upfront to be then applied in the recommendation process. In this sense, it remains as a future line the adaptation of this proposal to online scenarios, where models are created at real-time by consuming streams of input data (e.g., Twitter stream).

Although the FCA-enabled data representation accurately captures the latent conceptual structure of the data, which in turns results in state-of-the-art recommendations, the recommendation algorithm itself is far from wholly take advantage of this data representation. This algorithm is based on a basic navigation across the lattice, by applying the idea of the concept neighbourhood. Nevertheless, the concept lattice provides a rich interpretation of the conceptual structure inherent to the data represented in the lattice [Ganter et al., 2016]. In this sense, semantic-based similarity measures, as that studied in [Lastra-Díaz and García-Serrano, 2015] seems to be promising to take advantage of this interpretation power, thus enabling a more “intelligent” recommendation process. In this regard, as a future work remains the refining of the recommendation algorithm by including this idea.

In the scope of recommender systems, the inclusion of contextual features appears as a promising direction. Contextual features have been overlooked in this thesis; however, they might influence the recommendation process (i.e., users are more likely to consume certain types of items at certain environments). Contextual features can be easily included in the modelling process by generating context-based formal contexts (i.e., formal context only including user-item interactions at a given context) and then apply them in the recommendation process.

Finally, the experimentation in what refers to the recommendation task has been only focused on testing the accuracy of the recommendations. With this experimentation, we aimed to prove the suitability of our proposal from the most basic perspective; that is, it actually offers relevant recommendations. As a future work remains the follow-up of this experimentation, integrating other dimensions commonly considered in the evaluation of the recommendation performance: novelty, diversity, serendipity. As proven by the results in the literature of top-N recommender systems, there is still a large room for improvement.



# LIST OF PUBLICATIONS

---

- [Castellanos et al., 2013] Angel Castellanos, Juan Cigarrán, and Ana García-Serrano. Modelling techniques for twitter contents: A step beyond classification based approaches. In *Working Notes of the CLEF 2013*, 2013. URL <http://www.clef-initiative.eu/documents/71612/86f6c2c1-5148-4c18-a41e-a59a0d9f0fa1>.
- [Castellanos et al., 2014a] A. Castellanos, E.W. de Luca, J. Cigarrán, and A. García-Serrano. Improving data representation through a concept-based knowledge organization. In *Workshop on Knowledge Organization and Semantic Web, Leipzig*, 2014.
- [Castellanos et al., 2014b] Angel Castellanos, Ana García-Serrano, and Juan Cigarrán. Linked Data-based Conceptual Modelling for Recommendation: A FCA-based Approach. In Martin Hepp and Yigal Hoffner, editors, *E-Commerce and Web Technologies*, volume 188 of *Lecture Notes in Business Information Processing*, pages 71–76. Springer International Publishing, 2014. ISBN 978-3-319-10490-4. doi: [10.1007/978-3-319-10491-1\\_8](https://doi.org/10.1007/978-3-319-10491-1_8). URL [http://dx.doi.org/10.1007/978-3-319-10491-1\\_8](http://dx.doi.org/10.1007/978-3-319-10491-1_8).
- [Castellanos et al., 2014c] Angel Castellanos, Juan Cigarrán, and Ana García-Serrano. UNED @ CLEF NEWSREEL 2014. In *Working Notes of the CLEF 2014*, 2014. URL <http://www.clef-initiative.eu/documents/71612/86f6c2c1-5148-4c18-a41e-a59a0d9f0fa1>.
- [Castellanos et al., 2015] Angel Castellanos, Ernesto W De Luca, Juan Cigarrán, and Ana García-Serrano. Time, place and environment: Can conceptual modelling improve context-aware recommendation? In *Workshop on Context-awareness in Retrieval and Recommendation in conjunction with the 37th European Conference on Information Retrieval (ECIR '15). 29 March - 2 April, Vienna*, 2015.
- [Cigarrán et al., 2016] Juan Cigarrán, Ángel Castellanos, and Ana García-Serrano. A step forward for Topic Detection in Twitter: An FCA-based approach. *Expert Systems with Applications*, 57: 21 – 36, 2016. ISSN 0957-4174. doi: <http://dx.doi.org/10.1016/j.eswa.2016.03.011>. URL <http://www.sciencedirect.com/science/article/pii/S0957417416301038>.
- [Castellanos et al., 2016a] Angel Castellanos, Ana García-Serrano, and Juan C. Filling the gap between users and goods through a fca-based common representation space: Top-n recommendation experimentation and evaluation. *User Modelling User-Adaptation Interaction*, 2016. Submitted.
- [Castellanos et al., 2016b] Angel Castellanos, Ana García-Serrano, and Juan Cigarrán. Improving recommendation through a concept-driven common representation

space: An experimental study. *IEEE Transactions on Knowledge and Data Engineering*, 2016. Submitted.

[Castellanos et al., 2017] Angel Castellanos, Juan Cigarrán, and Ana García-Serrano. Formal concept analysis for topic detection: A clustering quality experimental analysis. *Information Systems*, 2017 (To appear).

[Meras et al., 2017] Adolfo Merás, Ana García-Serrano, Angel Castellanos. Extracción de información temporal de la DBpedia: propuesta e integración en un corpus semiestructurado. *Revista de Procesamiento del Lenguaje Natural*, N° 58. 2017. (To appear)

[Scheel et al., 2014] Christian Scheel, Angel Castellanos, Thebin Lee, and E.W. De Luca. *The Reason Why: A Survey of Explanations for Recommender Systems*. In *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation*. Springer International Publishing., 2014.

# BIBLIOGRAPHY

---

- [Abel et al., 2008] Fabian Abel, Ig Ibert Bittencourt, Nicola Henze, Daniel Krause, and Julita Vassileva. *A Rule-Based Recommender System for Online Discussion Forums*, pages 12–21. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-70987-9. doi: [10.1007/978-3-540-70987-9\\_4](https://doi.org/10.1007/978-3-540-70987-9_4). URL [http://dx.doi.org/10.1007/978-3-540-70987-9\\_4](http://dx.doi.org/10.1007/978-3-540-70987-9_4).
- [Abel et al., 2011] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In *Proceedings of the 3rd International Web Science Conference, WebSci '11*, pages 2:1–2:8, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0855-7. doi: [10.1145/2527031.2527040](https://doi.org/10.1145/2527031.2527040). URL <http://doi.acm.org/10.1145/2527031.2527040>.
- [Abello, 2004] James Abello. Hierarchical graph maps. *Computers & Graphics*, 28 (3): 345–359, 2004.
- [Adomavicius and Kwon, 2011] Gediminas Adomavicius and Y Kwon. Maximizing aggregate recommendation diversity: A graph-theoretic approach. In *Proc. of the 1st International Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011)*, pages 3–10, 2011.
- [Adomavicius and Tuzhilin, 2005] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17 (6): 734–749, 2005.
- [Agarwal et al., 2011a] Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media, LSM '11*, pages 30–38, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. ISBN 978-1-932432-96-1. URL <http://dl.acm.org/citation.cfm?id=2021109.2021114>.
- [Agarwal et al., 2011b] Deepak Agarwal, Bee-Chung Chen, and Bo Long. Localized factor models for multi-context recommendation. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11*, pages 609–617, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0813-7. doi: [10.1145/2020408.2020504](https://doi.org/10.1145/2020408.2020504). URL <http://doi.acm.org/10.1145/2020408.2020504>.
- [Aggarwal, 2016] Charu C. Aggarwal. *Recommender Systems: The Textbook*. Springer International Publishing, 2016. doi: [10.1007/978-3-319-29659-3](https://doi.org/10.1007/978-3-319-29659-3).
- [Aggarwal and Parthasarathy, 2001] Charu C Aggarwal and Srinivasan Parthasarathy. Mining massively incomplete data sets by conceptual reconstruction. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 227–232. ACM, 2001.
- [Agrawal et al., 2009] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Jeong. Diversifying search results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 5–14. ACM, 2009.

- [Ahern et al., 2007] Shane Ahern, Mor Naaman, Rahul Nair, and Jeannie Hui-I Yang. World explorer: Visualizing aggregate data from unstructured text in geo-referenced collections. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '07*, pages 1–10, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-644-8. doi: [10.1145/1255175.1255177](https://doi.org/10.1145/1255175.1255177). URL <http://doi.acm.org/10.1145/1255175.1255177>.
- [Alam and Napoli, 2014] Mehwish Alam and Amedeo Napoli. Lattice-based views over sparql query results. *Linked Data for Knowledge Discovery*, page 49, 2014.
- [Alam et al., Unpublished results] Mehwish Alam, Aleksey Buzmakov, Vctor Codocedo, and Amedeo Napoli. Completing rdf data in linked open data cloud using formal concept analysis. Unpublished results.
- [Albanese et al., 2011] M. Albanese, A. d’Acierno, V. Moscato, F. Persia, and A. Picariello. A multimedia semantic recommender system for cultural heritage applications. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on*, pages 403–410, Sept 2011. doi: [10.1109/ICSC.2011.47](https://doi.org/10.1109/ICSC.2011.47).
- [Alexandridis et al., 2013] Georgios Alexandridis, Georgios Siolas, and Andreas Stafylopatis. Improving social recommendations by applying a personalized item clustering policy. In *Proceedings of the 5th RecSys Workshop on Recommender Systems & the Social Web (RSWeb)*, 2013.
- [Allan et al., 1998] James Allan, Jaime G Carbonell, George Doddington, Jonathan Yamron, and Yiming Yang. Topic detection and tracking pilot study final report. In *In Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, pages 194–218, 1998.
- [Alqadah and Bhatnagar, 2011] Faris Alqadah and Raj Bhatnagar. Similarity measures in formal concept analysis. *Annals of Mathematics and Artificial Intelligence*, 61 (3): 245–256, 2011.
- [AlSumait et al., 2008] L. AlSumait, D. Barbara, and C. Domeniconi. On-line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking. In *Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on*, pages 3–12, 2008. doi: [10.1109/ICDM.2008.140](https://doi.org/10.1109/ICDM.2008.140).
- [Amigó et al., 2013a] Enrique Amigó, Jorge Carrillo de Albornoz, Irina Chugur, Adolfo Corujo, Julio Gonzalo, Tamara Martín, Edgar Meij, Maarten Rijke, and Damiano Spina. Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. In Pamela Forner, Henning MÅller, Roberto Paredes, Paolo Rosso, and Benno Stein, editors, *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, volume 8138 of *Lecture Notes in Computer Science*, pages 333–352. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-40801-4. doi: [10.1007/978-3-642-40802-1\\_31](https://doi.org/10.1007/978-3-642-40802-1_31). URL [http://dx.doi.org/10.1007/978-3-642-40802-1\\_31](http://dx.doi.org/10.1007/978-3-642-40802-1_31).
- [Amigó et al., 2013b] Enrique Amigó, Julio Gonzalo, and Felisa Verdejo. A general evaluation measure for document organization tasks. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13*,

- pages 643–652, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2034-4. doi: [10.1145/2484028.2484081](https://doi.org/10.1145/2484028.2484081). URL <http://doi.acm.org/10.1145/2484028.2484081>.
- [Anand et al., 2007] Sarabjot Singh Anand, Patricia Kearney, and Mary Shapcott. Generating semantically enriched user profiles for web personalization. *ACM Trans. Internet Technol.*, 7 (4), October 2007. ISSN 1533-5399. doi: [10.1145/1278366.1278371](https://doi.org/10.1145/1278366.1278371). URL <http://doi.acm.org/10.1145/1278366.1278371>.
- [Anta et al., 2012] Antonio Fernandez Anta, Luis Nunez Chiroque, Philippe Morere, and Agustin Santos. Sentiment Analysis and Topic Detection of Spanish Tweets: A Comparative Study of of NLP Techniques. *Procesamiento del Lenguaje Natural*, 50 (0), 2012. ISSN 1989-7553. URL <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4658>.
- [Anthes, 2010] Gary Anthes. Topic models vs. unstructured data. *Commun. ACM*, 53 (12): 16–18, December 2010. ISSN 0001-0782. doi: [10.1145/1859204.1859210](https://doi.org/10.1145/1859204.1859210). URL <http://doi.acm.org/10.1145/1859204.1859210>.
- [Arazy et al., 2009] Ofer Arazy, Nanda Kumar, and Bracha Shapira. Improving social recommender systems. *IT professional*, 11 (4): 38–44, 2009.
- [Armentano et al., 2012] Marcelo G Armentano, Daniela Godoy, and Analá Amandi. Topology-based recommendation of users in micro-blogging communities. *Journal of Computer Science and Technology*, 27 (3): 624–634, 2012.
- [Asmus et al., 2014] Josefine Asmus, Daniel Borchmann, Ivo F Sbalzarini, and Dirk Walther. Towards an fca-based recommender system for black-box optimization. In *Proceedings of the FCA4I Workshop in conjunction with European Conference on Artificial Intelligence (ECAI '14)*, pages 35–43, 2014.
- [Asur et al., 2011] Sitaram Asur, Bernardo A Huberman, Gabor Szabo, and Chunyan Wang. Trends in social media: Persistence and decay. *Available at SSRN 1755748*, 2011.
- [Aswanikumar and Srinivas, 2010] Ch. Aswanikumar and S. Srinivas. Concept lattice reduction using fuzzy k-means clustering. *Expert Syst. Appl.*, 37 (3): 2696–2704, 2010.
- [Auer et al., 2007] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *Proceedings of the 6th International The Semantic Web and 2Nd Asian Conference on Asian Semantic Web Conference, ISWC'07/ASWC'07*, pages 722–735, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 3-540-76297-3, 978-3-540-76297-3. URL <http://dl.acm.org/citation.cfm?id=1785162.1785216>.
- [Bagga and Baldwin, 1998] Amit Bagga and Breck Baldwin. Entity-based Cross-document Coreferencing Using the Vector Space Model. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1, ACL '98*, pages 79–85, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics. doi: [10.3115/980845.980859](https://doi.org/10.3115/980845.980859). URL <http://dx.doi.org/10.3115/980845.980859>.

- [Balabanovic and Shoham, 1997] Marko Balabanovic and Yoav Shoham. Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 40 (3): 66–72, 1997.
- [Bar et al., 2012] Ariel Bar, Lior Rokach, Guy Shani, Bracha Shapira, and Alon Schlar. Boosting simple Collaborative Filtering models using ensemble methods. *arXiv preprint arXiv:1211.2891*, 2012.
- [Basile et al., 2014] Pierpaolo Basile, Cataldo Musto, Marco De Gemmis, Pasquale Lops, Fedelucio Narducci, and Giovanni Semeraro. Aggregation strategies for linked open data-enabled recommender systems. In *European Semantic Web Conference (Satellite Events)*, 2014.
- [Basilico and Hofmann, 2004] Justin Basilico and Thomas Hofmann. Unifying collaborative and Content-based filtering. In *Proceedings of the Twenty-first International Conference on Machine Learning, ICML '04*, pages 9–, New York, NY, USA, 2004. ACM. ISBN 1-58113-838-5. doi: [10.1145/1015330.1015394](https://doi.org/10.1145/1015330.1015394). URL <http://doi.acm.org/10.1145/1015330.1015394>.
- [Becker et al., 2011] Hila Becker, Mor Naaman, and Luis Gravano. Beyond trending topics: Real-world event identification on twitter. *ICWSM*, 11: 438–441, 2011.
- [Bedi and Richa, 2015] Punam Bedi and Richa. User interest expansion using spreading activation for generating recommendations. In *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on*, pages 766–771, Aug 2015. doi: [10.1109/ICACCI.2015.7275703](https://doi.org/10.1109/ICACCI.2015.7275703).
- [Bedi and Vashisth, 2014] Punam Bedi and Pooja Vashisth. Empowering recommender systems using trust and argumentation. *Information Sciences*, 279: 569–586, 2014.
- [Beel et al., 2016] Joeran Beel, Corinna Breiter, Stefan Langer, Andreas Lommatzsch, and Bela Gipp. Towards reproducibility in recommender-systems research. *User Modeling and User-Adapted Interaction*, pages 1–33, 2016.
- [Beitzel et al., 2009] Steven M. Beitzel, Eric C. Jensen, and Ophir Frieder. *MAP*, pages 1691–1692. Springer US, Boston, MA, 2009. ISBN 978-0-387-39940-9. doi: [10.1007/978-0-387-39940-9\\_492](https://doi.org/10.1007/978-0-387-39940-9_492). URL [http://dx.doi.org/10.1007/978-0-387-39940-9\\_492](http://dx.doi.org/10.1007/978-0-387-39940-9_492).
- [Bell and Koren, 2007] Robert M. Bell and Yehuda Koren. Lessons from the netflix prize challenge. *SIGKDD Explor. Newsl.*, 9 (2): 75–79, December 2007. ISSN 1931-0145. doi: [10.1145/1345448.1345465](https://doi.org/10.1145/1345448.1345465). URL <http://doi.acm.org/10.1145/1345448.1345465>.
- [Belohlavek and Vychodil, 2009] R. Belohlavek and V. Vychodil. Formal concept analysis with background knowledge: Attribute priorities. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 39 (4): 399–409, 2009. ISSN 1094-6977. doi: [10.1109/TSMCC.2008.2012168](https://doi.org/10.1109/TSMCC.2008.2012168).
- [Belohlavek, 2008] RADIM Belohlavek. Introduction to formal concept analysis. *Olomouc, UPOL, Faculty of Science, Department of Computer Science*, 2008.
- [Bengel et al., 2004] Jason Bengel, Susan Gauch, Era Mittur, and Rajan Vijayaraghavan. ChatTrack: Chat Room Topic Detection Using Classification. In Hsinchun Chen, Reagan Moore, Daniel D. Zeng, and John Leavitt, editors, *Intelligence and Security Informatics*,

- volume 3073 of *Lecture Notes in Computer Science*, pages 266–277. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-22125-8. doi: [10.1007/978-3-540-25952-7\\_20](https://doi.org/10.1007/978-3-540-25952-7_20). URL [http://dx.doi.org/10.1007/978-3-540-25952-7\\_20](http://dx.doi.org/10.1007/978-3-540-25952-7_20).
- [Bengio and Courville, 2016] Ian Goodfellow Yoshua Bengio and Aaron Courville. Deep learning. Book in preparation for MIT Press, 2016. URL <http://www.deeplearningbook.org>.
- [Bengio, 2012] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In Grégoire Montavon, GenevieveB. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade*, volume 7700 of *Lecture Notes in Computer Science*, pages 437–478. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-35288-1. doi: [10.1007/978-3-642-35289-8\\_26](https://doi.org/10.1007/978-3-642-35289-8_26). URL [http://dx.doi.org/10.1007/978-3-642-35289-8\\_26](http://dx.doi.org/10.1007/978-3-642-35289-8_26).
- [Bengio et al., 2003] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3: 1137–1155, 2003.
- [Bennett and Lanning, 2007] James Bennett and Stan Lanning. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35, 2007.
- [Bentivogli et al., 2004] Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. Revising the wordnet domains hierarchy: semantics, coverage and balancing. In *Proceedings of the Workshop on Multilingual Linguistic Ressources*, pages 101–108. Association for Computational Linguistics, 2004.
- [Berners-Lee et al., 2009] Tim Berners-Lee, Christian Bizer, and Tom Heath. Linked data—the story so far. *International Journal on Semantic Web and Information Systems*, 5 (3): 1–22, 2009.
- [Berrocal et al., 2013] J. Berrocal, Luis, C. G. Figuerola, and Á. Z. Rodríguez. REINA at RepLab2013 Topic Detection Task : Community Detection. In *RepLab 2013, in CLEF Working Notes*, 2013.
- [Bidart et al., 2014] R. Bidart, A.C.M. Pereira, J.M. Almeida, and A. Lacerda. Where should i go? city recommendation based on user communities. In *Web Congress (LA-WEB), 2014 9th Latin American*, pages 50–58, Oct 2014. doi: [10.1109/LAWeb.2014.15](https://doi.org/10.1109/LAWeb.2014.15).
- [Billsus and Pazzani, 1998] Daniel Billsus and Michael J. Pazzani. Learning collaborative information filters. In *Proceedings of the Fifteenth International Conference on Machine Learning, ICML '98*, pages 46–54, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc. ISBN 1-55860-556-8. URL <http://dl.acm.org/citation.cfm?id=645527.657311>.
- [Billsus and Pazzani, 2000] Daniel Billsus and Michael J. Pazzani. User modeling for adaptive news access. *User Modeling and User-Adapted Interaction*, 10 (2): 147–180, 2000. ISSN 1573-1391. doi: [10.1023/A:1026501525781](https://doi.org/10.1023/A:1026501525781). URL <http://dx.doi.org/10.1023/A:1026501525781>.
- [Blanco-Fernández et al., 2008] Yolanda Blanco-Fernández, José J Pazos-Arias, Alberto Gil-Solla, Manuel Ramos-Cabrer, Martn López-Nores, Jorge Garcá-Duque, Ana Fernández-Vilas, Rebeca P Dáz-Redondo, and Jesús Bermejo-Muñoz. A flexible semantic inference

- methodology to reason about user preferences in knowledge-based recommender systems. *Knowledge-Based Systems*, 21 (4): 305–320, 2008.
- [Blei and McAuliffe, 2010] D. M. Blei and J. D. McAuliffe. Supervised Topic Models. *ArXiv e-prints*, March 2010.
- [Blei, 2012] David M Blei. Probabilistic topic models. *Communications of the ACM*, 55 (4): 77–84, 2012.
- [Blei et al., 2003] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of machine Learning research*, 3: 993–1022, 2003.
- [Blunsom et al., 2014] Phil Blunsom, Edward Grefenstette, Nal Kalchbrenner, et al. A Convolutional Neural Network for Modelling Sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 655–665. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, USA, June 23-25 2014.
- [Bobadilla et al., 2013] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46: 109–132, 2013.
- [Bonhard et al., 2007] Philip Bonhard, M. Angela Sasse, and Clare Harries. "the devil you know knows best": How online recommendations can benefit from social networking. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It - Volume 1*, BCS-HCI '07, pages 77–86, Swinton, UK, UK, 2007. British Computer Society. ISBN 978-1-902505-94-7. URL <http://dl.acm.org/citation.cfm?id=1531294.1531305>.
- [Boteanu et al., 2014] Bogdan Boteanu, Ionut Mironica, Anca-Livia Radu, and Bogdan Ionescu. Lapi@ 2014 retrieving diverse social images task: A relevance feedback diversification perspective. volume 1263 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [Brand, 2005] Matthew Brand. A random walks perspective on maximizing satisfaction and profit. In *SDM*, pages 12–19. SIAM, 2005.
- [Burke, 2002] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12 (4): 331–370, November 2002. ISSN 0924-1868. doi: [10.1023/A:1021240730564](https://doi.org/10.1023/A:1021240730564). URL <http://dx.doi.org/10.1023/A:1021240730564>.
- [Burke, 2007] Robin Burke. *Hybrid Web Recommender Systems*, pages 377–408. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-72079-9. doi: [10.1007/978-3-540-72079-9\\_12](https://doi.org/10.1007/978-3-540-72079-9_12). URL [http://dx.doi.org/10.1007/978-3-540-72079-9\\_12](http://dx.doi.org/10.1007/978-3-540-72079-9_12).
- [Cacheda et al., 2011] Fidel Cacheda, Vctor Carneiro, Diego Fernández, and Vreixo Formoso. Comparison of Collaborative Filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web (TWEB)*, 5 (1): 2, 2011.
- [Calinski and Harabasz, 1974] Tadeusz Calinski and Jerzy Harabasz. Á dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3 (1): 1–27, 1974.
- [Candillier et al., 2008] Laurent Candillier, Frank Meyer, and Françoise Fessant. Designing specific weighted similarity measures to improve Collaborative Filtering systems. In

- Proceedings of the 8th Industrial Conference on Advances in Data Mining: Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, ICDM '08, pages 242–255, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-70717-2. doi: [10.1007/978-3-540-70720-2\\_19](https://doi.org/10.1007/978-3-540-70720-2_19). URL [http://dx.doi.org/10.1007/978-3-540-70720-2\\_19](http://dx.doi.org/10.1007/978-3-540-70720-2_19).
- [Canny, 2002] John Canny. Collaborative Filtering with privacy via factor analysis. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '02, pages 238–245, New York, NY, USA, 2002. ACM. ISBN 1-58113-561-0. doi: [10.1145/564376.564419](https://doi.org/10.1145/564376.564419). URL <http://doi.acm.org/10.1145/564376.564419>.
- [Cantador, 2008] Iván Cantador. *Exploiting the conceptual space in hybrid recommender systems: a semantic-based approach*. PhD thesis, Universidad Autónoma de Madrid. Escuela Politécnica Superior. Departamento de Ingeniería Informática, 2008.
- [Cantador et al., 2008a] Iván Cantador, Alejandro Bellogín, and Pablo Castells. A multilayer ontology-based hybrid recommendation model. *AI Commun.*, 21 (2-3): 203–210, April 2008. ISSN 0921-7126. URL <http://dl.acm.org/citation.cfm?id=1460172.1460184>.
- [Cantador et al., 2008b] Iván Cantador, Alejandro Bellogn, and Pablo Castells. News@hand: A semantic web approach to recommending news. In *Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, AH '08, pages 279–283, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-70984-8. doi: [10.1007/978-3-540-70987-9\\_34](https://doi.org/10.1007/978-3-540-70987-9_34). URL [http://dx.doi.org/10.1007/978-3-540-70987-9\\_34](http://dx.doi.org/10.1007/978-3-540-70987-9_34).
- [Cao et al., 2015] Ziqiang Cao, Sujian Li, Yang Liu, Wenjie Li, and Heng Ji. A novel neural topic model and its supervised extension. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2210–2216, 2015.
- [Carpineto and Romano, 2004] Claudio Carpineto and Giovanni Romano. *Concept data analysis: Theory and applications*. John Wiley & Sons, 2004.
- [Carrer-Neto et al., 2012] Walter Carrer-Neto, María Luisa Hernández-Alcaraz, Rafael Valencia-García, and Francisco García-Sánchez. Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications*, 39 (12): 10990 – 11000, 2012. ISSN 0957-4174. doi: [http://dx.doi.org/10.1016/j.eswa.2012.03.025](https://doi.org/10.1016/j.eswa.2012.03.025). URL <http://www.sciencedirect.com/science/article/pii/S0957417412004952>.
- [Castellanos et al., 2014a] A. Castellanos, E.W. de Luca, J. Cigarrán, and A. García-Serrano. Improving data representation through a concept-based knowledge organization. In *Workshop on Knowledge Organization and Semantic Web, Leipzig*, 2014.
- [Castellanos et al., 2013] Angel Castellanos, Juan Cigarrán, and Ana García-Serrano. Modelling techniques for twitter contents: A step beyond classification based approaches. In *Working Notes of the CLEF 2013*, 2013. URL <http://www.clef-initiative.eu/documents/-71612/86f6c2c1-5148-4c18-a41e-a59a0d9f0fa1>.
- [Castellanos et al., 2014b] Angel Castellanos, Ana García-Serrano, and Juan Cigarrán. Linked Data-based Conceptual Modelling for Recommendation: A FCA-based Approach.

- In Martin Hepp and Yigal Hoffner, editors, *E-Commerce and Web Technologies*, volume 188 of *Lecture Notes in Business Information Processing*, pages 71–76. Springer International Publishing, 2014. ISBN 978-3-319-10490-4. doi: [10.1007/978-3-319-10491-1\\_8](https://doi.org/10.1007/978-3-319-10491-1_8). URL [http://dx.doi.org/10.1007/978-3-319-10491-1\\_8](http://dx.doi.org/10.1007/978-3-319-10491-1_8).
- [Castells et al., 2014] Pablo Castells, Jun Wang, Rubén Lara, and Dell Zhang. Introduction to the special issue on diversity and discovery in recommender systems. *ACM Trans. Intell. Syst. Technol.*, 5 (4): 52:1–52:3, December 2014. ISSN 2157-6904. doi: [10.1145/2668113](https://doi.org/10.1145/2668113). URL <http://doi.acm.org/10.1145/2668113>.
- [Cataldi et al., 2010] Mario Cataldi, Luigi Di Caro, and Claudio Schifanella. Emerging topic detection on twitter based on temporal and social terms evaluation. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining, MDMKDD '10*, pages 4:1–4:10, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0220-3. doi: [10.1145/1814245.1814249](https://doi.org/10.1145/1814245.1814249). URL <http://doi.acm.org/10.1145/1814245.1814249>.
- [Chakrabarti et al., 2008] Soumen Chakrabarti, Rajiv Khanna, Uma Sawant, and Chiru Bhattacharyya. Structured Learning for Non-smooth Ranking Losses. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, pages 88–96, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4. doi: [10.1145/1401890.1401906](https://doi.org/10.1145/1401890.1401906). URL <http://doi.acm.org/10.1145/1401890.1401906>.
- [Cheekula et al., 2015] Siva Kumar Cheekula, Pavan Kapanipathi, Derek Doran, Prateek Jain, and Amit Sheth. Entity recommendations using hierarchical knowledge bases. In *In Proceedings of 4th workshop on Knowledge Discovery and Data Mining Meets Linked Open Data at ESWC2015*, 2015.
- [Chen et al., 2013] Hong Chen, Mingxin Gan, and Mengzhao Song. A graph model for recommender systems. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*. Atlantis Press, 2013.
- [Chen et al., 2010] Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. Short and tweet: Experiments on recommending content from information streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 1185–1194, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: [10.1145/1753326.1753503](https://doi.org/10.1145/1753326.1753503). URL <http://doi.acm.org/10.1145/1753326.1753503>.
- [Chen et al., 2012] Kailong Chen, Tianqi Chen, Guoqing Zheng, Ou Jin, Enpeng Yao, and Yong Yu. Collaborative personalized tweet recommendation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12*, pages 661–670, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1472-5. doi: [10.1145/2348283.2348372](https://doi.org/10.1145/2348283.2348372). URL <http://doi.acm.org/10.1145/2348283.2348372>.
- [Chen et al., 2014] Lun-Chi Chen, Ping-Jen Kuo, and I-En Liao. Ontology-based library recommender system using mapreduce. *Cluster Computing*, pages 1–9, 2014. ISSN 1386-7857. doi: [10.1007/s10586-013-0342-z](https://doi.org/10.1007/s10586-013-0342-z). URL <http://dx.doi.org/10.1007/s10586-013-0342-z>.

- [Chen and Chen, 2012] Shu-jiun Chen and Hsueh-hua Chen. Mapping multilingual lexical semantics for knowledge organization systems. *Electronic Library, The*, 30 (2): 278–294, 2012.
- [Cheng et al., 2015] Dehua Cheng, Xinran He, and Yan Liu. Model selection for topic models via spectral decomposition. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, pages 183–191, 2015.
- [Cheung et al., 2011] Christy M. K. Cheung, Pui-Yee Chiu, and Matthew K. O. Lee. Online social networks: Why do students use facebook? *Comput. Hum. Behav.*, 27 (4): 1337–1343, July 2011. ISSN 0747-5632. doi: [10.1016/j.chb.2010.07.028](https://doi.org/10.1016/j.chb.2010.07.028). URL <http://dx.doi.org/10.1016/j.chb.2010.07.028>.
- [Cheung and Vogel, 2005] Karen S. Cheung and Douglas Vogel. Complexity reduction in lattice-based information retrieval. *Inf. Retr.*, 8 (2): 285–299, April 2005. ISSN 1386-4564. doi: [10.1007/s10791-005-5663-y](https://doi.org/10.1007/s10791-005-5663-y). URL <http://dx.doi.org/10.1007/s10791-005-5663-y>.
- [Cho et al., 2014] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [Christensen and Schiaffino, 2014] Ingrid Christensen and Silvia Schiaffino. A hybrid approach for group profiling in recommender systems. 20 (4): 507–533, apr 2014. [[http://www.jucs.org/jucs\\_20\\_4/a\\_hybrid\\_approach\\_for](http://www.jucs.org/jucs_20_4/a_hybrid_approach_for)].
- [Cigarrán, 2008] J. Cigarrán. *Agrupación de Resultados de Búsqueda Mediante Análisis Formal de Conceptos*. PhD thesis, UNED, 2008.
- [Cigarrán et al., 2016] Juan Cigarrán, Ángel Castellanos, and Ana García-Serrano. A step forward for Topic Detection in Twitter: An FCA-based approach. *Expert Systems with Applications*, 57: 21 – 36, 2016. ISSN 0957-4174. doi: <http://dx.doi.org/10.1016/j.eswa.2016.03.011>. URL <http://www.sciencedirect.com/science/article/pii/S0957417416301038>.
- [Cigarrán et al., 2004] Juan M. Cigarrán, Julio Gonzalo, Anselmo Pe nas, and Felisa Verdejo. Browsing Search Results via Formal Concept Analysis: Automatic Selection of Attributes. In *Concept Lattices, Second International Conference on Formal Concept Analysis, ICFCA*, pages 74–87. Springer-Verlag, 2004.
- [Cilibrasi and Vitanyi, 2007] Rudi L Cilibrasi and Paul MB Vitanyi. The google similarity distance. *Knowledge and Data Engineering, IEEE Transactions on*, 19 (3): 370–383, 2007.
- [Cimiano et al., 2004] Philipp Cimiano, Andreas Hotho, Gerd Stumme, and Julien Tane. Conceptual knowledge processing with formal concept analysis and ontologies. In *Concept Lattices*, pages 189–207. Springer, 2004.
- [Codocedo and Napoli, 2015] Victor Codocedo and Amedeo Napoli. Formal concept analysis and information retrieval—a survey. In *Formal Concept Analysis*, pages 61–77. Springer, 2015.

- [Codochedo et al., 2011] Vctor Codochedo, Carla Taramasco, and Hernán Astudillo. Cheating to achieve formal concept analysis over a large formal context. In Amedeo Napoli and Vilém Vychodil, editors, *CLA*, volume 959 of *CEUR Workshop Proceedings*, pages 349–362. CEUR-WS.org, 2011.
- [Colace et al., 2015] Francesco Colace, Massimo De Santo, Luca Greco, Vincenzo Moscato, and Antonio Picariello. A collaborative user-centered framework for recommending items in online social networks. *Computers in Human Behavior*, (0): –, 2015. ISSN 0747-5632. doi: <http://dx.doi.org/10.1016/j.chb.2014.12.011>. URL <http://www.sciencedirect.com/-science/article/pii/S0747563214007262>.
- [Cooper et al., 2010] Antony K. Cooper, Derrick G. Kourie, and Serena Coetzee. Thoughts on exploiting instability in lattices for assessing the discrimination adequacy of a taxonomy. In Marzena Kryszkiewicz and Sergei A. Obiedkov, editors, *CLA*, volume 672 of *CEUR Workshop Proceedings*, pages 338–343. CEUR-WS.org, 2010.
- [Cordobés et al., 2013] Héctor Cordobés, Antonio Fernández-Anta, Luis Felipe Núñez, Fernando Pérez, Teófilo Redondo, and Agustín Santos. Técnicas basadas en grafos para la categorización de tweets por tema. In *Proceedings of the TASS workshop at SEPLN 2013. Actas del XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.*, 2013.
- [Cramer et al., 2008] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. The effects of transparency on trust in and acceptance of a Content-based art recommender. *User Modeling and User-Adapted Interaction*, 18 (5): 455–496, November 2008. ISSN 0924-1868. doi: [10.1007/s11257-008-9051-3](http://dx.doi.org/10.1007/s11257-008-9051-3). URL <http://dx.doi.org/10.1007/s11257-008-9051-3>.
- [Cremonesi et al., 2010] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 39–46. ACM, 2010.
- [Cui et al., 2011] Peng Cui, Fei Wang, Shaowei Liu, Mingdong Ou, Shiqiang Yang, and Lifeng Sun. Who should share what?: Item-level social influence prediction for users and posts ranking. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11*, pages 185–194, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0757-4. doi: [10.1145/2009916.2009945](http://dx.doi.org/10.1145/2009916.2009945). URL <http://doi.acm.org/10.1145/2009916.2009945>.
- [Damljanovic et al., 2012] Danica Damljanovic, Milan Stankovic, and Philippe Laublet. Linked data-based concept recommendation: Comparison of different methods in open innovation scenario. In Elena Simperl, Philipp Cimiano, Axel Polleres, Oscar Corcho, and Valentina Presutti, editors, *The Semantic Web: Research and Applications*, volume 7295 of *Lecture Notes in Computer Science*, pages 24–38. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-30283-1. doi: [10.1007/978-3-642-30284-8\\_9](http://dx.doi.org/10.1007/978-3-642-30284-8_9). URL [http://dx.doi.org/10.1007/978-3-642-30284-8\\_9](http://dx.doi.org/10.1007/978-3-642-30284-8_9).

- [Dang-Nguyen et al., 2014] Duc-Tien Dang-Nguyen, Luca Piras, Giorgio Giacinto, Giulia Boato, and FGB De Natale. Retrieval of diverse images by pre-filtering and hierarchical clustering. volume 1263 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.
- [Dang-Nguyen et al., 2015] Duc-Tien Dang-Nguyen, Giulia Boato, Francesco GB De Natale, Luca Piras, Giorgio Giacinto, Franco Tuveri, and Manuela Angioni. Multimodal-based diversified summarization in social image retrieval. 2015.
- [Das et al., 2007] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: scalable online Collaborative Filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271–280. ACM, 2007.
- [Das et al., 2015] Rajarshi Das, Manzil Zaheer, and Chris Dyer. Gaussian lda for topic models with word embeddings. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, 2015.
- [Davies and Bouldin, 1979] David L Davies and Donald W Bouldin. A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2): 224–227, 1979.
- [De Francisci Morales et al., 2012] Gianmarco De Francisci Morales, Aristides Gionis, and Claudio Lucchese. From chatter to headlines: Harnessing the real-time web for personalized news recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM '12*, pages 153–162, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-0747-5. doi: [10.1145/2124295.2124315](https://doi.org/10.1145/2124295.2124315). URL <http://doi.acm.org/10.1145/2124295.2124315>.
- [De Luca, 2013] Ernesto William De Luca. Extending the linked data cloud with multilingual lexical linked data. *KNOWLEDGE ORGANIZATION. Journal of the International Society for Knowledge Organization*, Richard P. Smiraglia (Eds.), 2013.
- [De Luca and Nürnberger, 2005] Ernesto William De Luca and Andreas Nürnberger. Supporting information retrieval on mobile devices. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services, MobileHCI '05*, pages 347–348, New York, NY, USA, 2005. ACM. ISBN 1-59593-089-2. doi: [10.1145/1085777.1085859](https://doi.org/10.1145/1085777.1085859). URL <http://doi.acm.org/10.1145/1085777.1085859>.
- [De Luca and Nürnberger, 2006] Ernesto William De Luca and Andreas Nürnberger. Lexires: A tool for exploring and restructuring euwordnet for information retrieval. In: *Proceedings of the Workshop on Text-based Information Retrieval (TIR-06). In conjunction with the 17th Conference on Artificial Intelligence (ECAI'06)*. Riva del Garda, Italy, 2006.
- [Degemmis et al., 2007] Marco Degemmis, Pasquale Lops, and Giovanni Semeraro. A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation. *User Modeling and User-Adapted Interaction*, 17 (3): 217–255, 2007.
- [Delimitrou and Kozyrakis, 2013] Christina Delimitrou and Christos Kozyrakis. The netflix challenge: Datacenter edition. *IEEE Comput. Archit. Lett.*, 12 (1): 29–32, January 2013. ISSN 1556-6056. doi: [10.1109/L-CA.2012.10](https://doi.org/10.1109/L-CA.2012.10). URL <http://dx.doi.org/10.1109/L-CA.2012.10>.

- [Devi and Venkatesh, 2009] MK Kavitha Devi and P Venkatesh. An improved collaborative recommender system. In *Networks and Communications, 2009. NETCOM'09. First International Conference on*, pages 386–391. IEEE, 2009.
- [Devooght et al., 2015] Robin Devooght, Nicolas Kourtellis, and Amin Mantrach. Dynamic matrix factorization with priors on unknown values. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15*, pages 189–198, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3664-2. doi: [10.1145/2783258.2783346](https://doi.org/10.1145/2783258.2783346). URL <http://doi.acm.org/10.1145/2783258.2783346>.
- [Di Noia et al., 2012a] Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, and Davide Romito. Exploiting the web of data in model-based recommender systems. In *Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys '12*, pages 253–256, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1270-7. doi: [10.1145/2365952.2366007](https://doi.org/10.1145/2365952.2366007). URL <http://doi.acm.org/10.1145/2365952.2366007>.
- [Di Noia et al., 2012b] Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, and Markus Zanker. Linked open data to support Content-based recommender systems. In *Proceedings of the 8th International Conference on Semantic Systems*, pages 1–8. ACM, 2012.
- [Di Noia et al., 2014] Tommaso Di Noia, Iván Cantador, and Vito Claudio Ostuni. Linked open data-enabled recommender systems: Eswc 2014 challenge on book recommendation. In Valentina Presutti, Milan Stankovic, Erik Cambria, Iván Cantador, Angelo Di Iorio, Tommaso Di Noia, Christoph Lange, Diego Reforgiato Recupero, and Anna Tordai, editors, *Semantic Web Evaluation Challenge*, volume 475 of *Communications in Computer and Information Science*, pages 129–143. Springer International Publishing, 2014. ISBN 978-3-319-12023-2. doi: [10.1007/978-3-319-12024-9\\_17](https://doi.org/10.1007/978-3-319-12024-9_17). URL [http://dx.doi.org/10.1007/978-3-319-12024-9\\_17](http://dx.doi.org/10.1007/978-3-319-12024-9_17).
- [Diaz-Aviles et al., 2012] Ernesto Diaz-Aviles, Lucas Drumond, Lars Schmidt-Thieme, and Wolfgang Nejdl. Real-time top-n recommendation in social streams. In *Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys '12*, pages 59–66, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1270-7. doi: [10.1145/2365952.2365968](https://doi.org/10.1145/2365952.2365968). URL <http://doi.acm.org/10.1145/2365952.2365968>.
- [Díaz Esteban et al., 2013] Alberto Díaz Esteban, Iñaki Alegría, and Julio Villena Román, editors. *Proceedings of the TASS workshop at SEPLN 2013. Actas del XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.*, 2013.
- [dos Santos and Gatti, 2014] Ccero Nogueira dos Santos and Mara Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 69–78, Dublin, Ireland, August 23-29 2014.

- [du Boucher-Ryan and Bridge, 2006] Patrick du Boucher-Ryan and Derek Bridge. Collaborative recommending using formal concept analysis. *Knowledge-Based Systems*, 19 (5): 309–315, 2006.
- [Dunn, 1974] Joseph C Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4 (1): 95–104, 1974.
- [Eldén and Berry, 2008] Lars Eldén and Michael W Berry. Matrix methods in data mining and pattern recognition. *SIAM review*, 50 (1): 154, 2008.
- [Esslimani et al., 2011] Ilham Esslimani, Armelle Brun, and Anne Boyer. Densifying a behavioral recommender system by social networks link prediction methods. *Social Network Analysis and Mining*, 1 (3): 159–172, 2011.
- [Falk and Gardent, 2014] Ingrid Falk and Claire Gardent. Combining formal concept analysis and translation to assign frames and semantic role sets to french verbs. *Annals of Mathematics and Artificial Intelligence*, 70 (1-2): 123–150, 2014.
- [Fang and Zheng, 2009] Peici Fang and Siyao Zheng. A research on fuzzy formal concept analysis based Collaborative Filtering recommendation system. *Knowledge Acquisition and Modeling, International Symposium on*, 3: 352–355, 2009. doi: <http://doi.ieeecomputersociety.org/10.1109/KAM.2009.40>.
- [Fawcett, 2006] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27 (8): 861–874, 2006.
- [Fellbaum, 1998] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.
- [Feng et al., 2015] Shanshan Feng, Xutao Li, Yifeng Zeng, Gao Cong, Yeow Meng Chee, and Quan Yuan. Personalized ranking metric embedding for next new poi recommendation. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, pages 2069–2075. AAAI Press, 2015. ISBN 978-1-57735-738-4. URL <http://dl.acm.org/citation.cfm?id=2832415.2832536>.
- [Fernández et al., 2011] Miriam Fernández, Iván Cantador, Vanesa López, David Vallet, Pablo Castells, and Enrico Motta. Semantically enhanced information retrieval: An ontology-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9 (4): 434 – 452, 2011. ISSN 1570-8268. doi: <http://dx.doi.org/10.1016/j.websem.2010.11.003>. URL <http://www.sciencedirect.com/science/article/pii/S1570826810000910>. {JWS} special issue on Semantic Search.
- [Fiegerman, 2013] Seth Fiegerman. Netflix Knows You Better Than You Know Yourself. <http://mashable.com/2013/12/11/netflix-data/>, December 2013. Accessed: 2015-05-21.
- [Fiscus and Doddington, 2002] JonathanG. Fiscus and GeorgeR. Doddington. Topic Detection and Tracking Evaluation Overview. In James Allan, editor, *Topic Detection and Tracking*, volume 12 of *The Information Retrieval Series*, pages 17–31. Springer US, 2002. ISBN 978-1-4613-5311-9. doi: [10.1007/978-1-4615-0933-2\\_2](http://dx.doi.org/10.1007/978-1-4615-0933-2_2). URL [http://dx.doi.org/10.1007/978-1-4615-0933-2\\_2](http://dx.doi.org/10.1007/978-1-4615-0933-2_2).

- [Fouss et al., 2005] Francois Fouss, Alain Pirotte, and Marco Saerens. A novel way of computing similarities between nodes of a graph, with application to collaborative recommendation. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, pages 550–556. IEEE, 2005.
- [Gajdos et al., 2004] Petr Gajdos, Pavel Moravec, and Václav Snásel. Concept lattice generation by singular value decomposition. In Václav Snásel and Radim Belohlávek, editors, *CLA*, volume 110 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2004.
- [Gangemi et al., 2001] Aldo Gangemi, Nicola Guarino, and Alessandro Oltramari. Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-level. In *Proceedings of the International Conference on Formal Ontology in Information Systems - Volume 2001*, FOIS '01, pages 285–296, New York, NY, USA, 2001. ACM. ISBN 1-58113-377-4. doi: [10.1145/505168.505195](https://doi.org/10.1145/505168.505195). URL <http://doi.acm.org/10.1145/505168.505195>.
- [Ganter and Wille, 1997] Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematical foundations*. Springer-Verlag New York, Inc., 1997.
- [Ganter et al., 2016] Bernhard Ganter, Sergei Obiedkov, Sebastian Rudolph, and Gerd Stumme. *Conceptual Exploration*. Springer, 2016.
- [Gantner et al., 2011] Zeno Gantner, Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. MyMediaLite: A free recommender system library. In *5th ACM International Conference on Recommender Systems (RecSys 2011)*, 2011.
- [Garcia and Amatriain, 2010] Ruth Garcia and Xavier Amatriain. Weighted content based methods for recommending connections in online social networks. In *Workshop on Recommender Systems and the Social Web*, pages 68–71. Citeseer, 2010.
- [Garcin and Faltings, 2013] Florent Garcin and Boi Faltings. Pen recsys: A personalized news recommender systems framework. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 469–470, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2508218](https://doi.org/10.1145/2507157.2508218). URL <http://doi.acm.org/10.1145/2507157.2508218>.
- [Garcin et al., 2013] Florent Garcin, Christos Dimitrakakis, and Boi Faltings. Personalized news recommendation with context trees. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 105–112, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2507166](https://doi.org/10.1145/2507157.2507166). URL <http://doi.acm.org/10.1145/2507157.2507166>.
- [Gassler et al., 2012] W. Gassler, M. Zangerle, and G. Bürgler, M. and Specht. Snoopytagging: recommending contextualized tags to increase the quality and quantity of meta-information. In *Proceedings of the 21st international conference companion on World Wide Web, WWW '12 Companion*, pages 511–512, New York, NY, USA, 2012.
- [Gattiker, 2014] Anne Gattiker. Unstructured text: Test analysis techniques applied to non-test problems. In *VLSI Test Symposium (VTS), 2014 IEEE 32nd*, pages 1–4, April 2014. doi: [10.1109/VTS.2014.6818773](https://doi.org/10.1109/VTS.2014.6818773).

- [Ge et al., 2010] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, pages 257–260, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-906-0. doi: [10.1145/1864708.1864761](https://doi.org/10.1145/1864708.1864761). URL <http://dl.acm.org/10.1145/1864708.1864761>.
- [Geng et al., 2008] Liqiang Geng, Larry Korba, Yunli Wang, Xin Wang, and Yonghua You. Finding topics in email using formal concept analysis and fuzzy membership functions. In *Proceedings of the Canadian Society for Computational Studies of Intelligence, 21st Conference on Advances in Artificial Intelligence, Canadian AI'08*, pages 108–113, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 3-540-68821-8, 978-3-540-68821-1. URL <http://dl.acm.org/citation.cfm?id=1788714.1788725>.
- [Glorot et al., 2011a] Xavier Glorot, Yoshua Bengio, and Yann N Dauphin. Large-scale learning of embeddings with reconstruction sampling. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 945–952, 2011.
- [Glorot et al., 2011b] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 513–520, 2011.
- [Godin et al., 2013] Frédéric Godin, Viktor Slavkovikj, Wesley De Neve, Benjamin Schrauwen, and Rik Van de Walle. Using topic models for twitter hashtag recommendation. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13 Companion*, pages 593–596, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2038-2. doi: [10.1145/2487788.2488002](https://doi.org/10.1145/2487788.2488002). URL <http://doi.acm.org/10.1145/2487788.2488002>.
- [Golbeck, 2006] Jennifer Golbeck. Generating predictive movie recommendations from trust in social networks. In Ketil Stålen, WilliamH. Winsborough, Fabio Martinelli, and Fabio Massacci, editors, *Trust Management*, volume 3986 of *Lecture Notes in Computer Science*, pages 93–104. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-34295-3. doi: [10.1007/11755593\\_8](https://doi.org/10.1007/11755593_8). URL [http://dx.doi.org/10.1007/11755593\\_8](http://dx.doi.org/10.1007/11755593_8).
- [Golbeck and Hansen, 2011] Jennifer Golbeck and Derek L Hansen. A framework for recommending collections. In *Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011)*, page 35, 2011.
- [Goldberg et al., 1992] David Goldberg, David Nichols, Brian M Oki, and Douglas Terry. Using Collaborative Filtering to weave an information tapestry. *Communications of the ACM*, 35(12): 61–70, 1992.
- [Gori and Pucci, 2007] Marco Gori and Augusto Pucci. Itemrank: A random-walk based scoring algorithm for recommender engines. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI'07*, pages 2766–2771, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc. URL <http://dl.acm.org/citation.cfm?id=1625275.1625720>.

- [Gradmann, 2010] Stefan Gradmann. *Knowledge=Information in Context: on the Importance of Semantic Contextualisation in Europeana*. [The Hague]: Europeana Office, 2010.
- [Griffiths, 2002] Tom Griffiths. Gibbs sampling in the generative model of latent dirichlet allocation. 2002.
- [Gruber, 1995] Thomas R Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43 (5): 907–928, 1995.
- [Gulla et al., 2013] Jon Atle Gulla, Kevin C. Almeroth, Mozghan Tavakolifard, and Frank Hopfgartner, editors. *NRS '13: Proceedings of the 2013 International News Recommender Systems Workshop and Challenge*, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2302-4.
- [Guo et al., 2015] Guibing Guo, Jie Zhang, Zhu Sun, and Neil Yorke-Smith. Librec: A Java Library for Recommender Systems. In *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd International Conference on User Modeling, Adaptation and Personalization*, 2015.
- [Guo et al., 2013] Xin Guo, Yang Xiang, Qian Chen, Zhenhua Huang, and Yongtao Hao. LDA-based online topic detection using tensor factorization. *J. Inf. Sci.*, 39 (4): 459–469, August 2013. ISSN 0165-5515. doi: [10.1177/0165551512473066](https://doi.org/10.1177/0165551512473066). URL <http://dx.doi.org/10.1177/0165551512473066>.
- [Gupta et al., 2013] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. Wtf: The who to follow service at twitter. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13*, pages 505–514, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2035-1. URL <http://dl.acm.org/citation.cfm?id=2488388.2488433>.
- [Guy et al., 2009] Ido Guy, Inbal Ronen, and Eric Wilcox. Do you know?: Recommending people to invite into your social network. In *Proceedings of the 14th International Conference on Intelligent User Interfaces, IUI '09*, pages 77–86, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-168-2. doi: [10.1145/1502650.1502664](https://doi.org/10.1145/1502650.1502664). URL <http://doi.acm.org/10.1145/1502650.1502664>.
- [Hannon et al., 2010] John Hannon, Mike Bennett, and Barry Smyth. Recommending twitter users to follow using content and Collaborative Filtering approaches. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, pages 199–206, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-906-0. doi: [10.1145/1864708.1864746](https://doi.org/10.1145/1864708.1864746). URL <http://doi.acm.org/10.1145/1864708.1864746>.
- [Hannon et al., 2011] John Hannon, Kevin McCarthy, and Barry Smyth. Finding useful users on twitter: Twittomender the followee recommender. In *Proceedings of the 33rd European Conference on Advances in Information Retrieval, ECIR'11*, pages 784–787, Berlin,

- Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-20160-8. URL <http://dl.acm.org/-citation.cfm?id=1996889.1997003>.
- [Hassanzadeh and Consens, 2009] Oktie Hassanzadeh and Mariano Consens. M.: Linked movie data base. In *In: Workshop on Linked Data on the Web (LDOW 2009)*, 2009.
- [Heitmann and Hayes, 2010] Benjamin Heitmann and Conor Hayes. Using linked data to build open, collaborative recommender systems. In *AAAI Spring Symposium: Linked Data Meets Artificial Intelligence*, pages 76–81, 2010.
- [Hernando et al., 2014] A. Hernando, Ricardo Moya, F. Ortega, and J. Bobadilla. Hierarchical graph maps for visualization of collaborative recommender systems. *Journal of Information Science*, 40(1): 97–106, 2014.
- [Hernando et al., 2013] Antonio Hernando, Jesús Bobadilla, Fernando Ortega, and Abraham Gutiérrez. Trees for explaining recommendations made through Collaborative Filtering. *Information Sciences*, 239: 1–17, 2013.
- [Hinton et al., 2006] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18 (7): 1527–1554, 2006.
- [Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9 (8): 1735–1780, November 1997. ISSN 0899-7667. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735). URL <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- [Hofmann, 1999] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57. ACM, 1999.
- [Hofmann, 2004] Thomas Hofmann. Latent semantic models for Collaborative Filtering. *ACM Trans. Inf. Syst.*, 22 (1): 89–115, January 2004. ISSN 1046-8188. doi: [10.1145/963770.963774](https://doi.org/10.1145/963770.963774). URL <http://doi.acm.org/10.1145/963770.963774>.
- [Honarkhah and Caers, 2010] Mehrdad Honarkhah and Jef Caers. Stochastic Simulation of Patterns Using Distance-Based Pattern Modeling. *Mathematical Geosciences*, 42 (5): 487–517, 2010. ISSN 1874-8961. doi: [10.1007/s11004-010-9276-7](https://doi.org/10.1007/s11004-010-9276-7). URL <http://dx.doi.org/10.1007/s11004-010-9276-7>.
- [Hopfgartner et al., 2014] Frank Hopfgartner, Benjamin Kille, Andreas Lommatzsch, Till Plumbaum, Torben Brodt, and Tobias Heintz. Benchmarking news recommendations in a living lab. In *CLEF'14: Proceedings of the 5th International Conference of the CLEF Initiative*, LNCS, pages 250–267. Springer Verlag, 09 2014. doi: [10.1007/978-3-319-11382-1\\_21](https://doi.org/10.1007/978-3-319-11382-1_21).
- [Hotho et al., 2002] Andreas Hotho, Alexander Maedche, and Steffen Staab. Ontology-based text document clustering. *KI*, 16 (4): 48–54, 2002.
- [Hu and Pu, 2011] Rong Hu and Pearl Pu. Helping users perceive recommendation diversity. In *Workshop on novelty and diversity in recommender systems, DiveRS*. Chicago, 2011.
- [Hu et al., 2014] Zhengyin Hu, Shu Fang, and Tian Liang. Empirical study of constructing a knowledge organization system of patent documents using topic modeling. *Scientometrics*,

- 100 (3): 787–799, 2014. ISSN 0138-9130. doi: [10.1007/s11192-014-1328-1](https://doi.org/10.1007/s11192-014-1328-1). URL <http://dx.doi.org/10.1007/s11192-014-1328-1>.
- [Huang et al., 2012] Bo Huang, Yan Yang, Amjad Mahmood, and Hongjun Wang. Microblog Topic Detection Based on LDA Model and Single-Pass Clustering. In JingTao Yao, Yan Yang, Roman SÅ,owiÅ,ski, Salvatore Greco, Huaxiong Li, Sushmita Mitra, and Lech Polkowski, editors, *Rough Sets and Current Trends in Computing*, volume 7413 of *Lecture Notes in Computer Science*, pages 166–171. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-32114-6. doi: [10.1007/978-3-642-32115-3\\_19](https://doi.org/10.1007/978-3-642-32115-3_19). URL [http://dx.doi.org/10.1007/978-3-642-32115-3\\_19](http://dx.doi.org/10.1007/978-3-642-32115-3_19).
- [Huang et al., 2010] Jeff Huang, Katherine M. Thornton, and Efthimis N. Efthimiadis. Conversational tagging in twitter. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, HT '10, pages 173–178, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0041-4. doi: [10.1145/1810617.1810647](https://doi.org/10.1145/1810617.1810647). URL <http://doi.acm.org/10.1145/1810617.1810647>.
- [Huang and Bian, 2009] Yuxia Huang and Ling Bian. A bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the internet. *Expert Syst. Appl.*, 36 (1): 933–943, January 2009. ISSN 0957-4174. doi: [10.1016/j.eswa.2007.10.019](https://doi.org/10.1016/j.eswa.2007.10.019). URL <http://dx.doi.org/10.1016/j.eswa.2007.10.019>.
- [Huang and Bian, 2015] Yuxia Huang and Ling Bian. Using ontologies and formal concept analysis to integrate heterogeneous tourism information. *IEEE Transactions on Emerging Topics in Computing*, 2015. doi: [10.1109/TETC.2015.2417111](https://doi.org/10.1109/TETC.2015.2417111).
- [Hurley, 2013] Neil J. Hurley. Personalised Ranking with Diversity. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 379–382, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2507226](https://doi.org/10.1145/2507157.2507226). URL <http://doi.acm.org/10.1145/2507157.2507226>.
- [Ignatov et al., 2012] D.I. Ignatov, S.O. Kuznetsov, and J. Poelmans. Concept-based biclustering for internet advertisement. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 123–130, Dec 2012. doi: [10.1109/ICDMW.2012.100](https://doi.org/10.1109/ICDMW.2012.100).
- [Ignatov and Kuznetsov, 2009] Dmitry I. Ignatov and Sergei O. Kuznetsov. Concept-based recommendations for internet advertisement. *CoRR*, 2009.
- [Ignatov et al., 2013] DmitryI. Ignatov, AlexandraYu. Kaminskaya, AnastasyaA. Bezzubtseva, AndreyV. Konstantinov, and Jonas Poelmans. FCA-Based Models and a Prototype Data Analysis System for Crowdsourcing Platforms. In HeatherD. Pfeiffer, DmitryI. Ignatov, Jonas Poelmans, and Nagarjuna Gadiraju, editors, *Conceptual Structures for STEM Research and Education*, volume 7735 of *Lecture Notes in Computer Science*, pages 173–192. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-35785-5. doi: [10.1007/978-3-642-35786-2\\_13](https://doi.org/10.1007/978-3-642-35786-2_13). URL [http://dx.doi.org/10.1007/978-3-642-35786-2\\_13](http://dx.doi.org/10.1007/978-3-642-35786-2_13).
- [Ionescu et al., 2014a] Bodgan Ionescu, Adrian Popescu, Mihai Lupu, Alexandru Lucian Ginsca, and Henning Müller. Retrieving diverse social images at mediaeval 2014: Challenge, dataset

- and evaluation. In *Proceedings of MediaEval Benchmarking Initiative for Multimedia Evaluation.*, 2014.
- [Ionescu et al., 2015] Bogdan Ionescu, Alexandru Lucian Ginsca, B. Boteanu, Adrian Popescu, Mihai Lupu, , and Henning Müller. Retrieving Diverse Social Images at MediaEval 2015: Challenge, Dataset and Evaluation. In *Proceedings of MediaEval Benchmarking Initiative for Multimedia Evaluation.*, 2015.
- [Ionescu et al., 2014b] Bogdan Ionescu, Adrian Popescu, Anca-Livia Radu, and Henning Müller. Result diversification in social image retrieval: a benchmarking framework. *Multimedia Tools and Applications*, pages 1–31, 2014. ISSN 1380-7501. doi: [10.1007/s11042-014-2369-4](https://doi.org/10.1007/s11042-014-2369-4). URL <http://dx.doi.org/10.1007/s11042-014-2369-4>.
- [Ionescu et al., 2014c] Bogdan Ionescu, Anca-Livia Radu, Mará Menéndez, Henning Müller, Adrian Popescu, and Babak Loni. Div400: a social image retrieval result diversification dataset. In *Proceedings of the 5th ACM Multimedia Systems Conference*, pages 29–34. ACM, 2014.
- [Iyyer et al., 2014] Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III. A neural network for factoid question answering over paragraphs. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 633–644, 2014.
- [Jack and Duclayee, 2008] Kris Jack and F Duclayee. Improving explicit preference entry by visualising data similarities. In *Intelligent User Interfaces, International Workshop on Recommendation and Collaboration (ReColl). Spain*, 2008.
- [Jahrer and Töscher, 2012] Michael Jahrer and Andreas Töscher. Collaborative Filtering Ensemble for Ranking. In Gideon Dror, Yehuda Koren, and Markus Weimer, editors, *KDD Cup*, volume 18 of *JMLR Proceedings*, pages 153–167. JMLR.org, 2012. URL <http://dblp.uni-trier.de/db/journals/jmlr/jmlrp18.html#JahrerT12a>.
- [Jahrer et al., 2010] Michael Jahrer, Andreas Töscher, and Robert Legenstein. Combining predictions for accurate recommender systems. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pages 693–702, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0055-1. doi: [10.1145/1835804.1835893](https://doi.org/10.1145/1835804.1835893). URL <http://doi.acm.org/10.1145/1835804.1835893>.
- [Jain and Dhillon, 2013] Prateek Jain and Inderjit S. Dhillon. Provable inductive matrix completion. *CoRR*, abs/1306.0626, 2013. URL <http://arxiv.org/abs/1306.0626>.
- [Jain et al., 2010] Prateek Jain, Pascal Hitzler, Peter Z Yeh, Kunal Verma, and Amit P Sheth. Linked data is merely more data. *Linked Data Meets Artificial Intelligence*, pages 82–86, 2010.
- [Jannach et al., 2010a] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. An introduction to recommender systems. In *25th ACM Symposium on Applied Computing*, page 139, 2010.
- [Jannach et al., 2010b] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

- [Järvelin and Kekäläinen, 2002] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20 (4): 422–446, October 2002. ISSN 1046-8188. doi: [10.1145/582415.582418](https://doi.org/10.1145/582415.582418). URL <http://doi.acm.org/10.1145/582415.582418>.
- [Jay et al., 2008] Nicolas Jay, François Kohler, and Amedeo Napoli. Analysis of social communities with iceberg and stability-based concept lattices. In Raoul Medina and Sergei A. Obiedkov, editors, *ICFCA*, volume 4933 of *Lecture Notes in Computer Science*, pages 258–272. Springer, 2008. ISBN 978-3-540-78136-3.
- [Jeon et al., 2013] Seungwoo Jeon, Y. Khosiawan, and Bonghee Hong. Making a graph database from unstructured text. In *Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on*, pages 981–988, Dec 2013. doi: [10.1109/CSE.2013.144](https://doi.org/10.1109/CSE.2013.144).
- [Ji and Shen, 2015] Ke Ji and Hong Shen. Making recommendations from top-n user-item subgroups. *Neurocomputing*, (0): –, 2015. ISSN 0925-2312. doi: <http://dx.doi.org/10.1016/j.neucom.2015.03.013>. URL <http://www.sciencedirect.com/science/article/pii/S0925231215002830>.
- [Jin et al., 2004] Rong Jin, Joyce Y. Chai, and Luo Si. An automatic weighting scheme for Collaborative Filtering. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 337–344, New York, NY, USA, 2004. ACM. ISBN 1-58113-881-4. doi: [10.1145/1008992.1009051](https://doi.org/10.1145/1008992.1009051). URL <http://doi.acm.org/10.1145/1008992.1009051>.
- [Johnson and Zhang, 2014] Rie Johnson and Tong Zhang. Effective use of word order for text categorization with convolutional neural networks. *To appear NAACL-2015*, 2014.
- [Johnson and Zhang, 2015] Rie Johnson and Tong Zhang. Semi-supervised convolutional neural networks for text categorization via region embedding. 2015.
- [JÄ\_sang et al., 2007] Audun JÄ\_sang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43 (2): 618 – 644, 2007. ISSN 0167-9236. doi: <http://dx.doi.org/10.1016/j.dss.2005.05.019>. URL <http://www.sciencedirect.com/science/article/pii/S0167923605000849>. Emerging Issues in Collaborative Commerce.
- [Kabbur et al., 2013] Santosh Kabbur, Xia Ning, and George Karypis. FISM: Factored Item Similarity Models for top-N Recommender Systems. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 659–667, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2174-7. doi: [10.1145/2487575.2487589](https://doi.org/10.1145/2487575.2487589). URL <http://doi.acm.org/10.1145/2487575.2487589>.
- [Kapanipathi et al., 2011] Pavan Kapanipathi, Fabrizio Orlandi, Amit P. Sheth, and Alexandre Passant. Personalized filtering of the twitter stream. In *SPIM'11*, pages 6–13, 2011.
- [Kashnitsky and Ignatov, 2014] Yury Kashnitsky and Dmitry I Ignatov. Can fca-based recommender system suggest a proper classifier? In *Proceedings of the FCA4I Workshop*

- in conjunction with European Conference on Artificial Intelligence (ECAI '14)*, pages 17–26, 2014.
- [Kent, 2003] Robert E Kent. The IFF foundation for ontological knowledge organization. *Cataloging & classification quarterly*, 37 (1-2): 187–203, 2003.
- [Khrouf and Troncy, 2013] Houda Khrouf and Raphaël Troncy. Hybrid event recommendation using linked data and user diversity. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, pages 185–192, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2507171](https://doi.org/10.1145/2507157.2507171). URL <http://doi.acm.org/10.1145/2507157.2507171>.
- [Kille et al., 2014] Benjamin Kille, Torben Brodt, Tobias Heintz, Frank Hopfgartner, Andreas Lommatzsch, and Jonas Seiler. Overview of clef newsreel 2014: News recommendations evaluation labs. Working paper, September 2014. URL <http://eprints.gla.ac.uk/101285/>.
- [Kim and Kim, 2003] Choonho Kim and Juntae Kim. A recommendation algorithm using multi-level association rules. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*, pages 524–527, Oct 2003. doi: [10.1109/WI.2003.1241257](https://doi.org/10.1109/WI.2003.1241257).
- [Kim et al., 2011] Heung-Nam Kim, Inay Ha, Kee-Sung Lee, Geun-Sik Jo, and Abdulmotaleb El-Saddik. Collaborative user modeling for enhanced content filtering in recommender systems. *Decision Support Systems*, 51 (4): 772–781, 2011.
- [Kim and Compton, 2004] Mihye Kim and Paul Compton. Evolutionary document management and retrieval for specialized domains on the web. *International Journal of Human-Computer Studies*, 60 (2): 201–241, 2004.
- [Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1746–1751, 2014.
- [Kirchberg et al., 2012] Markus Kirchberg, Erwin Leonardi, Yu Shyang Tan, Sebastian Link, Ryan KL Ko, and Bu Sung Lee. Formal concept discovery in semantic web data. In *Formal Concept Analysis*, pages 164–179. Springer, 2012.
- [Kitisin and Neuman, 2006] Sukumal Kitisin and Clifford Neuman. Reputation-based trust-aware recommender system. In *Securecomm and Workshops, 2006*, pages 1–7. IEEE, 2006.
- [Kohavi et al., 2009] Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M. Henne. Controlled experiments on the web: Survey and practical guide. *Data Min. Knowl. Discov.*, 18 (1): 140–181, February 2009. ISSN 1384-5810. doi: [10.1007/s10618-008-0114-1](https://doi.org/10.1007/s10618-008-0114-1). URL <http://dx.doi.org/10.1007/s10618-008-0114-1>.
- [Kompan and Bieliková, 2010] Michal Kompan and Mária Bieliková. Content-based news recommendation. In Francesco Buccafurri and Giovanni Semeraro, editors, *E-Commerce and Web Technologies*, volume 61 of *Lecture Notes in Business Information Processing*, pages 61–72. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-15207-8. doi: [10.1007/978-3-642-15208-5\\_6](https://doi.org/10.1007/978-3-642-15208-5_6). URL [http://dx.doi.org/10.1007/978-3-642-15208-5\\_6](http://dx.doi.org/10.1007/978-3-642-15208-5_6).

- [Konstan et al., 2006] Joseph A Konstan, Sean M McNee, Cai-Nicolas Ziegler, Roberto Torres, Nishikant Kapoor, and John Riedl. Lessons on applying automated recommender systems to information-seeking tasks. In *AAAI*, volume 6, pages 1630–1633, 2006.
- [Konstan et al., 2011] Joseph A. Konstan, Ricardo Conejo, José-Luis Marzo, and Nuria Oliver, editors. *User Modeling, Adaption and Personalization - 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings*, volume 6787 of *Lecture Notes in Computer Science*, 2011. Springer. ISBN 978-3-642-22361-7. doi: [10.1007/978-3-642-22362-4](https://doi.org/10.1007/978-3-642-22362-4). URL <http://dx.doi.org/10.1007/978-3-642-22362-4>.
- [Koren, 2009a] Y. Koren. The bellkor solution to the netflix grand prize. *Netflix prize documentation*, 2009.
- [Koren, 2008] Yehuda Koren. Factorization meets the neighborhood: A multifaceted Collaborative Filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, pages 426–434, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4. doi: [10.1145/1401890.1401944](https://doi.org/10.1145/1401890.1401944). URL <http://doi.acm.org/10.1145/1401890.1401944>.
- [Koren, 2009b] Yehuda Koren. Collaborative Filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, pages 447–456, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-495-9. doi: [10.1145/1557019.1557072](https://doi.org/10.1145/1557019.1557072). URL <http://doi.acm.org/10.1145/1557019.1557072>.
- [Koren and Bell, 2011] Yehuda Koren and Robert Bell. Advances in Collaborative Filtering. In *Recommender Systems Handbook*, pages 145–186. Springer, 2011.
- [Krizhevsky et al., 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- [Kullback and Leibler, 1951] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22 (1): 79–86, 1951.
- [Kumaran and Allan, 2004] Giridhar Kumaran and James Allan. Text classification and named entities for new event detection. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '04*, pages 297–304, New York, NY, USA, 2004. ACM. ISBN 1-58113-881-4. doi: [10.1145/1008992.1009044](https://doi.org/10.1145/1008992.1009044). URL <http://doi.acm.org/10.1145/1008992.1009044>.
- [Kuznetsov et al., 2007] Sergei Kuznetsov, Sergei Obiedkov, and Camille Roth. Reducing the representation complexity of lattice-based taxonomies. In Uta Priss, Simon Polovina, and Richard Hill, editors, *Conceptual Structures: Knowledge Architectures for Smart Applications*, volume 4604 of *Lecture Notes in Computer Science*, pages 241–254. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-73680-6. doi: [10.1007/978-3-540-73681-3\\_18](https://doi.org/10.1007/978-3-540-73681-3_18). URL [http://dx.doi.org/10.1007/978-3-540-73681-3\\_18](http://dx.doi.org/10.1007/978-3-540-73681-3_18).

- [Kuznetsov, 2007] Sergei O. Kuznetsov. On stability of a formal concept. *Annals of Mathematics and Artificial Intelligence*, 49 (1-4): 101–115, 2007. ISSN 1012-2443. doi: [10.1007/s10472-007-9053-6](https://doi.org/10.1007/s10472-007-9053-6). URL <http://dx.doi.org/10.1007/s10472-007-9053-6>.
- [Kwak et al., 2010] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8. doi: [10.1145/1772690.1772751](https://doi.org/10.1145/1772690.1772751). URL <http://doi.acm.org/10.1145/1772690.1772751>.
- [Laboreiro et al., 2010] Gustavo Laboreiro, Lus Sarmiento, Jorge Teixeira, and Eugénio Oliveira. Tokenizing micro-blogging messages using a text classification approach. In *Proceedings of the fourth workshop on Analytics for noisy unstructured text data*, AND '10, pages 81–88, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0376-7. doi: [10.1145/1871840.1871853](https://doi.org/10.1145/1871840.1871853). URL <http://doi.acm.org/10.1145/1871840.1871853>.
- [Laniado and Mika, 2010] David Laniado and Peter Mika. Making sense of twitter. In *Proceedings of the 9th International Semantic Web Conference on The Semantic Web - Volume Part I*, ISWC'10, pages 470–485, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-17745-X, 978-3-642-17745-3. URL <http://dl.acm.org/citation.cfm?id=1940281.1940312>.
- [Larochelle and Lauly, 2012] Hugo Larochelle and Stanislas Lauly. A neural autoregressive topic model. In *Advances in Neural Information Processing Systems*, pages 2708–2716, 2012.
- [Larochelle and Murray, 2011] Hugo Larochelle and Iain Murray. The neural autoregressive distribution estimator. In *International Conference on Artificial Intelligence and Statistics*, pages 29–37, 2011.
- [Larson et al., 2014] Martha Larson, Bodgan Ionescu, Xavier Anguera, Pavel Eskevich, Maria adn Korshunov, Markus Schedl, Mohammad Soleymani, Georgios Petkos, Richard Sutcliffe, Jaeyoung Choi, and Gareth J.F. Jones. Mediaeval 2014 multimedia benchmark workshop. In *Working Notes Proceedings of the MediaEval 2014 Workshop*, 2014.
- [Lastra-Díaz and García-Serrano, 2015] Juan J. Lastra-Díaz and Ana García-Serrano. A new family of information content models with an experimental survey on wordnet. *Knowledge-Based Syst.*, 89 (C): 509–526, November 2015. ISSN 0950-7051. doi: [10.1016/j.knosys.2015.08.019](https://doi.org/10.1016/j.knosys.2015.08.019). URL <http://dx.doi.org/10.1016/j.knosys.2015.08.019>.
- [Lau et al., 2008] Raymond Y. K. Lau, Albert Y. K. Chung, Dawei Song, and Qiang Huang. Towards fuzzy domain ontology based concept map generation for e-learning. In *Proceedings of the 6th international conference on Advances in web based learning*, ICWL'07, pages 90–101, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 3-540-78138-2, 978-3-540-78138-7. URL <http://dl.acm.org/citation.cfm?id=2170285.2170297>.
- [LeCun et al., 1989] Yann LeCun, L. D. Jackel, B. Boser, J. S. Denker, H. P. Graf, I. Guyon, D. Henderson, R. E. Howard, and W. Hubbard. Handwritten digit recognition:

- applications of neural network chips and automatic learning. *IEEE Communications Magazine*, 27 (11): 41–46, Nov 1989. ISSN 0163-6804. doi: [10.1109/35.41400](https://doi.org/10.1109/35.41400).
- [Lerman, 2007] Kristina Lerman. Social information processing in news aggregation. *IEEE Internet Computing*, 11 (6): 16–28, November 2007. ISSN 1089-7801. doi: [10.1109/MIC.2007.136](https://doi.org/10.1109/MIC.2007.136). URL <http://dx.doi.org/10.1109/MIC.2007.136>.
- [Leskovec et al., 2009] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 497–506. ACM, 2009.
- [Levy and Goldberg, 2014] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In *Advances in Neural Information Processing Systems*, pages 2177–2185, 2014.
- [Levy et al., 2015] Omer Levy, Yoav Goldberg, and Ido Dagan. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3: 211–225, 2015.
- [Li and Murata, 2010] Xiaohui Li and T. Murata. A knowledge-based recommendation model utilizing formal concept analysis and association. In *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, volume 4, pages 221–226, 2010. doi: [10.1109/ICCAE.2010.5451721](https://doi.org/10.1109/ICCAE.2010.5451721).
- [Lin et al., 2013] Jovian Lin, Kazunari Sugiyama, Min-Yen Kan, and Tat-Seng Chua. Addressing cold-start in app recommendation: Latent user models constructed from twitter followers. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 283–292. ACM, 2013.
- [Linden et al., 2003] Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item Collaborative Filtering. *Internet Computing, IEEE*, 7 (1): 76–80, 2003.
- [Lipton, 2015] Zachary Chase Lipton. Does deep learning come from the devil? <http://www.kdnuggets.com/2015/10/deep-learning-vapnik-einstein-devil-yandex-conference.html>, 2015. Accessed: 2015-11-12.
- [Liu et al., 2010a] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI '10*, pages 31–40, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-515-4. doi: [10.1145/1719970.1719976](https://doi.org/10.1145/1719970.1719976). URL <http://doi.acm.org/10.1145/1719970.1719976>.
- [Liu et al., 2011] Xiaohua Liu, Shaodian Zhang, Furu Wei, and Ming Zhou. Recognizing named entities in tweets. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11*, pages 359–367, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. ISBN 978-1-932432-87-9. URL <http://dl.acm.org/citation.cfm?id=2002472.2002519>.
- [Liu et al., 2010b] Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, and Junjie Wu. Understanding of Internal Clustering Validation Measures. In *Data Mining (ICDM), 2010*

- IEEE 10th International Conference on, ICDM '10*, pages 911–916, Washington, DC, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4256-0. doi: [10.1109/ICDM.2010.35](https://doi.org/10.1109/ICDM.2010.35). URL <http://dx.doi.org/10.1109/ICDM.2010.35>.
- [Loftus and Scheff, 1971] Elizabeth F Loftus and Ronald W Scheff. Categorization norms for fifty representative instances. *Journal of Experimental Psychology*, 91 (2): 355, 1971.
- [Lops et al., 2011] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*, pages 73–105. Springer, 2011.
- [Lu et al., 2015] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. Recommender system application developments: A survey. *Decision Support Systems*, (0): –, 2015. ISSN 0167-9236. doi: <http://dx.doi.org/10.1016/j.dss.2015.03.008>. URL <http://www.sciencedirect.com/science/article/pii/S0167923615000627>.
- [Lü et al., 2012] Linyuan Lü, Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. Recommender systems. *Physics Reports*, 519 (1): 1 – 49, 2012. ISSN 0370-1573. doi: <http://dx.doi.org/10.1016/j.physrep.2012.02.006>. URL <http://www.sciencedirect.com/science/article/pii/S0370157312000828>. Recommender Systems.
- [Luo et al., 2014] Yang Luo, Boyi Xu, Hongming Cai, and Fenglin Bu. A hybrid user profile model for personalized recommender system with linked open data. In *Enterprise Systems Conference (ES), 2014*, pages 243–248. IEEE, 2014.
- [Luong et al., 2013] Minh-Thang Luong, Richard Socher, and Christopher D Manning. Better word representations with recursive neural networks for morphology. *CoNLL-2013*, 104, 2013.
- [Ma et al., 2007] Hao Ma, Irwin King, and Michael R. Lyu. Effective missing data prediction for Collaborative Filtering. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '07*, pages 39–46, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7. doi: [10.1145/1277741.1277751](https://doi.org/10.1145/1277741.1277751). URL <http://doi.acm.org/10.1145/1277741.1277751>.
- [Ma et al., 2008] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940. ACM, 2008.
- [Ma et al., 2009] Hao Ma, Irwin King, and Michael R Lyu. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 203–210. ACM, 2009.
- [Ma et al., 2011] Hao Ma, Tom Chao Zhou, Michael R Lyu, and Irwin King. Improving recommender systems by incorporating social contextual information. *ACM Transactions on Information Systems (TOIS)*, 29 (2): 9, 2011.
- [Maaloe et al., 2015] Lars Maaloe, Morten Arngren, and Ole Winther. Deep belief nets for topic modeling. *arXiv preprint arXiv:1501.04325*, 2015.
- [Maio et al., 2012] C. De Maio, G. Fenza, M. Gaeta, V. Loia, F. Orciuoli, and S. Senatore. Rss-based e-learning recommendations exploiting fuzzy {FCA} for knowledge modeling.

- Applied Soft Computing*, 12 (1): 113 – 124, 2012. ISSN 1568-4946. doi: <http://dx.doi.org/10.1016/j.asoc.2011.09.004>. URL <http://www.sciencedirect.com/-science/article/pii/S1568494611003826>.
- [Maio et al., 2016] Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, and Mimmo Parente. Time aware knowledge extraction for microblog summarization on twitter. *Information Fusion*, 28: 60 – 74, 2016. ISSN 1566-2535. doi: <http://dx.doi.org/10.1016/j.inffus.2015.06.004>. URL <http://www.sciencedirect.com/-science/article/pii/S156625351500055X>.
- [Manning et al., 2008] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge, 2008.
- [McAuley and Leskovec, 2013] Julian McAuley and Jure Leskovec. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, pages 165–172, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2507163](https://doi.org/10.1145/2507157.2507163). URL <http://doi.acm.org/10.1145/2507157.2507163>.
- [Mcauliffe and Blei, 2008] Jon D Mcauliffe and David M Blei. Supervised topic models. In *Advances in neural information processing systems*, pages 121–128, 2008.
- [McCallum, 2005] Andrew McCallum. Information Extraction: Distilling Structured Data from Unstructured Text. *Queue*, 3 (9): 48–57, November 2005. ISSN 1542-7730. doi: [10.1145/1105664.1105679](https://doi.org/10.1145/1105664.1105679). URL <http://doi.acm.org/10.1145/1105664.1105679>.
- [McNee et al., 2006] Sean M. McNee, John Riedl, and Joseph A. Konstan. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems, CHI EA '06*, pages 1097–1101, New York, NY, USA, 2006. ACM. ISBN 1-59593-298-4. doi: [10.1145/1125451.1125659](https://doi.org/10.1145/1125451.1125659). URL <http://doi.acm.org/10.1145/1125451.1125659>.
- [Meijer et al., 2014] Kevin Meijer, Flavius Frasinca, and Frederik Hogenboom. A semantic approach for extracting domain taxonomies from text. *Decision Support Systems*, 62: 78–93, 2014.
- [Meymandpour and Davis, 2015] Rouzbeh Meymandpour and Joseph G Davis. Enhancing recommender systems using linked open data-based semantic analysis of items. In *Proceedings of the 3rd Australasian Web Conference (AWC 2015)*, volume 27, page 30, 2015.
- [Michelson and Macskassy, 2010] Matthew Michelson and Sofus A. Macskassy. Discovering users' topics of interest on twitter: A first look. In *Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data, AND '10*, pages 73–80, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0376-7. doi: [10.1145/1871840.1871852](https://doi.org/10.1145/1871840.1871852). URL <http://doi.acm.org/10.1145/1871840.1871852>.

- [Mikolov et al., 2010] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernocky, and Sanjeev Khudanpur. Recurrent neural network based language model. In *INTERSPEECH*, volume 2, page 3, 2010.
- [Mikolov et al., 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [Mirowski et al., 2010] Piotr Mirowski, M Ranzato, and Yann LeCun. Dynamic auto-encoders for semantic indexing. In *Proceedings of the NIPS 2010 Workshop on Deep Learning*, pages 1–9, 2010.
- [Miyahara and Pazzani, 2000] Koji Miyahara and Michael J Pazzani. Collaborative Filtering with the simple bayesian classifier. In *PRICAI 2000 Topics in Artificial Intelligence*, pages 679–689. Springer, 2000.
- [Mobasher et al., 2004] Bamshad Mobasher, Xin Jin, and Yanzan Zhou. Semantically enhanced Collaborative Filtering on the web. In *Web Mining: From Web to Semantic Web*, pages 57–76. Springer, 2004.
- [Mobasher et al., 2007] Bamshad Mobasher, Robin Burke, Runa Bhaumik, and Chad Williams. Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness. *ACM Transactions on Internet Technology (TOIT)*, 7 (4): 23, 2007.
- [Moore et al., 2013] Joshua L Moore, Shuo Chen, Douglas Turnbull, and Thorsten Joachims. Taste over time: The temporal dynamics of user preferences. In *ISMIR*, pages 401–406, 2013.
- [Mossel and Roch, 2007] Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 128–134. ACM, 2007.
- [Musto et al., 2012] Cataldo Musto, Giovanni Semeraro, Pasquale Lops, Marco de Gemmis, and Fedelucio Narducci. Leveraging social media sources to generate personalized music playlists. In Christian Huemer and Pasquale Lops, editors, *E-Commerce and Web Technologies*, volume 123 of *Lecture Notes in Business Information Processing*, pages 112–123. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-32272-3. doi: [10.1007/978-3-642-32273-0\\_10](https://doi.org/10.1007/978-3-642-32273-0_10). URL [http://dx.doi.org/10.1007/978-3-642-32273-0\\_10](http://dx.doi.org/10.1007/978-3-642-32273-0_10).
- [Muthukrishnan, 2005] Shanmugavelayutham Muthukrishnan. *Data streams: Algorithms and applications*. Now Publishers Inc, 2005.
- [Naveed et al., 2011] Nasir Naveed, Thomas Gottron, Jérôme Kunegis, and Arifah Che Alhadi. Bad news travel fast: A Content-based analysis of interestingness on twitter. In *Proceedings of the 3rd International Web Science Conference*, page 8. ACM, 2011.
- [Nenova et al., 2013] Elena Nenova, Dmitry I Ignatov, and Andrey V Konstantinov. An fca-based boolean matrix factorisation for Collaborative Filtering. *FCAIR 2012-Formal Concept Analysis Meets Information Retrieval*, pages 57–82, 2013.
- [Ning and Karypis, 2012] Xia Ning and George Karypis. Sparse linear methods with side information for top-n recommendations. In *Proceedings of the Sixth ACM Conference on*

- Recommender Systems*, RecSys '12, pages 155–162, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1270-7. doi: [10.1145/2365952.2365983](https://doi.org/10.1145/2365952.2365983). URL <http://doi.acm.org/10.1145/2365952.2365983>.
- [Noulas et al., 2012] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. A random walk around the city: New venue recommendation in location-based social networks. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, SOCIALCOM-PASSAT '12, pages 144–153, Washington, DC, USA, 2012. IEEE Computer Society. ISBN 978-0-7695-4848-7. doi: [10.1109/SocialCom-PASSAT.2012.70](https://doi.org/10.1109/SocialCom-PASSAT.2012.70). URL <http://dx.doi.org/10.1109/SocialCom-PASSAT.2012.70>.
- [O'Banion et al., 2012] Shawn O'Banion, Larry Birnbaum, and Kristian Hammond. Social media-driven news personalization. In *Proceedings of the 4th ACM RecSys Workshop on Recommender Systems and the Social Web*, RSWeb '12, pages 45–52, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1638-5. doi: [10.1145/2365934.2365943](https://doi.org/10.1145/2365934.2365943). URL <http://doi.acm.org/10.1145/2365934.2365943>.
- [O'Donovan and Smyth, 2005] John O'Donovan and Barry Smyth. Trust in recommender systems. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, IUI '05, pages 167–174, New York, NY, USA, 2005. ACM. ISBN 1-58113-894-6. doi: [10.1145/1040830.1040870](https://doi.org/10.1145/1040830.1040870). URL <http://doi.acm.org/10.1145/1040830.1040870>.
- [Ostuni et al., 2013] Vito Claudio Ostuni, Tommaso Di Noia, Eugenio Di Sciascio, and Roberto Mirizzi. Top-n recommendations from implicit feedback leveraging linked open data. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 85–92, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2507172](https://doi.org/10.1145/2507157.2507172). URL <http://doi.acm.org/10.1145/2507157.2507172>.
- [Ozsoy, 2016] Makkbule Gulcin Ozsoy. From word embeddings to item recommendation. *CoRR*, abs/1601.01356, 2016. URL <http://arxiv.org/abs/1601.01356>.
- [Paisley et al., 2012] John Paisley, Chong Wang, David M Blei, and Michael I Jordan. Nested hierarchical dirichlet processes. *arXiv preprint arXiv:1210.6738*, 2012.
- [Park et al., 2012] Deuk Hee Park, Hyea Kyeong Kim, Il Young Choi, and Jae Kyeong Kim. A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39 (11): 10059 – 10072, 2012. ISSN 0957-4174. doi: [http://dx.doi.org/10.1016/j.eswa.2012.02.038](https://doi.org/10.1016/j.eswa.2012.02.038). URL <http://www.sciencedirect.com/science/article/pii/S0957417412002825>.
- [Park et al., 2007] Moon-Hee Park, Jin-Hyuk Hong, and Sung-Bae Cho. Location-based recommendation system using bayesian user preference model in mobile devices. In *Ubiquitous Intelligence and Computing*, pages 1130–1139. Springer, 2007.
- [Park and Chu, 2009] Seung-Taek Park and Wei Chu. Pairwise preference regression for cold-start recommendation. In *Proceedings of the Third ACM Conference on Recommender*

- Systems*, RecSys '09, pages 21–28, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-435-5. doi: [10.1145/1639714.1639720](https://doi.org/10.1145/1639714.1639720). URL <http://doi.acm.org/10.1145/1639714.1639720>.
- [Parra-Santander and Brusilovsky, 2010] D. Parra-Santander and P. Brusilovsky. Improving Collaborative Filtering in social tagging systems for the recommendation of scientific articles. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 136–142, Aug 2010. doi: [10.1109/WI-IAT.2010.261](https://doi.org/10.1109/WI-IAT.2010.261).
- [Passant, 2010a] Alexandre Passant. dbrec - music recommendations using dbpedia. In PeterF. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, Lei Zhang, JeffZ. Pan, Ian Horrocks, and Birte Glimm, editors, *The Semantic Web : ISWC 2010*, volume 6497 of *Lecture Notes in Computer Science*, pages 209–224. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-17748-4. doi: [10.1007/978-3-642-17749-1\\_14](https://doi.org/10.1007/978-3-642-17749-1_14). URL [http://dx.doi.org/10.1007/978-3-642-17749-1\\_14](http://dx.doi.org/10.1007/978-3-642-17749-1_14).
- [Passant, 2010b] Alexandre Passant. Dbrec: Music recommendations using dbpedia. In *Proceedings of the 9th International Semantic Web Conference on The Semantic Web - Volume Part II, ISWC'10*, pages 209–224, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-17748-4, 978-3-642-17748-4. URL <http://dl.acm.org/citation.cfm?id=1940334.1940349>.
- [Paterek, 2007] Arkadiusz Paterek. Improving regularized singular value decomposition for Collaborative Filtering. In *Proceedings of KDD cup and workshop*, volume 2007, pages 5–8, 2007.
- [Pazzani and Billsus, 1997] Michael Pazzani and Daniel Billsus. Learning and revising user profiles: The identification of interesting web sites. *Mach. Learn.*, 27 (3): 313–331, June 1997. ISSN 0885-6125. doi: [10.1023/A:1007369909943](https://doi.org/10.1023/A:1007369909943). URL <http://dx.doi.org/10.1023/A:1007369909943>.
- [Pazzani, 1999] Michael J Pazzani. A framework for collaborative, Content-based and demographic filtering. *Artificial Intelligence Review*, 13 (5-6): 393–408, 1999.
- [Pazzani and Billsus, 2007] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.
- [Peis et al., 2008] Eduardo Peis, JM Morales del Castillo, and Juan A Delgado-López. Semantic recommender systems. analysis of the state of the topic. *Hipertext. net*, 6: 1–5, 2008.
- [Pennington et al., 2014] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1532–1543, 2014. URL <http://www.aclweb.org/anthology/D14-1162>.
- [Pereira and Hruschka, 2015] Andre Luiz Vizine Pereira and Eduardo Raul Hruschka. Simultaneous co-clustering and learning to address the cold start problem in recommender systems. *Knowledge-Based Systems*, 2015.

- [Peska and Vojtas, 2015] Ladislav Peska and Peter Vojtas. Using implicit preference relations to improve content based recommending. In *E-Commerce and Web Technologies*, pages 3–16. Springer, 2015.
- [Petkos et al., 2014] Georgios Petkos, Symeon Papadopoulos, Luca Aiello, Ryan Skraba, and Yiannis Kompatsiaris. A soft frequent pattern mining approach for textual topic detection. In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14)*, WIMS '14, pages 25:1–25:10, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2538-7. doi: [10.1145/2611040.2611068](https://doi.org/10.1145/2611040.2611068). URL <http://doi.acm.org/10.1145/2611040.2611068>.
- [Pham et al., 2011] Manh Cuong Pham, Yiwei Cao, Ralf Klamma, and Matthias Jarke. A clustering approach for Collaborative Filtering recommendation using social network analysis. 17 (4): 583–604, feb 2011. [http://www.jucs.org/jucs\\_17\\_4/a\\_clustering\\_approach\\_for](http://www.jucs.org/jucs_17_4/a_clustering_approach_for)].
- [Phan et al., 2008] Xuan-Hieu Phan, Le-Minh Nguyen, and Susumu Horiguchi. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *Proceedings of the 17th international conference on World Wide Web, WWW '08*, pages 91–100, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2. doi: [10.1145/1367497.1367510](https://doi.org/10.1145/1367497.1367510). URL <http://doi.acm.org/10.1145/1367497.1367510>.
- [Phelan et al., 2011] Owen Phelan, Kevin McCarthy, Mike Bennett, and Barry Smyth. Terms of a feather: Content-based news recommendation and discovery using twitter. In *Proceedings of the 33rd European Conference on Advances in Information Retrieval, ECIR'11*, pages 448–459. Springer-Verlag, Berlin, Heidelberg, 2011. ISBN 978-3-642-20160-8. URL <http://dl.acm.org/citation.cfm?id=1996889.1996947>.
- [Phuvipadawat and Murata, 2010] Swit Phuvipadawat and Tsuyoshi Murata. Breaking news detection and tracking in twitter. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 03, WI-IAT '10*, pages 120–123, Washington, DC, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4191-4. doi: [10.1109/WI-IAT.2010.205](https://doi.org/10.1109/WI-IAT.2010.205). URL <http://dx.doi.org/10.1109/WI-IAT.2010.205>.
- [Pirasteh et al., 2015] Parivash Pirasteh, Dosam Hwang, and Jason J. Jung. Exploiting matrix factorization to asymmetric user similarities in recommendation systems. *Knowledge-Based Systems*, (0): –, 2015. ISSN 0950-7051. doi: <http://dx.doi.org/10.1016/j.knosys.2015.03.006>. URL <http://www.sciencedirect.com/science/article/pii/S0950705115000982>.
- [Pitsilis et al., 2011] Georgios Pitsilis, Xiangliang Zhang, and Wei Wang. Clustering recommenders in Collaborative Filtering using explicit trust information. In *Trust Management V*, pages 82–97. Springer, 2011.
- [Pla and Hurtado, 2013] Ferran Pla and Lluís F. Hurtado. Elirf-upv en tass-2013: Análisis de sentimientos en twitter. In *Proceedings of the TASS workshop at SEPLN 2013. Actas del XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural. IV Congreso Español de Informática. 17-20 September 2013, Madrid, Spain.*, 2013.

- [Poelmans et al., 2013] Jonas Poelmans, Dmitry I Ignatov, Sergei O Kuznetsov, and Guido Dedene. Formal concept analysis in knowledge processing: A survey on applications. *Expert systems with applications*, 40 (16): 6538–6560, 2013.
- [Porcel et al., 2012] C. Porcel, A. Tejada-Lorente, M.A. Martínez, and E. Herrera-Viedma. A hybrid recommender system for the selective dissemination of research resources in a technology transfer office. *Information Sciences*, 184 (1): 1 – 19, 2012. ISSN 0020-0255. doi: <http://dx.doi.org/10.1016/j.ins.2011.08.026>. URL <http://www.sciencedirect.com/science/article/pii/S0020025511004452>.
- [Priss, 1998] Uta Priss. The formalization of wordnet by methods of relational concept analysis. *WordNet: An electronic lexical database and some of its applications*, pages 179–196, 1998.
- [Priss, 2000] Uta Priss. Lattice-based information retrieval. *Knowledge Organization*, 27 (3): 132–142, 2000.
- [Qu and Liu, 2011] Zhonghua Qu and Yang Liu. Interactive group suggesting for twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 519–523, Portland, Oregon, USA, June 2011. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/P11-2091>.
- [Quercia et al., 2012] Daniele Quercia, Harry Askham, and Jon Crowcroft. Tweetlda: Supervised topic classification and link prediction in twitter. In *Proceedings of the 4th Annual ACM Web Science Conference, WebSci '12*, pages 247–250, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1228-8. doi: [10.1145/2380718.2380750](https://doi.org/10.1145/2380718.2380750). URL <http://doi.acm.org/10.1145/2380718.2380750>.
- [Rahman et al., 2013] M. Rahman, Q Zhu, and Y. E. Szu-Li. Trect: A hashtag recommendation system for twitter \*. In *Proceedings of the 2nd International Workshop on Recommender Systems meet Big Data & Semantic Technologies. In conjunction with 7th ACM Conference on Recommender Systems (RecSys' 13)*. Honk Kong, 2013.
- [Rahman and Chow, 2010] M.K.M. Rahman and Tommy W.S. Chow. Content-based hierarchical document organization using multi-layer hybrid network and tree-structured features. *Expert Systems with Applications*, 37 (4): 2874 – 2881, 2010. ISSN 0957-4174. doi: <http://dx.doi.org/10.1016/j.eswa.2009.09.002>. URL <http://www.sciencedirect.com/science/article/pii/S0957417409007830>.
- [Ramage et al., 2009] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D Manning. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 248–256. Association for Computational Linguistics, 2009.
- [Ramage et al., 2010] Daniel Ramage, Susan T Dumais, and Daniel J Liebling. Characterizing microblogs with topic models. *ICWSM*, 10: 1–1, 2010.

- [Rao, 2010] K Nageswara Rao. Application domain and functional classification of recommender systems—a survey. *DESIDOC Journal of Library & Information Technology*, 28 (3): 17–35, 2010.
- [Ren et al., 2011] Yuyan Ren, Yajun Du, Xiaoping Huang, and Yong Xu. Topic detection of news stories with formal concept analysis†. *Journal of Information and Computational Science*, 8 (9): 1675–1682, 2011.
- [Rendle et al., 2009] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 452–461, Arlington, Virginia, United States, 2009. AUAI Press. ISBN 978-0-9749039-5-8. URL <http://dl.acm.org/citation.cfm?id=1795114.1795167>.
- [Rendle et al., 2011] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 635–644, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0757-4. doi: [10.1145/2009916.2010002](https://doi.org/10.1145/2009916.2010002). URL <http://doi.acm.org/10.1145/2009916.2010002>.
- [Rendón et al., 2011] Eréndira Rendón, Itzel M. Abundez, Citlalih Gutierrez, Sergio Dáz Zagal, Alejandra Arizmendi, Elvia M. Quiroz, and H. Elsa Arzate. A comparison of internal and external cluster validation indexes. In *Proceedings of the 2011 American Conference on Applied Mathematics and the 5th WSEAS International Conference on Computer Engineering and Applications*, AMERICAN-MATH'11/CEA'11, pages 158–163, Stevens Point, Wisconsin, USA, 2011. World Scientific and Engineering Academy and Society (WSEAS). ISBN 978-960-474-270-7. URL <http://dl.acm.org/citation.cfm?id=1959666.1959695>.
- [Resnick and Varian, 1997] Paul Resnick and Hal R. Varian. Recommender systems. *Communications of the ACM*, 40 (3): 56–58, 1997.
- [Resnick et al., 1994] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for Collaborative Filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, CSCW '94, pages 175–186, New York, NY, USA, 1994. ACM. ISBN 0-89791-689-1. doi: [10.1145/192844.192905](https://doi.org/10.1145/192844.192905). URL <http://doi.acm.org/10.1145/192844.192905>.
- [Ricci and Shapira, 2011] Francesco Ricci and Bracha Shapira. *Recommender systems handbook*. Springer, 2011.
- [Ricci et al., 2011] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender Systems Handbook*, pages 1–35. Springer, 2011.
- [Ristoski et al., 2014] Petar Ristoski, Eneldo Loza Mencía, and Heiko Paulheim. A hybrid multi-strategy recommender system using linked open data. In Valentina Presutti, Milan Stankovic, Erik Cambria, Iván Cantador, Angelo Di Iorio, Tommaso Di Noia, Christoph

- Lange, Diego Reforgiato Recupero, and Anna Tordai, editors, *Semantic Web Evaluation Challenge*, volume 475 of *Communications in Computer and Information Science*, pages 150–156. Springer International Publishing, 2014. ISBN 978-3-319-12023-2. doi: [10.1007/978-3-319-12024-9\\_19](https://doi.org/10.1007/978-3-319-12024-9_19). URL [http://dx.doi.org/10.1007/978-3-319-12024-9\\_19](http://dx.doi.org/10.1007/978-3-319-12024-9_19).
- [Rohani et al., 2014] Vala Ali Rohani, Zarinah Mohd Kasirun, and Kuru Ratnavelu. An enhanced Content-based recommender system for academic social networks. In *Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on*, pages 424–431, Dec 2014. doi: [10.1109/BdCloud.2014.131](https://doi.org/10.1109/BdCloud.2014.131).
- [Rosenthal, 2016] Ethan Rosenthal. Explicit matrix factorization: Als, sgd, and all that jazz, 2016. URL <http://blog.ethanrosenthal.com/2016/01/09/explicit-matrix-factorization-sgd-als/>.
- [Roth et al., 2008] Camille Roth, Sergei Obiedkov, and Derrick Kourie. Towards concise representation for taxonomies of epistemic communities. In *Proceedings of the 4th international conference on Concept lattices and their applications, CLA'06*, pages 240–255, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 3-540-78920-0, 978-3-540-78920-8. URL <http://dl.acm.org/citation.cfm?id=1793623.1793644>.
- [Rousseeuw, 1987] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20: 53–65, 1987.
- [Rudinac et al., 2013] Stevan Rudinac, Alan Hanjalic, and Martha Larson. Generating visual summaries of geographic areas using community-contributed images. *IEEE Transactions on Multimedia*, 15 (4): 921–932, 2013.
- [Sabucedo et al., 2014] Luis M Alvarez Sabucedo, Roberto Soto Barreiros, Juan M Santos Gago, and Manuel J Fernández Iglesias. An adaptive and social-aware recommendation algorithm for administration services. *Journal of Universal Computer Science*, 20 (11): 1523–1542, 2014.
- [Said, A. Bellogín, A. and de Vries, A., 2013] Said, A. Bellogín, A. and de Vries, A. News Recommendation in the Wild: CWI’s Recommendation Algorithms in the NRS Challenge. In *Proceedings of the International News Recommender Systems Challenge (NRS 2013), at the 7th ACM Conference on Recommender Systems (RecSys 2013)*, 2013.
- [Salakhutdinov et al., 2007] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for Collaborative Filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798. ACM, 2007.
- [Salakhutdinov et al., 2013] Ruslan Salakhutdinov, Joshua B Tenenbaum, and Antonio Torralba. Learning with hierarchical-deep models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35 (8): 1958–1971, 2013.
- [Salton, 1971] Gerard Salton. The smart retrieval system—experiments in automatic document processing. 1971.
- [Salton and McGill, 1983] Gerard Salton and Michael J McGill. Introduction to modern information retrieval. 1983.

- [Sankaranarayanan et al., 2009] Jagan Sankaranarayanan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. Twitterstand: news in tweets. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 42–51, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-649-6. doi: [10.1145/1653771.1653781](https://doi.org/10.1145/1653771.1653781). URL <http://doi.acm.org/10.1145/1653771.1653781>.
- [Saranya and Sadhasivam, 2012] K. Saranya and G. Sudha Sadhasivam. Article: A personalized online news recommendation system. *International Journal of Computer Applications*, 57 (18): 6–14, November 2012. Published by Foundation of Computer Science, New York, USA.
- [Sarwar et al., 2000] Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedl. Application of dimensionality reduction in recommender system – a case study. In *In WebKDD Workshop at ACM SIGKDD Conference*, 2000.
- [Sayyadi et al., 2009] Hassan Sayyadi, Matthew Hurst, and Alexey Maykov. Event detection and tracking in social streams. In *ICWSM*, 2009.
- [Schafer et al., 2001] J. Ben Schafer, Joseph A. Konstan, and John Riedl. E-commerce recommendation applications. *Data Min. Knowl. Discov.*, 5 (1-2): 115–153, January 2001. ISSN 1384-5810. doi: [10.1023/A:1009804230409](https://doi.org/10.1023/A:1009804230409). URL <http://dx.doi.org/10.1023/A:1009804230409>.
- [Schandl and Blumauer, 2010] Thomas Schandl and Andreas Blumauer. Poolparty: Skos thesaurus management utilizing linked data. In *The Semantic Web: Research and Applications*, pages 421–425. Springer, 2010.
- [Schein et al., 2002] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '02, pages 253–260, New York, NY, USA, 2002. ACM. ISBN 1-58113-561-0. doi: [10.1145/564376.564421](https://doi.org/10.1145/564376.564421). URL <http://doi.acm.org/10.1145/564376.564421>.
- [Schröder et al., 2011] Gunnar Schröder, Maik Thiele, and Wolfgang Lehner. Setting goals and choosing metrics for recommender system evaluations. In *UCERSTI2 Workshop at the 5th ACM Conference on Recommender Systems, Chicago, USA*, volume 23, 2011.
- [Schwartz, 2004] Barry Schwartz. *The paradox of choice: Why more is less*. Ecco New York, 2004.
- [Sedhain et al., 2015] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. Autorec: Autoencoders meet Collaborative Filtering. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pages 111–112, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3473-0. doi: [10.1145/2740908.2742726](https://doi.org/10.1145/2740908.2742726). URL <http://doi.acm.org/10.1145/2740908.2742726>.
- [Semeraro et al., 2009] Giovanni Semeraro, Pierpaolo Basile, Marco de Gemmis, and Pasquale Lops. User profiles for personalizing digital libraries. In D.G.H. Lian J.C. Na Y.L. Theng, S. Foo, editor, *Handbook of Research on Digital Libraries: Design, Development and Impact*, pages 149–158. IGI Global, 2009.

- [Senatore and Pasi, 2013] Sabrina Senatore and Gabriella Pasi. Lattice navigation for Collaborative Filtering by means of (fuzzy) formal concept analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC '13*, pages 920–926, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1656-9. doi: [10.1145/2480486.2480538](https://doi.org/10.1145/2480486.2480538). URL <http://doi.acm.org/10.1145/2480486.2480538>.
- [Shani and Gunawardana, 2011] Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer, 2011.
- [Shaw et al., 2010] Gavin Shaw, Yue Xu, and Shlomo Geva. Using association rules to solve the cold-start problem in recommender systems. In *Advances in Knowledge Discovery and Data Mining*, pages 340–347. Springer, 2010.
- [Shen et al., 2014] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion*, pages 373–374, Republic and Canton of Geneva, Switzerland, 2014. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2745-9. doi: [10.1145/2567948.2577348](https://doi.org/10.1145/2567948.2577348). URL <http://dx.doi.org/10.1145/2567948.2577348>.
- [Shi, 2013] Lei Shi. Trading-off among accuracy, similarity, diversity, and long-tail: a graph-based recommendation approach. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 57–64. ACM, 2013.
- [Shi et al., 2010] Yue Shi, Martha Larson, and Alan Hanjalic. List-wise learning to rank with matrix factorization for Collaborative Filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, pages 269–272, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-906-0. doi: [10.1145/1864708.1864764](https://doi.org/10.1145/1864708.1864764). URL <http://doi.acm.org/10.1145/1864708.1864764>.
- [Shiri, 2014] Ali Shiri. Linked data meets big data: A knowledge organization systems perspective. *Advances in Classification Research Online*, 24 (1): 16–20, 2014.
- [Shoval et al., 2008] Peretz Shoval, Veronica Maidel, and Bracha Shapira. An ontology-Content-based filtering method. 2008.
- [Shyu et al., 2005] Mei-Ling Shyu, Choochart Haruechaiyasak, Shu-Ching Chen, and Na Zhao. Collaborative Filtering by mining association rules from user access sequences. In *Web Information Retrieval and Integration, 2005. WIRI'05. Proceedings. International Workshop on Challenges in*, pages 128–135. IEEE, 2005.
- [Simovici et al., 2012] DanA. Simovici, Paul Fomenky, and Werner Kunz. Polarities, axialities and marketability of items. In Alfredo Cuzzocrea and Umeshwar Dayal, editors, *Data Warehousing and Knowledge Discovery*, volume 7448 of *Lecture Notes in Computer Science*, pages 243–252. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-32583-0. doi: [10.1007/978-3-642-32584-7\\_20](https://doi.org/10.1007/978-3-642-32584-7_20). URL [http://dx.doi.org/10.1007/978-3-642-32584-7\\_20](http://dx.doi.org/10.1007/978-3-642-32584-7_20).
- [Singh and Gordon, 2008] Ajit P. Singh and Geoffrey J. Gordon. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD International*

- Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 650–658, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4. doi: [10.1145/1401890.1401969](https://doi.org/10.1145/1401890.1401969). URL <http://doi.acm.org/10.1145/1401890.1401969>.
- [Smyth and Cotter, 2000] B Smyth and P Cotter. A personalised {TV} listings service for the digital {TV} age. *Knowledge-Based Systems*, 13 (2â€“3): 53 – 59, 2000. ISSN 0950-7051. doi: [http://dx.doi.org/10.1016/S0950-7051\(00\)00046-0](http://dx.doi.org/10.1016/S0950-7051(00)00046-0). URL <http://www.sciencedirect.com/science/article/pii/S0950705100000460>.
- [Snásel et al., 2007] Václav Snásel, Martin Polovincak, and Hussam M. Dahwa Abdulla. Concept lattice reduction by singular value decomposition. In Sergei D. Kuznetsov, Andrey Fomichev, Boris Novikov, and Dmitry Shaporenkov, editors, *SYRCoDIS*, volume 256 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2007.
- [Sobhanam and Mariappan, 2013] Hridya Sobhanam and AK Mariappan. Addressing cold start problem in recommender systems using association rules and clustering technique. In *Computer Communication and Informatics (ICCCI), 2013 International Conference on*, pages 1–5. IEEE, 2013.
- [Socher and Fei-Fei, 2010] R. Socher and Li Fei-Fei. Connecting modalities: Semi-supervised segmentation and annotation of images using unaligned text corpora. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 966–973, June 2010. doi: [10.1109/CVPR.2010.5540112](https://doi.org/10.1109/CVPR.2010.5540112).
- [Socher et al., 2014] Richard Socher, Andrej Karpathy, Quoc V Le, Christopher D Manning, and Andrew Y Ng. Grounded compositional semantics for finding and describing images with sentences. *Transactions of the Association for Computational Linguistics*, 2: 207–218, 2014.
- [Spertus et al., 2005] Ellen Spertus, Mehran Sahami, and Orkut Buyukkokten. Evaluating similarity measures: a large-scale study in the orkut social network. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, KDD '05, pages 678–684, New York, NY, USA, 2005. ACM. ISBN 1-59593-135-X. doi: [10.1145/1081870.1081956](https://doi.org/10.1145/1081870.1081956). URL <http://doi.acm.org/10.1145/1081870.1081956>.
- [Spina et al., 2013] D. Spina, J. Carrillo de Albornoz, T. Mart'in, E. Amigó, J. Gonzalo, and F. Giner. UNED Online Reputation Monitoring Team at RepLab 2013. In *CLEF 2013 Eval. Labs and Workshop Online Working Notes*. CLEF, 2013.
- [Spina et al., 2014] Damiano Spina, Julio Gonzalo, and Enrique Amigó. Learning Similarity Functions for Topic Detection in Online Reputation Monitoring. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, pages 527–536, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2257-7. doi: [10.1145/2600428.2609621](https://doi.org/10.1145/2600428.2609621). URL <http://doi.acm.org/10.1145/2600428.2609621>.
- [Spyromitros-Xioufis et al., 2015] Eleftherios Spyromitros-Xioufis, Adrian Popescu, Symeon Papadopoulos, and Ioannis Kompatsiaris. Usemp: Finding diverse images at mediaeval 2015. 2015.

- [Sriram et al., 2010] Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. Short text classification in twitter to improve information filtering. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 841–842, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0153-4. doi: [10.1145/1835449.1835643](https://doi.org/10.1145/1835449.1835643). URL <http://doi.acm.org/10.1145/1835449.1835643>.
- [Srivastava and Salakhutdinov, 2012] Nitish Srivastava and Ruslan R Salakhutdinov. Multimodal learning with deep boltzmann machines. In *Advances in neural information processing systems*, pages 2222–2230, 2012.
- [Stankovic et al., 2010] Milan Stankovic, Matthew Rowe, and Philippe Laublet. Mapping tweets to conference talks: A goldmine for semantics. In *Workshop on Social Data on the Web, Shanghai, China*, volume 664, 2010.
- [Su and Khoshgoftaar, 2009] Xiaoyuan Su and Taghi M. Khoshgoftaar. A survey of Collaborative Filtering techniques. *Adv. in Artif. Intell.*, 2009: 4:2–4:2, January 2009. ISSN 1687-7470. doi: [10.1155/2009/421425](https://doi.org/10.1155/2009/421425). URL <http://dx.doi.org/10.1155/2009/421425>.
- [Sullivan et al., 2004] Derry O' Sullivan, Barry Smyth, and David Wilson. Preserving recommender accuracy and diversity in sparse datasets. *International Journal on Artificial Intelligence Tools*, 13 (01): 219–235, 2004.
- [Summers et al., 2008] Ed Summers, Antoine Isaac, Clay Redding, and Dan Krech. Lcsh, skos and linked data. *arXiv preprint arXiv:0805.2855*, 2008.
- [Sun et al., 2009] Aaron R Sun, Jiesi Cheng, and Daniel Dajun Zeng. A novel recommendation framework for micro-blogging based on information diffusion. In *19th Annual Workshop on Information Technologies & Systems (WITS'09)*, 2009.
- [Sutterer et al., 2008] M. Sutterer, O. Droegehorn, and K. David. Upos: User profile ontology with situation-dependent preferences support. In *Advances in Computer-Human Interaction, 2008 First International Conference on*, pages 230–235, Feb 2008. doi: [10.1109/ACHI.2008.23](https://doi.org/10.1109/ACHI.2008.23).
- [Takács and Tikk, 2012] Gábor Takács and Domonkos Tikk. Alternating Least Squares for Personalized Ranking. In *Proceedings of the Sixth ACM Conference on Recommender Systems*, RecSys '12, pages 83–90, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1270-7. doi: [10.1145/2365952.2365972](https://doi.org/10.1145/2365952.2365972). URL <http://doi.acm.org/10.1145/2365952.2365972>.
- [Takács et al., 2009] Gábor Takács, István Pilászy, Bottyán Németh, and Domonkos Tikk. Scalable Collaborative Filtering approaches for large recommender systems. *J. Mach. Learn. Res.*, 10: 623–656, June 2009. ISSN 1532-4435. URL <http://dl.acm.org/citation.cfm?id=1577069.1577091>.
- [Tan et al., 2002] Chade-Meng Tan, Yuan-Fang Wang, and Chan-Do Lee. The use of bigrams to enhance text categorization. *Information processing & management*, 38 (4): 529–546, 2002.

- [Tan et al., 2011] Shulong Tan, Jiajun Bu, Chun Chen, and Xiaofei He. Using rich social media information for music recommendation via hypergraph model. *Social media modeling and computing*, pages 213–237, 2011.
- [Tanase, 2015] Diana Tanase. *A model for information retrieval driven by conceptual spaces*. PhD thesis, University of Westminster, 2015.
- [Taneva et al., 2010] Bilyana Taneva, Mouna Kacimi, and Gerhard Weikum. Gathering and ranking photos of named entities with high precision, high recall, and diversity. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10*, pages 431–440, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-889-6. doi: [10.1145/1718487.1718541](https://doi.org/10.1145/1718487.1718541). URL <http://doi.acm.org/10.1145/1718487.1718541>.
- [Tao et al., 2012] Ke Tao, Fabian Abel, Qi Gao, and Geert-Jan Houben. Tums: Twitter-based user modeling service. In *Proceedings of the 8th International Conference on The Semantic Web, ESWC'11*, pages 269–283, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-25952-4. doi: [10.1007/978-3-642-25953-1\\_22](https://doi.org/10.1007/978-3-642-25953-1_22). URL [http://dx.doi.org/10.1007/978-3-642-25953-1\\_22](http://dx.doi.org/10.1007/978-3-642-25953-1_22).
- [Tavakolifard et al., 2013] Mozghan Tavakolifard, Jon Atle Gulla, Kevin C. Almeroth, Frank Hopfgartner, Benjamin Kille, Till Plumbaum, Andreas Lommatzsch, Torben Brodt, Arthur Bucko, and Tobias Heintz. Workshop and challenge on news recommender systems. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, pages 481–482, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2508004](https://doi.org/10.1145/2507157.2508004). URL <http://doi.acm.org/10.1145/2507157.2508004>.
- [Teh et al., 2006] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 101 (476), 2006.
- [Tejeda-Lorente et al., 2014] Álvaro Tejeda-Lorente, Carlos Porcel, Eduardo Peis, Rosa Sanz, and Enrique Herrera-Viedma. A quality based recommender system to disseminate information in a university digital library. *Information Sciences*, 261: 52–69, 2014.
- [Tian et al., 2016] F. Tian, B. Gao, D. He, and T.-Y. Liu. Sentence Level Recurrent Topic Model: Letting Topics Speak for Themselves. *ArXiv e-prints*, 2016.
- [Tian, 2006] Ye Tian. *A System for Automated Content Organization*. PhD thesis, Cleveland, OH, USA, 2006. AAI3216885.
- [Ticha et al., 2014] Sonia Ben Ticha, Azim Roussanally, Anne Boyer, and Khaled Bsaes. User semantic model for dependent attributes to enhance Collaborative Filtering. In *Proceedings of the International Conference on Web Information Systems and Technologies (WEBIST '14)*, pages 205–212, 2014.
- [Touzi et al., 2013] Amel Grissa Touzi, Hela Ben Massoud, and Alaya Ayadi. Automatic ontology generation for data mining using fca and clustering. *arXiv preprint arXiv:1311.1764*, 2013.

- [Ugander et al., 2012] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109 (16): 5962–5966, 2012.
- [Uysal and Croft, 2011] Ibrahim Uysal and W. Bruce Croft. User oriented tweet ranking: A filtering approach to microblogs. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11*, pages 2261–2264, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0717-8. doi: [10.1145/2063576.2063941](https://doi.org/10.1145/2063576.2063941). URL <http://doi.acm.org/10.1145/2063576.2063941>.
- [Vakali et al., 2012] Athena Vakali, Maria Giatsoglou, and Stefanos Antaris. Social networking trends and dynamics detection via a cloud-based framework design. In *Proceedings of the 21st international conference companion on World Wide Web, WWW '12 Companion*, pages 1213–1220, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1230-1. doi: [10.1145/2187980.2188263](https://doi.org/10.1145/2187980.2188263). URL <http://doi.acm.org/10.1145/2187980.2188263>.
- [Valverde-Albacete and Peláez-Moreno, 2007] Francisco J Valverde-Albacete and Carmen Peláez-Moreno. Galois connections between semimodules and applications in data mining. In *Proceedings of the 5th International Conference on Formal Concept Analysis, (ICFCA 2007) Clermont-Ferrand, France, February 12-16, 2007*, pages 181–196. Springer, 2007.
- [Valverde-Albacete and Peláez-Moreno, 2013] Francisco J Valverde-Albacete and Carmen Peláez-Moreno. Systems vs. methods: an analysis of the affordances of formal concept analysis for information retrieval. In *FCAIR 2012 Formal Concept Analysis Meets Information Retrieval Workshop co-located with the 35th European Conference on Information Retrieval (ECIR 2013) March 24, 2013, Moscow, Russia*, page 113, 2013.
- [Van Craenendonck and Blockeel, 2015] Toon Van Craenendonck and Hendrik Blockeel. Using internal validity measures to compare clustering algorithms. In *AutoML Workshop at ICML 2015*, pages 1–8, 2015.
- [Victor et al., 2011] Patricia Victor, Martine De Cock, and Chris Cornelis. *Trust and Recommendations*, pages 645–675. Springer US, Boston, MA, 2011. ISBN 978-0-387-85820-3. doi: [10.1007/978-0-387-85820-3\\_20](https://doi.org/10.1007/978-0-387-85820-3_20). URL [http://dx.doi.org/10.1007/978-0-387-85820-3\\_20](http://dx.doi.org/10.1007/978-0-387-85820-3_20).
- [Vossen, 2004] Piek Vossen. Eurowordnet: a multilingual database of autonomous and language-specific wordnets connected via an inter-lingual index. *International Journal of Lexicography*, 17 (2): 161–173, 2004.
- [Vozalis and Margaritis, 2004] Manolis Vozalis and Konstantinos G Margaritis. Enhancing Collaborative Filtering with demographic data: The case of item-based filtering. In *4th International Conference on Intelligent Systems Design and Applications*, pages 361–366, 2004.
- [Wallach, 2006] Hanna M. Wallach. Topic modeling: Beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*, pages 977–984, New York,

- NY, USA, 2006. ACM. ISBN 1-59593-383-2. doi: [10.1145/1143844.1143967](https://doi.org/10.1145/1143844.1143967). URL <http://doi.acm.org/10.1145/1143844.1143967>.
- [Wan et al., 2012] Li Wan, Leo Zhu, and Rob Fergus. A hybrid neural network-latent topic model. In *International Conference on Artificial Intelligence and Statistics*, pages 1287–1294, 2012.
- [Wang et al., 2013a] Beidou Wang, Can Wang, Jiajun Bu, Chun Chen, Wei Vivian Zhang, Deng Cai, and Xiaofei He. Whom to mention: Expand the diffusion of tweets by @ recommendation on micro-blogging systems. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13*, pages 1331–1340, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2035-1. URL <http://dl.acm.org/citation.cfm?id=2488388.2488505>.
- [Wang and Wang, 2014] Guangxin Wang and Yinglin Wang. Probabilistic attribute mapping for cold-start recommendation. In *Foundations and Applications of Intelligent Systems*, pages 421–431. Springer, 2014.
- [Wang and Yeung, 2016] H. Wang and D.-Y. Yeung. Towards Bayesian Deep Learning: A Survey. *ArXiv e-prints*, April 2016.
- [Wang et al., 2013b] Hao Wang, Manolis Terrovitis, and Nikos Mamoulis. Location recommendation in location-based social networks using user check-in data. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL'13*, pages 374–383, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2521-9. doi: [10.1145/2525314.2525357](https://doi.org/10.1145/2525314.2525357). URL <http://doi.acm.org/10.1145/2525314.2525357>.
- [Wang et al., 2015a] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15*, pages 1235–1244, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3664-2. doi: [10.1145/2783258.2783273](https://doi.org/10.1145/2783258.2783273). URL <http://doi.acm.org/10.1145/2783258.2783273>.
- [Wang et al., 2006] Jun Wang, Arjen P. de Vries, and Marcel J. T. Reinders. Unifying user-based and item-based Collaborative Filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '06*, pages 501–508, New York, NY, USA, 2006. ACM. ISBN 1-59593-369-7. doi: [10.1145/1148170.1148257](https://doi.org/10.1145/1148170.1148257). URL <http://doi.acm.org/10.1145/1148170.1148257>.
- [Wang et al., 2015b] Peng Wang, Jiaming Xu, Bo Xu, Cheng-Lin Liu, Heng Zhang, Fangyuan Wang, and Hongwei Hao. Semantic Clustering and Convolutional Neural Network for Short Text Categorization. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, volume 2, pages 352–357, 2015.

- [Wang and Manning, 2012] Sida Wang and Christopher D Manning. Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 90–94. Association for Computational Linguistics, 2012.
- [Wang et al., 2015c] Xibin Wang, Junhao Wen, Fengji Luo, Wei Zhou, and Haijun Ren. *Personalized Recommendation System Based on Support Vector Machine and Particle Swarm Optimization*, pages 489–495. Springer International Publishing, Cham, 2015. ISBN 978-3-319-25159-2. doi: [10.1007/978-3-319-25159-2\\_44](https://doi.org/10.1007/978-3-319-25159-2_44). URL [http://dx.doi.org/10.1007/978-3-319-25159-2\\_44](http://dx.doi.org/10.1007/978-3-319-25159-2_44).
- [Wayne, 2000] Charles L. Wayne. Multilingual Topic Detection and Tracking: Successful Research Enabled by Corpora and Evaluation. In *2nd International Conference on Language Resources & Evaluation (LREC 2000)*, 2000.
- [Weimer et al., 2008] Markus Weimer, Alexandros Karatzoglou, and Alex Smola. Improving maximum margin matrix factorization. *Machine Learning*, 72 (3): 263–276, 2008.
- [Weng et al., 2010] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twiterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10*, pages 261–270, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-889-6. doi: [10.1145/1718487.1718520](https://doi.org/10.1145/1718487.1718520). URL <http://doi.acm.org/10.1145/1718487.1718520>.
- [Wille, 1992] Rudolf Wille. Concept lattices and conceptual knowledge systems. *Computers & mathematics with applications*, 23 (6): 493–515, 1992.
- [Wille, 1995] Rudolf Wille. The basic theorem of triadic concept analysis. *Order*, 12 (2): 149–158, 1995. ISSN 0167-8094. doi: [10.1007/BF01108624](https://doi.org/10.1007/BF01108624). URL <http://dx.doi.org/10.1007/BF01108624>.
- [Wille, 2009] Rudolf Wille. *Restructuring Lattice Theory: An Approach Based On Hierarchies Of Concepts*, volume 5548 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-01814-5. doi: [10.1007/978-3-642-01815-2\\_23](https://doi.org/10.1007/978-3-642-01815-2_23). URL [http://dx.doi.org/10.1007/978-3-642-01815-2\\_23](http://dx.doi.org/10.1007/978-3-642-01815-2_23).
- [Wu et al., 2013] Xiang Wu, Qi Liu, Enhong Chen, Liang He, Jingsong Lv, Can Cao, and Guoping Hu. Personalized next-song recommendation in online karaokes. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, pages 137–140, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2409-0. doi: [10.1145/2507157.2507215](https://doi.org/10.1145/2507157.2507215). URL <http://doi.acm.org/10.1145/2507157.2507215>.
- [Wu et al., 2016a] Yao Wu, Christopher DuBois, Alice X. Zheng, and Martin Ester. Collaborative denoising auto-encoders for top-n recommender systems. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, WSDM '16*, pages 153–162, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3716-8. doi: [10.1145/2835776.2835837](https://doi.org/10.1145/2835776.2835837). URL <http://doi.acm.org/10.1145/2835776.2835837>.

- [Wu et al., 2016b] Yao Wu, Xudong Liu, Min Xie, Martin Ester, and Qing Yang. Cccf: Improving Collaborative Filtering via scalable user-item co-clustering. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, WSDM '16*, pages 73–82, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3716-8. doi: [10.1145/2835776.2835836](https://doi.org/10.1145/2835776.2835836). URL <http://doi.acm.org/10.1145/2835776.2835836>.
- [Xu et al., 2012] Bin Xu, Jiajun Bu, Chun Chen, and Deng Cai. An exploration of improving collaborative recommender systems via user-item subgroups. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12*, pages 21–30, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1229-5. doi: [10.1145/2187836.2187840](https://doi.org/10.1145/2187836.2187840). URL <http://doi.acm.org/10.1145/2187836.2187840>.
- [Yan et al., 2012] Rui Yan, Mirella Lapata, and Xiaoming Li. Tweet recommendation with graph co-ranking. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1, ACL '12*, pages 516–525, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=2390524.2390597>.
- [Yang, 2010] Sheng-Yuan Yang. Developing an ontology-supported information integration and recommendation system for scholars. *Expert Systems with Applications*, 37 (10): 7065–7079, 2010.
- [Yang et al., 2011] Shuang-Hong Yang, Bo Long, Alex Smola, Narayanan Sadagopan, Zhaohui Zheng, and Hongyuan Zha. Like like alike: Joint friendship and interest propagation in social networks. In *Proceedings of the 20th International Conference on World Wide Web, WWW '11*, pages 537–546, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0632-4. doi: [10.1145/1963405.1963481](https://doi.org/10.1145/1963405.1963481). URL <http://doi.acm.org/10.1145/1963405.1963481>.
- [Yang et al., 2014] Xiwang Yang, Yang Guo, Yong Liu, and Harald Steck. A survey of Collaborative Filtering based social recommender systems. *Comput. Commun.*, 41: 1–10, March 2014. ISSN 0140-3664. doi: [10.1016/j.comcom.2013.06.009](https://doi.org/10.1016/j.comcom.2013.06.009). URL <http://dx.doi.org/10.1016/j.comcom.2013.06.009>.
- [Yao and Van Durme, 2014] Xuchen Yao and Benjamin Van Durme. Information extraction over structured sata: Question answering with freebase. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 956–966, Baltimore, Maryland, USA, June 23-25 2014. Association for Computational Linguistics.
- [Yih et al., 2011] Wen-tau Yih, Kristina Toutanova, John C Platt, and Christopher Meek. Learning discriminative projections for text similarity measures. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, pages 247–256. Association for Computational Linguistics, 2011.
- [Yih et al., Baltimore, Maryland, USA, June 23-25 2014] Wen-tau Yih, Xiaodong He, and Christopher Meek. Semantic parsing for single-relation question answering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)*, page 643–648, Baltimore, Maryland, USA, June 23-25 2014.

- [Yu et al., 2014] Hsiang-Fu Yu, Cho-Jui Hsieh, Si Si, and Inderjit S Dhillon. Parallel matrix factorization for recommender systems. *Knowledge and Information Systems*, 41 (3): 793–819, 2014.
- [Yuan et al., 2013] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, pages 363–372, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2034-4. doi: [10.1145/2484028.2484030](https://doi.org/10.1145/2484028.2484030). URL <http://doi.acm.org/10.1145/2484028.2484030>.
- [Yuan et al., 2010] Weiwei Yuan, Donghai Guan, Young-Koo Lee, Sungyoung Lee, and Sung Jin Hur. Improved trust-aware recommender system using small-worldness of trust networks. *Knowledge-Based Systems*, 23 (3): 232 – 238, 2010. ISSN 0950-7051. doi: <http://dx.doi.org/10.1016/j.knosys.2009.12.004>. URL <http://www.sciencedirect.com/science/article/pii/S095070511000002X>.
- [Zaman et al., 2010] Tauhid R. Zaman, Ralf Herbrich, Jurgen Van Gael, and David Stern. Predicting information spreading in twitter. December 2010. URL <http://research.microsoft.com/apps/pubs/default.aspx?id=141866>.
- [Zanardi and Capra, 2008] V. Zanardi and L. Capra. Social ranking: uncovering relevant content using tag-based recommender systems. In *Proceedings of the 2008 ACM conference on Recommender systems (RecSys 08)*, pages 51–58, New York, NY, USA, 2008.
- [Zhang et al., 2016] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '16) San Francisco, California, USA.*, 2016.
- [Zhang et al., 2007] Li Zhang, Zheng Pei, and Honghua Chen. Extracting fuzzy linguistic summaries based on including degree theory and fca. In *Foundations of Fuzzy Logic and Soft Computing*, pages 273–283. Springer, 2007.
- [Zhang et al., 2006] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. Learning from incomplete ratings using non-negative matrix factorization. In *SDM*, volume 6, pages 548–552. SIAM, 2006.
- [Zhang et al., 2002] Yi Zhang, Jamie Callan, and Thomas Minka. Novelty and redundancy detection in adaptive filtering. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '02, pages 81–88, New York, NY, USA, 2002. ACM. ISBN 1-58113-561-0. doi: [10.1145/564376.564393](https://doi.org/10.1145/564376.564393). URL <http://doi.acm.org/10.1145/564376.564393>.
- [Zhang et al., 2013] Yongfeng Zhang, Min Zhang, Yiqun Liu, and Shaoping Ma. Improve Collaborative Filtering through bordered block diagonal form matrices. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, pages 313–322, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2034-4. doi: [10.1145/2484028.2484101](https://doi.org/10.1145/2484028.2484101). URL <http://doi.acm.org/10.1145/2484028.2484101>.

- [Zhang et al., 2014] Yuchen Zhang, Amr Ahmed, Vanja Josifovski, and Alexander Smola. Taxonomy discovery for personalized recommendation. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 243–252. ACM, 2014.
- [Zhao and Rosson, 2009] Dejin Zhao and Mary Beth Rosson. How and why people twitter: The role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work, GROUP '09*, pages 243–252, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-500-0. doi: [10.1145/1531674.1531710](https://doi.org/10.1145/1531674.1531710). URL <http://doi.acm.org/10.1145/1531674.1531710>.
- [Zheng et al., 2015] Hai-Tao Zheng, Yang-Hui Yan, and Ying-Min Zhou. Graph-based hybrid recommendation using random walk and topic modeling. In Reynold Cheng, Bin Cui, Zhenjie Zhang, Ruichu Cai, and Jia Xu, editors, *Web Technologies and Applications*, volume 9313 of *Lecture Notes in Computer Science*, pages 573–585. Springer International Publishing, 2015. ISBN 978-3-319-25254-4. doi: [10.1007/978-3-319-25255-1\\_47](https://doi.org/10.1007/978-3-319-25255-1_47). URL [http://dx.doi.org/10.1007/978-3-319-25255-1\\_47](http://dx.doi.org/10.1007/978-3-319-25255-1_47).
- [Zhou et al., 2005] Baoyao Zhou, SiuCheung Hui, and Kuiyu Chang. A formal concept analysis approach for web usage mining. In Zhongzhi Shi and Qing He, editors, *Intelligent Information Processing II*, volume 163 of *IFIP International Federation for Information Processing*, pages 437–441. Springer US, 2005. ISBN 978-0-387-23151-8. doi: [10.1007/0-387-23152-8\\_55](https://doi.org/10.1007/0-387-23152-8_55). URL [http://dx.doi.org/10.1007/0-387-23152-8\\_55](http://dx.doi.org/10.1007/0-387-23152-8_55).
- [Zhou et al., 2010] Tao Zhou, Zoltán Kuscik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107 (10): 4511–4515, 2010.
- [Zhou et al., 2012] Xujuan Zhou, Yue Xu, Yuefeng Li, Audun Josang, and Clive Cox. The state-of-the-art in personalized recommender systems for social networking. *Artificial Intelligence Review*, 37 (2): 119–132, 2012.
- [Zipf, 1949] George Kingsley Zipf. Human behavior and the principle of least effort. 1949.