Università degli Studi di Torino

Computer Science Department

Ph.D. Program in Computer Science

Cycle XXXII

Doctoral Thesis

Suggestion Models to Support Personalized Information Filtering

Noemi Mauro

Supervisor Prof. Liliana Ardissono

Reviewers Prof. Francesco Ricci
Faculty of Computer Science
Free University of Bozen-Bolzano

Prof. Julita Vassileva
Department of Computer Science
University of Saskatchewan

Ph.D. Program Coordinator Prof. Marco Grangetto

Noemi Mauro
Suggestion Models to Support Personalized Information Filtering
Advisor: Prof. Liliana Ardissono
Reviewers: Prof. Francesco Ricci and Prof. Julita Vassileva
Ph.D. Program Coordinator: Prof. Marco Grangetto

Università degli Studi di Torino
Computer Science Department
Corso Svizzera 185
10149, Torino, Italy

Abstract

During the last few decades, a lot of information has become available on the web. It is evident that with this large amount of available data, it is difficult for people to find what they are looking for, they can feel overloaded and it can become complex for them to solve their search task. To overcome these problems, search engines and recommender systems have become an important part of most of the online services available nowadays, since they are able to assist the users in the process of retrieving relevant information on the web and they help the users to discover new items that they were not aware of before. In this thesis, we analyze different aspects in the field of information retrieval, recommender systems, and human-computer interaction in order to improve the intelligence of a recommender system with the final goal of offering users a mixed-initiative model that helps them to explore the retrieved and suggested information in order to satisfy their information needs. From the algorithmic perspective, we describe different recommendation models that leverage several types of information at different granularity levels. Specifically, we analyze the impact of ratings, search queries, topics and categories and trust information on item recommendation by employing them in distinct recommendation models. However, we are not only interested in improving the performance of the recommender systems but we are also interested in investigating the use of an appropriate user interface that allows users to inspect and interact with the retrieved information. During the past few years, it emerged the need of offering to the user a mixed-initiative interactive model that mixes the intelligence coming from the recommender system with the possibility for the users to tune and interact with the retrieved information. Thus, from the human-computer interaction perspective, we investigate the use of a set of widgets to help the users to explore the retrieved information in a map-based web application. In the future, we plan to use the insights that we collected from the results of the works presented in this thesis to build a hybrid recommender system to improve the recommender system intelligence and to integrate it inside a web application in order to offer to the users a mixed-initiative interaction model.

Publications

The candidate contributed to the following publications by discussing the ideas and main concepts, conducting experiments and user tests and writing sections of the final paper.

Journal Papers:

- Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., and Voghera, A. (2017a). “Personalised Community Maps”. In: *Int. J. of Electronic Governance* 9.1-2, pp. 156–178.
- Ardissono, L. and Mauro, N. (2020). “A Compositional Model of Multi-faceted Trust for Personalized Item Recommendation”. In: *Expert Systems with Applications*, p. 112880.
- Ludewig, M., Mauro, N., Latifi, S., and Jannach, D. (Submitted). “Empirical Analysis of Session-based Recommendation Algorithms”. In: *User Modeling and User-Adapted Interaction*.
- Mauro, N., Ardissono, L., and Lucenteforte, M. (Submitted). “Faceted Search of Heterogeneous Geographic Information for Dynamic Map Projection”. In: *Information Processing & Management*.

Conference Papers:

- Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., and La Riccia, L. (2017c). “OnToMap: Semantic Community Maps for Knowledge Sharing”. In: *Proceedings of the 28th ACM Conference on Hypertext and Social Media*. HT ’17. Prague, Czech Republic: ACM, pp. 317–318.
- Mauro, N. and Ardissono, L. (2017a). “Concept-aware Geographic Information Retrieval”. In: *Proceedings of 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. Leipzig, Germany: ACM, pp. 34–41.
- Mauro, N. and Ardissono, L. (2018). “Session-based Suggestion of Topics for Geographic Exploratory Search”. In: *Proceedings of 23rd International Conference on Intelligent User Interfaces*. IUI ’18. Tokyo, Japan: ACM, pp. 341–352.

- Ardissono, L., Delsanto, M., Lucenteforte, M., Mauro, N., Savoca, A., and Scanu, D. (2018a). "Map-based Visualization of 2D/3D Spatial Data via Stylization and Tuning of Information Emphasis". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. AVI '18. Castiglione della Pescaia, Grosseto, Italy: ACM, 38:1–38:5.
- Ardissono, L., Delsanto, M., Lucenteforte, M., Mauro, N., Savoca, A., and Scanu, D. (2018b). "Transparency-based Information Filtering on 2D/3D Geographical Maps". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. AVI '18. Castiglione della Pescaia, Grosseto, Italy: ACM, 56:1–56:3.
- Mauro, N., Ardissono, L., Di Rocco, L., Bertolotto, M., and Guerrini, G. (2019a). "Impact of Semantic Granularity on Geographic Information Search Support". In: *Proceedings of 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. Santiago, Chile: IEEE, pp. 323–328.
- Mauro, N., Ardissono, L., and Hu, Z. F. (2019b). "Multi-faceted Trust-based Collaborative Filtering". In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '19. Larnaca, Cyprus: ACM, pp. 216–224.
- Mauro, N. and Ardissono, L. (2019). "Extending a Tag-based Collaborative Recommender with Co-occurring Information Interests". In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '19. Larnaca, Cyprus: ACM, pp. 181–190.
- Ludewig, M., Mauro, N., Latifi, S., and Jannach, D. (2019). "Performance Comparison of Neural and Non-Neural Approaches to Session-based Recommendation". In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys '19, pp. 462–466.

Workshop Papers:

- Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., and La Riccia, L. (2016). "Exploration of Cultural Heritage Information via Textual Search Queries". In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '16. Florence, Italy: ACM, pp. 992–1001.
- Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., and La Riccia, L. (2017d). "Semantic Interpretation of Search Queries for Personalization". In: *Adjunct Publication of the 25th Conference on User Modeling,*

Adaptation and Personalization. UMAP '17. Bratislava, Slovakia: ACM, pp. 101–102.

- Ardissono, L., Mauro, N., and Savoca, A. (2017e). "Supporting Knowledge Sharing and Learning via Semantic Geographical Maps". In: *Proceedings of the 2017 ACM Workshop on Intelligent Interfaces for Ubiquitous and Smart Learning*. SmartLearn '17. Limassol, Cyprus: ACM, pp. 3–6.
- Mauro, N. and Ardissono, L. (2017b). "Thematic Maps for Geographical Information Search". In: *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. UMAP '17. Bratislava, Slovakia: ACM, pp. 337–342.

The candidate is the sole author of the following publication:

- Mauro, N. (2017). "Intelligent and Personalized Community Maps". In: *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion*. IUI '17 Companion. Limassol, Cyprus: ACM, pp. 181–184.
- Mauro, N. (2018). "Suggestion Models in Geographic Exploratory Search". In: *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. IUI '18. Tokyo, Japan: ACM, pp. 669–670.
- Mauro, N. (2019). "Supporting the Exploration of Cultural Heritage Information via Search Behavior Analysis". In: *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. UMAP'19 Adjunct. Larnaca, Cyprus: ACM, pp. 371–376.

Contents

1 Introduction

In 1989, at the research center CERN in Switzerland, Tim Berners-Lee started the development of the World Wide Web. In 1991, he published the first website and this event has been central to the Information Age. Every day, billions of people interact on the web and the Internet has become an essential tool in our daily life. A lot of applications have been developed to support people to fulfill their tasks at work, for entertainment (e.g. music and video websites), for traveling (e.g. booking and maps websites), for shopping (e.g. e-commerce website) and for social relations (e.g. social networks and dating applications). Indeed, through the development of Search Engines, today people are able to instantly access a large amount of information. For instance, in 2019, Google received every second over 70,000 searches, people watched over 80,000 videos per second and every second on the Internet around 80,000 GB of data were moving. It is evident that with this large amount of available data, people can feel overloaded while looking for information to solve their search tasks. For instance, for each query, Google search engine retrieves millions of results. Thus, if all results were presented together to the users, people would probably feel confused and overloaded and they would never find what they are looking for. However, nowadays, Google ranks the results according to the users' preferences and their past behaviors, and only shows the first ten results on the first page. It is important to notice that it gives the possibility to the users to explore more results, if needed, and to interact with them. This example can be extended to multiple domains where people have to interact with a large amount of information.

Thus, in addition to search engines, recommender systems have become an important part of most of the online services available nowadays. Their goal is not only to help the users find what they are looking for, but also to suggest them new items in order to help people discover something that they don't already know. For instance, new clothes are recommended on Zalando, interesting videos on Youtube, events and new friends on Facebook, new songs on Spotify, etc. However, novelty is not enough, if it is not coupled with relevance to the user and, in order to reach this goal, search and recommendation engines are becoming increasingly more intelligent. When a user is looking for some product or information, these systems already know her/his tastes and they offer personalized suggestions and results to her/him. Since the final goal of a recommendation and a search engine is to satisfy the user's information needs, the two technologies are evolving toward each other and the boundary between them is becoming every day more blurred. On the search side, this is driven by the merging of question answering capabilities with search, led by systems like Google Now and Apple Siri that move search toward intelligent personal assistants. On the recommendation side, there has been a merging of techniques from not just keyword search but also faceted search, along with user-based and item-based collaborative filtering techniques and other more proactive recommenders (Chi, 2015). Even though these systems have become more personalized and social, search and recommendation engines have also become more interactive. In most of the online services, they offer the capability to enable users to tune the recommendation results instantly. It is, thus, important to offer to the user a mixed-initiative interaction model implemented through an appropriate user interface where (s)he can interact with the suggestions and results and tune them depending on her/his interests.

In this thesis we analyze different aspects in the field of information retrieval, recommender systems, and human-computer interaction in order to improve the intelligence of a recommender system with the final goal of offering users a mixed-initiative model that helps them to explore the retrieved and suggested information in order to satisfy their information needs. From the algorithmic perspective, we introduce different recommendation models that leverage several types of information at different granularity levels. Specifically, we analyze the impact of ratings, search queries, topics and categories and trust information on item recommendation by employing them in distinct recommendation models. However, we are not only interested in improving the performance of the recommender systems but we are also interested in investigating the use of an appropriate user interface that allows users to inspect and interact with the retrieved information. During the past few years, it emerged the need of offering to the user a mixed-initiative interactive model that mixes the intelligence coming from the recommender system with the possibility for the users to tune and interact with the retrieved information. Thus, from the human-computer interaction perspective, we investigate the use of a set of widgets to help the users explore the retrieved information in a map-based web application. In the future, we plan to use the insights that we collected to build a hybrid recommender system in order to improve the quality of suggestions. We also plan to build and integrate a recommender system in the OnToMap web application in order to offer to the stakeholders a mixed-initiative interaction model that helps them to explore the territory through the navigation of relevant suggestions. In this way we will also be able to carry out online evaluations and to test the model in a real scenario.

The remainder of this chapter provides a brief history and background about information retrieval models (Section 1.1) and recommender systems (Section 1.2) in order to make the reader aware of the context of the models that will be presented in this thesis.

1.1 Historical Development and Background on Information Retrieval

Taking inspiration from the paper (Sanderson and Croft, 2012), in the following we will briefly describe the history of the information retrieval models in order to offer to the reader some information useful as a background on the research and development of this field.

The history of information retrieval began before the Internet was born. However, with the growth of digitized unstructured information and, via high-speed networks, the only solution to find relevant items from large text databases was using search, thus IR systems became ubiquitous. An IR system locates information that is relevant to a user's query and it typically searches in collections of unstructured or semistructured data (e.g., web pages, documents, images, video, etc.). In the 1950s, two keys points developed in the first IR systems have been introduced:

- How to index documents: the first winning approach was the Uniterm System (Taube et al., 1952) that was using words to index the documents of an IR system.
- How to retrieve documents: the first developed approach was called Boolean retrieval. In this method, a query was structured as a logical combination of terms whose result was a set of those documents that exactly match the query. Subsequently, in the next decades, the term frequency weighting (tf) (Luhn, 1958) was introduced and developed. The idea behind this method was to assign a score of relevance to each document with respect to a query based on the terms belonging to the document.

In the next decades, many new approaches were introduced. Documents and queries started to be seen as vectors within an N-dimensional space (N being the number of unique terms in the collection that can be searched). This led to the definition of the similarity between a document and query vector as the cosine coefficient between the two vectors (Salton, 1968). Another significant innovation at that time was the introduction of relevance feedback (Rocchio, 1971). This was a process to support iterative search, where documents previously retrieved could be marked as relevant in an IR system. A subsequent user's query was automatically adjusted using information extracted from the relevant documents.

Another key development of IR systems was that the Luhn's term frequency (tf) weighting method (based on the occurrence of words within a document) was complemented with the introduction of the word occurrence across the documents belonging to a collection (Sparck Jones, 1988). This method, called inverse document frequency (idf), introduced the idea that the frequency of a word occurrence in a document collection is inversely proportional to its significance in retrieval: less common words tend to refer to more specific concepts, which are more important in retrieval. Later on, advances in the basic vector space model were also developed. The most well known is the latent semantic indexing (LSI), where the dimensionality of the vector space of a document collection is reduced though singular-value decomposition (Deerwester et al., 1990). Indeed to extend the set of documents that a query could match, these approaches started to consider not only the exact terms and keywords but also the semantics by addressing anaphora, ambiguity, and named entities. One technique that was found effective was stemming, the process of matching words to their lexical variants.

With the invention of the World Wide Web, new problems in IR have been studied. The use of links between web pages has been exploited and the PageRank algorithm has been introduced (Page et al., 1999). Adding link analysis and multiple text representations of documents to existing document ranking functions increased the complexity of the algorithm of an IR system. Thus, setting manually the parameters of the different features of the algorithm became a challenge. This led to the analysis of search logs of user interactions in order to exploit them to find the correct parameters for the IR system. The analysis of users' queries, click patterns, and queries reformulations enabled researchers to develop more effective query processing techniques based on understanding the user's intent.

The field of IR is continuing to evolve as the computing environment changes. One example of this type of change is the rapid growth of mobile devices and social media. One response from the IR community has been the development of social search (Dodds et al., 2003), which deals with search involving communities of users and informal information exchange. Since new types of information are exploited and the context of the users has become more important, IR systems started to employ personalization in ranking the results. This led the IR research field to merge with other fields such as conversation retrieval (Magnani et al., 2012), filtering and recommendation (Resnick et al., 1994), and collaborative search in order to provide effective new tools for managing personal and social information and to give personalized results and suggestions to the users.

1.2 Historical Development and Background on Recommender Systems

This section takes inspiration from the Recommender Systems Handbook introduction (Ricci et al., 2011). Recommender Systems (RSs) are software tools and techniques that provide suggestions of items that are expected to be useful to a user. Compared to the research in the information retrieval field, the study of recommender systems is relatively recent. Indeed, recommender systems emerged as an independent research area in the mid-1990s. For example, in 1994 Paul Resnick and John Riedl developed the first collaborative filtering architecture to suggest news articles to the users (Resnick et al., 1994). However, research in recommender systems showed a rapid growth with the Netflix Prize competition from 2006 to 2009 where Netflix awarded a million dollar prize to the team that first succeeded in improving substantially the performance of its recommender system (Koren et al., 2009).

Recommender systems have proved to be a valuable means to deal with the information overload problem. Thus, different techniques have been developed and several information sources have been employed. A recommender system addresses this problem by showing to the user new items that may be relevant to her/his current task. Upon a user's request, which can be articulated, depending on the recommendation approach, by the user's context and need, a recommender system generates recommendations using various types of knowledge and data about users, the available items, and the user's previous actions. The user can then browse the recommendations. (S)he may accept them or not and may provide, immediately or at a next stage, an implicit or explicit feedback. Recommender systems collect from users their preferences, which are either explicitly expressed, e.g., as ratings for products, or are inferred by interpreting user actions. All these user actions and feedback can be used for generating new recommendations in the next user interactions with the system.

Recommender systems can be based on different machine learning techniques. From the simple ones based on K-Nearest Neighbors (K-NN) to the ones based on Matrix

Factorization and, in recent years, also recommender systems research started to investigate the use of neural networks and deep learning in making personalized suggestions.

However, recommendation approaches can be classified into two main classes:

- **Content-based:** the system learns to recommend items that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items. For example, if a user has positively rated a Chinese restaurant, then the system can learn to recommend other restaurants related to Chinese cuisine.
- **Collaborative filtering:** this approach recommends to the active user the items that other users with similar tastes liked in the past (Schafer et al., 2007). The similarity in tastes of two users is computed based on the similarity in the rating history of the users. For example, if a person A and a person B like the restaurants X and the person A also likes the restaurant Y, there is more probability that the person B also likes the Y restaurant than any other random restaurants. Collaborative filtering is considered to be the most popular and widely implemented technique in RS.

Regardless of the adopted approach, a recommender system generates suggestions by relying on three main classes of data objects:

- **Items:** they are the objects that are recommended to the users. Recommender systems can leverage different properties and features of the items. For example in a music recommender system, the musical genre (such as jazz, indie, etc.), the composer, and the singer can be used to describe a song and to learn how the utility of an item depends on its features. Items can be represented using various information and representation approaches, e.g., in a simple way as a single id code, or in a richer form, as a set of attributes, but even as a concept (category) in an ontological representation of the domain.
- **Users:** in order to offer personalized suggestions to the users, recommender systems need to model and collect preferences and needs about the users and build their user profiles. However, if the suggestions are not personalized there is no need to build a user profile. This information can be structured in various ways and depends on the recommendation technique. For instance, in collaborative filtering, users are modeled as a simple list containing the ratings provided by the user for some items. Differently, in a content-based approach user profiles can be modeled with the features of the items that the user liked in the past. However, users can be also described by other information. For

instance, in a trust-based recommender system, user models are enriched with relations between users and in some case also with the trust level of these relations between users. A recommender system can use this information to recommend items to users that are preferred by similar or trusted users.

- **Transactions:** it is a recorded interaction between a user and the recommender system. Transactions are log-like data that store important information generated during the human-computer interaction and that are useful to generate recommendations. Ratings are the most popular form of transaction data that a recommender system can collect. Other information that can be collected are, for instance, the ones referring to the browsing behavior of the users (e.g. clicks, bookmarks, etc.).

In order to offer personalized suggestions to the users, recommender systems rely on the transactions that can be represented as users' feedback to understand which are the users' tastes. Two different techniques can be adopted for recording user's feedback (Lops et al., 2011):

- **Explicit feedback:** a user explicitly evaluate items. There are three main categories of explicit feedback: (1) like/dislike - items are classified as "relevant" or "not relevant" by adopting a simple binary rating scale, e.g. Facebook posts, (2) ratings - a discrete numeric scale is usually adopted to judge items, e.g. Tripadvisor and (3) text comments - comments about items are collected and presented to the users to help them in the decision-making process, e.g. Amazon.

- **Implicit feedback:** it does not require any active user involvement because the feedback is derived from monitoring and analyzing user's activities. Implicit feedback methods are based on assigning a relevance score to specific user actions on an item, such as clicking, saving, discarding, bookmarking, etc.

It is worth to notice that some models started to leverage anonymous feedback since there is a growing sensibility of users towards privacy protection.

1.3 Thesis Structure

The models presented in this thesis leverage different types of information (i.e. ratings, search queries, topics and categories, trust information) and group and combine them in different ways in order to make useful suggestions to the users. The content of this thesis can be summarized following the structure described below:

- Suggestion models that leverage topics or items categories taking into account information related to search sessions and ratings to make recommendations:
  - Suggestion model of geographical topics based on the analysis of search query logs (Mauro and Ardissono, 2018).
  - Analysis of the impact of semantic granularity on Geographic Information Search Support (Mauro et al., 2019a).
  - Category-based recommender system that leverages rating information about POIs and behavior patterns about topics exploration extracted from search query logs (Mauro and Ardissono, 2019).
- Visualization model that groups the items according to their attributes values in order to offer to the user a faceted-based exploration model:
  - User study about the analysis of a set of visualization widgets useful to tune the suggested information in a map-based web application that contains a faceted-based exploration model (Ardissono et al., 2018a; Mauro et al., Submitted).
- Recommendation model based on different groups of feedback useful to produce metrics of trust that are leveraged to make relevant suggestions:
  - Multi-faceted trust model integrated into a User-to-User Collaborative filtering based on K-Nearest Neighbours and Matrix Factorization models (Mauro et al., 2019b; Ardissono and Mauro, 2020).

2 Models Overview and Research Questions

This chapter shows an overview of the models presented in this thesis. Specifically, we analyze the problem of making recommendations to the users under different points of view and leveraging several types of information and feedback. The main research question underlying this thesis is:

RQ: *What is the impact of different types of information in the process of offering relevant suggestions to users?*

In order to answer to this research question, we are going to present in this thesis a set of models that leverage different types of information (i.e. ratings, search queries, topics and categories, trust information). These pieces of information are grouped and combined in different ways in order to make useful suggestions to the users. We are going to analyze the problem not only from the algorithmic perspective but also by presenting the results of a user study about the investigation of different graphical representations of the search context by means of alternative types of widgets to support an interactive data visualization of the recommendations. On one side we are interested to help the user to explore the information space and to discover new information (query expansion and suggestion), while on the other side our goal is to increase the capability of the user to explore the retrieved information by means of an efficient filtering and visualization model (user interface with widgets useful for faceted search).

2.1 Topics Information: Ratings and Search Sessions

This part of the thesis concerns suggestion models that leverage topics in making recommendation and are framed in the OnToMap project. The main research question about this part of the thesis is:

RQ1: Is it possible to leverage topics or categories extracted from search sessions to improve the recommendations offered to the users?

Thus, we are going to first give to the reader an overview of the project. Then, we are going to present a session-based suggestion model of geographical topics for exploratory search that is able to suggest to the user clusters of topics that could be relevant for her/him in the context of a geographical search task. This model will be integrated in the OnToMap web application. Furthermore, we are going to present an analysis of the impact of different domain conceptualizations by examining the performance of the session-based suggestion model using three geographical ontologies with different semantic granularities to model the information space. Then, we are going to introduce a model that suggests POIs leveraging rating information and behaviour patterns about topics exploration extracted from search query logs.

2.1.1 OnToMap Overview

OnToMap is a Web collaborative GIS that supports the management of interactive maps for information sharing and participatory decision-making; see (Ardissono et al., 2017c). OnToMap is an interesting testbed for the integration of mixed-initiative user interfaces and intelligent systems because it can expose the user to information overload. Specifically, the geographic map can be seen as a shared project between stakeholders that they enrich with geographical information and annotations to describe the territory. On one side, the map might become really cluttered since a large amount of information can be used inside a project. Thus, it is important to help the users to find the information that they are looking for by offering them the possibility to temporarily filter the retrieved information to have different perspectives of the territory. On the other side, we are also interested in helping the user to discover information that is novel to her/him. Thus, we developed a concept suggestion model to be integrated inside the OnToMap web application.

OnToMap is based on a semantic representation of domain knowledge based on an OWL ontology that defines the structure of the information space and enables data retrieval from heterogeneous sources by applying ontology mappings. The ontology currently makes it possible to query both a dataset of Public Open Data about Piedmont area in Italy, and the OpenStreetMap (OSM) server (OpenStreetMap Contributors, 2017). The ontology can also be used to inhibit the inclusion of attributes in the graphical widgets supporting faceted search; e.g., we exclude the geometry of items by design because users can more conveniently select the bounding box to be applied by interacting with the map through zoom and pan. Finally, the ontology provides graphical details for map visualization, such as the

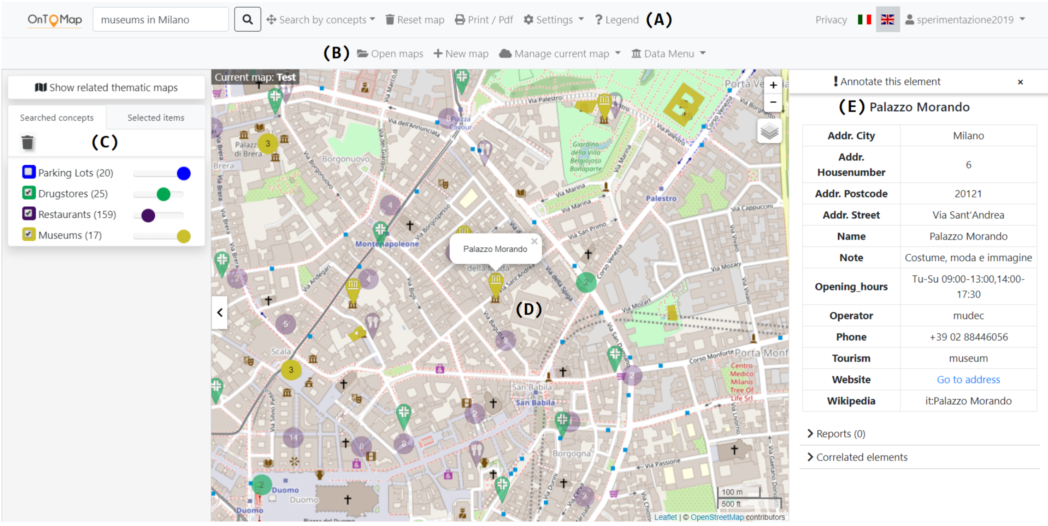


Fig. 2.1: OnToMap user interface showing the widgets based on transparency sliders. The top bar contains the control panel (A) supporting search (text input box and "Search by concepts" button), map management, user authentication and (B) other map management tools. The left sidebar (C) shows a graphical widget for each searched category. The right portion of the page (D) shows the geographic map and (E) the table of details of an item ("Palazzo Morando").

color and icon associated with each data category; e.g., drugstores are depicted in light green and they are represented as icons marked by a cross.

In summary, OnToMap offers the following functions:

- Creation of public and private custom geographic maps to help project design and group collaboration. The maps can be visualized in 2D and 3D modalities but in this thesis, we focus on 2D maps; see (Ardissono et al., 2018a; Ardissono et al., 2018b) for more details on this topic.
- Search support based on free-text queries and category browsing:
  - Textual queries are semantically interpreted using Natural Language Processing techniques with Word Sense Disambiguation (Moro et al., 2014). In this way, the user's information needs can be identified by abstracting from the specific terminology (s)he uses; see (Ardissono et al., 2016; Mauro and Ardissono, 2017a).
  - Data categories can be browsed by means of a simple alphabetical menu with auto-completion, or by navigating a graphical representation of the taxonomy defined in the domain ontology.

  The selection of data categories has a disjunctive semantics: each category selected by the user, either via textual query or by browsing, is separately used to populate the map. For instance, if the user submits query “restaurants in Torino" and (s)he also selects categories “Post Offices” and “Kindergartens” from the alphabetical menu, all these types of information are shown in the map.

- OnToMap also supports annotation and crowdsourcing of geographic elements within a map; see (Voghera et al., 2016) for details.

2.1.2 Session-based Suggestion of Topics for Exploratory Search

In the context of the OnToMap project, we developed a session-based suggestion model useful to suggest geographical concepts to the users during an exploratory search task. Exploratory information search can challenge users in the formulation of efficacious search queries. Moreover, complex information spaces can disorient people, making it difficult to find relevant data. In order to address these issues, we developed a session-based suggestion model that proposes concepts as a *“you might also be interested in”* function, by taking the user’s previous queries into account. Our model can be applied to incrementally generate suggestions in interactive search. It can be used for query expansion, and in general to guide users in the exploration of possibly complex spaces of data categories.

Our model is based on a concept co-occurrence graph that describes how frequently concepts are searched together in search sessions. Starting from an ontological domain representation, we generated the graph by analyzing the query log of a major search engine. Moreover, we identified clusters of ontology concepts that frequently co-occur in the sessions of the log via community detection on the graph.

With respect to the works available in the literature, we point out that our work is related to the contextual query suggestion model presented in (Cao et al., 2008), who suggests queries on the basis of the context provided by the user’s recent search history. However, we mine *ontology concepts* by interpreting search queries (we used the approach described in (Ardissono et al., 2016; Mauro and Ardissono, 2017a)), while the concepts defined in (Cao et al., 2008) are clusters of queries associated with similar sets of click results selected by users. Our model also relates to session-based term suggestion approaches such as the one presented in (Huang et al., 2003), and with term suggestion models used for web site advertisement, e.g., see (Chen et al., 2008). However, it differs from those works because we look for *concept co-occurrence*, which abstracts from the specific words used to refer to concepts, while they observe *term co-occurrence* for query expansion.

Specifically, we aim at answering the following research question:

$\mathbf{RQ1_1}$: *Can the data about the concepts frequently searched together by people within a search session be exploited to help the user explore the portions of an information space relevant to her/his information needs?*

Our model can be applied in different domains, but since our final goal is to integrate it in the OnToMap web application, we are interested in instantiating it in the context of Geographic Information Retrieval (Jones and Purves, 2008; Ballatore et al., 2016). In this context, several queries are performed to find the available items, per category, within a geographical area. Thus, guiding the user towards the exploration of information enables her/him to quickly generate custom maps reflecting individual information needs.

The details of this work will be presented in Chapter 4.

2.1.3 Impact of Semantic Granularity on Geographic Information Search Support

Information Retrieval research has used semantics to provide accurate search results, but the analysis of conceptual abstraction has mainly focused on information integration. Indeed, in Geographic Information Retrieval, ontologies have been used to improve geographic *features* extraction (Laurini, 2015). On the other hand, we employ them for geographic *concepts* extraction, in order to provide the user with topics to explore, rather than individual items. Specifically, we consider the model introduced in Section 2.1.2, and we investigate the impact of semantic granularity (i.e., the specificity of concepts representation) on the suggestion of relevant types of information to search for. We study how different levels of granularity in knowledge representation influence the capability of guiding the user in the exploration of a complex information space. We perform a comparative analysis of the performance of the query expansion model introduced in Section 2.1.2, using three spatial ontologies defined at different semantic granularity levels.

Notice that our aim is not only to measure the precision of the system's suggestions, but also to define a notion of "richness" based on the number of relevant suggested concepts, as this is important for catalog exploration. With this work, we are interested in giving to the reader a series of guidelines regarding the performance of a suggestion model considering different situations in which it is applied in domains whose representations have different levels of granularity. Specifically, we investigate the following research questions:

$\mathbf{RQ1_2}$: *What is the relationship between the semantic granularity of a domain conceptualization and the capability of suggesting types of information relevant to the user's needs during an exploratory search task?*

The details of this work will be presented in Chapter 5.

2.1.4 Extending a Tag-based Collaborative Recommender with Co-occurring Information Interests

Collaborative Filtering is largely applied to personalize item recommendation but its performance is affected by the sparsity of rating data. In order to address this issue, recent systems have been developed to improve recommendations by extracting latent factors from the rating matrices, or by exploiting different sources of information (e.g. tags, comments. etc.) according to (Brusilovsky and He, 2018). To the best of our knowledge, so far the research on recommender systems has not developed any models to combine rating data with information from search logs. In order to fill this gap, in this work, we aim at testing whether the integration of frequently co-occurring interests in information search logs can improve recommendation performance in User-to-User Collaborative Filtering (U2UCF). Our model does not perform a match between the user IDs of the two sources of information since they are identified in a different way. However, we propose the *Extended Category-based Collaborative Filtering (ECCF)* recommender, which enriches category-based user profiles derived from the analysis of rating behavior with data categories that are frequently searched together by people in search sessions. The categories co-occurrences are identified using part of the model introduced in 2.1.2. It can be noticed that interest co-occurrence can be learned by analyzing anonymous interaction sessions because it is aimed at describing general user behavior. Therefore, it can be applied to anonymized search logs, as long as search sessions can be identified.

Starting from a category-based representation of user preferences, based on the analysis of ratings and on items categorization, we propose the following research question:

$\mathbf{RQ1_3}$: *How does the integration of data about interest co-occurrence in information search influence the performance of a collaborative recommender system that manages category-based user profiles?*

The details of this work will be presented in Chapter 6.

2.2 Faceted Search of Heterogeneous Geographic Information

In this part of the thesis, we investigate the interaction aspects related to user interface management. When the system has retrieved the information that the user was looking for, it should be interesting to offer to the user the possibility to interact with the retrieved information in order to offer her/him the option of satisfying temporary information needs. In general, we are interested in developing an efficient user interface for faceted search in the Geographic context that offers the possibility to the user to better explore the information space. More specifically, the use of geographic maps for information sharing can challenge users with the presentation of large amounts of heterogeneous data in the presence of diverse and temporary information needs. For instance, within a project, as well as in small-scale group collaboration such as the organization of a tourist trip, people can focus on different aspects of the shared data during the execution of activities. Therefore, highly flexible visualization techniques are needed to reduce visual complexity by supporting dynamic map projection on a relatively stable information space that represents the overall set of items to be managed.

Similar to the works in the literature, our model exposes metadata derived from semantic knowledge representation. However, it enables users to work on maps populated with multiple data categories, i.e., with heterogeneous information, as well as to focus the maps on temporary interests without losing the overall set of data they contain because the map is interpreted as a long-term content sharing tool. This is useful to answer information needs in long-lasting user activities. Notice also that the OnToMap system, in which this work in integrated, does not assume to work on RDF data in order to comply with more general data sources, like public crowdmapping platforms, thanks to the mediation of its domain ontology.

In our work, we developed a faceted information exploration model that supports coarse-grained and fine-grained focusing of geographic maps by offering a graphical representation of data attributes within interactive widgets. The proposed approach enables (i) a multi-category projection of long-lasting geographic maps, based on the proposal of efficient facets for data exploration in sparse and noisy datasets, and (ii) an interactive representation of the search context based on widgets that support data visualization, faceted exploration, category-based information hiding and transparency of results at the same time. On this topic, we carried out an online experiment to understand how to visualize the retrieved information and give to the user the possibility of making dynamic projections on the information space using different widgets (i.e. checkboxes, treemaps and sunburst). The model is applied within the OnToMap web application and supports the exploration of information retrieved from heterogeneous data sources, such as OpenStreetMap.

Thus, we pose the following research questions:

$\mathbf{RQ2_1}$: *How does a compact, graphical view of the visualization constraints applied to a map, impact on her/his efficiency and experience in data exploration?*

$\mathbf{RQ2_2}$: *How much does the user's familiarity with the widgets for faceted exploration impact on her/his efficiency in search and on her/his appreciation of the exploration model they offer?*

The details of this work will be presented in Chapter 7.

2.3 Multi-faceted Trust for Personalized Item Recommendation

In this part of the thesis, we leverage another type of information with respect to the works presented above. We are interested in grouping and combining different types of feedback to produce metrics of trust that are leveraged to make relevant suggestions to users. Specifically, with the growth of social networks, social data have become an important information to be used by recommender systems. This trend led to the introduction of a new category of recommenders called trust-based recommender systems. Specifically, trust-based recommender systems improve rating prediction with respect to Collaborative Filtering by leveraging the additional information provided by a trust network among users to deal with the cold start problem. However, they focus on explicit relations between users and they leave out other types of information that can contribute to determine users' global reputation; e.g., public recognition of reviewers' quality. As the exploitation of information about social relations is challenged by recent studies according to which people generally perceive the usage of data about social relations as a violation of their own privacy (Burbach et al., 2018), we propose a novel model, which can work with or without this type of information, and only relies on public and anonymous information provided by users for the analysis of trust. In order to address the issues described above, we take inspiration from the LOCABAL trust-based recommender system (Tang et al., 2013) and we extend it with additional evidence about trust, based on public anonymous information, and we make them configurable with respect to the data that can be used in the given application domain:

1. We propose the Multi-faceted Trust Model (MTM) to define trust among users in a compositional way, possibly including or excluding the types of

information it contains. MTM flexibly integrates social links with public anonymous feedback received by user profiles and user contributions in social networks.

2. We propose LOCABAL+, based on MTM, which extends the LOCABAL trust-based recommender system with multi-faceted trust and trust-based social regularization.

In the Multi-faceted Trust Model we integrate diverse facets of trust and, in a specific application domain, one or more of them might not be available or usable. Therefore, besides assessing their overall value in improving Top-N recommendation, we separately study their impact on recommendation performance. We thus formulate the following research question:

$\mathbf{RQ3_1}$: *Can multi-faceted trust be used to improve the performance of a trust-based recommender system with respect to the standard state-of-the-art trust models that only rely on social links and rating similarity among users?*

In order to answer these questions, we carry out experiments to measure the performance of LOCABAL+ on a spectrum of MTM configurations that tune in different ways the influence of the facets of trust we consider.

The details of this work will be presented in Chapter 8.

2.4 Empirical Analysis of Session-based Recommendation Algorithms

In collaboration with University of Klagenfurt and TU Dortmund, a project concerning the evaluation of session-based recommender systems has been carried out. Since the details of the work are out of scope of this thesis we are not going to show the results of the analysis that we carried out. However, the knowledge acquired with this collaboration can be used to benchmark and compare the models described in the following chapters of this thesis.

This work is interesting because it sheds light on the state-of-the-art in the area of session-based recommendation and on the progress that is made with neural approaches. For this purpose, we compare twelve algorithmic approaches, among them six recent neural methods, under identical conditions on various datasets. We find that the progress in terms of prediction accuracy that is achieved with neural methods is still limited. In most cases, our experiments show that simple heuristic methods based on nearest-neighbors schemes are preferable over conceptually and computationally more complex methods. Observations from a user study furthermore indicate that recommendations based on heuristic methods were also well accepted by the study participants.

The details of this work can be found in the following papers:

- Ludewig, M., Mauro, N., Latifi, S., and Jannach, D. (2019). “Performance Comparison of Neural and Non-Neural Approaches to Session-based Recommendation”. In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys ’19, pp. 462–466.
- Ludewig, M., Mauro, N., Latifi, S., and Jannach, D. (Submitted). “Empirical Analysis of Session-based Recommendation Algorithms”. In: *User Modeling and User-Adapted Interaction*.

We plan to use the knowledge acquired from this work to benchmark the models described in the following chapters in order to have uniform evaluation and understand how to build the final hybrid recommender system to be integrated inside the OnToMap web application.

3 Related Work

In this chapter, we overview the related work concerning the models that will be presented in the next chapters.

3.1 Related Work on Information Search

Since we developed a suggestion model that proposes topics to the users in the context of geographical search, we aim to first explore the related work concerning information search, terms and concepts suggestion.

Various semantic information retrieval models employ concept networks to identify the meaning of queries and propose query expansions aimed at finding information, regardless of the terminology used by the user. E.g., both (Qiu and Frei, 1993) and (Grootjem and T.P. van der Weide, 2006) use a local thesaurus inferred from the source pool of documents to identify the concepts referred by the queries. Moreover, (Mandala et al., 1999) shows that the integration of different types of thesauri (linguistic, domain specific, etc.) improves the performance of query expansion techniques. The co-occurrence of word phrases in documents is also discussed in (J. Van Den Berg and Schuemie, 1999) and (Hoeber et al., 2005) to automatically generate associative conceptual spaces, and in (Joshi and Motwani, 2006) and (Akaishi et al., 2004) for term suggestion based on the documents returned by search engines. Differently, (Wang et al., 2012) classify queries in patterns according to their syntactic components and match them to a knowledge base to generate the answers. Finally, a knowledge-based approach is adopted in (Wang et al., 2017) for concept interpretation, and in (Fernández et al., 2011) to enhance information retrieval in the Semantic Web. On a different perspective, (Molina and Bayarri, 2011) proposes domain-specific ontologies for the interpretation of queries, assuming that users find it easier to specify what they want to do, rather than the concepts they are interested in.

Some recent work on information filtering and in recommender systems attempts to acquire relations among information items from the observation of users' behavior, and is complementary to our work. E.g., Google search engine manages the Knowledge Graph (Google, 2017) to relate facts, concepts and entities depending on their co-occurrence in queries. Moreover, (Oramas et al., 2015) use a knowledge graph for personalized item recommendation in the music domain. Furthermore, CoSeNa (Candan et al., 2009) employs keyword co-occurrence in the corpus of documents to be retrieved, and ontological knowledge about the domain concepts, to support the exploration of text collections using a keywords-by-concepts graph.

Similarly to these works, we use an ontology, and linguistic information, to interpret search queries at the conceptual level. However, we offer a *"you can also be interested in"* function to propose complementary concepts, i.e., topics, for expanding queries in order to satisfy the user's information needs, in a serendipitous way. For this purpose, we propose an associative information retrieval model based on the work described in (Giuliano and Jones, 1962) that concerns the observation of concepts co-occurrence in search sessions. Our work is related to the contextual query suggestion model presented in (Cao et al., 2008), who suggests queries on the basis of the context provided by the user's recent search history. However, we mine *ontology concepts* by interpreting search queries (we used the approach described in (Ardissono et al., 2016; Mauro and Ardissono, 2017a)), while the concepts defined in (Cao et al., 2008) are clusters of queries associated with similar sets of click results selected by users.

Our model also relates to session-based term suggestion approaches such as the one presented in (Huang et al., 2003), and with term suggestion models used for web site advertisement, e.g., see (Chen et al., 2008). However, it differs from those works because we look for *concept co-occurrence*, which abstracts from the specific words used to refer to concepts, while they observe *term co-occurrence* for query expansion.

Our work differs from works based on knowledge graphs because we exploit a knowledge graph to predict further concepts that the user might be interested in, i.e., we suggest topics, not individual items.

Working at the conceptual level, our work is also complementary to the research on exploration vs. exploitation in information retrieval, which provides models to recognize the type of search that the user is performing, and/or to adapt the search results accordingly; e.g., (Porrini et al., 2014; Athukorala et al., 2016; Medlar et al., 2017).

3.2 Related Work on Geographical Information Retrieval

The session-based recommender system of topics is applied in the domain of Geographical Information Retrieval. After the development of the model, we analyzed the impact of ontologies granularity on topic suggestion by testing the model using three different levels of conceptualization of the domain. Thus, in this section, we explore the related work concerning geographical information retrieval and the use of ontologies in the geographic domain.

Geographical Information Retrieval mainly focuses on the interpretation of the spatial component of a search task (Palacio et al., 2015; Henrich and Lüdecke, 2007). However, (Ballatore et al., 2016) identify research themes and open questions, which include aspects related to search behavior models and semantic aspects of spatial search. In this context, we focus on geospatial knowledge and aim to identify the potential of semantic granularity in search support.

In Geographic Information Retrieval, ontologies have been used to improve geographic *features* extraction (Laurini, 2015). On the other hand, we employ them for geographic *concepts* extraction, in order to provide the user with topics to explore, rather than individual items.

Researchers developed specific geographic ontologies, such as GeoNames. Moreover, they attempted to semantify geographical data targeted to specific tasks; e.g., the LinkedGeoData ontology links OpenStreetMap information to DBpedia, GeoNames, and others ontologies (Janowicz et al., 2012). (Fonseca et al., 2002a) propose an ontology to classify geographic elements with respect to geometrical features and attribute values; i.e., semantic features. Finally, a relevant amount of research has been devoted to semantifying geospatial Open Data and crowdsourced data. For instance, OSMonto (Codescu et al., 2011) attempts to structure the shallow implicit ontology of OpenStreetMaps tags. Moreover, in (Ballatore et al., 2013) a Semantic Web resource extracted from the OpenStreetMap Wiki website is proposed, to semantify crowdsourced data.

Differently, we aim at proposing concepts by using common patterns of concept exploration in search query logs. Furthermore, for our analysis, we use three geographical ontologies: a portion of the Geonames ontology, the ontology underlying the OnToMap web application and an ontology that structures OpenStreetMap tags and refers to geographic concepts in the context of cities.

3.3 Background and Related Work on Category-based Recommender Systems

In the work introduced in Section 2.1.2, we used information coming from search sessions to extract patterns from general search behavior in order to suggest topics to the users. In this work, we introduce the concept of long-term user models in the suggestion process and we leverage the concept co-occurrence graph employed in the work introduced in Section 2.1.2 to improve the suggestion capability of a category-based recommender system. This section reviews the literature related to category-based/tag-aware recommender systems (Zhang et al., 2011; Kim and Kim, 2014) highlighting the differences with our work.

Cross-domain recommendation has received the researchers' attention as a way to employ multiple information sources to contrast data sparsity; e.g., (Fernández-Tobías et al., 2016). Moreover, holistic user models have been developed that jointly analyze different types of user behavior to enhance the recognition of the user's needs; e.g., (Teevan et al., 2005; Musto et al., 2018b). However, the fusion of personal information from different applications is problematic, unless it is done within a tightly integrated software environment. For instance, most people operate anonymously (Greenstein-Messica et al., 2017) or have multiple identities (Doychev et al., 2014); moreover, most user activity logs are anonymized for privacy preservation purposes. It is thus interesting to consider other types of knowledge integration that do not require user identification across applications. Our work investigates this path of research.

Collaborative Filtering generates suggestions by analyzing item ratings to identify similar users or similar items. Several algorithms have been developed, from K-Nearest Neighbors (KNN) to more recent ones such as Matrix Factorization (Desrosiers and Karypis, 2011; Koren and Bell, 2011). In our work we adopt KNN because it has nice explanation capabilities and has proved to achieve good performance in a comparison with other approaches (Herlocker et al., 2000; Jannach and Ludewig, 2017; Ludewig and Jannach, 2018). By analyzing the characteristics of the user's neighbours found by the algorithm, it is possible to give explanations to her/him about the recommended items. In our future work, we plan to implement the model into more advanced techniques (e.g. Matrix factorization).

Ontological user profiles model preferences at the semantic level. (Sieg et al., 2007; Sieg et al., 2010) propose to exploit a taxonomy whose concepts represent item types, and to infer user interests on the basis of the observed ratings to the instances of such concepts. The neighborhood for rating estimation is then identified by measuring the semantic similarity between ontological user profiles. The category-based user similarity we propose is close to this approach. This type of extension also differentiates our work from that of (Ronen et al., 2016), who propose to extend the preferences of the individual user by analyzing her/his behavior in search logs: that work assumes that the user's activities can be tracked across applications and extends the user profile by analyzing her/his overall behavior.

Sen et al. define tag-aware recommender systems as "recommender algorithms that predict user's preferences for tags". In (Sen et al., 2009) they describe different signs of interest; e.g., searching or applying a tag, and so forth. (Gemmel et al., 2012) present a linear-weighted hybrid framework for resource recommendation that models different scenarios, among which tag-specific item recommendation. They propose to match users and items on the basis of their tag profiles. Differently, we match users on the basis of category-based profiles learned from rating behavior. The same kind of difference holds between our work and the one of (Nakamoto et al., 2007).

While TagiCoFi (Zhen et al., 2009) employs user similarities defined from tagging information to regularize Matrix Factorization, we use tags in a KNN algorithm. (Tso-Sutter et al., 2008) extend the ratings matrix using tagging information. They reduce the three-dimensional correlations $< user, tag, item >$ to two-dimensional correlations $< user, tag >$, $< item, tag >$ and $< user, item >$. Then, they apply a fusion method to combine the correlations for rating prediction.

However, we go one step forward in the identification of preferences by extending the user profiles with frequently co-occurring information interests (i.e. categories). We extend user preferences by analyzing anonymous data about general search behavior. Our work relates to tag-aware recommender systems because we analyze rating behavior on items associated to categories expressed as tags. However, we do not consider any other types of interaction with tags for estimating user preferences.

Recently, rating information has been combined with other types of data to improve recommendation. For instance, item reviews are used, possibly in combination with ratings, in (Chen et al., 2015; Musat and Faltings, 2015; Muhammad et al., 2015; Lu et al., 2018). Moreover, trust relations and reputation are used to steer recommendation on the basis of the feedback on items provided by trusted parties; e.g., (Kuter and Golbeck, 2007; Liu and Lee, 2010; Tang et al., 2013; Alotaibi and Vassileva, 2016; Mcnally et al., 2014; Du et al., 2017; Yang et al., 2017). In Chapter 8, we investigate multi-faceted trust for personalized recommendation (Mauro et al., 2019b; Ardissono and Mauro, 2020). However, in this work, we focus on rating information to assess the potential improvement of Collaborative Filtering, when combined with general preference co-occurrence.

3.3.1 Graph-based Information Filtering

Knowledge graphs describe item features and relations among entities, supporting the analysis of item relatedness, as well as similarity for information filtering and top-N recommendation. In several works, these graphs are extracted from document pools and/or from the Linked Data Cloud. For instance, CoSeNa (Candan et al., 2009) employs keyword co-occurrence in the corpus of documents to be retrieved, and ontological knowledge about the domain concepts, to support the exploration of text collections using a keywords-by-concepts graph. Moreover, (Di Noia et al., 2016) create a relatedness graph by analyzing external data sources such as DBpedia in order to support the evaluation of semantic similarity between items. Analogously, item features have been extracted from the Linked Data Cloud to improve recommendation performance in (Musto et al., 2017a; Ragone et al., 2017; Musto et al., 2017b; Musto et al., 2018a).

Some works attempt to extend the relations among information items by integrating data derived from the observation of different types of user behavior. E.g., Google search engine manages the Knowledge Graph (Google, 2017) to relate facts, concepts and entities depending on their co-occurrence in queries. Moreover, entity2rec learns user-item relatedness from knowledge graphs by analyzing data about users' feedback and item information from Linked Open Data (Palumbo et al., 2017). Furthermore, (Oramas et al., 2015) propose a hybrid recommender that integrates users' implicit feedback into a knowledge graph describing item information, enriched with semantic data extracted from external sources. Finally, (Vahedian et al., 2017) generalize graph-based approaches by simultaneously taking into account multiple types of relations among entities: they introduce meta-paths to represent patterns of relations and apply random-walk along such paths to identify relevant entities to suggest.

Our work has analogies to the above listed ones because we employ a graph-based type of knowledge representation. However, we work at the conceptual level: our knowledge graph relates item categories instead of individual users and/or items. Moreover, we do not compute similarity or relatedness by means of the knowledge graph: we use the graph to extend category-based user profiles. In turn, those profiles are employed in neighborhood identification. The separation between how preferences are inferred and how they are used for recommendation makes it possible to extend both types of activities in a modular way.

3.4 Background and Related Work on User Interfaces for Faceted Search

This section explores the background and related work concerning the use of user interfaces for faceted search. It is related to the user study that we conducted in order to understand which graphical widget (i.e.: checkboxes, treemaps and sunburst) is able to better support the user to find the information that (s)he is looking for in a map-based web application (OnToMap).

Exploratory search of large information spaces challenges users in the specification of efficient queries because, as most people are hardly familiar with the search domain, their information goals are often ill-defined (Marchiorini, 2006; White and Roth, 2006).

Starting from the pioneer filtering model proposed by (Ahlberg and Shneiderman, 1994), both (Sacco, 2000)'s Dynamic Taxonomies and (Hearst, 2006)'s faceted search model propose to use dynamic filters extracted from items metadata as constraints that the system can suggest to help the user identify relevant terms for information filtering and visualization of results. Specifically, (Hearst et al., 2002) present the Flamenco framework in which facet-based filtering is based on the exposure of hierarchical faceted metadata that describes the items of the search domain, i.e., apartments, or images (Yee et al., 2003).

Researchers also investigate ways to support the specification of the facets to filter results, as well as the access to Semantic Web information and Linked Data (W3C, 2018). As far as facet specification is concerned, new types of elements are proposed to filter the set of results; e.g., keywords or terms extracted from textual queries, as in HotMap (Hoeber and Yang, 2006), and concepts extracted from a document pool, as in Concept Highlighter (Hoeber and Yang, 2006), or terms extracted from a thesaurus as in Thesaurus-Results Browser (Sutcliffe et al., 2000). FacetLens (Lee et al., 2009) visualizes clickable facets in matrix-based bubbles, each one associated with a different search filter. Moreover, FacetZoom (Dachselt et al., 2008) proposes a stack-based visualization of hierarchical facets, also applied in Mambo (Dachselt and Frisch, 2007) as a model to combine faceted browsing with zoomable user interfaces. SearchLens (Chang et al., 2019) enables users to define long-lasting composite facet specifications (denoted as lenses) to support information filtering on multiple search sessions. In SearchLens, the user can specify the importance of the selected facets; thus, filtering is based on soft constraints used to rank search results.

In the faceted exploration of semantic data (Tzitzikas et al., 2017), search interfaces expose rich metadata that support browsing the information space through semantic relations. For instance, in the /facet browser, (Hildebrand et al., 2006) propose to combine hierarchical faceted exploration with keyword-based search. Moreover, (Petrelli et al., 2009) enables the user to search for heterogeneous types of information about items (e.g., texts and images) linked according to semantic relations, by extracting facets to guide exploration. Hippalus (Papadakos and Tzitzikas, 2014) introduces the Faceted and Dynamic Taxonomies to manage both hard and soft constraints in faceted filtering of semantic data and PFSgeo (Lionakis and Tzitzikas, 2017) extends Hippalus to geographic information management. Finally, focusing on geographic information, (Stadler et al., 2014) propose a semantic navigation method for SPARQL-accessible data (W3C, 2017; OCG, 2017) in the Facete browser.

Some works propose interactive graphical presentations of keywords to support sensemaking in the exploration of document sets. For instance, (Peltonen et al., 2017) propose the Topic-Relevance Map to summarize on a radial basis the keywords (filters) characterizing the result set, using distance from the center to represent relevance to the search query and angle between keywords to denote their topical similarity. Moreover, FacetAtlas (Cao et al., 2010) relates topics in a 3D diagram supporting the representation of multi-dimensional relations among them, and SolarMap (Cao et al., 2011) combines topic-based document clustering with a radial representation of facets to support a two-level, topic-based document filtering.

Similar to the cited works, our model exposes metadata derived from semantic knowledge representation. However, it enables users to work on maps populated with multiple data categories, i.e., with heterogeneous information, as well as to focus the maps on temporary interests without losing the overall set of data they contain because the map is interpreted as a long-term content sharing tool. This is useful to answer information needs in long-lasting user activities. Notice also that the OnToMap system, in which this work in integrated, does not assume to work on RDF data in order to comply with more general data sources, like public crowdmapping platforms, thanks to the mediation of its domain ontology. Moreover, it supports: (i) a browsing-based exploration guided by the structure of the domain ontology, which makes it possible to search for information following both IS-A and semantic relations; (ii) the semantic interpretation of free text queries to identify the data categories (ontology concepts) of interest by abstracting from the specific words occurring in the queries, via Natural Language Processing (Ardissono et al., 2016; Mauro and Ardissono, 2017a). More generally, OnToMap enables search support over a configurable set of data categories; in this way, it enables complex map development on different information domains. In contrast, most of the previous systems work on a single data type or on a pre-defined set of data categories, as in (Petrelli et al., 2009). In this work, we focus on faceted search as an alternative, or a complement, to query typing in order to use browsing-based navigation as a proactive guide to information exploration, given the structure of the information space.

Furthermore, the dynamic extraction of facets can challenge the user with a large number of browsing options. (Oren et al., 2006) focus on the efficiency of exploration and they promote the facets that enable the user to split the set of results in balanced subsets in order to minimize navigation steps. In comparison, we propose a facet selection policy suitable for sparse and highly unbalanced result sets, such as those typically retrieved from crowdsourced data sources, in which very few properties of items split results in subsets having similar cardinality.

3.4.1 Faceted Search in Recommender Systems

Recent work on recommender systems (Ricci et al., 2011) employs graphical visualization to enhance their transparency. Specifically, when hybrid recommenders are used, the user should be allowed to choose the recommendation algorithms (s)he prefers. Moreover, in order to increase trust, (s)he should be enabled to analyze the suggested items and to assess the rationale behind results. In this research area, recommenders are thus mapped to facets on which the user can express soft constraints. While most work adopts a list of sliders, one for each algorithm, to let the user specify her/his preferences, systems differ in the visualization of results. For instance, MyMovieFinder (Loepp et al., 2015) adopts a ranked-list model; by clicking on items, it is possible to see which recommendation criteria they meet. Moreover, IntersectionExplorer (Cardoso et al., 2019) uses the UpSet matrix (Lex et al., 2014) to visualize the amount of intersection of the suggestions provided by the recommenders that the user has selected. The amount of intersection is expressed by means of a meter and the details are visualized on demand, by explicitly selecting the intersection list to be shown. Finally, Scatterviz and RelevanceTuner (Tsai and Brusilovsky, 2019) display results in a customizable scatterplot (depending on the pair of facets selected by the user), or using a stackable score bar, respectively, given the degree of importance that the user gives to facets.

Differently, OnToMap displays search results on a geographic map because the location of geo-data is a primary dimension for information visualization. However, it is interesting to consider other models as possible extensions. In this respect, a finer-grained comparison reveals that the explanation of criteria met by individual items adopted in MyMovieFinder is similar to that of OnToMap, but graduated, as (Loepp et al., 2015) deal with soft constraints. Interestingly, the UpSet matrix used in IntersectionExplorer might be considered as a possible explanation tool for faceted search, separately applied to each data category selected by the user. However, different from (Cardoso et al., 2019), where we expect that the user selects a small number of recommenders, geographic information may have a large number of facets, increasing the dimensions of the matrices to be visualized. Therefore, the real benefit of a detailed matrix-based analysis of results, with respect to dynamically selecting/deselecting items by applying faceted visualization constraints, should be investigated. For our future work, we plan this kind of study and the analysis of cartographic techniques designed to support visual thinking in geographic domains, such as those described by (Andrienko et al., 2007).

3.5 Background and Related Work on Trust-based Recommender Systems

In the work introduced in Section 2.1.4, we used information coming from search session to improve the suggestion capability of a recommender system. However, other types of information can be employed to enhance recommender systems' performance. In this part of the thesis, we investigate the use of reputation and trust indicators in recommender systems. This section reviews the literature related to trust-based recommender systems highlighting the differences with our work.

3.5.1 Basic Concepts: Trust and Reputation

Trust is generally described as a positive expectation that an agent has about other agents' behavior, from a subjective perspective. (Gambetta, 1988) defines it as "a particular level of the subjective probability with which an agent or group of agents will perform a particular action, both before [we] can monitor such action (or independently of [our] capacity of ever to be able to monitor it) and in a context in which it affects [our] own action". Moreover, both (Gambetta, 1988) and (Golbeck and Hendler, 2004) specify that a user trusts another one in a social network if (s)he believes that any future transaction with her/him will be rewarding rather than detrimental. On a more general perspective, (Mui et al., 2002) (and similarly (Misztal, 1996)) elect "subjective probability" to subjective expectation, or degree of belief, to highlight that, more than a statistical probability, trust represents a belief status that an agent A has about another agent B's future behavior, given B's past behavior and her/his *reciprocity* of action within a society.

Different from trust, *reputation* describes a general "expectation about an agent's behavior based on information about or observations of its past behavior" (Abdul-Rahman and Hailes, 2000). Reputation has a global perspective and (Mui et al., 2002) describe it as the "perception that an agent creates through past actions about its intentions and norms". According to (Misztal, 1996), reputation "helps us to manage the complexity of social life by singling out trustworthy people - in whose interest it is to meet promises".

While the previously described works analyze trust and reputation from the global viewpoint of agent-to-agent interaction, a few ones contextualize it in online collaboration systems and social networks, which are the scope of our present work. Noticeably, (Mcnally et al., 2014) generalize trust relations by analyzing the occurrence of collaboration events that involve users; this makes it possible to link users because they have downloaded or bookmarked contents provided by other users, and so forth. These authors explain that reputation can derive from direct user-to-user interaction (e.g., when users are rated) or from indirect one; e.g., when they interact by virtue of some item. Moreover, it can derive from explicit trust statements, such as ratings, or from implicit ones like follower relations.

Before describing the state of the art on trust-based recommender systems, it is worth briefly discussing the main issues affecting them. Specifically, it may be questioned whether relying on social relations and global feedback about users is a safe approach to evaluate trustworthiness. Some trust-based recommender systems focus on user-to-user relations and ignore the feedback on user actions because there is a general opinion that the latter could be biased. While we obviously agree that this may be true, we point out that *any* type of action that brings evidence about trust, including the establishment of friend relations, ratings, etc., could be performed with the aim of manipulating the reputation of some user. Therefore, data reliability assessment is a general pre-requisite for the development of recommender systems. Indeed, the weaknesses of some models adopted in e-commerce and collaboration sites have been analyzed to suggest how to improve the robustness of Reputation Management Systems; e.g., see (Resnick and Zeckhauser, 2002). However, Jøsang et al. point out that these systems are challenged by strategic manipulation and by various types of attacks that cannot always be detected by statistical analyses (Jøsang et al., 2007; Jøsang and Goldbeck, 2009). Therefore, (Jøsang, 2012) ultimately highlights the importance of strengthening legislation as a barrier to discourage malicious behavior.

3.5.2 Trust-based Recommender Systems

The homophily (McPherson et al., 2001) and social influence (Marsden and Friedkin, 1993) theories associate social links to user similarity. On this basis, social and trust-based recommender systems (Richthammer et al., 2017) exploit social networks as additional sources of information to complement rating data. These systems estimate user preferences by relying on the known social links existing between people; e.g., friend, follower and/or trust relations according to different inference techniques:

- AVG: average rating of selected (e.g., trusted) social links (Golbeck and Hendler, 2004; Golbeck and Hendler, 2006; Liu and Lee, 2010; Parvin et al., 2019).
- KNN: K-Nearest Neighbors on social links (O'Donovan and Smyth, 2005; Massa and Avesani, 2007; Groh and Ehmig, 2007; Moradi and Ahmadian, 2015; Ardissono et al., 2017b).
- MF: Matrix Factorization (in some cases with Random Walk) on the matrices of ratings and social links (Jamali and Ester, 2009; Jamali and Ester, 2010; Ma et al., 2011b; Yang et al., 2012; Tang et al., 2013; Deng et al., 2014; Guo et al., 2015; Yang et al., 2017).
- PMF: Probabilistic Matrix Factorization on the matrices of ratings and social links (Ma et al., 2011a; Ma et al., 2011c; Jiang et al., 2012; Liu and Aberer, 2013; Chaney et al., 2015).
- Probabilistic approaches on trust networks (Kuter and Golbeck, 2007; Li et al., 2014).
- Co-clustering of ratings and trust matrices (Du et al., 2017).

Some research about recommender systems studies the differences between trust and friends networks. (Guo et al., 2015), (Ma et al., 2011b) and (Li et al., 2018) find out that, different from explicit trust relations (such as those among Epinions users (Epinions, 2019)), friendship does not strictly imply preference similarity: user preferences are strongly correlated among trusted neighbors but they are only slightly positively correlated among "trust-alike" neighbors such as friends in social networks (Guo et al., 2015). Several authors recognize the importance of limiting the social context to the user's local proximity; for instance, (Massa and Avesani, 2007) and (Yuan et al., 2011) prove that recommendation accuracy decreases when indirect social connections (i.e., paths of social links) are used to estimate user preferences. Moreover, (Yang et al., 2012) point out that users may trust different subsets of friends regarding different domains. In order to deal with this issue, authors propose various methods to filter the neighborhood used for rating prediction; e.g., (Yang et al., 2017) use category-specific circles and (Yuan et al., 2011) use thematic groups to steer Matrix Factorization. Moreover, KNN and AVG systems select neighbors by ranking the users directly linked to the current user in the social/trust network on the basis of their rating similarity (Massa and Avesani, 2007; Liu and Lee, 2010; Li et al., 2014; Moradi and Ahmadian, 2015; Ardissono et al., 2017b; Parvin et al., 2019). Analogously, social regularization is used to increase the impact of like-minded users in Matrix Factorization: e.g., TrustMF (Yang et al., 2017) applies social regularization to users' direct social links and (Yuan et al., 2011) applies it to the members of thematic groups; RSTE (Ma et al., 2011a) and SOREG (Ma et al., 2011b) integrate trust and rating similarity in Probabilistic Matrix Factorization, and (Ma et al., 2011c) use tag-based similarity to build a larger social context for regularization. Finally, SocialMF (Jamali and Ester, 2010) employs rating similarity to regularize the impact of users who are reachable through a short path of social links in Random Walk. Other systems achieve similar filtering results by combining trust-based and item-based recommendation, as in TrustWalker (Jamali and Ester, 2009), or by filtering the users of the trust networks according to rating similarity, as in TCFACO (Parvin et al., 2019) and RelevantTrustWalker (Deng et al., 2014). Finally, (Du et al., 2017) apply co-clustering to the matrices of ratings and social relations in order to identify like-minded users within social connections.

Building on social influence theories, (Guo et al., 2015) propose TrustSVD that extends SVD++ (Koren, 2008) to jointly factorize the rating and trust matrices: they learn a truster model that describes how people are influenced in item evaluation by their parties' opinions. TrustMF (Yang et al., 2017) learns both the truster and trustee models to consider the fact that, in a social network, people mutually influence each other. (Jiang et al., 2012) investigate the relationship between social influence and personal preferences. Finally, some researchers leverage the local and global perspectives of social influence building on the observation that in the physical world humans ask for opinions from both local friends and highly reputable people; e.g., (Qian et al., 2016; Tang et al., 2013; Hu et al., 2018). Specifically, in LOCABAL, (Tang et al., 2013) combine rating similarity and social links with users' global reputation, which is based on the PageRank (Page et al., 1999) score as a measure of importance in the social network.

Some trust-based recommender systems assume the existence of both positive and negative evidence about trust as, e.g., in the social networks where users can rate other users positively or negatively (Li et al., 2011; Victor et al., 2011; Rafailidis and Crestani, 2017). In our work, we start from the consideration that the trust models provided by several social networks are only based on the expression of "likes". Therefore, we propose a model that can also work on positive-only feedback to comply with them.

Some works associate users' reputation to rating conformity; e.g., (O'Donovan and Smyth, 2005) and (Li et al., 2013) base reputation on the percentage of ratings provided by a user that agrees with those of the other people. (Su et al., 2017) cluster users on the basis of rating similarity and consider the largest cluster as the "honest" group. In the SoRS recommender system (Qian et al., 2016) derive reputation by iteratively calculating the correlation of the historical ratings provided by a user and the quality of items emerging from the rating scores they receive. However, review conformity does not fully characterize quality; e.g., (Victor et al., 2011) point out that controversial reviews must be considered and matched to individual preferences. We thus leave this aspect for our future work.

In comparison, in the work that will be presented in Chapter 8 in which we integrate our reputation model in a Matrix Factorization algorithm, we extend social regularization by tuning the impact of users on the Matrix Factorization process on the basis of both rating similarity and global multi-dimensional reputation. In other words, we select neighbors for rating prediction by privileging users who are trustworthy *and* like-minded. This approach improves prediction accuracy because it enhances the quality of the rating information used to estimate preferences.

Furthermore, we generate personalized recommendations by relying on a compositional, multi-faceted trust model that includes complementary data about user behavior: i.e., not only local trust among users inferred from social relations, but also quality of user contributions (derived from the anonymous global feedback they receive) and multi-dimensional global reputation derived from diverse types of information, among which anonymous endorsements to user profiles. The integration of these facets of trust supports a rich computation of reputation based on complementary aspects of user behavior. Moreover, it makes it possible to compensate trust evidence in application domains in which some types of information are not available or cannot be used. In particular, our model works with or without using social links. As previously noticed, this makes it possible to tune our model on the basis of the contributions of trustworthy, like-minded people, using public, anonymous information.

To the best of our knowledge, the only other work that employs global feedback about users is LGTR by (De Meo et al., 2018), which defines global reputation on the basis of the feedback collected by user actions and of a local context depending on social relations. However, in neighbor identification, LGTR discards rating similarity, which is very useful to select like-minded users for preference prediction. Moreover, LGTR does not take review quality and endorsements to user profiles into account. Our model is thus more general than this one.

4 Session-based Suggestion of Topics for Exploratory Search

The model described in this section has been developed with the aim to integrate it into the OnToMap web application. This because, after developing a model that was able to understand geographical textual search queries and retrieve the required information, we decide to take a step forward and also suggest to the users relevant clusters of concepts that (s)he can eventually explore. The model is able to generate suggestions by analyzing general behavior patterns in search sessions.

The work described in this section has been published in:

- Mauro, N. and Ardissono, L. (2018). "Session-based Suggestion of Topics for Geographic Exploratory Search". In: *Proceedings of 23rd International Conference on Intelligent User Interfaces*. IUI '18. Tokyo, Japan: ACM, pp. 341–352.

4.1 Introduction

During an information search task, various issues challenge the specification of efficacious queries. Firstly, as discussed by Belkin in (Belkin, 1980), users are often unable to use the appropriate keywords because they are looking for data they do not know about. Secondly, complex information spaces, organized in several topic categories, can overload and disorient people, preventing them from finding the information they need; e.g., see (Ratzan, 2004). In order to address these issues, several systems present lists, or hierarchies, of information categories to search for. However, in this way, they expose users to possibly large sets of options to choose from.

Our work focuses on exploratory search, which is affected by the above issues because the users' information goals are ill-defined; see (Marchiorini, 2006; White and Roth, 2006). We aim at helping users orientate themselves by suggesting relevant topics (concepts) to be explored, given their observed search behavior. More specifically, we aim at guiding data exploration by proposing a small set of concepts that the user might be interested in, given the search context, as a *"you might also be interested in"* function that helps her/him complete the search. We adopt an associative information retrieval model (Giuliano and Jones, 1962): our hypothesis is that, by analyzing the first query(ies) of a search session, and by taking into account which types of data are often searched together by people, the system can help the user identify further topics s(he) might be interested in, by suggesting terms for query expansion. E.g., if the user looks for kindergartens in a town, (s)he might also be interested in play and sports areas, as well as in other data related to children activities. Our work aims at supporting the suggestion of such concepts. Specifically, we aim at answering the following research question:

$\mathbf{RQ1_1}$: *Can the data about the concepts frequently searched together by people within a search session be exploited to help the user explore the portions of an information space relevant to her/his information needs?*

In order to answer our research question, we developed a session-based concept suggestion model which, starting from the queries submitted by the user, proposes a set of possibly relevant data categories, complementing the types of information that (s)he has focused on. Our model is based on an ontological representation of the information space, and on a Natural Language approach to the interpretation of search queries that identifies the referred concepts in a flexible way, by considering synonyms and by taking the ambiguity of words into account. For the development of our model, we analyzed the log of a major search engine and we generated a weighted co-occurrence graph that represents how often concepts are searched together in the sessions of the log. Then, we extracted the clusters of concepts that most frequently co-occur by applying a community detection algorithm to the graph. Those clusters are the basis for query expansion, starting from the concepts referred by the user in the observed part of the search session. We defined a few heuristics for recommending relevant types of information and we tested them on the log, achieving satisfactory accuracy results.

The main contributions of our work are: (i) a session-based query expansion model that suggests complementary concepts for satisfying the user's information needs in geographic exploratory search, and (ii) evaluation results of the model.

Section 4.2 describes the dataset we used for our experiments. Section 4.3 describes the approach we adopted to identify the co-occurrences of concepts in search sessions. Section 4.4 proposes some query expansion strategies and the subsequent section evaluates the strategies. The last section concludes the work and outlines some future work.

AnonID	Query	QueryTime	Rank	ClickURL
67910	las vegas sports teams	2006-04-30 18:14:57	1	http://www.vegas.com
67910	las vegas transportation	2006-04-30 18:19:59	1	http://www.vegas.com
67910	mccarran international airport	2006-04-30 18:22:30	3	http://en.wikipedia.org
67910	mccarran international airport	2006-04-30 18:22:30	1	http://www.mccarran.com
67910	hub airports in the united states	2006-04-30 18:25:28	9	http://www.airportcodes.us
67910	hub airports in the united states	2006-04-30 18:25:28	9	http://www.airportcodes.us
67910	black las vegas itineraries	2006-04-30 18:30:03		
67910	educational facilities in las vegas	2006-04-30 18:30:53		
67910	medical facilities in las vegas nv	2006-04-30 18:31:44	1	http://lasvegas.citysearch.com
67910	medical facilities in las vegas nv	2006-04-30 18:31:44	3	http://www.lasvegasnevada.gov
67910	medical facilities in las vegas nv	2006-04-30 18:31:44	9	http://www.lasvegasrelocating.com
67910	unique architecture in las vegas nv	2006-04-30 18:35:02	10	http://www.guggenheim.org
67910	unique architecture in las vegas nv	2006-04-30 18:35:02	2	http://travel.yahoo.com
67910	architecture in las vegas nv	2006-04-30 18:40:28	4	http://lasvegas.citysearch.com
67910	architecture in las vegas nv	2006-04-30 18:40:28	2	http://www.library.unlv.edu
67910	architecture in las vegas nv	2006-04-30 18:40:28	3	http://local.yahoo.com
67910	religious sites in lasvegas	2006-04-30 18:50:35	1	http://www.lasvegas.worldweb.com
67910	religious sites in lasvegas	2006-04-30 18:50:35	2	http://www.lasvegas.worldweb.com
67910	religious sites in lasvegas	2006-04-30 18:56:09	12	http://travel2.nytimes.com
67910	religious sites in lasvegas	2006-04-30 18:56:09	12	http://travel2.nytimes.com

Tab. 4.1: Sample session from the AOL log.

4.2 Dataset

We defined our concept suggestion model, and evaluated its accuracy, by using as a dataset the AOL query log[1]. Although that log was involved in an information leak issue, we decided to use it for two reasons: firstly, our analysis is ethically correct, because we analyze and model general search behavior, by aggregating individual users' actions in order to abstract from the search histories of particular users. The preference co-occurrence clusters we extract represent aggregate data about people's behavior, and they abstract from the search histories of particular users. Secondly, to the best of our knowledge, the AOL log is the only public, large dataset that reports detailed data about textual search queries, and that can be used for linguistic interpretation. Furthermore, we analyzed some public datasets but they did not meet our requirements. E.g., the Excite query dataset[2] contains fewer queries (about 1M queries against the 20M of AOL log). In the Yahoo dataset[3] queries are coded; thus, it is not possible to extract any linguistic information to learn the concept co-occurrence clusters. Moreover, it is worth noting that, even though the AOL query log dates back to 2006, it can be considered as a good information source as long as it is analyzed from the viewpoint of the concepts expressed by the users. In other words, while the specific information items mentioned in the log might not exist anymore, the topics referred in the queries are general and long-lasting.

[1] We retrieved the AOL query log in June 2016 from https://archive.org/details/AOL_search_data_leak_2006.

[2] `https://svn.apache.org/repos/asf/pig/trunk/tutorial/data/`

[3] `https://webscope.sandbox.yahoo.com/catalog.php?datatype=l&did=50`

The AOL log is composed of 10 files; however, we excluded one of them because it seemed to be corrupted by queries likely performed by a bot: that file included a set of 20,000 queries spawning over 6 days, or click through, submitted by the same user.

Each line of the AOL log represents either a query, or a click-through event (in which the user clicks on one of the search results of the previously submitted query). The lines contain the following fields:

- *AnonID*: ID that represents a user in an anonymous way.
- *Query*: search query submitted by the user.
- *QueryTime*: date and hour when the query was submitted.
- *Rank*: given the list of results (links to web pages) returned by the search engine, this field represents the relative position of the link on which the user has clicked. If the user has not selected any search results, this field is empty.
- *ClickURL*: URL of the search result selected by the user. Similar to field *ItemRank*, this field can be empty.

Table 4.1 shows a session from the AOL log. It can be noticed that the user looks for various types of information, concerning sports, transportation, medical facilities, and others. This substantiates the need for a suggestion model that helps her/him by proposing different topics for exploration. We decided to use the approach of splitting the dataset in sessions following the one adopted in the information retrieval community (see (White et al., 2007)) that is based on dividing the session according to their timestamp and not by looking at the context of search. In our future work, we plan to improve the model by recognizing the change of context in search sessions.

Since we aim at developing an intelligent search support function for OnToMap (Ardissono et al., 2017d; Ardissono et al., 2017c), we pre-processed the AOL log to work on a smaller dataset, focused on the search sessions relevant to geographical ontology underlying the system. The ultimate goal is to train a query expansion model suitable for improving information search in OnToMap.

In order to build the dataset of our experiments, we first identified the search sessions from the log: we aggregated the queries performed by the same user according to their temporal proximity, following the widely applied rule that two consecutive queries belong to different sessions if the time interval between them exceeds half an hour; see (White et al., 2007). Then, we selected the sessions including at least one

rdfs:label	"Kindergarten"@en
rdfs:comment	"An educational institution for young children, usually before they go to primary school."@en
:keywords	"child"@en, "educational"@en, "young"@en
:lemma	"kindergarten"@en
:synonyms	"0th_grade"@en, "all-day_kindergarten"@en, "didactics"@en, "childcare"@en, "educability"@en, "education_program"@en, "education_system"@en, "junior_kindergarten"@en, "kindergarden"@en, "nursery"@en, "pre-primary"@en, "pre-school"@en, "preschoolar"@en

Tab. 4.2: Ontology definition of concept "Kintergarden". Values are tagged with the reference language (@en, for English).

query that refers to the concepts of the OnToMap ontology.[4] For this purpose, our main task was the identification of the concepts referred by the terms of the queries, which we carried out using the approach described in (Ardissono et al., 2016; Mauro and Ardissono, 2017a). Concept identification is a particularly important task: as reported in (Wang et al., 2012), the analysis of the log of a major search engine proved that about the 62% of the queries contain at least one conceptual class term.

4.2.1 Reference Ontology

The OnToMap ontology describes concepts by merging linguistic, encyclopedic and technical knowledge, focusing on concepts related to the territory that are typical of Participatory Processes in which the population is involved to discuss solutions and projects about a specific geographic area; see (Ardissono et al., 2016; Voghera et al., 2016). Each concept has a textual description and a set of lemmatized synonyms and keywords extracted from the description. E.g., Table 4.2 shows the specification of concept "Kintergarden". It reports the concept name (rdfs:label), its textual description (rdfs:comment), a few lemmatized keywords (:keywords), the lemma of

[4] We did not consider any broader ontologies, such as WordNet (WordNet, 2017) or ProBase (Wu et al., 2012), because we aimed at obtaining a focused dataset to be analyzed.

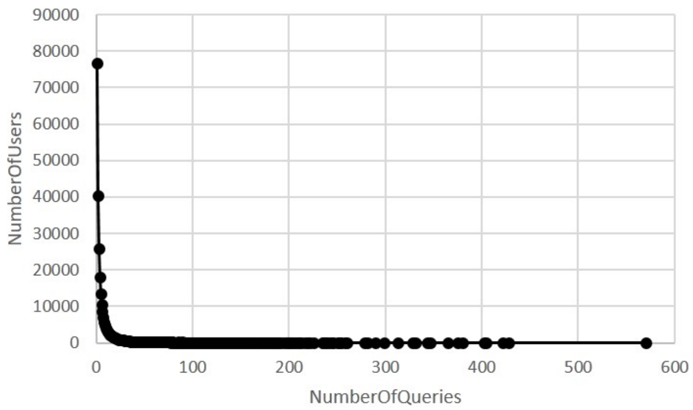


Fig. 4.1: Distribution of the number of queries per user.

Minimum number of queries per user	1
Maximum number of queries per user	428
Mean number of queries per user	6.38
Median number of queries per user	3
Standard Deviation	11.37

Tab. 4.3: Measures about the distribution of the queries per user.

the concept (:lemma), and a set of lemmatized synonyms of the terms occurring in the description (:synonyms).

4.2.2 Creation of the Dataset for the Experiments

We identified the AOL queries relevant to our experiments by matching the knowledge about concepts (i.e., the lemmatized synonyms and keywords extracted from the textual description of the ontology concepts) to the lemmatized words that compose the queries. If there was at least one match between a concept and a query, we considered the query as relevant, and we included the session to which it belonged in our dataset.

Each query can refer to one or more ontology concepts: the mean number of concepts per query of AOL-reduced is 1.502. This is due to the following reasons:

- The query might refer to independent concepts, each one identified in an unambiguous way. For instance, in a sample query like "public school and transportation in New York" the identified concepts would be *School* and *Local Public Transportation*. Notice that, similar to the findings reported in (Beitzel

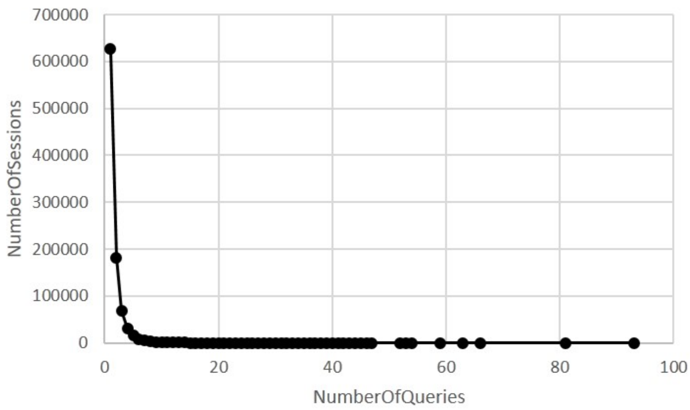


Fig. 4.2: Distribution of queries in sessions.

Minimum number of queries per session	1
Maximum number of queries per session	93
Mean number of queries per session	1.67
Median number of queries per session	1
Standard Deviation	1.45

Tab. 4.4: Measures about the length of the search sessions.

et al., 2005), most queries of the AOL log are short and refer to a single concept, as in the examples of Table 4.1.

- Because of ambiguity issues, more than one concept could be identified. For instance, given query "missouri child support" from the AOL log, the concepts identified from word "child" are *Childcare Service*, *Play Area* and *Kindergarten*, both including term "child" in their descriptions. As the query is short, it is difficult to understand which topic is the most probable one. Therefore, we assume that the query matches all concepts, with uncertainty.
  In the analysis of the log, we could not exploit the search results selected by users to disambiguate the queries, because we found out that, in most cases, they refer to the root pages of large web sites (e.g., see the URLs in Table 4.1), or they are obsolete (e.g., the link reported as a search result of query "missouri child support", http://www.dss.state.mo.us). Thus, they help in a few cases.

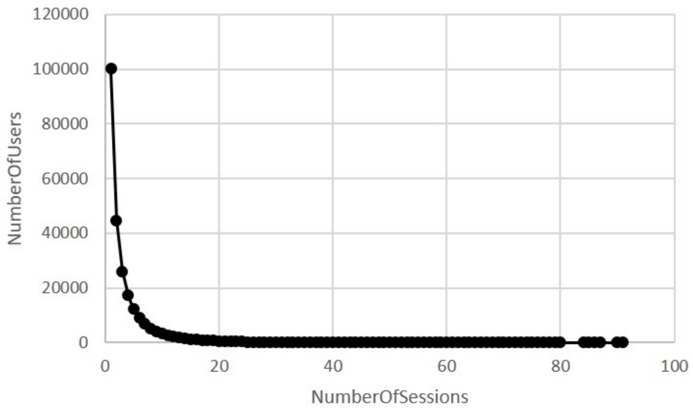


Fig. 4.3: Distribution of sessions per users.

4.2.3 Characteristics of the Dataset

The dataset obtained after the pre-processing phase, henceforth denoted as *AOL-reduced*, is composed of 1,581,817 queries submitted by 247,868 users. The chart in Figure 4.1 shows the distribution of users with respect to the number of queries they specified. It can be noticed that there is a long tail distribution: most users submitted a very small set of queries, whereas a small group of users submitted several queries. As shown in Table 4.3, the mean number of queries per user is 6, but the distribution differs from this value for about 11 queries (standard deviation).

The dataset includes 945,945 sessions, on which we performed two types of analyses:

- Firstly, we analyzed the distribution of sessions w.r.t. the number of queries they include. Table 4.4 shows the measures related to the length of sessions considering the number of queries. The mean length of a session is 1.67. The data follow a long tail distribution, as shown in Figure 4.2.
- Secondly, we analyzed the distribution of sessions in the dataset with respect to users in order to compute the mean number of sessions that the users started; see Figure 4.3. We found out that users engaged in a mean number of about 3.82 sessions (median = 2). Moreover, the data follow a long tail distribution, with several people engaging in few search sessions, and a small number of users starting a larger number of sessions.

4.3 Identifying clusters of frequently co-occurring concepts

As described in the introduction, starting from the interpretation of the search queries that the user submits, we aim at suggesting further concepts that could be relevant to her/his needs. The idea is to provide an adaptive list of pointers to the types of information provided by the system, which are useful in the context of the user's search. We did this by identifying, from the concepts referred by the queries of the AOL-reduced dataset, other concepts that are related to the former from the viewpoint of commonly shared interests.

4.3.1 Step 1: Creation of the Concept Co-occurrence Graph

Starting from the AOL-reduced dataset, we built a graph that represents concepts co-occurrence in search sessions. By co-occurrence we mean the fact that two or more concepts are referred by the queries belonging to the same session. We adopt the following approach because we are interested in finding co-occurrence of concepts starting from the queries terms by taking into account the lexical ambiguity in a lighter way. We are not interested in performing an heavy query processing to understand to which concept the query is referring to. Thus, we adopt a lighter approach that only intervenes on the edges weights in case of ambiguity.

The graph is composed as follows:

- Each node represents a concept;
- Each edge represents the co-occurrence weight of the two connected concepts. This weight is computed by summing up the evidence of co-occurrence observed in all the sessions of the dataset.

The idea is that, each time two concepts are identified within the same session, the weight of the edge connecting them must increase, in order to capture and reinforce the hypothesis that people frequently search these concepts together.

Given two concepts c_i and c_j, the weight of the edge that connects them is defined as:

$$w_{c_i c_j} = \sum_{S=1}^{nSessions} Freq_{ij}(S) \tag{4.1}$$

where $Freq_{ij}(S)$ represents the contribution provided by session S to the co-occurrence frequency of c_i and c_j.

The contribution of the sessions to the weights of the co-occurrence graph (CG) is represented by a local weighted graph created by interpreting the queries of S. Specifically, $Freq_{ij}(S)$ is obtained by considering the evidence of co-occurrence provided by the queries that compose S. However, within a session, we avoid summing up the evidence provided by multiple occurrences of the same concepts, because they could derive from a query reformulation (Rieh and Xie, 2006) or from the repetition of the queries in click-through events; see Table 4.1.

Given $S = \{Q_1, \ldots, Q_n\}$, $Freq_{ij}(S)$ is thus computed as follows:

$$Freq_{ij}(S) = Max_{k=1}^{n}(Freq_{ijQ_k}, ev_{ijQ_{k-1}}) \tag{4.2}$$

where $Freq_{ijQ_k}$ is the co-occurrence evidence (i.e. weight of the local co-occurence graph) of concepts c_i and c_j provided by query Q_k, and $ev_{ijQ_{k-1}}$ is the one estimated during the interpretation of queries $Q_1, \ldots, Q_{k-1}$.

Finally, the contribution of a query Q is computed as follows:

- If Q contains n terms ($n >= 0$), each one identifying a non-ambiguous concept: $T_1 \Rightarrow c_1, \quad T_2 \Rightarrow c_2, \quad \ldots, \quad T_n \Rightarrow c_n$, then, Q generates nodes $c_1, \ldots, c_n$ of the local graph (if they do not exist yet). For all pairs of concepts c_a and c_b referred by Q, $Freq_{abQ} = 1$. Moreover, $Freq_{vwQ} = 1$ for all the edges connecting c_a, or c_b, to some non-ambiguous concept of the local graph. $Freq_{vwQ} = 0$ for the other edges of the graph.

  For instance, suppose that Q ="public school and transportation in New York" is the first query of a session. The terms and concepts referred by Q are:

  - school $\Rightarrow$ *School* - denoted as concept x;

  - transportation $\Rightarrow$ *Local Public Transportation* - denoted as concept y.

  Then, the local graph is composed of the x and y nodes. Moreover, $Freq_{xyQ} = 1$.

- If Q contains a term t that refers to m concepts $\{c_1, \ldots, c_m\}$, the interpretation is ambiguous. Therefore, the evidence brought by t to the concepts is $\frac{1}{m}$, in order to take into account the possible interpretations of the query, and spread the weight to the ambiguous concepts. Thus, for each pair c_a, c_b in $\{c_1, \ldots, c_m\}$, $Freq_{abQ} = \frac{1}{m}$. $Freq_{vwQ} = 1/m$ for all the edges connecting c_a, or c_b, to the other concepts of the local graph. Finally, $Freq_{vwQ} = 0$ for all the edges of the local graph that are not outgoing arcs of c_a or c_b.

  For example, if Q = "missouri child support" is the first query of a session:

- The concepts identified from word "child" are: *Childcare Services* - denoted as x, *Play Areas* - y, and *Kindergartens* - z.
- Then, the local graph is composed of the x, y and z nodes. Moreover, $Freq_{xy_Q} = Freq_{xz_Q} = Freq_{yz_Q} = \frac{1}{3}$.

We now sketch the generation of the local co-occurrence graph for a sample search session s. We recall that the contribution of multiple sessions to the overall co-occurrence graph (GC) is incremental: i.e., we sum up the weights provided by the individual sessions. Let's suppose that the terms of s refer to the following concepts:
$Q_1 : t1 \Rightarrow c_1$
$Q_2 : t2 \Rightarrow c_2$
$Q_3 : t3 \Rightarrow c_3, c_4$
$Q_4 : t4 \Rightarrow c_2$
$Q_5 : t5 \Rightarrow c_3$

Figure 4.4 shows the local co-occurrence graph generated by s.

- Q_1 adds node c_1 to the local graph.
- Q_2 adds c_2 and assigns a weight $= 1$ to the edge between c_1 and c_2, because there is no ambiguity:
  $Max(Freq_{c_1 c_2 Q_2}, ev_{c_1 c_2 Q_1}) = Max(1, 0) = 1.$
- Q_3 adds c_3 and c_4. Moreover, it assigns a weight $= \frac{1}{2}$ to the edge connecting c_3 and c_4 (the ambiguity concerns two concepts). Furthermore, it assigns the same weight to the edges connecting c_3, and c_4, to the concepts of the previous queries (i.e., the edges between c_3 and c_1, c_3 and c_2, c_4 and c_1, c_4 and c_2).
- Q_4 does not add any nodes, nor does it modify the weights in the local graph because the evidence of c_2 doesn't solve the ambiguity between c_3 and c_4:
  $Max(Freq_{c_2 c_3 Q_4}, ev_{c_2 c_3 Q_3}) = Max(0.5, 0.5) = 0.5$
  $Max(Freq_{c_2 c_4 Q_4}, ev_{c_2 c_4 Q_3}) = Max(0.5, 0.5) = 0.5$
  $Max(Freq_{c_1 c_2 Q_4}, ev_{c_1 c_2 Q_3}) = Max(1, 1) = 1.$
- Q_5 does not add any nodes, but it solves the ambiguity between c_3 and c_1, and c_3 and c_2, respectively, because it provides a non ambiguous evidence of c_3. Thus, the respective weights of the local graph are updated:
  $Max(Freq_{c_1 c_3 Q_5}, ev_{c_1 c_3 Q_4}) = Max(1, 0.5) = 1.$
  $Max(Freq_{c_2 c_3 Q_5}, ev_{c_2 c_3 Q_4}) = Max(1, 0.5) = 1.$
  Notice that we maintain the ambiguity between c_3 and c_4 (in a conservative approach) because, in general, we cannot assume that Q_5 is a reformulation of Q_3. Thus, we do not use it to exclude c_4 from the interpretation of Q_3.

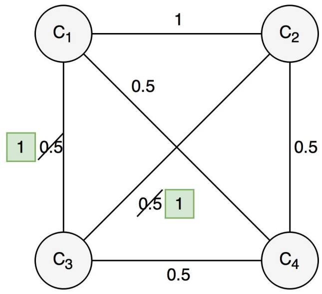


Fig. 4.4: Construction of the graph describing the concepts co-occurrence frequency for a session.

Fig. 4.5: Distribution of the weight of edges in the co-occurrence graph.

4.3.2 Step 2: Pruning the Graph

The co-occurrence graph of the AOL-reduced dataset is strongly connected: almost all of the ontology concepts are linked to each other by an edge whose weight is > 0. We thus decided to analyze the distribution of weights in the graph, in order to understand the strength of the correlation between concepts. Figure 4.5 shows this distribution: the x-axis represents the edges, and the y-axis represents their weights, which take values in [0, 68707.5].

The distribution highlights the fact that there is a large number of weakly connected concepts, i.e., of candidate items for query expansion. As, different from keyword suggestion in search engines, we aim at proposing few concepts, we decided to delete the edges having low weight, assuming that they represent weak associations between concepts, and that they capture less commonly shared interests.

We selected a threshold for pruning the graph in order to optimize the prediction accuracy of the resulting concept co-occurrence clusters; i.e., the degree of matching between the set of concepts composing the clusters and the concepts identified in the sessions of AOL-reduced. The selected threshold is 2200. Details about the evaluation of accuracy follow.

4.3.3 Step 3: Creation of the Concept Co-occurrence Clusters

Starting from a co-occurrence graph (CG) pruned with a threshold, the clusters are created by applying a community detection algorithm that identifies the sets of strictly correlated concepts. Our hypothesis is that they correspond to sets of concepts that are frequently searched together.

For the identification of clusters, we analyzed various algorithms. We selected COPRA (Gregory, 2010), which works on weighted graphs and detects overlapping communities. In this way, a concept can belong to different clusters at the same time. Our idea was that, within a session, the user might focus on more than one set of highly correlated concepts, and that, starting from the same query, (s)he might explore different paths of the information space.

4.3.4 Step 4: Validation of the Clusters

We tested the COPRA algorithm on different versions of the pruned graph (using different weight thresholds) in order to estimate how close they reflected user behavior in the AOL-reduced dataset and to find an optimal threshold to prune the graph. We considered the AOL-reduced sessions as the ground-truth and we computed standard accuracy measures for the evaluation.

We treated the clusters as unordered sets and we did not care about the observed order of exploration of concepts in the sessions. The reason for this is the fact that, after each search query, we plan to suggest concepts as a set of selectable items, in a multi-choice box. This set is a projection on the information categories managed by the system, based on the search context.

For each threshold, we tested the accuracy of the clusters by applying 10-fold cross-validation, after having randomly distributed the sessions of the dataset on folders. We used 90% of the sessions as learning set (to extract the clusters) and 10% as test set (to test the extracted clusters from the training set).

Given a session $S = \{Q_1, \dots, Q_i, \dots, Q_z\}$ that contains z queries and each query contains a set of concepts, we can infer that $S = \{c_{s_1}, \dots, c_{s_n}\}$ is a set of concepts. Thus, given a session $S = \{c_{s_1}, \dots, c_{s_n}\}$ and a cluster $CL = \{c_{cl_1}, \dots, c_{cl_m}\}$, we evaluated the following measures:

	Eval1	Eval2
Precision	0.659	0.626
Recall	0.794	0.794
F1	0.720	0.700

Tab. 4.5: Evaluation of the clusters accuracy.

- Precision, describing the rate of concepts of CL that also occur in S. These concepts represent correct predictions if CL is used for query expansion:

$$precision = \frac{|\{c_{s_1}, \ldots, c_{s_n}\} \cap \{c_{cl_1}, \ldots, c_{cl_m}\}|}{|\{c_{cl_1}, \ldots, c_{cl_m}\}|} \tag{4.3}$$

- Recall, representing the rate of relevant concepts of S which CL contains, and thus can suggest, if used for query expansion:

$$recall = \frac{|\{c_{s_1}, \ldots, c_{s_n}\} \cap \{c_{cl_1}, \ldots, c_{cl_m}\}|}{|\{c_{s_1}, \ldots, c_{s_n}\}|} \tag{4.4}$$

- F1 score, computed as:

$$2 * \frac{precision * recall}{precision + recall} \tag{4.5}$$

Moreover, we considered two types of evaluation:

- Eval1: accuracy of the clusters that best represent search behavior within a session. The aim of this test was to measure the optimal adherence of the clusters to users' behavior, assuming to be able to identify the best cluster from the observed queries. In this case, for each session S, we computed the precision, recall and F1 score of all the clusters. Then, we selected the accuracy of the best performing one as the representative value for S. Finally, we computed the mean accuracy of such best clusters.
- Eval2: mean accuracy of the clusters that include at least one concept referred in the search session. In this case, we aimed at testing the accuracy of a broader set of clusters, related to search sessions in a looser way. As above, we computed the mean precision, recall and F1 score.

Table 4.5 reports the accuracy values we obtained using the threshold that maximizes the mean F1 ($threshold = 2200$) for pruning the concept co-occurrence graph:

- The precision values show that, during the search sessions, users refer to a substantial portion of the concepts included in the clusters. However, the clusters contain some concepts that are not explored: by using all the clusters that have at least one concept in common with the session, we obtain a precision of about 0.626.
- The recall shows that the suggested clusters largely cover the search sessions: in both evaluations, the recall is about 0.794. Thus, the clusters help propose relevant concepts.

Overall, we can say that, by selecting clusters that have some concepts in common with those referred in the observed portion of a search session, we have good chances to suggest types of information that the user will be interested in. These results helped us in the identification of the query expansion strategies described in the following section.

In the computation of these measures, we did not take into account the occurrence of ambiguities in the interpretation of search queries. Basically, we considered all the concepts referred (either ambiguously or non ambiguously) in the queries as concepts belonging to the search sessions. This might introduce some noise in the evaluation results (e.g., it could increment the number of concepts relevant to the search sessions, reducing the precision results). However, we could not retrieve more precise information about the search interests of the AOL users because, as already discussed, most queries of the dataset are short, and the dataset provides little information to disambiguate them.

The COPRA algorithm, applied to the concept co-occurrence graph pruned with threshold 2200, returned 23 clusters having a minimum and a maximum number of concepts equal to 1 and 6, respectively.

Three sample clusters are:

1. {Play Area, Sport Area, Kindergarten, Law Enforcement, Hospital}.
2. {National Park, Provincial Park, Regional Park, Urban Park, Furnished Green}.
3. {Play Area, Sport Area, Library, Childcare service, Law Enforcement, School}.

4.4 Session-based query expansion

Let's consider a search session $S = \{Q_1, \ldots, Q_i, \ldots, Q_n\}$, and suppose that we have observed the first i queries of s. We denote the sets of concepts identified by interpreting the queries of $S@i$ as $C@i = \{C_1, \ldots, C_i\}$ where each component C_k contains a set of concepts $C_k = \{c_{k_1}, \ldots, c_{k_m}\}$ extracted from query Q_k.
Thus, it can be represented as $C@i = \{c1, \ldots, \{c_{k_1}, \ldots, c_{k_m}\}, \ldots, \{c_{t_1}, \ldots, c_{t_s}\}, \ldots, c_n\}$. The identification of the concepts referred in the queries has been done using the approach described in (Ardissono et al., 2016; Mauro and Ardissono, 2017a).

In $C@i$, the ambiguities in concept identification are represented by including the tuples of ambiguous concepts in subsets of $C@i$. For instance, in the above example, $\{c_{k_1}, \ldots, c_{k_m}\}$ and, respectively, $\{c_{t_1}, \ldots, c_{t_s}\}$ are ambiguous. In the following, we will refer to the tuples of ambiguous concepts as *ambiguity sets*.

We consider three concepts suggestion strategies, which differ from each other in the method for selecting the clusters to be used for query expansion. We assume that they could be applied immediately after the interpretation of the first query of a session, or incrementally, in order to support interactive information search.

- **SLACK:**

  1. Select a set of clusters $\{CL_{x_1}, ..., CL_{x_y}\}$ that contain *at least* one concept c such that $c \in C@i$; i.e., c has been referenced in the observed portion of the search session, $S@i$, either ambiguously or unambiguously.

  2. Propose the concepts that belong to at least one of the selected clusters, and the user has not yet explored. For each cluster CL_{x_j}, we denote the set of concepts included in CL_{x_j} as $Sugg_{x_j}$. The set of concepts to be suggested is computed as follows:
     $Sugg@i = \{Sugg_{x_1} \cup \cdots \cup Sugg_{x_y}\} - C@i$.

  For instance, given $C@i = \{c_1, \{c_3, c_4\}\}$, and two selected clusters, $CL_1 = \{c_1, c_7\}$ and $CL_2 = \{c_2, c_3, c_5, c_8\}$, $Sugg@i = \{c_2, c_5, c_7, c_8\}$.

- **SLACK-selective:** Same as SLACK, but only the clusters that best match $C@i$ are used to compute $Sugg@i$. For our experiments, we selected the best matching cluster. The evaluation of the concept suggestion strategies, described later on, provided satisfactory accuracy results for this setting of the strategy.

We compute the degree of matching between a cluster CL and $C@i$ as the cardinality of $CL \cap C@i$ (i.e., the number of concepts they have in common), taking the ambiguity in $C@i$ into account. Specifically:

- Each concept c occurring both in CL and, as a non-ambiguous element, in $C@i$, contributes to the computation of the degree of matching with a value = 1.

- Each concept c that occurs in CL, and is part of an ambiguity set AMB in $C@i$, contributes to the computation with $\frac{1}{|AMB|}$; i.e., the cardinality of the ambiguity set mitigates its contribution to the computation of the degree of matching.

For instance, given $C@i = \{c_1, \{c_3, c_4\}, c_5\}$, the degree of matching with cluster $CL_1 = \{c_1, c_7\}$ is 1 and the one with $CL_2 = \{c_2, c_3, c_5, c_8\}$ is 1.5.

- **STRICT:**

  1. Select the set of clusters $\{CL_{x_1}, \ldots, CL_{x_y}\}$ containing *all* of the concepts c such that $c \in C@i$. These are the clusters covering the whole portion of the search session observed so far (regardless of the fact that the concepts are ambiguous or not).

  2. As above, the set of concepts suggested for query expansion is: $Sugg@i = \{Sugg_{x_1} \cup \cdots \cup Sugg_{x_y}\} - C@i$.

We did not propose a STRICT-selective strategy because STRICT generates very small candidate sets.

4.5 Evaluation of concept suggestion strategies

We tested the accuracy of our strategies by applying 10-fold cross-validation on the AOL-reduced dataset, after having randomly distributed the sessions. This means that for each fold we generated a concept co-occurence graph (using the session in the training set) from which we extracted a set of clusters and we tested the different suggested strategies on the test set. We applied 10-fold cross-validation in order to have an average behaviour of the strategies. For the evaluation, we compared the concepts suggested by the strategies to those explored by users in the search sessions (ground-truth). Specifically, for each search session, we tested the strategies after having interpreted the first query alone ($S@1$), the first two queries ($S@2$), and so forth. In each case, we computed the mean precision, recall and F1 score of

Indicator	SLACK	SLACK-selective	STRICT
Min number of candidate clusters	0	0	0
Max number of candidate clusters	**2**	N=1	2
Mean number of candidate clusters	**1.192**	1	1.188
Min number of suggested concepts	0	0	0
Max number of suggested concepts	**7**	5	**7**
Mean number of suggested concepts	**2.688**	2.309	2.681
Precision	0.614	**0.621**	0.615
Recall	**0.791**	0.783	0.790
F1	**0.692**	**0.692**	**0.692**
Success rate (%)	**51.5**	51.2	**51.5**

Tab. 4.6: Statistical measures of concept suggestion strategies applied to the first query of the search sessions ($S@1$).

the strategies by comparing the concepts of the suggested clusters to the concepts referred in the remainder of the sessions.

4.5.1 Evaluation Results

Table 4.6 summarizes the results that we obtained in the suggestion of concepts immediately after having interpreted the first query of the search sessions ($S@1$):

- SLACK: the minimum, mean, and maximum number of candidate clusters identified by this strategy are 0, 1.192, and 2, respectively. The mean number of concepts suggested for query expansion is about 2.688. This number goes up to 7 in some sessions.
- SLACK-selective with number of best matching clusters N=1. The minimum, mean and maximum number of selected clusters are 0, 1 and N=1, respectively. The strategy suggests a lower mean number of concepts: about 2.309, and it proposes at most 5 concepts.
- The STRICT strategy proposes a maximum of 2 candidate clusters per session, with about 1.188 clusters in average, and it generates at most 7 concept suggestions, with a mean value of about 2.681 concepts.

The three strategies have almost the same accuracy (see Table 4.6): the F1 measure can be approximated to 0.692 in all the cases. However, they differ from each other if we look at finer-grained measures:

- **Precision:** SLACK-selective outperforms the other two, i.e., it suggests the highest number of concepts that users are observed to explore in the remainder of the search session. The lower precision of SLACK can be explained with the fact that it selects a superset of the clusters picked by SLACK-selective, and thus can introduce more noise in the suggestions. Similarly, STRICT can select more than one cluster (covering all the concepts referred in the first query), thus incrementing noise in some cases.
- **Recall:** the best strategy is SLACK, which outperforms SLACK-selective because it selects a larger pool of clusters and thus has better chances to guess the concepts explored by users. Interestingly, however, STRICT achieves almost the same recall, having a stricter cluster selection policy. We explain this finding with the fact that, by selecting all the clusters that cover the observed query, the strategy has good chances to guess the user's exploration path, among the possible ones (remember that the concept co-occurrence clusters mirrored common search behavior in a fairly accurate way).
- **Success rate:** as proposed by Huang et al. in (Huang et al., 2003), this metric measures the percentage of sessions for which our strategies propose at least one concept that is referred in the portion of the search session to be observed. Looking at the table, we can see that all of the strategies are successful in more than 50% of cases. However, the best performing ones are SLACK and STRICT. In other words, SLACK is as good as STRICT if we aim at suggesting at least one relevant concept.

Table 4.7 shows the evaluation results after the interpretation of the first two queries of the search sessions. It can be seen that, for all the strategies, the minimum, mean and maximum number of selected clusters are constant. Moreover, the minimum, mean and maximum number of suggested concepts are constant or increase.
Regarding accuracy, we notice that for all the three strategies, the recall increases, while the precision and F1 scores decrease to about 0.6, but the SLACK-selective strategy is the most precise one. SLACK has the best recall and STRICT a rather similar one; as a result, both strategies have the best F1 value. The higher recall of these strategies suggests that the selection of a single cluster, best representing the observed portion of the search session, is not general enough to cover all the possible developments of the search session. Thus, a less restrictive approach, that

Indicator	SLACK	SLACK-selective	STRICT
Min number of candidate clusters	0	0	0
Max number of candidate clusters	**2**	N=1	2
Mean number of candidate clusters	**1.223**	1	1.192
Min number of suggested concepts	0	0	0
Max number of suggested concepts	**7**	5	**7**
Mean number of suggested concepts	**2.861**	2.446	2.799
Precision	0.594	**0.598**	0.596
Recall	**0.819**	0.801	0.818
F1	**0.689**	0.685	**0.689**
Success rate (%)	**39.7**	39.2	39.6

Tab. 4.7: Statistical measures of concept suggestion strategies applied to the first two queries of the search sessions (S@2).

makes it possible to consider more than one development of a search session (as the one adopted in both SLACK and STRICT) achieves better results.

We measured the trend for longer portions of search sessions and we discovered that it is consistent, i.e., the precision decreases, and the recall increases, when more queries are interpreted. See Figures 4.6 and 4.7, which report the values of these measures after the interpretation of 1 to 5 queries[5]. However, the three strategies behave slightly differently:

- SLACK-selective is the most precise strategy when observing the first two queries, and is always better than SLACK, but it is outperformed by STRICT after the observation of 3 queries. We explain this with the fact that, when more information about the search context is available, STRICT is able to select the clusters that best represent the user's search interests, and thus to mirror her/his behavior.
- Regarding recall, we believe that it increases when more information about a search session is available because, if more concepts are identified from the search queries, there are more chances to select good candidates and thus

[5]We did not consider any longer sequences because very few search sessions are composed of more than 5 queries.

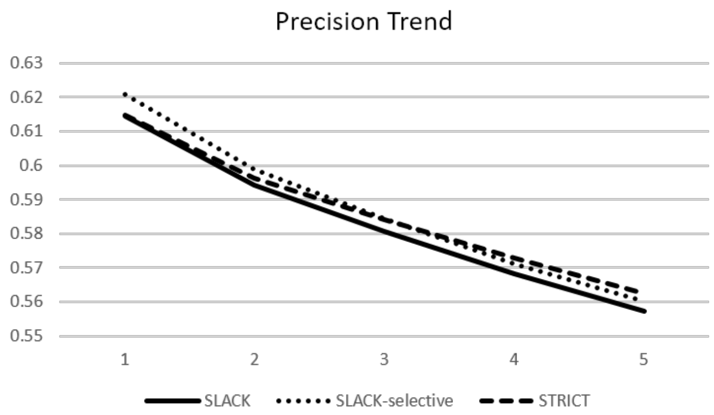


Fig. 4.6: Precision of the concept suggestion strategies in function of the number of interpreted queries ($S@i$).

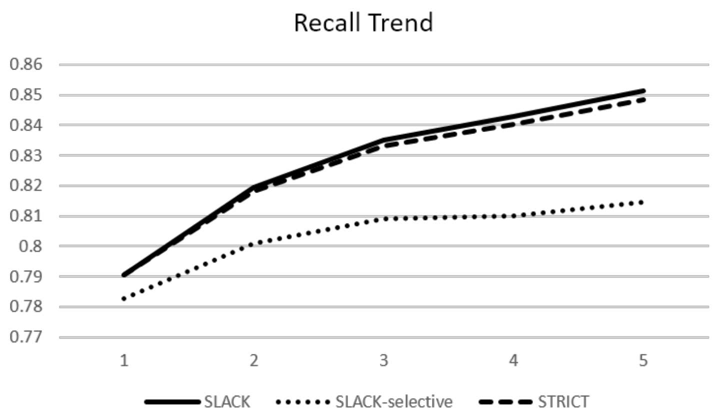


Fig. 4.7: Recall of the concept suggestion strategies in function of the number of interpreted queries ($S@i$).

to propose relevant concepts. However, the improvement of this accuracy measure varies depending on the selectivity of the strategy:

- SLACK selects the highest number of clusters, and thus suggest a larger set of concepts. This increases the chances that the suggested concepts are identified in the session.
- SLACK-selective chooses a single cluster that best matches the search session, at the expense of recall.
- STRICT selects the clusters matching the whole portion of the session observed so far. These are the best representative of the session and thus have very good changes to predict correct concepts, being used together for concept prediction.

4.5.2 Discussion

The results of our experiments enable us to positively answer our research question:

RQ1$_1$: *Can the data about the concepts frequently searched together by people within a search session be exploited to help the user explore the portions of an information space relevant to her/his information needs?*

The results show that our concept suggestion strategies have high recall, i.e., they are able to suggest most of the concepts that the users explore in the AOL-reduced dataset. However, their accuracy differs:

- The SLACK-selective strategy is the best one from the viewpoint of precision, at least when the first queries of a session are analyzed in order to suggest concepts for query expansion.
- The SLACK and the STRICT ones achieve the best recall.

Looking at these findings, we believe that the decision of which strategy best suits concept suggestion depends on the way how we want to implement this function:

- If we can suggest a larger number of concepts, then the SLACK strategy is the best one, because (i) it achieves better recall and (ii) the mean number of suggested concepts is just a bit higher than that of the other two.
- Differently, if we aim at precision, SLACK-selective is the preferred one at the beginning of a search session, while STRICT is more suitable for concepts suggestion after the first queries.

For our current work, SLACK is the most convenient strategy because it is fairly precise and, at the same time, it suggests more relevant concepts than the other ones. Specifically, we aim at providing the user with an overview of the sets of concepts to choose from, given the search context provided by the previous queries. As the user will be enabled to choose relevant concepts in a multiple-choice box, the presence of very few irrelevant concepts, as is the case of SLACK, is not a problem.

It should be noticed that other strategies might be considered. For instance, the past search history of the individual user might be analyzed to identify the concepts (s)he most frequently searched for, assuming that these represent interesting types of information. Then, this might be used to select candidate clusters for query expansion. In this work, we left these strategies apart for two main reasons:

- Firstly, as discussed by (Greenstein-Messica et al., 2017) and as introduced in Section 2.3 and in Section 2.4, session-based recommendation is of primary importance because users frequently interact with systems in an anonymous way, or there might be privacy or technical reasons for avoiding the tracking of their identities. Therefore, we decided to focus on this function before exploring other scenarios.
- Secondly, the introduction of historic information about user behavior deserves a separate research; e.g., see (White et al., 2010; Sontag et al., 2012; Smyth et al., 2005). Specifically, (Bennett et al., 2012) showed that historic information about the search behavior of an individual user helps re-ranking of search results at the beginning of a search session, while local information about the session is more effective later on, because it provides information about the actual context of search. As we deal with concept suggestion, these results cannot be immediately transferred to our case. We will thus carry out this type of analysis in a broader perspective of ontology-based personalized search support and ontology-based user models; e.g., see (Teevan et al., 2005; Jiang and Tan, 2009; Sieg et al., 2010; Gauch et al., 2003; Liu et al., 2002; Leung and D.L. Lee, 2010).

5 Impact of Semantic Granularity on Geographic Information Search Support

After the development of the session-based suggestion model of concepts we were interested to investigate the impact of the semantic granularity of the domain on the performance of the model. We thus took into account three different spatial ontologies with different characteristics and granularities to filter the search logs and analyze the performance of the model on the three filtered logs. The aim of this work was to offer to the reader a series of guidelines regarding the performance of a suggestion model considering different situations in which it is applied in domains that have different levels of granularity.

The work described in this section has been done in collaboration with University College Dublin (UCD) and University of Genova and has been published in:

- Mauro, N., Ardissono, L., Di Rocco, L., Bertolotto, M., and Guerrini, G. (2019a). "Impact of Semantic Granularity on Geographic Information Search Support". In: *Proceedings of 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. Santiago, Chile: IEEE, pp. 323–328.

5.1 Introduction

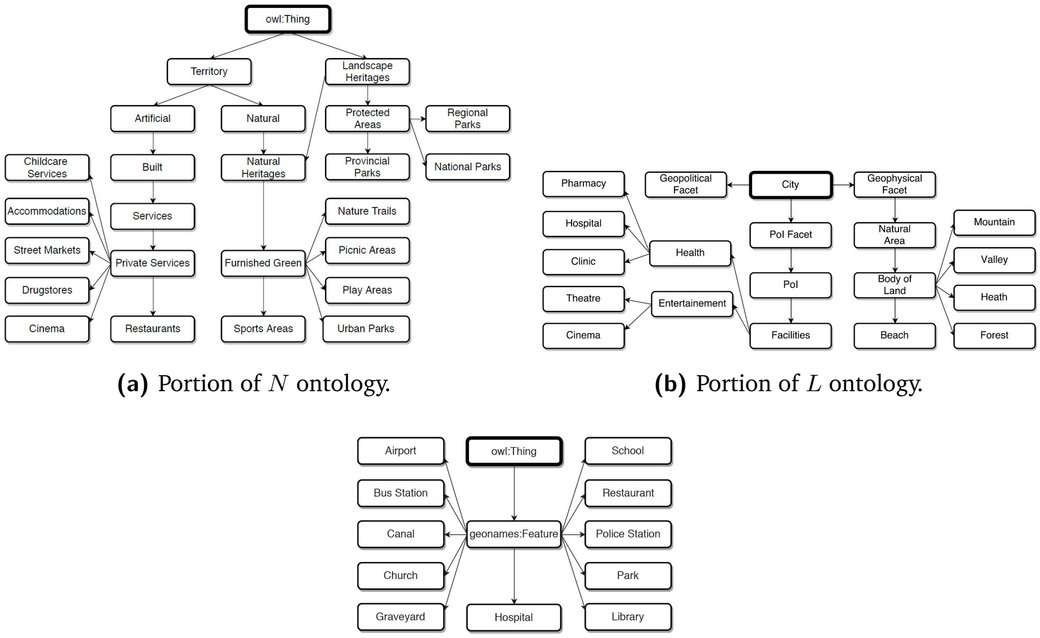
Several researchers have used semantic knowledge to support query expansion and reformulation in information search support; e.g., (Wang et al., 2017). Moreover, in Ontology-Driven Geographic Information Systems, abstraction has been analyzed to understand its "potential for information retrieval at different levels of granularity" (Fonseca et al., 2002b). However, to the best of our knowledge, the impact of different domain conceptualizations on the accuracy of information search support has not been fully investigated yet. We focus on Geographical Information Retrieval, and aim at studying the influence of semantic granularity (i.e., the degree of specificity in concept representation) on query expansion. Specifically, we aim at measuring the extent to which, by combining general information about search behavior with a more generic, or a more detailed domain conceptualization, an automated system can learn regularities useful in the suggestion of concepts relevant to the user's information needs. For instance, by analyzing general search behavior in a fine-grained domain conceptualization we might discover that, if a user looks for kindergartens in a town, (s)he might also be interested in information related to other children's activities, such as play or sport areas. While, in a coarse-grained domain conceptualization we might find out that looking for kindergartens is related to be interested in parks that is more general than play or sport areas.

Notice that our aim is not only to measure the precision of the system's suggestions, but also to define a notion of "richness" based on the number of relevant suggested concepts (i.e. recall and mean number of relevant suggested concepts), as this is important for catalog exploration. Specifically, we investigate the following research questions:

$\mathbf{RQ1_2}$: *What is the relationship between the semantic granularity of a domain conceptualization and the capability of suggesting types of information relevant to the user's needs during an exploratory search task?*

In order to answer these questions, we need to know which concepts people frequently focus on during a search session, as a source of evidence of co-occurring information needs. Thus, we apply the session-based concept suggestion model presented in Chapter 4 to extract concepts co-occurrence, and we compare its performance by using different ontologies to train it and to interpret the search queries: a fine-grained ontology (see Figure 5.1a), a less detailed one (5.1b) developed for OpenStreetMap (`www.openstreetmap.org`) and the GeoNames Mappings ontology (see Figure 5.1c - `www.geonames.org/ontology/mappings_v3.01.rdf`). The results of an experiment based on a large query log reveal that a finer-grained semantic granularity improves recall and richness of results: with the first ontology the concept suggestion model achieves the best recall and supports the generation of more suggestions than with the other two. However, the model achieves the best precision and accuracy (F1) using the second ontology, which is based on crowdsourced data.

In the following, Sections 5.2 and 5.3 present the ontologies and the concept suggestion model we employed. Section 5.4 describes our empirical evaluation.



(a) Portion of N ontology.

(b) Portion of L ontology.

(c) Portion of $GeoNamesMappings$ ontology.

Fig. 5.1: Ontologies selected for the empirical evaluation. The root nodes of the ontologies have thick borders for readability purposes.

5.2 The Used Ontologies

In this section we empirically analyze and compare the three geographical ontologies used for the experiments. We are interested in classifying them from the finest-grained one to the coarser-grained one. We focus on:

1. Semantic granularity, i.e., the level of detail at which geographic objects are described (Fonseca et al., 2002b). This is different from spatial granularity, which refers to the granularity of geographic toponyms; e.g., see (Palacio et al., 2015).

2. Ontology structure, quantified in terms of number of concepts and number of subclass relations.

***N* ontology** (see Figure 5.1a) contains fine-grained concepts about cities (Voghera et al., 2016). This ontology is the one underlying the OnToMap web application. This ontology provides a multi-faceted specification of classes and relations that characterize the information space considering three high-level aspects: natural (e.g., parks), artificial (e.g., infrastructures) and normative (e.g., administrative boundaries). These aspects are specialized in classes at more than one level of detail.

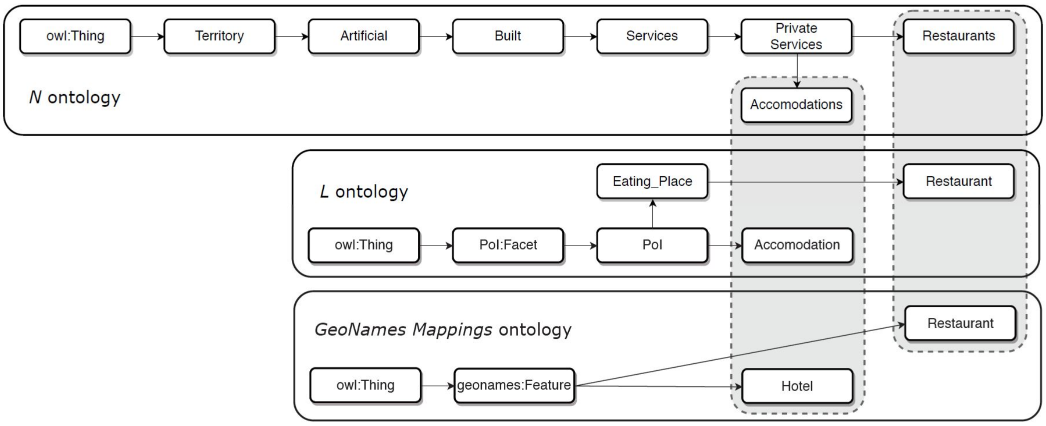


Fig. 5.2: Portion of the three ontologies with an example of shared concepts (in gray areas).

Metric	***N***	***L***	**GeoNames**
Total number of classes	195	97	150
Total number of subclass relations	268	94	155
Longest path to leaves (graph nodes)	6	5	3
Number of used classes	72	70	136
Number of used subclass relations	81	69	135
Longest path to used leaves (graph nodes)	6	5	2
Mean degree of used classes (graph nodes)	4.5	3.63	67.5

Tab. 5.1: Structural comparison of the GeoNames, L and N ontologies.

For instance, as partially reported in the figure, Landscape Heritage is specified in a deep hierarchy of classes in order to distinguish various types of Furnished Green from Protected Areas. The longest path between owl:Thing and the leaf classes has length = 6.

***L* ontology** (see Figure 5.1b) structures OpenStreetMap tags that refer to geographic concepts in the context of cities (Di Rocco, 2016). It is multi-faceted and has a similar structure, but coarser granularity, than N; moreover, being derived from the largely-used OpenStreetMap repository, it reflects a conceptualization that comes from the general public. The geographic information about a city is described using three facets: Point of Interest (PoI), geoPolitical and geoPhysical. The longest path between the owl:Thing class and the leaf classes of L has length = 5.

GeoNames Mappings ontology (see Figure 5.1c) maps the GeoNames taxonomy to external ontologies, such as schema.org and linkedgeodata.org. GeoNames derives from the geonames.org database and describes more than 9,000,000 features categorized in 645 alphanumeric feature codes. The Mappings ontology maps 108 of its codes and describes knowledge at a coarse semantic granularity: information is organized in a flat structure including a top-level OWL class (owl:Thing), and a child class (geonames:Feature) that has all the GeoNames codes as its children. Only two subclasses of geonames:Feature, representing toponyms, have children themselves. Thus, even though the longest path between owl:Thing and the leaves of the ontology has length = 3, the path from owl:Thing to most classes has length = 2.

As a reference for the comparison among ontologies, we selected GeoNames Mappings (henceforth, GeoNames for brevity), which is largely adopted in the geospatial community. In each ontology, we only considered geospatial classes related to natural features of the territory (e.g, rivers, mountains) and artificial features (e.g., malls, buildings, streets). We excluded administrative boundaries, which concern spatial, rather than semantic, granularity and are abstract concepts. The ontologies used in our analysis are thus subsets of the original ones.

Table 5.1 shows a set of metrics to describe the complete ontologies, and the subsets focused on natural and artificial features. As shown in row "Number of used classes", the used subset of GeoNames is the largest one, followed by the N and the L ones. The three ontologies share 8 concepts; GeoNames and L share 23 ones; GeoNames and N share 17 ones.

The subsets of N, L and GeoNames have different semantic granularity. A rough measure of this can be obtained by jointly considering the length of the longest path between owl:Thing and the leaves of an ontology, and the mean degree of the nodes of the ontology graph. Among the three, N has the maximum longest path (length = 6) and its mean degree is 4.5, which denotes the fact that several classes are specified into a relatively small set of subclasses. L has few subclasses per node as well (mean degree = 3.63) but it defines concepts at a slightly more superficial level (length of longest path = 5). Finally, GeoNames is flat (length = 2) and has a very high mean node degree (67.5). We report some examples to compare the ontologies. Class Park in GeoNames is represented in N as Regional_Park, Provincial_Park and National_Park. Therefore, we can say that Park is shared between N and GeoNames but they differ in granularity and thus the two representations can support concept identification at different levels of detail. Similarly, Shopping of L corresponds to Shopping_mall, Market and Shop in N. Finally, in N, concept Restaurant is associated with a much longer and descriptive path than in GeoNames. The corresponding path in L is descriptive but not as exhaustive as in N; see Figure 5.2.

5.3 Concept Suggestion Model (CS Model)

In order to compare suggestion accuracy using different reference ontologies, we used the Concept Suggestion Model described in Chapter 4. This model suits our needs because it extracts concept co-occurrence (w.r.t. term, or query co-occurrence, returned by other algorithms) from the search sessions of a query log, and thus enables us to identify information needs from a semantic point of view.

5.3.1 Creation of Concept Co-occurrence Clusters

The process to create the concept co-occurrence clusters, starting from the query log and the ontology conceptualizing the domain knowledge, is the one described in Section 4.3.1. The output of this phase is a set of clusters, each one consisting of an unordered set $CL = \{c_1, \ldots, c_n\}$ of concepts c_x that highly co-occur in the sessions of the reduced dataset.

5.3.2 Concept Suggestion Strategies

In order to evaluate the impact of granularity on concept suggestion we analyze the performance of the strategies SLACK, SLACK-SELECTIVE and STRICT. Specifically, given a search session $S = \{Q_1, \ldots, Q_i, \ldots, Q_n\}$, let $S@i$ be the observed portion of S, i.e., $\{Q_1, \ldots, Q_i\}$. Moreover, let $C@i = \{c_1, \ldots, c_k\}$ be the set of concepts identified by interpreting the queries of $S@i$. We aim at generating suggestions immediately after the first query of a search session; thus, we apply the strategies to $S@1$.

5.3.3 AOL-reduced Datasets

We used the AOL query log as a source of search sessions. To compare the impact of domain conceptualization on concept suggestion, we reduced the AOL query log using the N, L and GeoNames ontologies. We generated three *AOL-reduced* datasets (AOL_N, AOL_L, AOL_{GN}) by filtering the sessions of the AOL log according to the concepts defined in each ontology; see Section 5.3.1. For this task, we pre-processed the ontologies by automatically annotating their classes with linguistic knowledge (synonyms and linguistic definitions), using BabelFy multilingual Entity Linking and Word Sense Disambiguation service (Moro et al., 2014), and Stanford CoreNLP lemmatizer (Manning et al., 2014).

Measure	AOL$_N$	AOL$_L$	AOL$_{GN}$
Number of queries	1,581,817	1,486,122	1,443,448
Number of sessions	945,945	911,399	864,869
Number of users	247,868	240,474	232,931

Tab. 5.2: Size of *AOL-reduced* datasets.

Measure	AOL$_N$	AOL$_L$	AOL$_{GN}$
Min number of queries per session	1	1	1
Max number of queries per session	93	96	103
Mean number of queries per session	1.67	1.63	1.67
Median number of queries per session	1	1	1
Standard deviation	1.45	1.38	1.44

Tab. 5.3: Length of the search sessions of the *AOL-reduced* datasets.

As shown in Table 5.2, the AOL$_N$ dataset contains the highest number of queries, sessions and users. AOL$_L$ contains the second highest number of queries, sessions and users, and AOL$_{GN}$ is the smallest dataset, even though the GeoNames ontology is larger than the other two. We explain this finding using semantic granularity: N is the most detailed ontology and supports the selection of the largest number of queries from the log. However, in GeoNames, the coarse-grained classes limit the identification of concepts in search queries, and make the selection of relevant sessions more constrained than using L. The datasets have some similarities. As reported in Table 5.3, the mean length of their sessions is almost the same: 1.67 queries in AOL$_N$ and AOL$_{GN}$, 1.63 in AOL$_L$. Moreover, the distribution of sessions w.r.t. their length follows a Power Law (most sessions are short, few contain many queries). Thus, we can use them to compare the concept suggestion strategies.

5.3.4 Concept Co-occurrence Clusters

For each ontology X, and AOL-reduced dataset AOL_X, we created a concept co-occurrence graph G_X and from each graph G_X we extracted a set of clusters. The algorithm returned 23 clusters for AOL$_N$, 19 for AOL$_L$ and 23 for AOL$_{GN}$. Table 5.4 shows two sample clusters for each dataset.

The clusters generated from AOL$_N$ contain a larger number of concepts than the others (about 21% more than those produced using AOL$_L$, and about 31% more than those derived from AOL$_{GN}$). This finding is in line with the granularity of the

Dataset	Sample clusters
AOL$_N$	{Play Areas, Sport Areas, Kindergartens, Law Enforcement, Hospitals} {Play Areas, Sport Areas, Libraries, Childcare services, Law Enforcement, Schools}
AOL$_L$	{Educational_institution, Kindergarten, University} {Bar, Cafe, Eating_place, Pub, Restaurant}
AOL$_{GN}$	{Library, Museum, Police_Station, University} {Beach, Resort}

Tab. 5.4: Sample concept co-occurrence clusters.

ontologies: being more specific, N makes it possible to identify a more diverse set of co-occurring concepts in the search sessions. In contrast, GeoNames supports the identification of few concepts, and thus produces small clusters.

The clusters generated using AOL$_L$ are more topic-centered than the other ones: the concepts of each AOL$_L$ cluster have the same parent concept in the L ontology. As L is fairly fine-grained, it describes different topics as non-leaf concepts of the ontology. Thus, having the same parent means being strictly related from a semantic point of view. Indeed, AOL$_N$ generates a few topic-centered clusters, as well. However, most of its clusters include semantically distant concepts that typically descend from different high-level concepts of the ontology. For instance, in Table 5.4, Play Areas and Sport Areas descend from Natural, a high-level class of the N ontology. The other concepts of the cluster descend from Artificial (another high-level class). Finally, the clusters generated from AOL$_{GN}$ cover a broad range of topics, given the flat structure of the GeoNames ontology.

5.4 Empirical Evaluation

We measured the performance of the CS Model on the AOL-reduced datasets, using the respective concept co-occurrence clusters for concept suggestion. For each dataset, we measured the richness and the accuracy of the suggestions that the SLACK, SLACK-SELECTIVE and STRICT strategies generated, by matching the concepts they proposed, given the first query of a session S (S@1), against those referenced in the remainder of S; i.e., those actually explored by users. For the evaluation we applied 10-fold cross-validation on each dataset, after having randomly

	AOL$_N$	*AOL*$_L$	*AOL*$_{GN}$
Min number of suggested concepts	0	0	0
Max number of suggested concepts	**7**	4	4
Mean number of suggested concepts	**2.688**	0.496	1.371
Recall	**0.791**	0.783	0.768
Precision	0.614	**0.882**	0.712
F1	0.691	**0.829**	0.739

(a) SLACK performance (S@1).

	AOL$_N$	*AOL*$_L$	*AOL*$_{GN}$
Min number of suggested concepts	0	0	0
Max number of suggested concepts	**5**	4	3
Mean number of suggested concepts	**2.309**	0.455	1.054
Recall	**0.783**	0.759	0.761
Precision	0.621	**0.875**	0.725
F1	0.692	**0.813**	0.743

(b) SLACK-SELECTIVE performance (S@1).

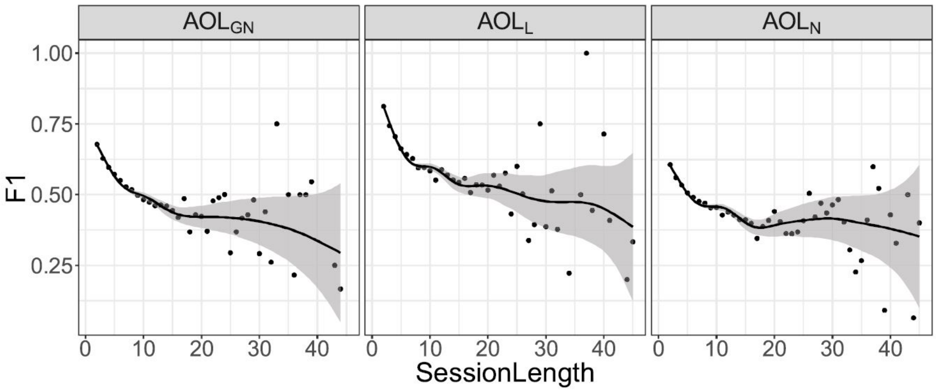
Tab. 5.5: Performance of the concept suggestion strategies applied to the first query (S@1) of the sessions of the *AOL-reduced* datasets.

distributed its sessions on folders. We used 90% of the sessions as learning set and 10% as test set.

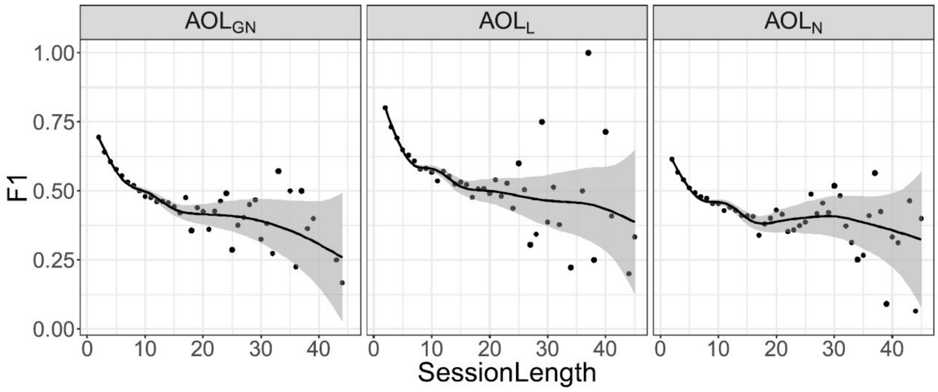
Table 5.5 presents the evaluation results obtained by applying SLACK and SLACK-SELECTIVE to the *AOL-reduced* datasets, considering only the sessions that include at least two queries. We omit the results of STRICT due to space limitations.

Henceforth, C_S denotes the set of concepts explored by the user in a session S; $C_{Sugg_{S@i}}$ is the set of concepts suggested by a strategy after analyzing the $S@i$ portion ($i = 1$).

Richness of suggestions. The first 3 lines of Tables 5.5a and 5.5b compare the numbers of relevant concepts suggested by the strategies, per dataset. These numbers measure the suggested concepts that are explored by the users, and thus represent useful suggestions. It can be seen that SLACK recommends more relevant concepts than SLACK-SELECTIVE. However, both strategies obtain the highest values when they are applied to AOL$_N$, which is the most detailed ontology.



(a) SLACK



(b) SLACK-SELECTIVE

Fig. 5.3: Distribution of F1 score with respect to the length of sessions.

Recall ($R@i$). This measure represents the percentage of concepts explored in search sessions that are suggested by the strategies. It represents the system capability to cover the topics that users actually explore in their searches. Tables 5.5a and 5.5b show that strategies achieve the best $R@1$ with AOL$_N$.

Precision ($P@i$) represents the percentage of suggested concepts that are explored in the remainder of the sessions. Tables 5.5a and 5.5b show that the strategies achieve the best $P@1$ with AOL$_L$, and this value is much higher than those obtained on the other datasets.

Accuracy ($F1@i$) integrates precision and recall. Both strategies obtain the highest $F1@1$ score on AOL$_L$. Figure 5.3 shows the variation of $F1@1$ w.r.t. the length of search sessions (length $>= 2$). The points denote the mean value of F1 for each length value. The black lines represent the local regression on this data and show the trend of F1: at first, it decreases; then the regression line stabilizes. For longer sessions, the values of F1 do not correspond to any trend. Thus, the confidence of the regression line (in gray) decreases.

The results of the evaluation contribute to answer our research question $\mathbf{RQ1_2}$. Results support the hypothesis of a dependency between the semantic granularity of a domain conceptualization and the accuracy in the suggestion of concepts relevant to a user search. Specifically, a finer-grained semantic granularity enhances recall by supporting the generation of larger concept co-occurrence clusters that have a broad scope (they can refer to semantically distant concepts of the ontology). Thus, given a set of concepts identified in the observed query, the selection of one or more clusters that match it provides a larger pool of concepts to be suggested.

Precision increases when using an ontology that covers a broad range of concepts, like GeoNames. However, it depends on how close the ontology is to the way people conceptualize and refer to geographic information. The evaluation shows that the suggestions obtained with the crowdsourced L ontology, which is more specific than GeoNames, but less than N, are more precise than those obtained with the other ontologies.

Moreover, the results support the hypothesis that the semantic granularity of a domain conceptualization impacts on the richness of concept suggestions. A finer-grained ontology supports the suggestion of a larger set of relevant concepts because it is able to match a larger variety of concepts in the query logs. Moreover, it is associated with the largest clusters; thus, given a search context, it has higher suggestion capability.

6 Extending a Tag-based Collaborative Recommender with Co-occurring Information Interests

This chapter describes a work that aims to integrate information coming from search sessions with rating information in order to suggest objects at a fine-grained granularity (Point of Interests) leveraging behavior patterns about topics exploration. Specifically, we propose the *Extended Category-based Collaborative Filtering (ECCF)* recommender, which enriches category-based user profiles derived from the analysis of rating behavior with data categories that are frequently searched together by people in search sessions. The categories co-occurrences in search sessions are identified with the algorithm used to build the concept co-occurence graph described in Chapter 4.

This work has been published in:

- Mauro, N. and Ardissono, L. (2019). "Extending a Tag-based Collaborative Recommender with Co-occurring Information Interests". In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '19. Larnaca, Cyprus: ACM, pp. 181–190.

6.1 Introduction

Recommender systems research has employed item ratings, bookmarking actions and other user activities as primary sources of information to generate personalized suggestions because they provide evidence about user preferences. In particular, User-to-User Collaborative Filtering (Desrosiers and Karypis, 2011) (henceforth, denoted as U2UCF) analyzes the ratings of items provided by users in order to identify "like-minded" people for preference prediction. However, the sparsity of the rating matrices affects recommendation performance. Thus, recent algorithms have been proposed to improve the recognition of preference similarity from rating data (e.g., Matrix Factorization algorithms (Koren and Bell, 2011) such as SVD++ (Koren, 2008)), possibly combined with trust information derived from the establishment of social links among users; e.g., (Tang et al., 2013; Yang et al., 2017). While these algorithms achieve good accuracy and coverage, they challenge the explanation of recommendation results because the policies applied to rank items can hardly be described in an intuitive way.

In the present work, we are interested in assessing whether U2UCF, which has nice explanation properties, can be improved by using other types of information that are complementary to rating data. Specifically, we investigate whether the identification of frequently co-occurring interests in information search can be used to improve recommendation performance. We start from the observation that, if the people who search for items tagged with a certain information category typically also search for items tagged with another category, the two categories might represent related interests. Therefore, even though we ignore the reasons behind this relatedness, we might leverage the strength of the association in preference estimation. In this perspective, we propose to build rich user profiles by extending the preferences for categories of items identified from rating behavior with frequently co-occurring interests for item categories, extracted from the logs of search engines. It can be noticed that interest co-occurrence can be learned by analyzing anonymous interaction sessions because it is aimed at describing general user behavior. Therefore, it can be applied to anonymized search logs, as long as search sessions can be identified. It should be notice that we combine different source of information to improve the recommendation performance. However, our work is different from hybrid recommender systems in which several recommendation models are combined.

Starting from a category-based representation of user preferences, based on the analysis of ratings and on items categorization, we propose the following research question:

$\mathbf{RQ1_3}$: *How does the integration of data about interest co-occurrence in information search influence the performance of a collaborative recommender system that manages category-based user profiles?*

In order to answer this question, we start from a *Simple Category-based Collaborative Filtering (SCCF)* algorithm which infers a user's preferences on the basis of the distribution of her/his ratings on item categories: a category-based user profile provides a conceptual view on preferences, so that user similarity can be computed by abstracting from item ratings, thus contrasting data sparsity; see (Sieg et al., 2007; Sieg et al., 2010). Then, we propose the *Extended Category-based Collaborative Filtering (ECCF)* algorithm that enriches category-based user profiles with evidence about interests that frequently co-occur in information search. ECCF employs the extended user profiles for rating estimation.

In order to evaluate the recommendation performance of ECCF, we extract information about co-occurring interests by analyzing the query log of a largely used search engine. Then, we test our algorithm by applying it to the Yelp Dataset (Yelp, 2019b), which stores user ratings of various types of businesses.

We analyze a few settings of ECCF in order to integrate different amounts of information about co-occurring preferences with rating data. In our experiments, we evaluate performance by taking U2UCF and SCCF as baselines: these algorithms differ in neighbor identification but are based on the same rating estimation approach. Therefore, they are a good basis to assess the impact of extended category-based user profiles on preference prediction. We also compare these algorithms with SVD++ to evaluate whether preference extension challenges the capability of recommending relevant items. The results of our experiments show that ECCS outperforms U2UCF and SCCF in accuracy, MRR, diversity of recommendations and user coverage; moreover, it outperforms SVD++ in accuracy and diversity of the generated suggestion lists. We thus conclude that preference co-occurrence information can positively contribute to the identification of good neighbors for rating estimation.

In summary, the main contributions of this work are:

- The integration of data about frequently co-occurring information interests (inferred by observing general search behavior) with category-based user preferences, in order to acquire rich individual user profiles.
- The ECCF category-based recommendation algorithm, which extends User-to-User Collaborative Filtering to take both frequently co-occurring information interests and preference similarity into account in neighbor identification.
- Evaluation results aimed at proving the benefits of frequently co-occurring interests to Collaborative Filtering.

In the following, Section 6.2 presents ECCF. Section 6.3 describes the experiments we carried out to validate ECCF and discusses the evaluation results.

6.2 Extended Category-based Collaborative Filtering

We describe ECCF incrementally, starting from U2UCF that provides the basic matchmaking approach for rating estimation.

6.2.1 User-to-User Collaborative Filtering

In (Ricci et al., 2011), they define U2UCF as follows: “the simplest and original implementation of this approach recommends to the active user the items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of the users". Given:

- U as the set of users and I as the set of items;
- $r : UXI \Rightarrow \mathbb{R}$ as a map of ratings;
- $R \in \mathbb{R}^{UXI}$ as the users-items rating matrix, where each value is a rating $r_{ui} = R[u, i]$ given by a user $u \in U$ to an item $i \in I$.

The recommender system estimates u's rating of i ($\hat{r}_{ui}$) as follows:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum\limits_{v \in N_i(u)} \sigma(u, v)(r_{vi} - \bar{r}_v)}{\sum\limits_{v \in N_i(u)} |\sigma(u, v)|} \tag{6.1}$$

where $N_i(u)$ is the set of neighbors of u that rated item i and $\sigma(u, v)$ is the similarity between user u and user v ($v \in N_i(u)$). The similarity among users is computed by applying a distance metric, e.g., Cosine or Pearson similarity, to their rating vectors.

6.2.2 Simple Category-based Collaborative Filtering (SCCF)

SCCF manages user profiles in which the user's interest in each item category is represented as a positive number; the higher is the value, the stronger is the interest. We define:

- U, I, r and R as above;
- C as the set of item categories;
- $f : UXC \Rightarrow \mathbb{N}$ as a map between users and categories;
- $UC \in \mathbb{N}^{UXC}$ as the Users-Categories matrix. For each $u \in U$ and $c \in C$, $UC[u, c]$ represents the interest of u in c. We take as evidence of interest the *frequency of exploration* of a category, i.e., the frequency of interaction of the user with items associated with the category.

Category exploration can be mapped to different types of user behavior; e.g., tagging items and searching for items by tag. We map exploration to rating behavior and we define $UC[u, c]$ as the number of ratings that u has given to the items associated with c. We believe that a user can be interested in a category even if s(he) gave a negative rating to the item.

SCCF computes user similarity on the basis of the estimated user preferences for item categories. Specifically, $\sigma(u, v)$ is defined as the Cosine similarity of the users' vectors in the UC matrix and it is used in Equation (6.1) to estimate ratings. Thus, $\hat{r}_{ui}$ is computed on the basis of the ratings r_{vi} provided by the users $v \in U$ whose preferences for categories are similar to those of u.

6.2.3 Acquisition of Preferences Co-occurrence

In order to learn the strength of the associations between item categories in search behavior, we analyze their co-occurrence in the search sessions of a query log. To build the Category Co-occurrence Graph (CCG) we adopted the approach described in Section 4.3.1. The Category Co-occurrence Graph (CCG) represents category co-occurrence: in the CCG, nodes represent the data categories referenced in the analyzed queries and the weight of edges represents the co-occurrence frequency of the connected categories; i.e., how many times the categories have been identified within the same search sessions.

6.2.4 Extended Category-based Collaborative Filtering (*ECCF*)

In this recommendation model we employ frequent co-occurring information interests to extend category-based user profiles. We reinforce the preferences for item categories learned by analyzing rating behavior (stored in the Users-Categories matrix UC) with interest co-occurrence associations (stored in the CCG graph) in order to acquire an extended set of user preferences for neighbor identification.

The idea behind preference extension is that, the more the user has appreciated the items of a category, the more interest co-occurrence makes sense. Therefore, starting from the category-based user profiles stored in the UC matrix, we increment user preferences with the contribution of the strongest co-occurrence relations of the CCG graph, depending on the number of positive ratings available in the users-items matrix R. The output of this process is stored in the Extended Preferences matrix EP, which is used to compute $\sigma(u, v)$ in Equation 6.1.

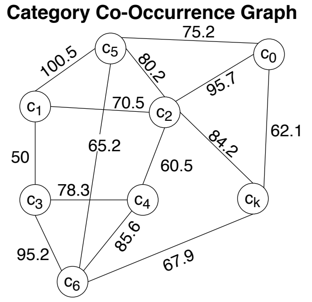
R Matrix

	i_0	i_1	i_2	i_3	i_z
u_0	2	0	0	4	1
u_1	0	3	0	0	0
u_2	0	0	1	0	0
u_3	5	0	0	3	0
u_4	2	0	4	0	0
u_5	0	1	1	4	0
u_n	0	5	3	0	0

UC Matrix

	c_0	c_1	c_2	c_3	c_4	c_5	c_k
u_0	20	0	5	15	0	0	1
u_1	0	7	23	8	3	0	0
u_2	4	0	0	16	2	10	4
u_3	16	5	0	0	4	8	3
u_4	1	0	0	0	18	5	2
u_5	13	18	0	3	0	0	0
u_n	6	3	0	0	0	0	12

Category Co-Occurrence Graph



Extended Preferences Matrix EP

	c_0	c_1	c_2	c_3	c_4	c_5	c_k
u_0	25	0	5	18	3	0	1
u_1	0	10	28	8	5	0	1
u_2	6	0	0	18	2	5	6
u_3	20	5	0	1	4	9	3
u_4	3	0	2	0	20	5	3
u_5	16	22	0	3	0	0	2
u_n	7	3	0	0	2	0	15

Fig. 6.1: Extension of Category-based User Profiles.

Figure 6.1 provides a graphical view of the computation of EP: the information stored in UC is combined with that stored in the CCG to set the values of this matrix. In this process, the users-ratings matrix R is used to limit the reinforcement of preferences to the categories of the positively rated items. Moreover, the CCG is used to propagate preference information according to the strongest co-occurrence of interests. In detail, we compute the values of EP as follows:

- let Cat_i be the set of categories associated to item i;
- let $CatSet_i$ be the set of categories directly connected to any category $c \in Cat_i$ in the CCG through the heaviest outbound arcs. These are the categories which most frequently co-occur with some categories of Cat_i in search sessions.

Then:

$$EP[u, c] = UC[u, c] + \sum_{i \in |I|} f(u, i, c) \tag{6.2}$$

where

$$f(u, i, c) = \begin{cases} 1 & \text{if } R[u, i] \in \textit{PositiveRatings} \wedge c \in CatSet_i \\ 0 & \text{otherwise} \end{cases} \tag{6.3}$$

In Equation 6.3 $PositiveRatings$ denotes the set of ratings that are considered as positive in the dataset; e.g., {5}, or {4, 5} in a [1, 5] Likert scale.

Yelp	Number of users	26,600
	Number of businesses	76,317
	Number of ratings	1,326,409
AOL	Number of sessions	1,248,803
	Number of queries	2,136,029

Tab. 6.1: Statistics about the Filtered Datasets.

6.3 Validation of ECCF

6.3.1 Dataset of Item Ratings

As a source of rating data we exploit the Yelp Dataset (Yelp, 2019b), which contains information about a set of businesses, users and reviews and is available for academic purposes. In the dataset, item ratings take values in a [1, 5] Likert scale where 1 is the worst value and 5 is the best one. Moreover, each item is associated with a list of categories describing the kind of service it offers.

The full list of Yelp categories[1] is organized in a taxonomy to specify businesses at different levels of detail. The taxonomy includes a large set of first-level categories, representing broad types of businesses; e.g., "Active life", "Arts & entertainment", "Automotive", . . . , "Food", "Restaurants", and many others. In turn, the first-level categories are specialized into sub-categories; e.g., "Restaurants" includes many types of restaurants such as "Indian", "Chinese" and the like. We apply two filters to the dataset:

1. We select all the Yelp categories that are subclasses of "Restaurants" or "Food": e.g., "Indian", "Chinese", "Cafes", "Kebab", "Pizza", and so forth; the total number of categories is 254. Then, we project the Yelp dataset on the set of items associated with at least one of these categories. In the rest of this chapter we refer to this set of categories as *CATS*.
2. We further filter the dataset on the users who rated at least 20 items.

The higher portion of Table 6.1 summarizes the number of users, businesses and ratings of the filtered Yelp dataset.

[1] `www.yelp.com/developers/documentation/v3/category_list`

6.3.2 Dataset of Search Sessions

For the generation of the Category Co-occurrence Graph we use the AOL query log. Each line of the log represents either a query or a click-through event on one of the search results of a query. The line contains various fields, among which the submitted query and the submission date and hour.

In order to build a graph that is thematically related to the items of the filtered Yelp dataset, we select from the log the search sessions relevant to the categories $c \in CATS$ enriched with the following two types of external knowledge. The enrichment is useful to abstract from the specific category names used in Yelp and to take into account semantically related information:

1. *Lemmatized knowledge:* we enrich each element $c \in CATS$ with a set of keywords and synonyms from WordNet (WordNet, 2017) lexical database.
2. *Relevant terms from the Probase (Wu et al., 2012) taxonomy:*
   - For each element $c \in CATS$, we enrich c with the $< concept, instance >$ pairs of ProBase such that $concept$ has at least 85% WordNet similarity with any term of the lemmatized knowledge of c, and the WordNet similarity between the two components of the pair is 85%.
   - ProBase, recently called Microsoft Concept Graph, is a large concept network harnessed from web pages and search logs. It is organized as a list of $< instance, concept >$ pairs related by a subclass relation and it contains 5,376,526 classes and 12,501,527 instances.

For the selection of relevant search queries in the AOL log we match the lemmatized words occurring in the queries to the enriched categories of $CATS$. If there is at least one match between a term and a query, we consider the query as relevant and we include its parent session in the filtered log.

6.3.3 Category Co-occurrence Graph

We instantiate the *CCG* with the interests that co-occur in the sessions of the filtered AOL dataset by applying the procedure described in Section 6.2.3. The resulting graph is strongly connected: almost all of the categories are linked to each other by an edge having weight > 0. However, the distribution of weights in the graph shows that there is a large number of weakly connected categories and a very small number of strongly associated ones. The "heavy" edges identify the interests that co-occur very frequently in search sessions and suggest to select the arcs having maximum weight in the *CCG* for the extension of the user profiles, as done in Section 6.2.4.

6.3.4 Test Methodology

We evaluate the recommendation performance of ECCF by comparing it to U2UCF and SCCF, which we consider as baselines. Moreover, we compare these algorithms with SVD++ in order to assess the improvement in the suggestion of relevant items given by frequently co-occurring interests.

The SCCF and ECCF recommendation algorithms are developed by extending the Surprise library (Hug, 2017), while we use the default Surprise implementations of U2UCF and SVD++.

We test the algorithms by applying a 10-fold cross-validation on the filtered Yelp dataset, after having randomly distributed ratings on folds: we use 90% of the ratings as training set and 10% as test set. In all the tests, we configure the KNN algorithms to work with 50 neighbors.

In order to analyze the impact on recommendation performance of a looser, or stricter extension of user preferences with category co-occurrence, we validate ECCF on different settings of $PositiveRatings$ in Equation 6.3, i.e., on different interpretations of what is a good rating. For each fold we generate three versions of the Extended Preferences matrix EP having set $PositiveRatings$ to $\{3, 4, 5\}$, $\{4, 5\}$, and $\{5\}$ respectively.

We evaluate Top-k recommendation performance with k=10 and k=20 by taking the ratings observed in the Yelp dataset as ground truth. For the evaluation we consider the following metrics: Precision, Recall, F1, RMSE, MRR, Diversity and User Coverage.

Diversity describes the mean intra-list diversity of items in the suggestion lists @k; see (Bradley and Smyth, 2001). In this work, we interpret diversity from the viewpoint of item classification. Therefore, we measure the diversity of a recommendation list as follows:

$$\text{intra-list diversity@k} = \frac{\sum_{i=1}^{k}\sum_{j=i}^{k}(1 - sim(i,j))}{\frac{k*(k+1)}{2}} \tag{6.4}$$

where $sim(i, j)$ is the cosine similarity between the lists of categories associated to items i and j in the ratings dataset.

Metrics	U2UCF	SCCF	ECCF {3,4,5}	ECCF {4,5}	ECCF {5}
Precision	~~0.7823~~	**0.786**	0.7857	0.7855	0.7859
Recall	~~0.7473~~	0.7526	0.7536	**0.755**	0.7529
F1	~~0.7644~~	0.7689	0.7693	**0.7699**	0.769
RMSE	~~1.0001~~	0.9899	0.9897	0.9893	**0.9892**
MRR	~~0.733~~	0.7367	0.737	**0.7391**	0.7384
Diversity	~~0.3042~~	0.3053	**0.3056**	0.3053	0.3049
User cov.	~~0.8497~~	0.8521	0.8526	**0.8542**	0.8534

Tab. 6.2: Performance evaluation @10; the best values are in boldface, the worst ones are strikethrough.

6.3.5 Results

Table 6.2 shows the performance results of the KNN recommenders we compared, by taking into account a maximum of 10 suggested items (performance@10).

- **Precision:** similar to previous results described in (Sieg et al., 2007), all of the category-based recommenders outperform U2UCF. This can be explained by the fact that the matrices describing preferences for item categories are denser than the ratings one. Thus, they improve recommendations by supporting a better identification of neighbors for Equation 6.1. However, SCCF outperforms all of the ECCF variants. The second best recommender is ECCF$\{5\}$ that extends user profiles in the strictest way: it only considers as pivots for extension the categories associated to the items that the user has rated 5 stars. Notice also that the precision of ECCF decreases when $PositiveRatings$ is lax. The reason is that the extension of user profiles with frequently co-occurring interests can increase the estimated interest in some noisy categories with respect to the pure observation of ratings distribution on categories. In particular, noise grows when the policy applied to extend preferences is less restrictive.
- **Recall:** ECCF outperforms the baselines in all the settings of $PositiveRatings$. Specifically, ECCF{4,5} achieves the best result, while recall is lower in ECCF{3,4,5} and further decreases in ECCF{5}. We explain this finding as follows: an extension of user profiles based on the categories of highly rated items supports the identification of a richer set of user preferences, and a more efficacious identification of neighbors, than only considering rating distribution on categories. However, if we restrict $PositiveRatings$ too much, the user

profiles are not extended enough to sensibly improve Recall. Moreover, as noticed for Precision, if $PositiveRatings$ is lax, noise in the estimation of user preferences challenges neighbor selection.

- **F1:** ECCF outperforms the baselines. In detail, ECCF{4,5} achieves the best F1 = 0.7691; moreover, F1 varies consistently with Recall, depending on $PositiveRatings$.
- **RMSE:** SCCF reduces the mean error between estimated and observed ratings with respect to the baseline, showing the benefits of category-based user profiles. Moreover, consistently with the variation of Precision, the best results are obtained by ECCF{5}, i.e., with a strict extension of user profiles. RMSE progressively increases (i.e., gets worse) for $PositiveRatings = \{4, 5\}$ and {3, 4, 5}.
- **MRR:** ECCF outperforms the baselines. Specifically, ECCF{4,5} obtains the best MRR = 0.7391. The second best value corresponds to a more selective extension of user profiles in ECCF{5}; moreover, if $PositiveItems = \{3, 4, 5\}$ results get worse.
- **Diversity**: both SCCF and ECCF outperform U2UCF. In this case, the best results are obtained with a lax extension of user preferences (ECCF{3,4,5}) and Diversity decreases while the preference extension policy becomes stricter. We explain these findings with the fact that category-based user profiles improve the estimation of user preferences concerning a variegate set of item categories, with respect to a flat recommendation based on ratings. However, the stricter is the extension of user preferences, the less item categories are used in neighbor identification.
- **User coverage:** ECCF outperforms the baselines, confirming the usefulness of preference extension. However, the selection of the ratings for the extension influences coverage: ECCF{4,5} achieves the best results by suggesting at least one relevant item to 85.42% of the users, against 84.97% of U2UCF. The second best is ECCF{5} and ECCF{3,4,5} has the worst results.

In the described experiments the EP Matrix is defined by only taking into account positive ratings. In order to get a broader view on the performance of ECCF, we also consider its application to all the user ratings; i.e., we set $PositiveRatings$ to $\{1, \dots, 5\}$. With respect to the previous results, in this case the algorithm achieves similar Precision but lower Recall (0.7524), MRR (0.7369) and User coverage (0.8155).

Metrics	U2UCF	SCCF	ECCF {3,4,5}	ECCF {4,5}	ECCF {5}
Precision	~~0.7806~~	**0.7842**	0.7839	0.7838	**0.7842**
Recall	~~0.757~~	0.7624	0.7634	**0.7649**	0.7626
F1	~~0.7686~~	0.7731	0.7735	**0.7742**	0.7732
RMSE	~~0.9935~~	0.9838	0.9835	**0.9832**	**0.9832**
MRR	~~0.733~~	0.7369	0.7372	**0.7391**	0.7384
Diversity	~~0.3059~~	0.307	**0.3073**	0.307	0.3067
User cov.	~~0.8497~~	0.8521	0.8526	**0.8542**	0.8534

Tab. 6.3: Performance evaluation @20.

Table 6.3 shows the results obtained by comparing performance@20. These results confirm the usefulness of category-based user profiles and of their extension with frequently co-occurring information interests:

- Also in this case, ECCF{4,5} is the best recommendation algorithm. It outperforms the others in Recall, F1, MRR and User coverage. Moreover both ECCF{5} and ECCF{4,5} achieve the best RMSE in comparison with the other recommenders.
- However, while SCCF has the best Precision@10, both SCCF and ECCF{5} achieve the best Precision@20.

With respect to k=10, Precision@20 is lower while Recall@20 and F1@20 take higher values; this makes sense because we are considering longer suggestion lists. Moreover, RMSE@20 is lower, which tells us that the longer lists contain proportionally less errors in the estimation of ratings. Differently, most algorithms obtain the same MRR for k=10 and k=20 (except for SCCF and ECCF{3,4,5}): this shows that the first relevant item is almost always placed in the first 10 positions of the suggestion lists. Furthermore, the Diversity@20 has the highest values for all the recommenders: this might be due to the fact that the longer suggestion lists have more chances to include items belonging to different categories. Finally, User coverage@10 = User coverage@20 because we interpret coverage as the percentage of users who receive at least one suggestion.

Figures 6.2 and 6.3 depict the accuracy @10 and @20:

- All of the category-based recommenders outperform U2UCF, confirming the benefits of the introduction of category-based preferences in KNN Collaborative Filtering. The conceptual representation of user preferences generally improves performance because the matrices describing user preferences (*UC* and *EP*)

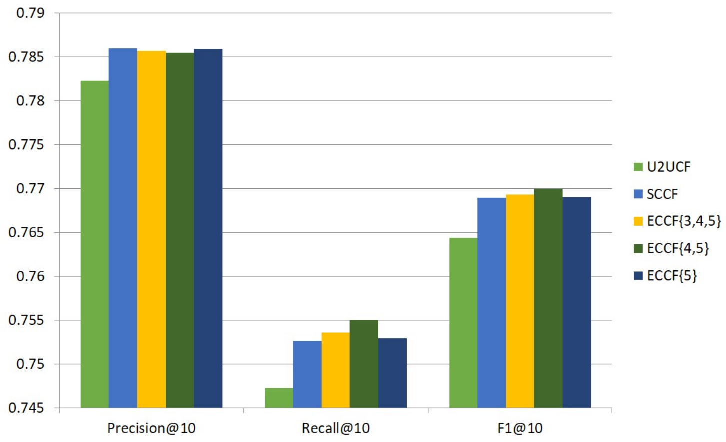


Fig. 6.2: Graphical Representation of Accuracy@10.

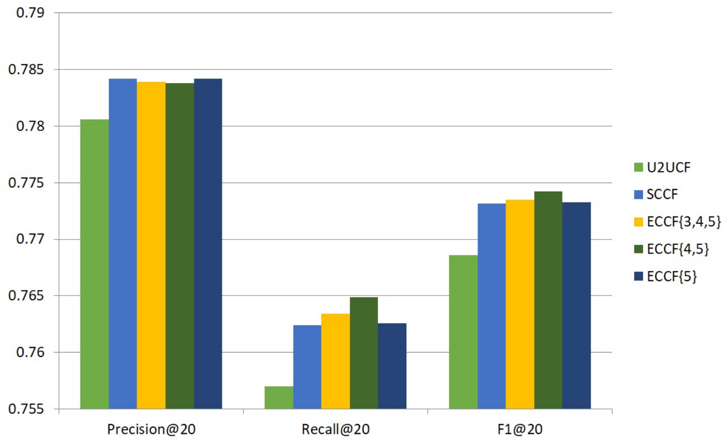


Fig. 6.3: Graphical Representation of Accuracy@20.

are denser than the users-items matrix storing ratings (R). Therefore, better neighbors can be identified for the computation of Equation 6.1.

- A comparison between category-based algorithms shows that the best performance results are obtained by extending user profiles on the basis of the items that users have rated very well, i.e., with 4 or 5 stars in a [1, 5] Likert scale. If the items that received middle ratings are considered as well, accuracy decreases.

- The category-based representation of user profiles has positive impact on the Diversity of recommendation lists. Conversely, the extension of user profiles does not further help this aspect, unless user profiles are extended in a lax way. However, a lax extension is not convenient because it decreases other measures.

In order to assess the usefulness of preference extension in Top-k recommendation, we also compare the previously described algorithms with SVD++ (Koren, 2008), which adopts Matrix Factorization to learn latent user and item factors, basing rating prediction on the sole analysis of user ratings. The comparison results show that:

- SVD++ is more accurate than U2UCF and SCC. On the filtered Yelp dataset, SVD++ obtains F1@10 = 0.7696. This finding shows that the management of category-based user profiles helps recommendation but it can be outperformed by a deeper understanding of the features of items and users.
- SVD++ achieves similar accuracy results with respect to ECCF but it is outperformed by ECCF{4, 5}. Therefore, the extension of user profiles with frequently co-occurring information interests, integrated into a KNN recommender, improves accuracy and makes it comparable or higher than that of Matrix Factorization algorithms.
- ECCF outperforms SVD++ as far as the diversity of the recommendation lists is concerned: SVD++ has Diversity@10 = 0.3041; this is comparable to the diversity achieved by U2UCF and lower than that of all the category-based recommenders we presented.
- In contrast, SVD++ has the highest User coverage of all the algorithms (0.8709), showing its superior capability to contrast data sparsity.

6.3.6 Discussion

In summary, the evaluation results show that ECCF outperforms U2UCF, SCCF and SVD++ in accuracy and intra-list diversity. Moreover, it outperforms U2UCF and SCCF in MRR and user coverage, while SVD++ excels in the latter metric. The results also show that ECCF achieves the best results when applied to positive ratings, while its performance slightly decreases when the user profiles are extended by taking both positive and negative ratings.

These results support the hypothesis that preference extension, based on frequently co-occurring information interests, improves the accuracy of the suggestions generated by a KNN recommender system. However, since the results are limited, research has to be carried out to improve other performance metrics, possibly also investigating the integration of preference co-occurrence in Matrix Factorization algorithms. We also plan to test the model on other datasets in order to better understand the results.

It might be questioned whether extending user profiles with general interest co-occurrence data might provide less personalized recommendations than, e.g., focusing the extensions on the user's neighborhood. In this respect, we point out that we aim at developing a model that does not depend on cross-domain user identification. However, an investigation of this issue can be interesting to deal with the cases in which user information can be shared among the applications, or public information about the users can be connected to the local profiles; e.g., public data on social networks.

7 Faceted Search of Heterogeneous Geographic Information for Dynamic Map Projection

The work described in this chapter has a different focus with respect to the previous one because, instead of supporting information exploration by suggesting relevant content, it is aimed at empowering the user in the exploration of information by offering advanced interactive tools in the user interface. The described work has been conducted in the context of the OnToMap web application. As mentioned in Section 2.1.1 a geographic map can be seen as a container of information for content sharing. Thus, users can explore the information about the territory and they can analyze it under different perspectives. In order to help the user to filter and interact with the retrieved information, we developed and integrated four different widgets for faceted search in the OnToMap web application. Specifically, we carry out a user study in order to understand which widget is more helpful for users in the context of geographic faceted search.

This work has been published in:

- Ardissono, L., Delsanto, M., Lucenteforte, M., Mauro, N., Savoca, A., and Scanu, D. (2018a). "Map-based Visualization of 2D/3D Spatial Data via Stylization and Tuning of Information Emphasis". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. AVI '18. Castiglione della Pescaia, Grosseto, Italy: ACM, 38:1–38:5.

and submitted to:

- Mauro, N., Ardissono, L., and Lucenteforte, M. (Submitted). "Faceted Search of Heterogeneous Geographic Information for Dynamic Map Projection". In: *Information Processing & Management*.

7.1 Introduction

Several works promote the development of faceted search interfaces (Hearst, 2006) to reduce information overload and keep users in control of the search process in exploratory search (Marchiorini, 2006). However, most of the existing approaches support an individual user who inspects a single data category, e.g., documents or movies, while pursuing a short-term information goal. Similarly, most research on geographic information search focuses on helping individual users retrieve relevant data for particular short-term goals; e.g., finding the available routes between two Points of Interest (Quercia et al., 2014), identifying the 2-star hotels in a specific area (Lionakis and Tzitzikas, 2017), or studying the relations among the items of an information category (Andrienko et al., 2007).

Indeed, map management can go farther than that in order to provide long-term representations of shared projects to users having different information interests. For instance, in participatory decision-making (Coulton et al., 2011; Brown and Weber, 2012), (Hu et al., 2015) point out that 2D maps and 3D virtual environments can facilitate participants' learning and understanding, especially as far as spatial decision-making processes are concerned. Moreover, maps can support information sharing and collaboration in simpler and less formal scenarios. For example, if somebody is planning a holiday, a custom map including selected places to visit, hotels, and so forth, would provide a personalized projection of the area to be visited that the user can consult and annotate before, during and after the trip, possibly in collaboration with the other people traveling with her/him to gain a common view of the vacation.

These scenarios suggest the development of custom maps that define Personal Information Spaces (Ardito et al., 2013; Ardito et al., 2016) useful to organize individual and group activities. For this purpose, maps should be adapted to reflect temporary information goals while persistently storing data in order to facilitate a quick projection and resumption of the collaboration context.

In this work, we present a faceted exploration model for the management of this type of map. Our model supports a flexible, map-based visualization of heterogeneous data and it enables map focusing to satisfy specific information needs by offering two graphical interactive exploration functions:

- The former enables coarse-grained map projection on data categories via opacity tuning, without taking the facets of items into account; all the items of a category are subject to the same visualization policy. The user can visualize or hide all the items of a specific category at the same time but (s)he cannot filter them according to their facets.

- The latter combines opacity tuning with fine-grained faceted search support to enable map projection at different granularity levels, by taking the properties of information items into account.

In both cases, the projection is only visual and the information stored in the map is preserved. The work we present has the following innovative aspects:

- *Efficient multi-category faceted projection of long-lasting custom maps to answer temporary information needs in sparse and noisy datasets*. Our model suggests information visualization constraints based on attributes of data that support an efficient exploration of the information items stored in the maps.
- *Representation of the search context by associating each data category to a compact graphical widget that supports interactive data visualization, faceted exploration, category-based information hiding and transparency of results.* The widgets of the categories searched by the user are located in a sidebar of the user interface and play the role of breadcrumbs, representing the types of information that (s)he has explored during the interaction with the system and the applied visualization constraints.

Our model supports geographic information search within the OnToMap collaborative Web GIS. We tested the model in a user study to assess user experience and performance in exploratory search. For the experiments, we compared different graphical widgets supporting faceted exploration, from traditional ones such as checkboxes, to advanced ones based on treemaps and sunburst diagrams. The study showed that, when working on geographic maps populated with heterogeneous information, our model outperforms simple category-based map projection and traditional faceted search tools such as checkboxes. Specifically, we obtained the best user performance and experience results using the widget based on the sunburst diagram, which displays visualization criteria in a compact structure.

This work builds on the work described in (Ardissono et al., 2018a), which presents our first opacity tuning model. With respect to that paper, the present work introduces graphical widgets that support fine-grained data management and a novel approach to the selection of efficient facets for information exploration in sparse and noisy datasets. The widgets extend the category hiding function provided by the previous model with faceted data exploration to enhance information search and visualization. The present work also provides an extensive evaluation of the faceted exploration model.

In the following, Section 7.2 presents our research questions. Section 7.3 describes our faceted search model. Section 7.4 presents the experiments we carried out and Section 7.6 discusses the evaluation results.

7.2 Research Questions

We designed the faceted search model presented in this work after two preliminary experiments with users carried out in the urban planning domain; see (Voghera et al., 2016; Voghera et al., 2018). In those experiments, the projection of long-lasting maps on specific types of information emerged as a useful feature to support data interpretation during project development. This feature was also requested in the final analysis phase, in which human planners analyzed complex maps obtained by integrating the students' projects to identify the most recurring represented territorial elements.

The present work describes the faceted search support offered by the current version of OnToMap and investigates its usefulness to data search and interpretation in a project map. We pose the following research questions:

$\mathbf{RQ2_1}$: *How does a compact, graphical view of the visualization constraints applied to a map, impact on her/his efficiency and experience in data exploration?*

$\mathbf{RQ2_2}$: *How much does the user's familiarity with the widgets for faceted exploration impact on her/his efficiency in search and on her/his appreciation of the exploration model they offer?*

The experiments described in Section 7.4 are aimed at answering these questions.

7.3 Information Exploration Model

Our information exploration model is integrated in the OnToMap system to support information search and it includes two main types of functions, implemented as interactive graphical widgets.

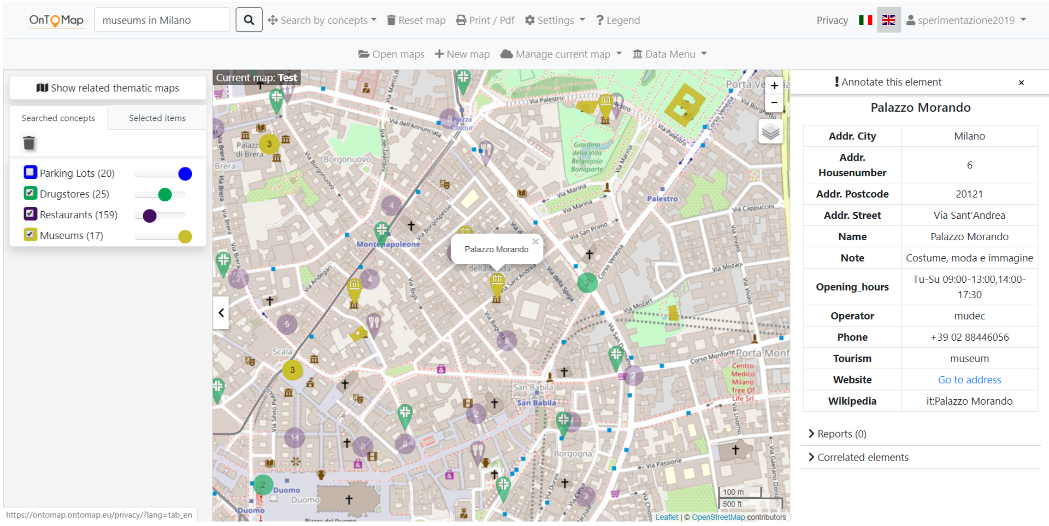


Fig. 7.1: User interface showing the widgets based on sliders.

7.3.1 Exploration Function 1: Coarse-grained Map Projection by Means of Transparency Sliders

This function, introduced in (Ardissono et al., 2018a), supports map projection via opacity tuning: for each searched category, a transparency slider enables the user to assign different levels of opacity to its items; the widget also has a checkbox to temporarily hide information by means of a click, without changing the degree of opacity selected for the category.

The sidebar of Figure 7.1 shows the widgets based on transparency sliders. In the map, museums are visualized in full color because the slider of the "Museums" category is selected and tuned to maximum opacity. Differently, drugstores and restaurants are semi-transparent and the map hides the items of the "Parking Lots" category because its slider is de-selected.

The transparency slider does not enable the specification of constraints on facet values; i.e., it works at the granularity level of the represented category and it uniformly tunes the opacity of all its items. Nevertheless, this widget supports visual simplification by enabling the user to temporarily hide information by type. Basically, opacity tuning enables her/him to highlight the information in focus while maintaining an overview of what has been searched on the map. This model is inspired by (Colby and Sholl, 1991)'s work on layers visualization but it separately handles the opacity of items belonging to different categories; moreover, it supports the visualization of multiple layers, as a generalization of Translucent Overlay (Lobo et al., 2015).

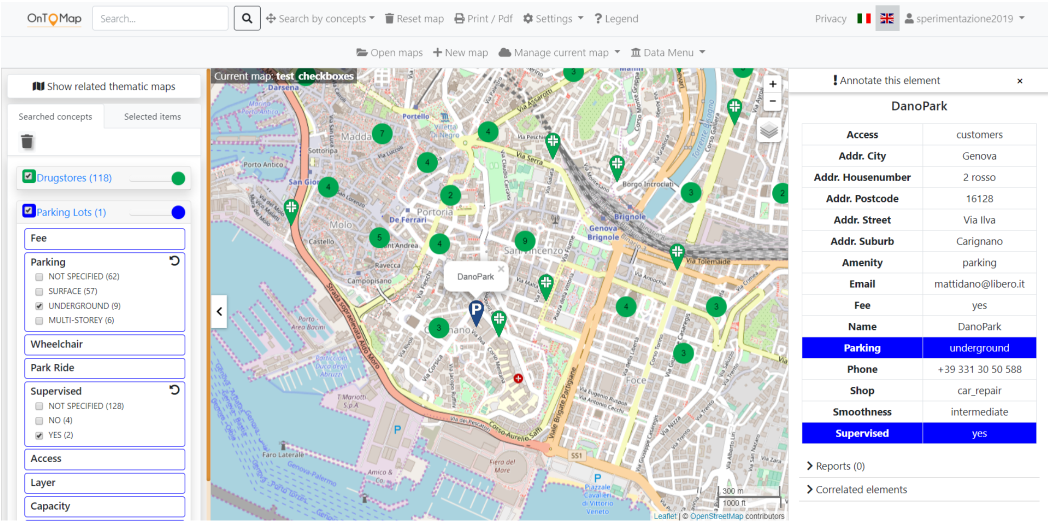


Fig. 7.2: User interface showing the widgets based on checkboxes.

7.3.2 Exploration Function 2: Faceted Approach

This function combines coarse-grained and fine-grained specification of visualization constraints by integrating transparency sliders with faceted information exploration. The widgets implementing this function include a transparency slider and an internal component showing the facets of the represented category. The internal component can be a set of checkboxes, a treemap or a sunburst diagram, depending on the layout selected for the user interface, and it enables the user to specify visualization constraints based on facet values. The transparency slider works in combination with facet selection and tunes the opacity of the visualized items. The widgets are interactive and they can be opened/closed by clicking on them; a closed widget only shows its own transparency slider; e.g., see "Drugstores" in Figure 7.2 which shows the layout based on checkboxes.

Let us consider a facet f of a category C and the set of retrieved items that belong to C, henceforth denoted as E_C, i.e., extension of C. The visualization of the values of f in the widget is aimed at providing the user with a preview of the corresponding items in the map. For this purpose, we adopt a standard approach to facet suggestion (Oren et al., 2006; Hearst, 2006):

- The widget only displays the values $\{v_1, \ldots, v_m\}$ of f that have at least one item in E_C to prevent the user from following links to zero solutions.
- The values of f are sorted from the most frequent to the least frequent ones in E_C. Moreover, the widget shows, or makes available on mouse over, the number of items corresponding to each value. Notice that the widget may also

show a “NOT SPECIFIED” value to represent the subset of items in which f is not defined. This is aimed at providing the user with a visual representation of the coverage of the facet in the results.

- In order to limit visual complexity, long lists of values are dropped, making their tails available on demand by providing a “More...” link or a “+” symbol, depending on the layout of the widget.

By default, none of the facets in the widget of a category C is selected. If the user picks one or more values of the same facet, this is interpreted as an OR constraint because (s)he has specified that all those values are eligible for visualization. Conversely, the selection of values that belong to distinct facets of C generates an AND constraint because it identifies the items having more than one property restricted to specific values. For instance, if the user chooses $f_i = v_{i1}$, $f_i = v_{i2}$ and $f_j = v_{j1}$, items $\{x \in E_C \mid f_i(x) \in \{v_{i1}, v_{i2}\} \wedge f_j(x) = v_{j1}\}$ are shown and the other items are hidden.

We use color coding to link visualization constraints to map content: the tables showing the details of items highlight the facets corresponding to the selected visualization constraints in the color associated to the category. In this way, the user can quickly identify the characteristics that make items eligible for being displayed. For instance, the table of “DanoPark” in Figure 7.2 has the “Parking” and “Supervised” facets highlighted in blue because they correspond to the visualization constraints imposed on the “Parking Lots” widget.

7.3.2.1 Layouts of the Widgets for Faceted Information Exploration

Before providing details about how we select the facets to be included in the widgets we present the layouts we developed.

- The **widgets based on checkboxes** contain a rimmed rectangle for each facet to be shown. By clicking on a rectangle, the user expands (or closes) the corresponding facet. An expanded facet shows values as checkboxes and offers a “More...” link to show the hidden ones. For instance, Figure 7.2 shows eight facets of widget “Parking Lots": “Fee”, . . . , “Capacity”. Users can select the checkboxes to impose visualization constraints on the map. In the figure, the user has expanded the “Parking” and “Supervised” facets and (s)he has selected values “UNDERGROUND” and “YES”.
- **The widgets based on treemaps** include facets in rimmed rectangles as well. However, when a facet is expanded, its values are displayed as components

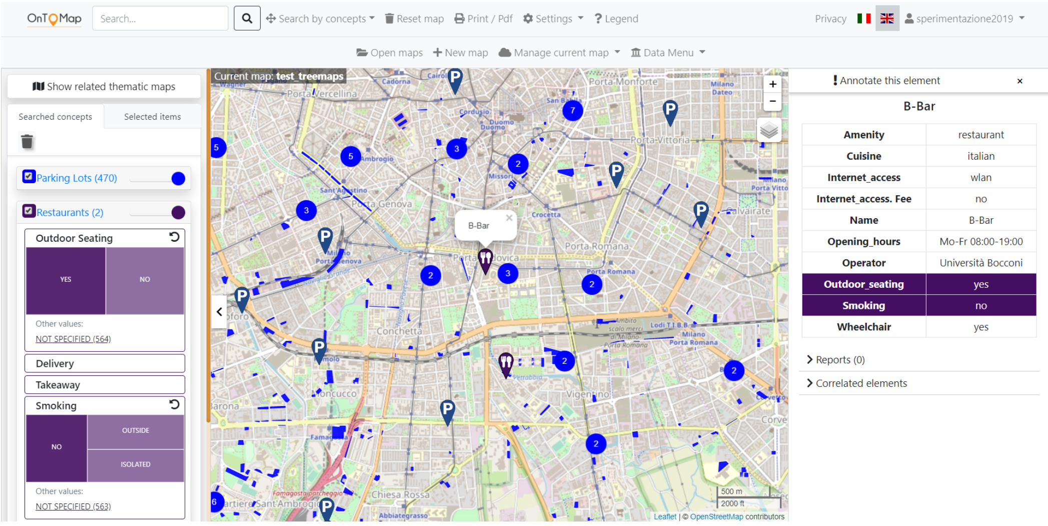


Fig. 7.3: User interface showing the treemaps as faceted exploration widgets.

of a treemap whose size depends on the cardinality of the corresponding set of items (larger size means larger cardinality); see Figure 7.3. Long values are shortened but they can be visualized, together with the cardinality of the corresponding sets of items, on mouse over. Only the most frequent values are included in the treemap; the other ones are available below it or on demand ("More...") and the user can add them to the treemap by means of a click. The user can (de)select values by clicking on them. The selected values take the color of the category (e.g., "YES" in "Outdoor Seating" and "NO" in "Smoking"); the other ones have a pale tone of the same color.

- The **widgets based on the sunburst diagram** show the facets of the represented category C in a ring having the color associated to C. The diagram is visualized in a pop-up window that the user can open or close by means of a click, and the sidebar of the user interface only displays the thumbnails of the sunbursts; see Figure 7.4. The user can expand each facet by clicking on the portion of ring representing it: values are shown in a second level, sorted clockwise by decreasing frequency in the extension of C. Only the most frequent values are shown but the user can view and add the other ones by clicking on the "+" button located in each portion of the internal ring; the sunburst is extended by adapting the size of the displayed values. The user can (de)select values by clicking on them and color coding is applied to link visualization constraints to map content.

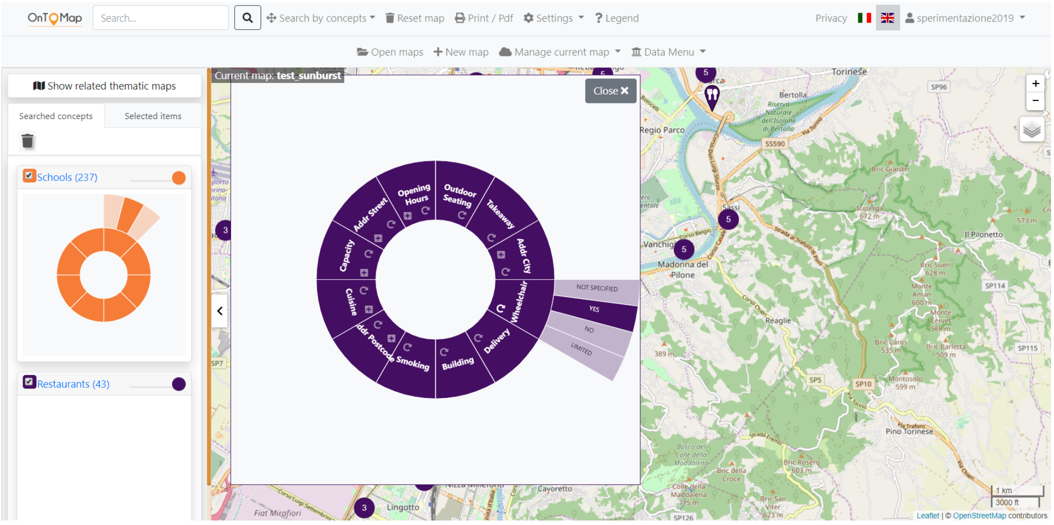


Fig. 7.4: User interface showing the sunburst as faceted exploration widgets.

7.3.3 Selection of Facets to be Included in the Information Exploration Widgets

The dynamic generation of widgets for the exploration of search results retrieved from open data sources is challenged by the amount and variability of the information items to be managed. Facets have thus to be analyzed in order to identify the most convenient ones for map content analysis.

7.3.3.1 Navigation Quality in Semantic Data Repositories

(Oren et al., 2006) introduce the *navigation quality* of a facet f to describe its efficiency in supporting information browsing of RDF data repositories. This measure takes values in [0, 1] (where 1 is the best value) and is based on the product of three metrics, which take values in [0, 1] as well:

1. The *balance* of f, i.e., its capability to split results in subsets having similar cardinality; equally distributed facets have maximum balance.
2. An inverse measure of the *number of distinct values* of f occurring in the results, denoted as “*object cardinality*”. The authors consider as acceptable the facets that have between 2 and 20 values because they can be displayed in a search interface without overloading the user.
3. The *frequency* of f in the results, i.e., the percentage of retrieved items in which the value of f is specified.

Cuisine	Count
PIZZA	111
ITALIAN	74
REGIONAL	37
CHINESE	28
JAPANESE	17
ITALIAN; PIZZA	15
SUSHI	11
ITALIAN; REGIONAL	10
PIZZA; ITALIAN	10
MEXICAN	9
KEBAB	7
INDIAN	4
ASIAN	3
CHINESE; JAPANESE	3
FISH	3
INTERNATIONAL	3
ITALIAN; PIZZA; REGIONAL	3
ITALIAN_PIZZA	3
PERUVIAN	3
STEAK_HOUSE	3
AMERICAN	2
CHINESE; PIZZA	2
GREEK	2
ITALIAN_PIZZA; PIZZA	2
KEBAB; PIZZA	2
LOCAL	2
MEDITERRANEAN	2
PIZZA; KEBAB	2
REGIONAL; ITALIAN	2
AFRICAN	1
...	
56 more values with Count=1.	

Outdoor Seating	Count
NO	59
YES	33

Takeaway	Count
YES	62
NO	10
ONLY	4

Tab. 7.1: Value distribution of facets “Cuisine”, “Outdoor Seating” and “Takeaway” (OSM tag: “amenity=restaurant”) in Torino city bounding box; retrieved using Overpass Turbo on Sept. 20th, 2019. The results include 719 items, out of which 432 specify the value of “Cuisine”, 92 specify the value of “Outdoor Seating” and 76 specify the value of “Takeaway”.

Navigation quality cannot be applied in OnToMap because of its assumptions: firstly, statistics about OSM data provided by TagInfo (OSM Contributors, 2019) show that most of the tags are hardly used.[1] This can be explained because crowdmappers tend to underspecify the items they map; moreover, they sometimes define new tags instead of using the existing ones, thus generating a plethora of synonyms which increase in an uncontrolled way the number of distinct facets and values. This phenomenon is so widespread that several efforts try to systematize OpenStreetMap

[1]For instance, by invoking `https://taginfo.openstreetmap.org/tags/amenity=restaurant#combinations` it is possible to learn that “amenity=restaurant" has 178 different tags, only 36 of which occur in more than 2% of the items mapped in OSM worldwide. Moreover, the most frequent tag is “name”, which is only defined in 90.92% of items, in spite of its importance as a POI identifier.

through semantic knowledge representation; e.g., (Codescu et al., 2011; Ballatore et al., 2013). We also notice that several results retrieved from OSM are unbalanced and can be split into (i) a large set of items in which the facet is not available, (ii) a few values identifying sets of items with reasonable cardinality, and (iii) a long tail of values represented by one or two items. For instance, Table 7.1 shows the distribution of three facets retrieved from OSM by searching for "amenity=restaurant" (which corresponds to the "Restaurants" category of the OnToMap domain ontology) on Torino city bounding box. The facets have fairly poor coverage and they are unbalanced: "Cuisine" is specified in 432 items out of 719 and it exhibits a long tail distribution; "Outdoor Seating" and "Takeaway" are specified in 92 and 76 items respectively; "Name", not shown, is balanced but it only occurs in 675 items.

We thus define a novel approach to the computation of facet efficiency that suits these types of distribution and is robust towards information lack. The idea is that (i) coverage has to be taken into account as a separate factor to select useful facets, and (ii) balance and number of values have to be controlled by the cardinality of the sets of items identified by the facet.

7.3.3.2 Our Approach: Evaluating Exploration Cost in Sparse, Unbalanced Datasets

When searching for information in crowdsourced data sources, the suggestion of the most representative facet values in a result set is a primary goal because it enables the system to provide the user with a relevant number of items to choose from. Moreover, it can be complemented by free text queries that let the user express specific information needs; e.g., in OnToMap free text queries support the retrieval of very specific items, such as "*Pediatric* hospitals in Torino". It thus makes sense to propose facets that, regardless of balance, identify some fairly large subsets of items, possibly leaving the long tail apart or making it available on demand; e.g., consider the first values of "Cuisine" in Table 7.1. Given these premises, we propose a two-step evaluation of facets efficiency to exploration support.

1) In the first step, we consider *frequency* as a pre-filtering metric to exclude from any further computation the facets that appear very rarely in the results. Having sampled a set of queries to OSM and taking Taginfo statistics as a baseline, we empirically set to 3% the minimal frequency threshold under which a facet is considered as useless. Only the facets over this threshold are considered for the evaluation of their efficiency.

2) In the second step, given the highly variable distribution of facets, we consider balance and number of values in combination. We are interested in facets that split

E_C in at least some portions having significant cardinality because they identify homogeneous, relatively large sets of items to be analyzed. These facets enable the system to propose visualization criteria that significantly reduce the search space by showing the most representative values, leaving the other ones on demand. Differently, in a small result set, as those typically retrieved when the selected bounding box is strict, there are few items; therefore, the efficiency in splitting results is less important because the user can easily analyze items one by one. In order to capture this intuition, we compute the *cost of exploring the extension* E_C of a category C by means of a facet f that takes values in $\{v_1, \ldots, v_m\}$ as follows:

$$explorationCost(f) = \frac{-\sum_{j=1}^{m} p(v_j) log_2 p(v_j)}{meanCard(f)} \tag{7.1}$$

$explorationCost(f)$ takes values in $\mathbb{R}^+$; $p(v_j)$ is the probability of v_j in E_C, computed by considering the values $v_j \neq$ “NOT SPECIFIED”; $meanCard(f)$ is the mean cardinality of the subsets of results identified by f:

$$meanCard(f) = \frac{|E_C|}{m} \tag{7.2}$$

It should be notice that since Equation 7.1 represents a cost, the facets will be ranked in ascending order. The components of Equation 7.1 have the following roles:

- The numerator represents the (not normalized) entropy of f, which takes values in $\mathbb{R}^+$. The entropy of an information source is an average measure of the amount of uncertainty of its own m symbols; it is positively influenced by both the number of values that the source can take and by the balance of the corresponding subsets of items. For instance, given two balanced facets f_1 and f_2, if f_2 has more values than f_1, f_2 also has higher entropy than f_1. Moreover, if two facets have the same number of values, the most balanced one has the highest entropy. Finally, if all the items of E_C have the same value of f (e.g., all the schools located in the bounding box are primary ones), the entropy is 0, meaning that the facet does not help discriminate among the items of E_C. It should be notice that since the entropy is not normalized it grows with the number of values.

- The denominator of Equation 7.1 captures our interest in the facets that split results in fairly large subsets: even though a facet f has high entropy (e.g., because it has several values), its cost is smoothed if the subsets of items it identifies have high cardinality, because f enables the user to browse a large portion of results in few steps.

	Exploration cost	(1 - navigation quality)
Outdoor Seating	0.0205	0.8924
Takeaway	0.0335	0.9497
Ex1	0.1500	0.8949
Cuisine	0.8738	1.0000
Ex2	1.0000	0.9842
Name	9.3987	1.0000

Fig. 7.5: Exploration cost and complement of navigation quality of a set of facets. The color scale varies from the lowest cost values, depicted in green, to the highest ones, in red. Notice that colors are tuned to the values observed in this example; i.e., [0, 10] for exploration cost and [0, 1] for the complement of navigation quality.

	entropy	meanCard	exploration cost
Outdoor Seating	0.9416	46.0000	0.0205
Takeaway	0.8482	25.3333	0.0335
Ex1	3.0000	20.0000	0.1500
Cuisine	4.3895	5.0233	0.8738
Ex2	3.0000	3.0000	1.0000
Name	9.3987	1.0000	9.3987

Tab. 7.2: Entropy, mean cardinality and exploration cost of the facets displayed in Figure 7.5.

Figure 7.5 graphically compares the exploration cost of Equation 7.1 with (Oren et al., 2006)'s navigation quality on a few facets; see Tables 7.2 and 7.3 in the Appendix for details. We consider "Cuisine", "Takeaway", "Outdoor Seating" and "Name", based on the data described in Table 7.1, and two toy examples:

- "Ex1", specified in 160 items, has 8 distinct balanced values, with $meanCard = 20$.
- "Ex2", specified in 24 items, has 8 distinct balanced values, with $meanCard = 3$.

Notice that Oren and colleagues compute a quality measure, i.e., the highest values are the preferred ones; conversely, we compute a cost function that has the opposite interpretation. In order to facilitate the comparison, Figure 7.5 graphically shows the *complement of navigation quality* in the [0, 1] interval and it tunes the color scale

	balance	object cardinality	frequency	navigation quality
Outdoor Seating	0.8587	0.9794	0.1280	0.1076
Takeaway	0.5175	0.9201	0.1057	0.0503
Ex1	1.0000	0.4725	0.2225	0.1051
Cuisine	0.3663	4.54E-66	0.6008	9.99E-67
Ex2	1.0000	0.4725	0.0334	0.0157
Name	1.0000	0.0000	0.9388	0.0000

Tab. 7.3: Balance, object cardinality, frequency and navigation quality of the facets shown in Figure 7.5, according to (Oren et al., 2006) with $\mu = 2$ and $\sigma = 4.9$. We remind that the colors of facets in Figure 7.5 correspond to the complement of the values reported in the present table.

to the values observed in this example; i.e., [0, 10] for exploration cost and [0, 1] for the complement of navigation quality.[2]

- In both approaches "Name" has a very high cost, which is desirable because this facet identifies hundreds of subsets of items to be browsed one by one, thus it does not help the user to reduce the amount of information on the map.
- According to (Oren et al., 2006), "Outdoor Seating", "Ex1" and "Takeaway" are moderately inefficient, and "Takeaway" has higher cost than the other ones; the reason is the low coverage of these facets and, with the exception of "Ex1", their lack of balance. Differently, our model attributes low cost to these facets because they have few values which represent non-elementary sets of solutions to be inspected.
- The main disagreement is in the evaluation of "Cuisine" and "Ex2". According to (Oren et al., 2006), "Cuisine" is totally inefficient because of its partial coverage of items, lack of balance and high number of values. Moreover, "Ex2" is penalized by the lack of coverage of results. In our approach "Cuisine" has moderate cost, in spite of the many values it can take, because it identifies a few large subsets that deserve attention when browsing results, and the long tail of the facet can be ignored. "Ex2" has higher cost than "Cuisine" because it identifies very small sets of items.

In summary, our approach supports the identification of facets which are not "perfect" from the *divide et impera* viewpoint because they only occur in a subset of results

[2](Oren et al., 2006)'s model introduces the σ and μ parameters for the computation of balance and object cardinality metrics but we could not find the exact values that they applied in their experiments. We reproduced the expected behavior, following the indications given in the paper, by setting $\mu = 2$ and $\sigma = 4.9$.

and/or they split data in an unbalanced way. However, it works on realistic cases in which balanced, frequent facets are extremely rare. For example, if a user looks for takeaway restaurants in Torino, there is a big set of them that offer takeaway (62) and a small set (10) that does not offer this characteristic. However, this facets should be ranked in the top part of the list because it is a relevant filter for the user even if it is not balanced. By analyzing some OpenStreetMap data we noticed that it is really difficult to find balanced facets. Moreover, it promotes facets that split results in subsets having a significant cardinality because they are valuable for browsing results.

7.3.4 Selection of Facets to be Included in the Widgets

In order to select the facets to be shown in the widgets, we first exclude those having $cost(f) = 0$ because this means that they have a single value in E_C. Then, we sort facets by increasing cost and we include them in the widget up to a maximum number of 12 to avoid cluttering the user interface.

By applying Equation 7.1 to the results of query "amenity=restaurant" on Torino city bounding box, we obtain the following sorted list of facets: "Outdoor Seating", "Takeaway", "Wheelchair", "Delivery", "Addr city", "Smoking", "Building", "Addr postcode", "Cuisine", "Capacity", "Addr street" and "Opening hours"; see Figure 7.4. Almost all these facets correspond to semantically relevant dimensions. Only "Addr city" seems useless because the query is bounded in Torino city; however, according to the geocoder we use, the area of the map that is considered includes Torino and a few small cities in its boundary. Other facets, such as "Name", "Phone" and "URL", are excluded from the sunburst because they have very high cost (they are identifiers) and thus take the final positions in the ranked list. Facets such as "Cuisine 1", which is redundant with respect to "Cuisine", are excluded because they are below the minimum coverage threshold. Indeed, "Cuisine 1" is the typical tag that somebody has duplicated instead of using the main "Cuisine" one.

7.4 Validation of our Faceted Exploration Model

7.4.1 Study Design

We conducted a user study to evaluate the four types of information exploration widgets described in Section 7.3, as far as data interpretation in a geographic map is concerned. Specifically, we were interested in comparing:

- The exploration model based on transparency sliders (which supports information hiding at the granularity level of the data category) to the more expressive one that also supports faceted exploration.
- The alternative graphical models that we defined for faceted information exploration in order to understand which ones are more effective to help users in the exploration of an information space via map projection.

We prepared four maps, each one focused on a different geographic area and populated with multiple data categories to simulate a friendly project planning context (a tourist trip). We investigated participants' performance and user experience in four map learning tasks, each one using a different type of widget:

- *Task1*: question answering using checkboxes in combination with transparency sliders.
- *Task2*: question answering using treemaps in combination with transparency sliders.
- *Task3*: question answering using sunburst in combination with transparency sliders.
- *Task4*: question answering using transparency sliders.

The study was a within-subjects design one. We considered each treatment condition as an independent variable and every participant received the 4 treatments; we counterbalanced the order of tasks to minimize the impact of result biases and the effects of practice and fatigue. People participated in the user study on a voluntary basis, without any compensation. All participants signed a privacy consensus according to GDPR. University research ethical committee approval was obtained for the study. The participation to the user study took place live, i.e., we did not perform any online interviews.

7.4.2 The Experiment

One person at a time performed the study which lasted about 30 minutes. Before starting the user study, the participant watched a video describing the widgets and showing how they work. After that, (s)he interacted with OnToMap on a sample map to get acquainted with the user interface of the system. We did not impose any restrictions on this activity and we allowed the participant to take as much time as (s)he needed in order to comply with diverse backgrounds and levels of confidence with technology. Then, we asked her/him to answer a pre-test questionnaire designed to assess demographic information, cultural background, as well as familiarity with map-based online applications.

During the study, we asked the participant to use OnToMap in the context of the organization of a trip. For each task (s)he had to look at the associated map, which included some categories of items, and (s)he had to answer two questions which required counting elements that have certain properties, or identifying items given their descriptions. As far as counting is concerned, we forced the participant to analyze the map by asking her/him to answer the questions in a geographic area delimited by an orange border. In this way, (s)he could not simply read the cardinality information provided by the faceted exploration widgets, which work by taking the bounding box of the map as a reference to specify how many items satisfy the selected visualization constraints. The questions proposed to the participants had the following templates:

- How many *category name* having $characteristic_1$ and/or ... and/or $characteristic_n$ are visualized within the area delimited by the orange line in the map?

  For instance, "How many Christian churches accessible to wheelchairs are visualized within the area delimited by the orange line in the map?". In the question, "Christian" is a value of facet "Religion" and wheelchair accessibility corresponds to value "YES" of facet "Wheelchair".

- Find *category name* having $characteristic_1$ and/or ... and/or $characteristic_n$ within the orange line in the map, and list them.

  E.g., find restaurants serving pizza or Italian food (values of "Cuisine").

In *Task1*, *Task2* and *Task3*, we proposed selective questions because we wanted to understand whether the widgets helped participants satisfy specific information needs by exploring the metadata of the searched categories and by projecting the maps accordingly. Differently, the questions of *Task4* did not require the imposition of any visualization constraints because participants only used the transparency sliders; in this task, we assessed the general usefulness of category-based map projection in reducing the visual complexity of a map that includes diverse types of information. This function was appreciated by users in a previous experiment (Ardissono et al., 2018a) but we wanted to evaluate it extensively.

While the participant carried out a task, the experimenter took notes about how much time (s)he used to answer the questions, sitting at some distance from her/him. We did not put any time restrictions on question answering and we allowed checking the answers multiple times.

#	Question
1	How much familiar are you with the widget that you just used?
2	How much did the widget help you find the information that you were looking for in the map?
3	How much did the widget help you save effort in answering the questions we asked you?
4	Please, rate the ease of use of the widget you just used.
5	Please, rate the novelty of the widget you just used.
6	Did you encounter any difficulties in finding the information that you were looking for?
7	Is there any aspect of the widget you used that you particularly appreciated?

Tab. 7.4: Post-task questionnaire (translated from the Italian language).

As objective performance indicators, we measured task completion time and the percentage of correctly answered questions. As a subjective measure, we analyzed user experience: after the completion of each task, the participant filled in a post-task questionnaire to evaluate the type of widget (s)he had just used; see Table 7.4. For questions 1-5 (s)he had to provide values in a 5-point Likert scale from 1, the worst value, to 5, the best one; questions 6 and 7 were open to free text comments.

After the completion of the four tasks the participant filled in a post-test questionnaire to compare the widgets. We also asked her/him to provide feedback to improve the user experience in OnToMap. For the experiments we used a set of laptops with 15.6" display and 1920x1080 resolution.

7.5 Results

7.5.1 Demographic Data and Background

For the user study, we recruited 62 participants (32.3% females, 66.1% males and 1.6% not declared). Their age is between 20 to 70 years, with a mean value of 33.45. They are part of the University staff (researchers, professors and secretaries) and students, as well as people working in the industry or retired. In the pre-test questionnaire we analyzed their background and familiarity with technology: 41.9% of participants have a scientific background, 29% a technical one, 21% humanities and linguistics, 6.5% economics and law, 1.6% arts. Regarding the education level, 46.8% of them attended the high school, 45.2% the university, 6.5% have a Ph.D and 1.6% attended the middle school. It should be noticed that the target population

Widget type	Min time	Max time	Mean time**	Correct answers*
1: Checkboxes	33	184	94.26	**100.00**%
2: Treemaps	33	180	77.39	98.39%
3: Sunburst	**20**	149	**55.94**	**100.00**%
4: Transparency sliders	23	**146**	57.05	95.16%

Tab. 7.5: Participants' performance during the execution of individual tasks. Time is expressed in seconds and the best values are in boldface. Statistically significant results are marked with ** ($p < 0.001$) or * ($p < 0.002$).

for the OnToMap web application is related to adult people with a middle-high education level and that are familiar in using online platforms. Indeed, 41.9% of people declared that they use e-commerce platforms or online booking services monthly, 38.7% said one or two times per year and 19.4% weekly. Moreover, 56.9% declared that they often use online services based on geographic maps, 17.7% sometimes and 25.8% every day.

7.5.2 User Performance

Table 7.5 shows the results concerning participants' execution time and percentage of correct answers for each task. A Friedman test is used to compare the performance among the four tasks. We choose the Friedman test since the data do not follow a normal distribution. A Friedman test on execution times among the four tasks showed that there is a statistically significant difference between them: $\chi^2(3) = 207.57$, $p < 0.001$. The percentages of correct answers is statistically significant, as well: $\chi^2(3) = 14.14$, $p < 0.002$.

As shown in the table, people achieved the lowest mean execution time and they correctly answered 100% of the questions when they used the widget based on the sunburst diagram. In comparison, when they used the checkboxes, they correctly answered all the questions but they spent the longest time to complete the task. By using the treemaps, participants spent a long time to perform the tasks (almost as long as with checkboxes) but they correctly answered 98.39% of the questions. Finally, they spent relatively little time with transparency sliders but they provided 95.16% correct answers. The high number of correct answers should not surprise because people could check them more than once.

We observed that, in *Task1*, *Task2* and *Task3*, almost all the participants removed some irrelevant data categories using the transparency sliders to reduce map cluttering; then, they used faceted exploration to analyze data. However, they leaned to

Question #	1^{**}	2	3	$4^{\diamond}$	5^{**}
		Task1: Checkboxes			
Mean	**3.90**	4.03	3.77	4.02	2.94
Variance	1.40	1.08	1.39	0.84	1.31
St. Dev.	1.18	1.04	1.18	0.91	1.14
		Task2: Treemap			
Mean	3.32	4.00	**3.95**	3.98	3.48
Variance	1.21	0.66	0.87	0.84	0.84
St. Dev.	1.10	0.81	0.93	0.91	0.92
		Task3: Sunburst			
Mean	2.95	**4.11**	3.84	3.87	**4.10**
Variance	1.62	0.72	1.22	0.84	0.97
St. Dev.	1.27	0.85	1.10	0.91	0.99
		Task4: Transparency sliders			
Mean	3.79	3.85	3.69	**4.31**	3.02
Variance	1.28	1.21	1.20	0.87	1.52
St. Dev.	1.13	1.10	1.10	0.93	1.23

Tab. 7.6: Results of the post-task questionnaire. The best values are shown in boldface. Statistically significant values are marked with $**$ ($p < 0.001$) or $\diamond$ ($p < 0.03$).

use the checkboxes embedded in the transparency sliders instead of using the sliders to tune the opacity of items.

7.5.3 User Experience - Post-task Questionnaire

Table 7.6 shows the results of questions 1-5 of the post-task questionnaire and Table 7.7 shows the results of a Friedman significance test applied to these results.

- **Question 1 (familiarity):** participants were most familiar with the widgets based on checkboxes and, in second position, with the transparency sliders. They were less familiar with the treemaps and much less with the sunburst diagrams ($p < 0.001$).
- **Question 2 (helpfulness):** the results are not statistically significant but the generally high ratings prove that participants perceived all the widgets as helpful to find information items in the maps. The transparency sliders received the lowest ratings.

Question #	Friedman Test
1	$\chi^2(3) = 25.038$, $p < 0.001$
2	$\chi^2(3) = 4.6779$, $p = 0.197$
3	$\chi^2(3) = 1.4063$, $p = 0.7041$
4	$\chi^2(3) = 9.4442$, $p < 0.03$
5	$\chi^2(3) = 43.611$, $p < 0.001$

Tab. 7.7: Statistical significance of the post-task questionnaire results.

- **Question 3 (effort saving)**: the results are not statistically significant; however, similar to Question 2, the transparency sliders are evaluated worse than the other widgets. In this case, ratings show that participants felt that the widgets helped them to save efforts during task execution but values are a bit lower than those of Question 1.
- **Question 4 (ease of use):** participants perceived transparency sliders as the easiest tool, followed by the checkboxes, treemaps and sunburst diagram ($p < 0.03$). This finding is in line with the results of Question 1: even though sliders and checkboxes are in a different preference order, the Pearson Correlation between the answers to Question 1 and Question 4 shows that they are positively correlated both on checkboxes ($\rho = 0.5015$) and on transparency sliders ($\rho = 0.4802$). It should be noticed that transparency sliders are considered easy to use but people do not perceived them as a good tool for effort saving since they only allow to filter the data at the level of data categories.
- **Question 5 (novelty):** participants perceived the widgets based on the treemaps and sunburst diagrams as more innovative than the other ones; they also evaluated the checkboxes as the least innovative one ($p < 0.001$).

About a quarter of the participants answered the free text questions; the percentages reported below refer to this set of people.

- **Question 6 (difficulties):** 50% of the participants who answered this question declared that, due to the amount of textual information displayed in the checkboxes, they had difficulties in the identification of the widgets representing the categories of interest in the sidebar. Some people pointed out that the treemap and the sunburst were new visualization models; thus, they initially had some difficulties in understanding how they worked. A few participants complained about the shortening of facet values in the treemaps because they had to move the mouse over their components to read the information. The

#	Question/statement
1	The widget was familiar to me.
2	The widget helped me find the information I needed.
3	The widget helped me to save effort in answering the questions.
4	The widget was easy to use.
5	The widget is novel.
6	Do you think that using transparency sliders in combination with checkboxes, treemaps or sunburst diagram is useful?
7	Which information exploration widget would you use again in the future?
8	Why?
9	Which information exploration widget did you like the least?
10	Why?

Tab. 7.8: Post-test questionnaire (translated from the Italian language): free text questions.

only observed limitation of the sunburst was that it is visualized in a separate window, partially covering the map.

- **Question 7 (appreciations):** some participants liked the graphics of the treemaps and declared that the size of the components representing facet values provides an intuitive visualization of the cardinality of the corresponding sets of items. About 25% of people perceived the sunburst as good to compactly visualize all the facets and values of a data category. They also appreciated the fact that the sunburst reduces the vertical expansion of the sidebar; thus, it limits the scrolling to reach the widgets of interest. Some participants specified that they liked the correspondence of colors between sliders and items in the map; i.e., color coding. In general, the transparency slider was perceived as useful to reduce information overload by imposing visualization constraints on whole data categories.

7.5.4 User Experience - Post-test Questionnaire

After participants completed the four tasks, we asked them to fill in a post-test questionnaire to capture their overall experience with the widgets.

In the first part of this test we asked them to select the widgets which better matched familiarity, helpfulness, effort saving, easy of use and novelty; see Table 7.8. People could check multiple options in case more than one widget satisfied them; therefore, the percentages reported below may be over 100%. The results of this part of the

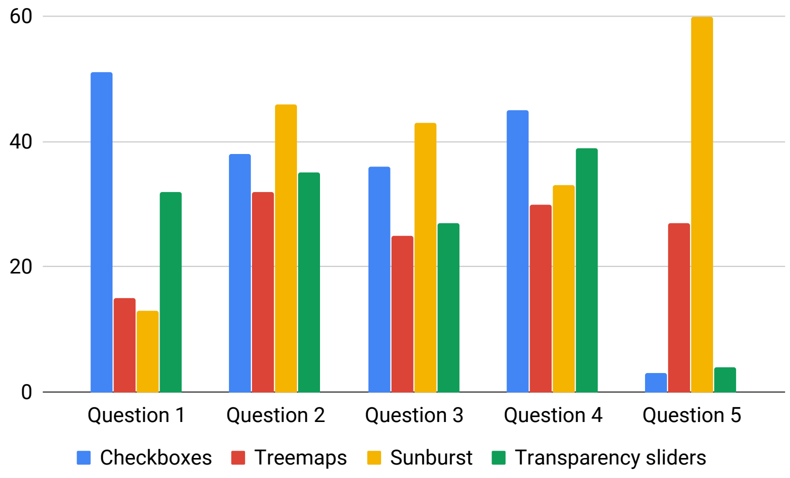


Fig. 7.6: Post-test: evaluations of the questions listed in Table 7.4.

test, shown in Figure 7.6, are consistent with those of the post-task questionnaires. Specifically, they confirm that:

- **Question 1 (familiarity):** people considered the checkboxes as the most familiar widget and they placed transparency sliders in second position.
- **Question 2 (helpfulness):** the sunburst was perceived as more helpful than the other widgets as far as information finding is concerned.
- **Question 3 (effort saving):** the sunburst, followed by the checkboxes, was the preferred widget from the viewpoint of effort saving. This is different from the results of the post-task questionnaires but it should be noticed that those results are not statistically significant.
- **Question 4 (ease of use):** the checkboxes were evaluated as the easiest widget to use, slightly easier than transparency sliders. In the post-task results were reversed but the two widgets were anyway the best rated ones.
- **Question 5 (novelty):** the sunburst was perceived as the most novel widget.

Regarding the second part of the post-test questionnaire (see Table 7.8):

- **Question 6 (transparency sliders with facet-based widget)**: 53.2% of participants declared that the joint usage of transparency sliders with checkboxes, treemaps or sunburst efficiently helps information exploration. They found it convenient to organize search in two steps: (i) visual simplification of maps by hiding the data categories irrelevant to the questions, using transparency sliders; (ii) identification of the items of the category of interest on the basis of their properties, using faceted exploration widgets.

- **Questions 7 and 8 (future usage of widgets and why)**: 60% of people stated that they would use the widget based on the sunburst again because it offers a complete view of each data category. Moreover, 56% declared that they would use the checkboxes again because this is a widespread way to search for information.

- **Questions 9 and 10 (least preferred widget and why):** 34% of participants evaluated the treemaps as the least preferred widget because they are not intuitive and they are difficult to use; 37% did not like the transparency sliders either because they poorly help to solve complex search tasks. 21% of people did not like the sunburst, mostly because it covers part of the map instead of being displayed within the sidebar. Finally, 15% declared that they would not use the checkboxes in the future because they carry a large amount of textual information and it's difficult to identify the relevant values out of it.

7.6 Discussion

The user performance and experience results consistently suggest that the sunburst is the best widget for faceted information exploration. User experience can be explained as follows:

- Regarding the familiarity with the types of widget (Question 1), we expected that people would be more familiar with checkboxes and transparency sliders because they are used to support faceted search in several e-commerce and booking applications while treemaps and sunburst are rarely used outside scientific contexts.

- Question 2 provides some evidence that participants perceived the widget based on the sunburst as the most helpful one (post-test), while transparency sliders were suitable to solve simple search problems because they do not support facet-based map projection (post-task and post-test). People also considered the transparency slider as useful within a facet-based widget (Question 6 - post-test).

- As far as effort saving is concerned (Question 3), the moderate appreciation and the mixed ratings given by participants might be explained by considering that, even though all the widgets support map projection, they require some interaction, which could be perceived as an effort.

- Participants' familiarity with the widgets can explain the fact that they evaluated transparency sliders and checkboxes as the easiest to use tools (Question 4), and treemaps and sunburst as the most novel ones (Question 5). Moreover, the moderate ease of use attributed to treemaps and sunburst can partially depend on the fact that people had to learn how to use them (Question 6 - post-task).

Interestingly, in the answers to the free text questions (Question 6 - post-task and Questions 9 and 10 - post-test) a relevant number of participants criticized the amount of textual information visualized in the checkboxes, complaining that it challenges the identification of the relevant widgets or values in the sidebar. Actually, all the faceted widgets include the same information, generated as described in Section 7.3. Therefore, this comment can be interpreted in a different way, in relation to the lack of compactness of the layout provided by the checkboxes (and presumably also by the treemaps).

The widget based on the treemaps was the least preferred one because it was not particularly intuitive and it was difficult to use (Questions 9 and 10 - post-test). Despite the appeal of its graphics (Question 7 - post-task), it challenges the user with readability issues. Moreover, similar to the checkboxes, it occupies a fairly relevant amount of vertical space in the sidebar (Question 6 - post-task), thus increasing the amount of scrolling needed to inspect the other treemaps.

We conclude that the experimental results help us answer our research questions, which we repeat here for the reader's convenience:

$\mathbf{RQ2_1}$: *How does a compact, graphical view of the visualization constraints applied to a map, impact on her/his efficiency and experience in data exploration?*

The results of the experiment show that not all the facet-based widgets equally helped participants while executing the tasks of the experiment; the reason for this difference is in the capability of the widgets to clearly and compactly describe the search context.

Specifically, the widget based on the sunburst was considered as particularly useful and effective, and it supported the best user performance. This can be explained by the fact that it provides a compact representation of the facets of a category and it supports the readability of their values. Its compactness also supports a concise visualization of widgets in the sidebar; in turn, this reduces scrolling needed to inspect it during faceted search. Conversely, the treemaps challenged participants because their graphical layout hampers readability and their vertical extension excessively increases the length of the sidebar.

Participants appreciated both sunburst and checkboxes; however, when using the latter they completed the tasks slower than with the former. As the main difference between these two widgets is in their vertical extension (much more compact in the sunburst), we can say that this is the main dimension determining the difference in performance.

The transparency sliders achieved the lowest user performance results because they fail to support the specification of fine-grained visualization constraints. However, using the sliders in combination with the other widgets was perceived as a very convenient approach because it enables to first focus the map on the categories of interest, and then further project it by imposing detailed visualization constraints on the remaining items.

$\mathbf{RQ2_2}$: *How much does the user's familiarity with the widgets for faceted exploration impact on her/his efficiency in search and on her/his appreciation of the exploration model they offer?*

The results of the experiment suggest that the familiarity with the widgets does not influence users' efficiency in search: the best performance in task execution was achieved by using the sunburst, which most participants considered as moderately easy to use and they did not know before interacting with OnToMap. Moreover, familiarity positively influences people's disposition towards the faceted exploration widgets and how they perceive the ease of use; see the case of the checkboxes. However, participants appreciated the sunburst as well because it efficiently supports exploration at the expense of some initial learning effort. We can thus conclude that, if the widget is not too difficult to use, the functionality it provides can override the effect of its familiarity on user appreciation.

8 A Compositional Model of Multi-faceted Trust for Personalized Item Recommendation

In this chapter we analyze the impact on the performance of a recommender system by integrating another type of information that aims to represent the concept of multi-faceted users reputation. Differently from the previous chapters, we leverage another type of information. We are interested in grouping and combining different types of feedback to produce metrics of trust that are leveraged to make relevant suggestions to users. First, we introduce a preliminary work done to define the multi-faceted reputation of a user and integrate it into a collaborative filtering model based on K-Nearest Neighbors. Then, we describe the work done to refine and improve the multi-faceted reputation model defined in the preliminary work and we integrated it into a collaborative filtering model based on matrix factorization by extending the LOCABAL algorithm developed by (Tang et al., 2013).

The work described in this section has been published in:

- Mauro, N., Ardissono, L., and Hu, Z. F. (2019b). "Multi-faceted Trust-based Collaborative Filtering". In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '19. Larnaca, Cyprus: ACM, pp. 216–224.

- Ardissono, L. and Mauro, N. (2020). "A Compositional Model of Multi-faceted Trust for Personalized Item Recommendation". In: *Expert Systems with Applications*, p. 112880.

8.1 Introduction

Trust-based recommender systems improve rating prediction with respect to Collaborative Filtering (Desrosiers and Karypis, 2011) by combining rating similarity with the additional information provided by a trust network among users to deal with the cold start problem. Most of these systems predict the rating scores that a person would attribute to items by relying on the observed preferences of the users who are linked to her/him by social relations, directly or through a short path of links, as in SocialMF (Jamali and Ester, 2010). Moreover, having observed that, in the physical world, people are likely to seek advice from both local friends and highly reputable users, some systems also take global reputation into account to improve recommendation performance. For instance, LOCABAL (Tang et al., 2013) computes users' reputation as a function of their importance in the social network and exploits this data to weight the impact of ratings in Matrix Factorization.

Despite the good recommendation results achieved by trust-based recommender systems, recent studies show that they are hardly accepted by people, who are concerned about the storage of personal information and the access to social relations (Burbach et al., 2018). It is thus vital to define trust models that can use data that is not perceived as personal.

For this purpose, we propose a compositional trust model and recommender system which rely on complementary information sources to obtain a twofold objective: (i) collecting rich evidence about user trust to improve Top-N recommendation, and (ii) adapting to possible restrictions on the information that can be used in the application domain of interest.

Indeed, various signs of trust can be used to compute users' global reputation without relying on sensitive information. For instance, social networks such as Booking.com, 2019, Expedia.com, 2001 and Yelp, 2019a publish anonymous feedback about users (expressed as endorsements to their profiles) and about their contributions (e.g., helpfulness of reviews) that can be used to assess reputation by ignoring the identity of the people who provided it. It is thus interesting to define a model that supports the interpretation of these types of feedback as an overall trustworthiness measure.

In this chapter we present the Multi-faceted Trust Model (MTM) as a framework to fuse *local trust between users* (inferred from direct social relations) with the following sources of information:

- The *quality of individual reviews* derived from the explicit feedback they receive from the social network.
- *Multi-dimensional user reputation* derived from the analysis and integration of different types of endorsements that users can receive with the quality of the reviews they author.

MTM makes it possible to separately include or exclude the components of trust to assess their relative impact on recommendation. This supports the evaluation of performance, e.g., when different types of anonymous feedback are considered, and when social relations are ignored.

We integrate MTM into a novel trust-based recommendation algorithm, denoted as LOCABAL+, which combines local trust and multi-dimensional global reputation in preference estimation. We take inspiration from the LOCABAL recommender system because we are interested in modelling the local and global reputation of users in a social network as the authors of LOCABAL. LOCABAL+ extends the LOCABAL recommender system, from which we take inspiration, as follows:

- It tunes Matrix Factorization by exploiting multi-faceted trust, which takes multiple aspects of user behavior into account, instead of only relying on social links.
- It regularizes social relations by means of rating similarity and multi-dimensional global reputation to exploit both properties in the selection of the like-minded users for rating prediction.
- It can be configured to use a subset of the facets of trust.

Experimental results show that LOCABAL+ achieves the best accuracy, error minimization and ranking results when it uses both global trust feedback and social relations. However, it also outperforms state-of-the-art recommender systems based on Matrix Factorization and on K-Nearest Neighbors when it ignores social relations. We thus conclude that multi-faceted trust enhances recommendation performance in trust-based recommenders and it makes them more flexible with respect to the types of information that can be used in a specific application domain.

In the remainder of this chapter, Section 8.2 introduces a preliminary work conducted to understand the impact of MTM in a K-Nearest Neighbours model. Section 8.3 presents our research questions and outlines the experiments to answer them. Sections 8.4 and 8.5 present MTM and LOCABAL+. Section 8.6 describes the datasets used for the experiments and the instantiation of MTM on the available types of information. Section 8.7 presents the validation methodology we used and the evaluation results. Section 8.8 discusses the evaluation results and outlines our future work.

8.2 Preliminary work

In a preliminary work (Mauro et al., 2019b), we proposed a multi-faceted trust model that integrates local trust with global reputation evidence available in social networks and e-commerce sites. We defined four general classes of evidence, which can be mapped to different types of information published by social networks. Specifically, we estimated the multi-faceted global reputation of a user by analyzing the trust statements provided by the other users; e.g., endorsements to her/his public profile and feedback on item reviews. Moreover, we modeled local trust between users by taking into account both direct friends and the relations depending on the existence of implicit user groups, which can be revealed by the presence of relevant numbers of common friends and might denote preference similarity. We tested our model on collaborative recommendation by applying it to a variant of User-to-User Collaborative Filtering (U2UCF) which we denote as Multi-faceted Trust-based Recommender (MTR). MTR can be configured to combine in different ways rating similarity, local trust derived from social relations and multi-faceted global reputation for neighbor detection and rating estimation. MTR also makes it possible to separately include or exclude different facets of trust; thus, it helps understand their impact on accuracy. We compared MTR with U2UCF, and with LOCABAL and TrustMF trust-based recommender systems, which combine rating similarity and trust information in Matrix Factorization. For the experiments, we worked on two datasets: the Yelp one (Yelp, 2019b) and the LibraryThing dataset (Zhao et al., 2019) that stores fewer types of feedback but it publishes more selective friend relations aimed at content sharing. The evaluation provided the following results:

- On the Yelp dataset, multi-faceted trust information helps recommendation performance, probably by complementing the social ties defined by friend relations. The MTR configurations that combine social relations and global multi-faceted reputation outperform U2UCF in accuracy, MRR and diversity of recommendations. Moreover, they outperform LOCABAL and TrustMF in RMSE and MAE. Furthermore, MTR obtains the best accuracy when local trust includes both direct social links and implicit groups of users having a relevant number of common friends. In contrast, in the LibraryThing dataset, LOCABAL obtains the best performance results by inferring users' reputation on the basis of friend relations, and by combining it with rating similarity.
- Profile endorsements and feedback on users' contributions improve recommendation performance: the configurations of MTR ignoring these types of

information (and especially the former) have lower accuracy, MRR and diversity than the other MTR configurations. However, taking profile endorsements into account reduces user coverage.

From this preliminary work, we understood that multi-faceted trust can help recommendation, especially when social relations are weak trust predictors, because it complements them with global reputation data. However, before using this type of information in a social network, an analysis of the type and amount of available trust evidence, as well as of the meaning of social relations, is needed to assess its real impact on recommendation performance. We thus decided to refine the multi-faceted trust model and to integrate it in a Matrix Factorization model in order to improve its performance. This step lead us to the development of the MTM model, described in the following.

8.3 Research Questions and Experimental Plan

In the Multi-faceted Trust Model we integrate diverse facets of trust and, in a specific application domain, one or more of them might not be available or usable. Therefore, besides assessing their overall value in improving Top-N recommendation, we separately study their impact on recommendation performance. We thus formulate the following research questions:

$\mathbf{RQ3_1}$: *Can multi-faceted trust be used to improve the performance of a trust-based recommender system with respect to the standard state-of-the-art trust models that only rely on social links and rating similarity among users?*

In order to answer these questions, we carry out experiments to measure the performance of LOCABAL+ on a spectrum of MTM configurations that tune in different ways the influence of the facets of trust we consider. We compare the performance of the algorithm when using or ignoring social links and trust statements about users and their contributions. We also compare the algorithm with the following baselines: the LOCABAL (Tang et al., 2013) and SocialMF (Jamali and Ester, 2010) trust-based recommender systems based on Matrix Factorization; SVD++ (Koren, 2008) and User-to-User Collaborative Filtering (Desrosiers and Karypis, 2011) which only use rating similarity; a user-to-user Collaborative Filtering algorithm (henceforth denoted as U2USocial) that employs friend relations to estimate ratings in a K-Nearest Neighbors approach, instead of using rating similarity.

We carry out the experiments on two large subsets the Yelp dataset (Yelp, 2019b). The first one, Yelp-Hotel, concerns accommodation facilities; the second one, Yelp-Food, is focused on restaurants.

We evaluate Top-k recommendation performance of algorithms with k=10 by taking the rating scores of the dataset as ground truth. Following the recent trends in the evaluation of recommender systems described in (Jannach et al., 2016), we measure their accuracy, error minimization, ranking capability and user coverage @k; see Section 8.7.1 for details.

8.4 Multi-faceted Trust Model (MTM)

MTM is aimed at computing users' trustworthiness in the context of recommender systems. It integrates local trust between users (inferred from social relations, in line with trust-based recommender systems research) with the public, anonymous feedback received by users and by their contributions in a social network. MTM is compositional and supports the inclusion or exclusion of facets of trust to comply with the requirements of the application domain of interest. We identified the classes of evidence about trust of MTM by analyzing the information publicly provided by social networks and e-commerce sites such as Yelp, 2019a, Booking.com, 2019, Expedia.com, 2001, LibraryThing, 2019, Amazon.com, 2006, Ciao (Danetsoft, 2019) and Epinions, 2019. However, we generalized those indicators to enhance the applicability of our model to heterogeneous domains. In the following we describe each class in detail.

8.4.1 Quality of Individual Contributions on an Item

By individual contribution on an item we mean a piece of information that a user provides about it. A contribution is usually a review associated with a rating score but we describe our model at a more general level because in some online services users can post different types of content; e.g., the Yelp social network allows to write both reviews and tips about items.

Let $\mathcal{U}$ be the set of users and $\mathcal{I}$ the set of items of a service. Given $v \in \mathcal{U}$ and $i \in \mathcal{I}$ we denote an individual contribution provided by v on i as $contr_{vi}$. Then, we define the quality of $contr_{vi}$ ($fContr_{vi}$, in [0, 1]) on the basis of the amount of feedback that $contr_{vi}$ has received. We measure quality in a relative way with respect to the most popular contributions on the same item in order to be robust with respect to item-specific biases:

- In the social networks that provide both positive and negative feedback about contributions we take inspiration from the definition of gold standard helpfulness of reviews defined in (Kim et al., 2006; Raghavan et al., 2012; O'Mahony and Smyth, 2018). In those works, helpfulness is defined as the ratio between the number of positive evaluations ($positiveVotes_{contr_{vi}}$) and the total number of evaluations ($positiveVotes_{contr_{vi}} + negativeVotes_{contr_{vi}}$) that a contribution receives:

$$helpfulness_{contr_{vi}} = \frac{positiveVotes_{contr_{vi}}}{positiveVotes_{contr_{vi}} + negativeVotes_{contr_{vi}}} \tag{8.1}$$

  We define the quality of $contr_{vi}$ in a relative way with respect to the best contribution on the same item as follows:

$$fContr_{vi} = \frac{helpfulness_{contr_{vi}}}{\max\limits_{a \in \mathcal{U}} helpfulness_{contr_{ai}}} \tag{8.2}$$

- In the social networks that only support positive feedback we compute quality as the ratio between the overall number of appreciations obtained by $contr_{vi}$ ($appreciations_{contr_{vi}}$) and the maximum number of appreciations received by the other contributions on the same item:

$$fContr_{vi} = \frac{appreciations_{contr_{vi}}}{\max\limits_{a \in \mathcal{U}} appreciations_{contr_{ai}}} \tag{8.3}$$

To overcome the problem of unpopular items in which the number of feedback that a contribution receives could be low or absent, we plan mas a future work, to define a set of features extracted from the review's text useful to estimate through a regression model the quality of the contributions.

8.4.2 Multi-dimensional Global Reputation

We define *multi-dimensional global reputation* building on heterogeneous types of information about users to capture different aspects of their behavior. Let $\mathcal{U}$ be the set of users, $\mathcal{I}$ the set of items, $v \in \mathcal{U}$ and $i \in \mathcal{I}$. We define the following sub-classes of trust evidence:

(P) *Importance of the user in the social network* (imp_v in [0, 1]), based on her/his social connections. Similar to LOCABAL, we use PageRank (Page et al., 1999) to model this type of indicator. PageRank estimates the relative importance of nodes in a graph by counting the number and quality of the links that enter them, under the assumption that being referenced by others is a quality sign. We compute imp_v as:

$$imp_v = \frac{1}{1 + log(rank_v)} \tag{8.4}$$

where $rank_v \in [1, |U|]$ is the PageRank value of v and the most important user is ranked with 1; see (Tang et al., 2013).

(U) *Global feedback about the user's profile:*

- *User profile endorsements and public recognition* ($fEndors_v$, in [0, 1]). This metric represents the global types of feedback that user profiles receive from the social network. It may have different instances representing individual trust indicators. We consider the appreciations that a user *v* receives from the other members of $\mathcal{U}$ (e.g., "likes"), public assessments of reputation which some social networks grant to their best contributors, and the number of friends, fans, or followers in the social networks that disclose the number but not the identity of users. Similar to the evaluation of the feedback on user contributions, we compute the value of each trust indicator as the ratio between the number of appreciations received by *v*, denoted as $appreciations_v$, and the maximum number of appreciations received by a user $a \in \mathcal{U}$:

$$fEndors_v = \frac{appreciations_v}{\max\limits_{a \in \mathcal{U}} appreciations_a} \tag{8.5}$$

  In this way we are able to assign a value that indicates the importance of each user profile with respect to the profiles of the other users of the community, on the basis of public or anonymous data.

- *Visibility* (vis_v in [0, 1]). This class is aimed at estimating how popular v becomes, thanks to her/his contributions. Intuitively, the visibility describes the impact of the user's contributions in the social network as observed from the feedback they receive. We compute vis_v as the ratio between the number of appreciations received by v and the total number of contributions provided by her/him, normalized by the maximum number of appreciations acquired by the other members of $\mathcal{U}$:

$$vis_v = \frac{appreciations_v}{\max\limits_{a \in \mathcal{U}} appreciations_a * |Contributions_v|} \tag{8.6}$$

where $Contributions_v$ is the set of contributions authored by v.

(Q) *Quality of the user as a contributor* (q_v, in [0, 1]) with respect to the other members of $\mathcal{U}$.

This class is aimed at providing an overall evaluation of the user by considering the feedback received by all her/his contributions. As in the previous cases, we compute quality in a relative way with respect to the best contributor of the social network. Specifically:

- In the social networks that only provide positive feedback about contributions we define q_v as follows:

$$q_v = \frac{\sum\limits_{c_1 \in Contributions_v} appreciations_{c_1}}{\max\limits_{a \in \mathcal{U}} \sum\limits_{c_2 \in Contributions_a} appreciations_{c_2}} \tag{8.7}$$

  where $appreciations_{c_1}$ is the number of appreciations received by contribution c_1 and c_1 is authored by user v (analogously for c_2).

- In the social networks that provide positive and negative feedback about reviews, we apply Equation 8.7 by replacing $appreciations_{c_n}$ with $fContr_{c_n}$ computed according to Equation 8.2; i.e., we compute the relative quality of the user's contributions by taking both the positive and negative votes they receive into account.

The previously described classes of trust evidence are generic and most of them could be mapped to multiple indicators. For instance, Yelp supports different types of endorsements to user profiles, such as "thanks" and "Elite" recognition. In other cases, the social relation between users might be mapped to friends, follower and trust links. In order to obtain a single value representing a user v's multi-dimensional reputation, we fuse these indicators by computing their average, assuming that they additively contribute to increasing *v*'s trustworthiness. Let's consider a set

$\mathcal{F}$ of indicators that are instances of the P, U and Q classes. We define the *multi-dimensional global reputation* of *v*, denoted as mgr_v (in [0, 1]) as:

$$mgr_v = \frac{\sum_{l=1}^{|\mathcal{F}|} C_l * indicator_l}{\sum_{l=1}^{|\mathcal{F}|} C_l} \tag{8.8}$$

where C_l can be set to 1 to take the trust indicator (e.g., social links) into account, 0 otherwise. We assume that each indicator is computed according to the method defined for the class to which it belongs.

It is worth noting that in Equation 8.8 all the indicators have the same weight because, for simplicity, we assume that they equally contribute to *v*'s reputation. In our future work, we plan to carry out a deeper analysis to understand the impact of different weighting schemes on recommendation performance. For this purpose, we will carry out experiments with LOCABAL+ by setting these weights to different values in [0, 1].

8.4.2.1 Multi-faceted Trust

The multi-faceted trust, mft_{vi} (in [0, 1]), describes the overall trust in the rating provided by a user v on an item i, given *v*'s multi-dimensional reputation and the quality of her/his contribution about i. We use multi-faceted trust values in LOCABAL+ to tune the influence of rating scores in the Matrix Factorization process used to learn the latent user and item vectors; see Section 8.5.3. We define mft_{vi} as follows:

$$mft_{vi} = \beta * mgr_v + C(1 - \beta) fContr_{vi} \tag{8.9}$$

where

- mgr_v is v's multi-dimensional reputation; see Equation 8.8.
- $fContr_{vi}$ is the quality of the contribution provided by v on item i and is computed according to Equations 8.2 or 8.3, depending on the type of feedback (positive/negative) that contributions can receive.
- β takes values in the [0, 1] interval and balances the relative weight of mgr_v and $fContr_{vi}$ in the computation of mft_{vi}. The higher β, the stronger is the impact of multi-dimensional reputation on trust.
- C can be either 0 or 1 and is used to ignore or use the feedback on contributions in the evaluation of mft_{vi}; by default, $C = 1$.

As discussed below in Section 8.7, the best configuration of the β parameter in Equation 8.9 depends on the dataset to which the trust model has to be applied and it can be empirically found by using the MTM model within a recommender system (LOCABAL+ in our case) and checking its performance. In the datasets we have considered, the best values are somehow low (e.g., 0.1 or 0.3), which means that the global feedback on user contributions, represented by $fContr_{vi}$, is very useful to steer recommendation.

8.5 Recommendation Model

We describe LOCABAL+ incrementally, starting from the main concepts that characterize the LOCABAL trust-based recommender system.

8.5.1 Basic Collaborative Filtering with Matrix Factorization

Basic Collaborative Filtering builds on the assumption that, if people rated items similarly in the past, they will do it again in the future. Thus, it uses rating similarity in preference estimation. The algorithms based on Matrix Factorization assume that a few latent patterns influence rating behavior and they perform a low-rank matrix factorization on the users-items rating matrix; e.g., see SVD++ (Koren, 2008). Given the following notation:

- $\mathcal{U} = \{u_1, \ldots, u_n\}$ is the set of users and $\mathcal{I} = \{i_1, \ldots, i_m\}$ is the set of items.
- $\mathbf{R} \in \mathbb{R}^{n \times m}$ is the users-items rating matrix.
- $\mathbf{R}_{xy}$ is the rating score given by user $u_x \in \mathcal{U}$ to item $i_y \in \mathcal{I}$, if any:
  - $\mathcal{O} = \{< u_x, i_y > | \mathbf{R}_{xy} \neq 0\}$ is the set of known ratings (ground truth)
  - $\mathcal{T} = \{< u_x, i_y > | \mathbf{R}_{xy} = 0\}$ is the set of unknown ratings

Assuming K latent factors, $\mathbf{u}_x \in \mathbb{R}^K$ denotes the user preference vector of u_x and $\mathbf{i}_y \in \mathbb{R}^K$ denotes the item characteristic vector of i_y.

In order to learn these vectors, the recommender system solves the following optimization problem:

$$\min_{\mathbf{U},\mathbf{I}} \sum_{<u_x,i_y>\in\mathcal{O}} (\mathbf{R}_{xy} - \mathbf{u}_x^T \mathbf{i}_y)^2 + \lambda(||\mathbf{U}||_F^2 + ||\mathbf{I}||_F^2) \tag{8.10}$$

where

- $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \in \mathbb{R}^{K \times n}$ and $\mathbf{I} = [\mathbf{i}_1, \ldots, \mathbf{i}_m] \in \mathbb{R}^{K \times m}$.
- $||.||_F$ denotes the Frobenius Norm and $||\mathbf{U}||_F^2 + ||\mathbf{I}||_F^2$ are the regularization terms to avoid over-fitting.
- $\lambda > 0$ controls the impact of **U** and **I** on regularization.

8.5.2 LOCABAL

LOCABAL (Tang et al., 2013) extends Collaborative Filtering based on Matrix Factorization in two ways:

1. It exploits a user's local social context to learn her/his preference vector by considering both rating similarity and social relations, regularized on the basis of the former. In this way, rating estimation can benefit from the contribution of users who are socially related to the current user but, at the same time, have similar preferences as her/him.
2. It relies on the user's global social context, represented by her/his reputation, to weight the contribution of rating similarity in Matrix Factorization. Global reputation is computed using PageRank as described in Section 8.4.2, Equation 8.4, page 120.

In detail:

- Let $\mathbf{T} \in \mathbb{R}^{n \times n}$ be the users-users social relation matrix. $\mathbf{T}_{uz} \neq 0$ denotes the existence of a direct social link between $u_x \in \mathcal{U}$ and $u_z \in \mathcal{U}$. Zero values mean that users are not socially related.
- Let $\mathcal{N}_x = \{u_z \mid \mathbf{T}_{xz} = 1\}$ be the set of u_x's direct social links.
- Let $\mathbf{S} \in \mathbb{R}^{n \times n}$ be a users-users trust matrix whose cells represent the strength of the social relations between users, depending on their rating similarity. For $u_z \in \mathcal{N}_x$, $\mathbf{S}_{xz} = \sigma(u_x, u_z)$, where $\sigma(u_x, u_z)$ is the Cosine similarity of u_x and u_z's rating vectors. Homophily indicates that users with similar tastes are more likely to be socially connected, and social influence suggests that users that are socially connected are more likely to share similar tastes. Since users with strong ties are more likely to share similar tastes than those with weak ties, treating all social relations equally is likely to lead to degradation in recommendation performance (Tang et al., 2012). These observations suggest that we should consider heterogeneous strengths when exploiting local social context for recommendation. In our model, we simply use the rating cosine

similarity to measure the social relation strength, although there are other more sophisticated measures in (Xiang et al., 2010).

LOCABAL solves the following optimization problem:

$$\min_{\mathbf{U},\mathbf{I},\mathbf{H}} \sum_{<u_x,i_y>\in\mathcal{O}} w_x(\mathbf{R}_{xy} - \mathbf{u}_x^T\mathbf{i}_y)^2 + \alpha \sum_{x=1}^{n} \sum_{u_z\in N_x} (\mathbf{S}_{xz} - \mathbf{u}_x^T\mathbf{H}\ \mathbf{u}_z)^2 + \lambda(||\mathbf{U}||_F^2 + ||\mathbf{I}||_F^2 + ||\mathbf{H}||_F^2) \tag{8.11}$$

where

- w_x in [0, 1] is u_x's global reputation computed by applying Equation 8.4. This weight tunes the contribution given by rating similarity so that highly reputable users influence the Matrix Factorization process more strongly than the other ones.
- $\alpha >= 0$ tunes the contribution given by u_x's local social context.
- $\mathbf{H}\in \mathbb{R}^{K\times K}$ captures user preference correlation: if u_x and u_z are strongly connected in $\mathbf{S}_{xz}$, then their preferences should be tightly correlated via **H**. We remind that K is the number of latent factors.
- $\lambda >= 0$ controls the impact of **U**, **I** and **H** on regularization.

As discussed in Section 3, this algorithm estimates trust by relying on social links. If this information is not available, LOCABAL cannot be applied or it reduces to SVD (Koren et al., 2009), by setting $w_x = 1$ and ignoring the trust matrix that cannot be computed. Our MTM model is aimed at providing a more general solution, which can be applied to complementary types of evidence about trust as can be found in social networks.

8.5.3 LOCABAL+

LOCABAL+ extends LOCABAL in two ways:

1. It models global context by taking multi-faceted trust into account.
2. It tunes social regularization on the basis of both rating similarity and multi-dimensional global reputation.

We consider the following optimization problem to be solved in order to learn the user preference and item characteristic vectors:

$$\min_{\mathbf{U},\mathbf{I},\mathbf{H}} \sum_{<u_x,i_y>\in\mathcal{O}} mft_{xy}(\mathbf{R}_{xy} - \mathbf{u}_x^T\mathbf{i}_y)^2 + \alpha\sum_{x=1}^{n}\sum_{u_z\in N_x} mgr_z(\mathbf{S}_{xz} - \mathbf{u}_x^T\mathbf{H}\,\mathbf{u}_z)^2 + \lambda(||\mathbf{U}||_F^2 + ||\mathbf{I}||_F^2 + ||\mathbf{H}||_F^2) \quad (8.12)$$

where:

- mft_{xy} represents the multi-faceted trust towards user u_x in the context of item i_y; see Equation 8.9 in page 122. This weight tunes the estimation of ratings in the Matrix Factorization process by taking users' global reputation and quality of contributions into account; i.e., by looking at users from a broad perspective on their behavior. We assume that we can estimate missing ratings more precisely by giving more importance to the ratings authored by users whose multi-faceted trust is high.

- $\mathbf{S} \in \mathbb{R}^{n\times n}$ is a users-users trust matrix such that, for $u_z \in \mathcal{N}_x$, $\mathbf{S}_{xz}$ is set to the Pearson Correlation similarity (PC) of u_x and u_z's rating vectors, limited to the set of items rated by both users:

$$PC(u_x, u_z) = \frac{\sum\limits_{i_y\in\mathcal{I}_{xz}} (r_{xy} - \bar{r}_x)(r_{zy} - \bar{r}_z)}{\sqrt{\sum\limits_{i_y\in\mathcal{I}_{xz}} (r_{xy} - \bar{r}_x)^2 \sum\limits_{i_y\in\mathcal{I}_{xz}} (r_{zy} - \bar{r}_z)^2}} \quad (8.13)$$

  where $\mathcal{I}_{xz}$ is the set of items rated by both u_x and u_z, r_{xy} is the rating given by u_x to i_y (and analogously for r_{zy}), $\bar{r}_x$ ($\bar{r}_z$) is the mean value of u_x's (u_z's) ratings.

  As suggested in (Ricci et al., 2011), we use Pearson Correlation similarity, instead of the Cosine similarity used in LOCABAL, because the latter does not consider the differences in the mean and variance of the ratings made by u_x and u_z. Pearson similarity removes the effects of mean and variance.

- mgr_z is the multi-dimensional global reputation of u_z and tunes preference correlation in the **H** matrix (which depends on rating similarity, given **S**) on the basis of u_z's multi-dimensional global reputation. By adding the mgr_z factor we impose that, the more reputable are u_x's friends, the higher impact they have in the estimation of their own similarity with u_x. Therefore, highly trustworthy users influence social regularization more than the others.

As mft_{xy} and mgr_z are based on a compositional model, they can be computed by using a subset of the trust facets considered so far; e.g., by ignoring social links, feedback on user profiles or feedback about user contributions. In those cases, LOCABAL+ runs with a lower amount of information about users but it can still work as a trust-based recommender system. Specifically, in the experiments we carried out, the algorithm reaches satisfactory performance results also with partial evidence about trust; see Section 8.7.

Obviously, the flexibility of MTM comes with a cost, i.e., the effort needed to map the facets of trust to the types of information available in the application domain in which the recommender system is used. This effort consists of understanding the semantics of the evidence about trust (e.g., types of feedback that are provided by users) and choosing the corresponding classes of trust in MTM. However, as previously discussed, we defined these classes by analyzing several social networks to abstract from the particular types of information they offer and to model trust in a general way. The next section describes the datasets we used for our experiments and the mappings we defined to apply LOCABAL+ to these datasets. Moreover, it sketches the work that should be done to map MTM to a different type of social network in order to give the reader a broader idea of the work to be done.

8.6 Datasets

For our experiments we use two subsets of the (Yelp, 2019b) dataset to analyze data about user behavior in different domains: accommodations versus restaurants.

The Yelp dataset contains information about the users of the social network and about a large set of businesses including food, accommodation, transportation, health, education and so forth. Yelp members can establish bidirectional friend relations to share posts; moreover, they can establish stricter unidirectional fan (follower) relations to get access to the contributions provided by other users. The dataset is structured as follows:

- Each item (business) is associated with a list of tags representing the categories[1] to which it belongs; e.g., a restaurant might be associated with the "Indian" tag to specify the type of cuisine it offers.
- Each item is associated with the rating scores and with textual reviews and tips provided by the members of Yelp. Every user can post a contribution

[1]The full list of Yelp categories is available at `https://www.yelp.com/developers/documentation/v3/category_list`.

(including review+rating, and possibly tip) on the same item. Item ratings take values in a [1,5] Likert scale where 1 is the worst value and 5 is the best one.

- User contributions are associated with the appreciations they receive from Yelp members; i.e., “useful”, “funny” and “cool” for reviews, “like” for tips.
- The dataset publishes explicit friend relations but it only provides the number of fans of each user (i.e. differently from friend relations). Therefore, only the former data can be used to infer direct *trust-alike* relations among users; see Section 3.5.2.
- The dataset publishes various types of endorsement that user profiles can receive: e.g., every year Yelp rewards its most valuable contributors by attributing them the status of Elite users. Moreover, each user profile can receive *compliments* by other Yelp users; e.g., “write more”, “thanks” and “great writer”.

Notice that both compliments and appreciations represent positive feedback about users and contributions; moreover, the dataset reports the number of compliments and appreciations but not the identities of the people who provided them. This type of feedback thus represents an important anonymous source of trust information that can be used by a recommender system.

8.6.1 Yelp-Hotel

Yelp-Hotel is obtained by filtering the complete Yelp dataset on users who provided at least 10 ratings and on businesses tagged with at least one category associated with accommodation facilities. The tags used to filter the dataset are: Hotels, Mountain Huts, Residences, Rest Stops, Bed & Breakfast, Hostels, Resorts.

Table 8.1 provides information about this dataset. It can be noticed that user profiles receive various types of feedback; e.g., the median number of Elite years is 4 and the median number of compliments to user profiles is 110. Also anonymous fans contribute to global reputation (median = 41). Moreover, the dataset contains a relatively high amount of feedback about user contributions: the median number of appreciations is 53 for reviews and 0 for tips. The number of compliments, fans, appreciations, etc. reaches very high values in some cases: for each type of feedback, the distribution of individuals (users or contributions) has a long tail. Both the users-items and the users-users friends matrices are sparse.

Measure				Value
#Users				654
#Items (businesses)				1081
#Ratings				10081
#Friend relations				11554
Sparsity of users-items rating matrix				0.9857
Sparsity of users-users friends matrix				0.9729
Measure	**Min**	**Max**	**Mean**	**Median**
#Elite years of individual users	0	13	4.4052	4
#Compliments received by individual users	0	45018	724.1177	110
#Fans of individual users	0	1803	91.9740	41
#Appreciations on reviews provided by ind. users	0	5194	112.7064	53
#Appreciations on tips provided by ind. users	0	152	2.5107	0
#Appreciations received by individual reviews	0	559	3.9897	3
#Friends of individual users	0	224	17.6667	7

Tab. 8.1: Statistics about the Yelp-Hotel dataset.

8.6.2 Yelp-Food

Yelp-Food, about 10 times larger than Yelp-Hotel, is obtained by filtering the complete Yelp dataset on users who provided at least 10 ratings and on businesses located in the cities of Phoenix, Toronto, Pittsburgh which are tagged with at least one category describing a type of restaurant (e.g., "Indian" and "Italian") for a total of 85 categories[2].

As shown in Table 8.2, the median number of compliments, fans and appreciations is very low and it reaches higher values only for the number of appreciations on reviews provided by individual users. Moreover, for each type of feedback, the distributions of users and reviews have long tails.

[2]The selection of businesses to define the Yelp-Food is based on the following tags: American, Argentine, Asian Fusion, Australian, Austrian, Bangladeshi, Belgian, Brasseries, Brazilian, British, Cambodian, Cantonese, Catalan, Chinese, Conveyor Belt Sushi, Cuban, Czech, Delis, Empanadas, Falafel, Filipino, Fish & Chips, French, German, Greek, Hawaiian, Himalayan/Nepalese, Hot Pot, Hungarian, Iberian, Indian, Indonesian, Irish, Italian, Japanese, Japanese Curry, Korean, Latin American, Lebanese, Malaysian, Mediterranean, Mexican, Middle Eastern, Modern European, Mongolian, New Mexican Cuisine, Noodles, Pakistani, Pan Asian, Persian/Iranian, Peruvian, Piadina, Pizza, Poke, Polish, Polynesian, Portuguese, Ramen, Russian, Salad, Scandinavian, Scottish, Seafood, Shanghainese, Sicilian, Singaporean, Soup, Southern, Spanish, Sri Lankan, Steakhouses, Sushi Bars, Syrian, Tacos, Tapas Bars, Tapas/Small Plates, Teppanyaki, Tex-Mex, Thai, Turkish, Ukrainian, Vegan, Vegetarian, Vietnamese, Wraps.

Measure				Value
#Users				8432
#Items (businesses)				8157
#Ratings				198759
#Friend relations				160891
Sparsity of users-items rating matrix				0.9971
Sparsity of users-users friends matrix				0.9977
Measure	**Min**	**Max**	**Mean**	**Median**
#Elite years of individual users	0	13	1.4154	0
#Compliments received by individual users	0	24635	55.0733	4
#Fans of individual users	0	1803	10.1950	2
#Appreciations on reviews provided by ind. users	0	9023	81.4163	27
#Appreciations on tips provided by ind. users	0	154	0.3876	0
#Appreciations received by individual reviews	0	642	3.4539	1
#Friends of individual users	0	1231	19.0810	4

Tab. 8.2: Statistics about the Yelp-Food dataset.

8.6.3 Trust indicators for both datasets

Let $\mathcal{U}$ and $\mathcal{I}$ be the sets of users and items of the dataset; let $u, v \in \mathcal{U}$ and $i \in \mathcal{I}$. We define the following trust indicators:

- *Quality of individual contributions on an item* ($fContr_{vi}$, in [0, 1]).

  For this class of trust evidence we apply Equation 8.3, which is suitable for positive-only feedback. We map $appreciations_{contr_{vi}}$ to the possibly different types of feedback that a contribution $contr_{vi}$ can receive:

  - $appreciations_{contr_{vi}} = useful_{contr_{vi}} + fun_{contr_{vi}} + cool_{contr_{vi}}$ for item reviews
  - $appreciations_{contr_{vi}} = like_{contr_{vi}}$ for tips

  where $useful_{contr_{vi}}$ is the number of "useful" appreciations received by $contr_{vi}$ (a review), fun is a shortener for "funny" and so forth.

- *Multi-dimensional global reputation* (mgr_v in [0, 1]).

Following the approach described in Section 8.4.2, we compute the multi-dimensional global reputation of a user v by fusing in Equation 8.8 the indicators described in the remainder of this section:

$$mgr_v = \frac{C_1 imp_v + C_2 elite_v + C_3 lup_v + C_4 opLeader_v + C_5 vis_v + C_6 q_v}{C_1 + C_2 + C_3 + C_4 + C_5 + C_6} \tag{8.14}$$

(P) *Importance of the user in the social network* (imp_v in [0, 1]).

In order to compute the PageRank score of users, we transform each bidirectional friend relation into two unidirectional social links. In this way, we can apply the approach described in Section 8.4.2 and Equation 8.4 to compute reputation on the basis of the connections among the users of the social network.

(U) *Global feedback on the user's profile*.

We consider the following trust indicators:

- $elite_v$ (in [0, 1]). We map the number of years in which v has the Elite status to $appreciations_v$ in Equation 8.5:

$$elite_v = \frac{\#EliteYears_v}{\max\limits_{a \in \mathcal{U}} \#EliteYears_a} \tag{8.15}$$

- lup_v (degree of liking of user profile, in [0, 1]). We map the number of compliments ("more", "thanks" - $thks$, "great writer" - gw) received by v to $appreciations_v$ in Equation 8.5:

$$lup_v = \frac{more_v + thks_v + gw_v}{\max\limits_{a \in \mathcal{U}}(more_a + thks_a + gw_a)} \tag{8.16}$$

where $more_v$ is the number of "more" compliments received by v, and similar for the other variables.

- $opLeader_v$ (opinion leader degree, in [0, 1]). The number of anonymous fans of a user v, $fans_v$, can be interpreted as a global recognition of her/his profile. We thus map this number to $appreciations_v$ in Equation 8.5:

$$opLeader_v = \frac{fans_v}{\max\limits_{a \in \mathcal{U}} fans_a} \tag{8.17}$$

– vis_v (visibility, in [0, 1]). We map the number of compliments received by v to $appreciations_v$, and the reviews and tips ($Revs_v \cup Tips_v$) authored by v to $Contributions_v$ in Eq. 8.6:

$$vis_v = \frac{more_v + thks_v + gw_v}{\max\limits_{a \in \mathcal{U}}(more_a + thks_a + gw_a)|Revs_v \cup Tips_v|} \tag{8.18}$$

(Q) *Quality of the user as a contributor* (q_v in [0, 1]).

We assume that the quality of a contributor depends on both the reviews and tips authored by her/him. Therefore, for this indicator, we map $Contributions_v$ to the reviews and the tips provided by v. Moreover, we map $appreciations_c$ to the amount of feedback obtained by these contributions:

$$q_v = \frac{\sum\limits_{c_1 \in Revs_v \cup Tips_v} useful_{c_1} + fun_{c_1} + cool_{c_1} + like_{c_1}}{\max\limits_{a \in \mathcal{U}} \sum\limits_{c_2 \in Revs_a \cup Tips_a} useful_{c_2} + fun_{c_2} + cool_{c_2} + like_{c_2}} \tag{8.19}$$

8.6.4 Instantiation of MTM in a different application domain

Let's consider, as a further example of instantiation of MTM, the LibraryThing, 2019 social network that publishes information about books. LibraryThings enables its members to create their own virtual libraries and to tag and review books. Users can establish friend relations to watch and take inspiration from the libraries created by other people; moreover, they can visualize the reviews published in the social network and they can express positive-only feedback about the helpfulness of each review. Users are not enabled to endorse other users' profiles. LibraryThing discloses the social relations among users and the number of helpfulness votes received by each review. Trust indicators can be mapped to MTM trust classes as follows:

- *Quality of individual contributions on an item* ($fContr_{vi}$). For this class of trust evidence we apply Equation 8.3 of page 119, which is suitable for positive-only feedback. For each review r published in the social network, we thus map $appreciations_r$ to the number of helpful votes that r has received.
- *Multi-dimensional global reputation* (mgr_v). We compute mgr_v by fusing in Equation 8.8 the imp_v and q_v indicators respectively describing *v*'s importance in the social network and her/his quality as a contributor:

$$mgr_v = \frac{C_1 imp_v + C_6 q_v}{C_1 + C_6} \tag{8.20}$$

We can compute imp_v as v's PageRank score by transforming bidirectional friend relations to pairs of unidirectional social links. Moreover, q_v can be defined as the ratio between the total number of helpful votes received by *v*'s reviews and the maximum number of helpful votes received by the other members of the social network.

8.7 Validation of LOCABAL+

8.7.1 Evaluation Metrics

As mentioned in Section 8.3, we evaluate recommendation algorithms on the basis of *accuracy* and *error minimization* (i.e., the ability to provide correct results), *ranking capability* (i.e., the ability to correctly sort items depending on their ground truth relevance to the user) and *user coverage* (i.e., the percentage of users for whom the recommender is able to find items that are likely to be relevant). This is in line with the recent trends in the evaluation of recommender systems, which do not exclusively focus on accuracy to provide a broader view on performance; e.g., see (Jannach et al., 2016). Before describing the evaluation metrics in detail we introduce the notation we use:

- U is the set of users and I the set of items; R is the set of ground truth ratings and $\hat{R}$ the set of estimated ones.
- r_{ui} is the rating score that $u \in U$ has given to $i \in I$ and $\hat{r}_{ui}$ is the rating score estimated by the recommender system.
- $Relevant_u$ is the set of items that u has positively rated; in a [1, 5] Likert scale we define $Relevant_u = \{i \in I \mid r_{ui} > 3\}$.
- $Recommended_u$ is the set of items that the system suggests to u: $Recommended_u = \{i \in I \mid \hat{r}_{ui} > 3\}$.

We evaluate recommendation accuracy and error minimization by means of the following metrics where k represents number of the top-k recommendation taken into account to compute the metrics:

- Precision: P@k $= \frac{1}{|U|} \sum\limits_{u \in U}$ P$_u$@k, where P$_u$@k $= \frac{|Recommended_u \cap Relevant_u|}{|Recommended_u|}$
- Recall: R@k $= \frac{1}{|U|} \sum\limits_{u \in U}$ R$_u$@k, where R$_u$@k $= \frac{|Recommended_u \cap Relevant_u|}{|Relevant_u|}$
- Accuracy: F1@k $= 2 * \frac{\text{P@k * R@k}}{\text{P@k + R@k}}$

- Root Mean Squared Error: RMSE@k $= \sqrt{\frac{1}{|\hat{R}@\mathrm{k}|} \sum_{\hat{r}_{ui} \in \hat{R@\mathrm{k}}} (r_{ui} - \hat{r}_{ui})^2}$
- Mean Absolute Error: MAE@k $= \frac{1}{|\hat{R}@\mathrm{k}|} \sum_{\hat{r}_{ui} \in \hat{R@\mathrm{k}}} |r_{ui} - \hat{r}_{ui}|$

As far as ranking capability is concerned we use the following metrics:

- Mean Reciprocal Rank, which measures the placement of the first relevant items in recommendation lists:

  MRR@k $= \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u}$, where $rank_u$ is the position of the first relevant item in the list generated for user u.

- Mean Average Precision, which measures the average correct positioning of items in the recommendation lists:

  MAP@k $= \frac{1}{|U|} \sum_{u \in U} \frac{1}{|Relevant_u|} \sum_{x=1}^{\mathrm{k}} \mathrm{P}_u@\mathrm{x} * Rel_u(x)$

  where $Rel_u(x) = 1$ if the item in position x of the list for u is relevant to her/him, 0 otherwise.

Finally, we measure User Coverage (shortened to UCov in the tables showing the evaluation results) as the percentage of users of the dataset for whom the algorithm finds at least one item $i \in I$ such that $\hat{r}_{ui} > 3$, i.e., an item that the system evaluates as relevant to the user.

8.7.2 Methodology Applied in the Experiments

We consider various configurations of MTM to evaluate the performance of LOCABAL+ when using all the facets of trust available in the YELP-Hotel/Yelp-Food datasets, or a subset of them. We are interested in understanding whether the algorithm can provide good recommendation results when we omit different sources of evidence in order to assess its applicability to social networks that disclose different types of information about users. Specifically, we consider the following cases, summarized in Table 8.3:

- LOCABAL+. This is the algorithm applied to the complete information available in the dataset (social relations, feedback about users and feedback about contributions). It computes multi-dimensional global reputation with $C_1 = \cdots = C_6 = 1$ in Equation 8.14, and multi-faceted trust for item i with $C = 1$ in Equation 8.9.

Configuration	Social relations	Feedback on user profiles	Feedback on user contributions
LOCABAL+	Yes	Yes	Yes
LOC+noF	Yes	Yes	-
LOC+noE	Yes	-	Yes
LOC+noS	-	Yes	Yes

Tab. 8.3: Configurations of LOCABAL+ used in the experiments.

- LOC+noF. This configuration ignores the feedback about reviews and tips; thus, it only relies on multi-dimensional global reputation, which is computed by taking social links and global feedback on user profiles into account. In detail, LOC+noF is obtained by switching off the quality of the user as a contributor (q_v) in the computation of multi-dimensional global reputation ($C_6 = 0$ in Equation 8.14) and the feedback received by the specific contribution ($fContr_{vi}$) in the computation of multi-faceted trust for item i ($C = 0$ in Equation 8.9). LOC+noF is useful to understand whether, by only using social information and anonymous feedback on user profiles, the recommender system is able to generate useful suggestions.
- LOC+noE. This configuration ignores the global feedback on user profiles, i.e., the trust indicators of class U in Section 8.4.2 (user profile endorsements and public recognition, visibility) in the computation of multi-dimensional global reputation (i.e., $C_2 = C_3 = C_4 = C_5 = 0$ in Equation 8.14). LOC+noE is particularly interesting because not all of the social networks manage profile endorsements; e.g., we mentioned in Section 8.6 that LibraryThing does not support this type of feedback. Therefore we are interested in understanding whether the recommender system can achieve good performance by only relying on social links and feedback on user contributions.
- LOC+noS. This configuration ignores social relations; it is obtained by switching off the importance of users (imp_v) in the computation of multi-dimensional global reputation ($C_1 = 0$ in Equation 8.14) and the social regularization component of LOCABAL+ ($\alpha = 0$ in Equation 8.12). LOC+noS helps understand whether, thanks to the exploitation of public, anonymous feedback about users and user contributions, LOCABAL+ can generate good recommendations in the application domains where the information about social links is unavailable. As previously discussed, this is an important aspect for the applicability of trust-based recommender systems, given the growing sensibility of users towards privacy protection.

	α	β	P	R	F1	MAP	RMSE	MAE	MRR	UCov
Significance	-	-	0.01	-	0.01	†	0.02	0.02	-	⋄
LOCABAL+	0.9	0	0.7919	0.7389	**0.7645**	**0.5303**	0.8922	0.671	**0.6125**	0.7543
LOC+noF	0.1	0.3	0.7923	0.7381	0.7642	0.5288	0.8927	0.6708	0.6087	0.7523
LOC+noE	0.3	0.3	0.7923	0.7377	0.764	0.5285	0.8938	0.6716	0.6086	0.7517
LOC+noS	-	0.1	**0.7931**	0.7368	0.7639	0.5274	**0.8916**	**0.6702**	0.6082	0.7513
U2UCF	-	-	0.76	**0.7399**	0.7498	0.5215	0.9582	0.7264	0.5982	0.6127
SocialMF	-	-	0.7757	0.7261	0.7501	0.5116	0.9238	0.6954	0.6055	0.7655
LOCABAL	0.1	-	0.7732	0.7259	0.7488	0.5112	0.9281	0.6994	0.6078	0.7698
SVD++	-	-	0.7595	0.717	0.7376	0.4994	0.976	0.7383	0.5993	**0.7755**
U2USocial	-	-	0.7503	0.7233	0.7366	0.4798	1.0085	0.773	0.5336	0.2589

Tab. 8.4: Performance@10 on Yelp-Hotel dataset (the best results are in boldface). The "⋄" symbol means that results of each configuration of LOCABAL+ are significant at $p < 0.05$ with respect to all the baselines except for SocialMF; the "†" means that results are significant at $p < 0.01$ with respect to all the baselines except for U2UCF.

In the experiments we use the Surprise (Hug, 2017) implementation of U2UCF and SVD++ and the RecQ implementation of LOCABAL (Coder-Yu, 2019). LOCABAL+ and U2USocial are developed by extending the implementations of LOCABAL and U2UCF respectively. All the algorithms are integrated in Surprise to uniformly evaluate their performance.

On each dataset we organize the evaluation as follows: we first validate the algorithms on 90% of the dataset by running Grid Search to find the best configuration of parameters with respect to MAP, using 5 cross-fold validation. All the executions are performed having set 50 latent factors. Then we additionally test the best configuration obtained from Grid Search on the remaining 10% of the dataset to measure the performance of the algorithms on new data in order to check their impact in a dynamic environment where new ratings are continuously provided.

8.7.3 Evaluation Results

8.7.3.1 Yelp-Hotel

Table 8.4 compares the performance achieved by each configuration of LOCABAL+ to that of the baselines (U2UCF, SocialMF, LOCABAL, SVD++ and U2USocial) on Yelp-Hotel. In this table, as well as in the following ones, the best values are in boldface. The significance level of results, reported in the second row, is obtained by separately comparing each configuration with all the baselines or viceversa in the case that the baselines have better performance. The rows describing performance are sorted by MAP, from the best one to the worst one, in order to highlight the ranking capabilities of the algorithms.

As shown in the central portion of the table, in this dataset the LOCABAL+ configurations outperform the baselines in all measures except for (i) Recall that is dominated by U2UCF, and (ii) User Coverage, where SVD++ is the best algorithm, followed by LOCABAL and SocialMF. The loss in user coverage is however compensated by higher accuracy and ranking capability as LOCABAL+ is the best algorithm in terms of F1, MAP and MRR. Noticeably, LOC+noS is the most precise algorithm, excelling in P, RMSE and MAE. In the following we analyze the LOCABAL+ configurations.

LOCABAL+ obtains its best performance with $\alpha = 0.9$ and $\beta = 0$. The value of α shows that the algorithm strongly relies on users' multi-dimensional reputation to steer social regularization: instead of minimizing the impact of the local social context of users ($\mathbf{S}_{xz}$ in Equation 8.12), as done in LOCABAL (where $\alpha = 0.1$ uniformly flattens the impact of the local social context across users), LOCABAL+ tunes social regularization on the basis of the preferences of the most similar and reputable friends. Differently, $\beta = 0$ means that the multi-faceted trust mtf_{xy} that tunes the impact of ratings in Matrix Factorization is computed by ignoring users' reputation; therefore, for this purpose, the algorithm only relies on the feedback ($fContr_{vi}$) received by the reviews and tips associated with the ratings. This is different from LOCABAL, which tunes the impact of ratings on the basis of users' PageRank score. In summary, in Yelp-Hotel, LOCABAL+ steers social regularization by multi-dimensional reputation and weights the impact of ratings on the basis of the publicly recognized value of user contributions, which emerges as a good source of information to identify reliable ratings.

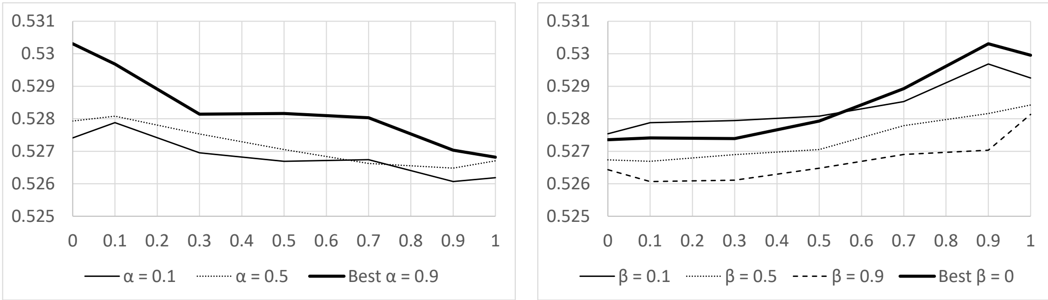
It should, however, be noticed that, in LOCABAL+, multi-dimensional global reputation is computed by taking multiple types of trust evidence into account, i.e., PageRank score, user profile endorsements and quality of the user as a contributor that, in turn, derives from the feedback on contributions. Therefore, it is difficult to say which type of evidence brings the most useful information. In order to clarify the situation we analyze the other configurations of LOCABAL+.

LOC+noF ignores the feedback on user contributions and is optimized with $\alpha = 0.1$ and $\beta = 0.3$. It has lower performance than LOCABAL+ but it outperforms all the baselines in Precision, F1, MAP, RMSE, MAE and MRR. The value of α dramatically weakens the role of social regularization in the Matrix Factorization process with respect to LOCABAL+ (it is flattened to 10% as in LOCABAL, but it is much weaker due to the presence of the mgr_z term within the nested summation of Equation 8.12). Moreover, $\beta = 0.3$ means that the values of multi-faceted trust computed by the algorithm are reduced to 30% of the multi-dimensional reputation. However, given the weak role of social regularization, reputation is central to learning the user preference and item characteristic vectors. These findings are coherent with the hypothesis that, in this dataset, the feedback on user contributions is a very useful type of information to learn user preferences but show that, even by only employing social links and the feedback on user profiles, the algorithm can achieve satisfactory results.

LOC+noE ignores the feedback on user profiles. It obtains its best performance with $\alpha = \beta = 0.3$: with respect to LOCABAL+, the algorithm weights social regularization much less but it partially takes multi-dimensional reputation into account (30%) in the computation of multi-faceted trust. With respect to LOC+noF, LOC+noE increases a little bit the role of social regularization in the Matrix Factorization process. The algorithm outperforms the baselines and is generally worse than LOCABAL+. Moreover, it performs slightly worse than LOC+noF: it has the same precision and very similar F1, MAP and MRR but it has lower recall, RMSE, MAE and User Coverage. We explain these findings with the fact that, as LOC+noE ignores the trust feedback received by user profiles, it generally misses useful information for preference prediction. In the dataset, user endorsements have high median values (e.g., the median number of compliments received by individual users is 110 and the median number of fans is 41). Therefore, when the algorithm ignores them, it has fewer chances to recognize highly reputable users. The value of β also shows that the feedback on user contributions determines the value of multi-faceted trust by 70%. Once more, it looks like the feedback on user contributions has an important role in defining trust; however, the feedback on user profiles is useful as well.

LOC+noS ignores social relations among users: it only employs anonymous trust statements and anonymous social information (number of fans) to learn user preferences. This means that it is not possible to compute the importance of users in the social network (imp_v) and that social regularization does not make sense. We obtain LOC+noS by forcing parameter $C_1 = 0$ in Equation 8.14 and $\alpha = 0$ in Equation 8.12. This algorithm outperforms the baselines in all measures except for User Coverage and Recall; moreover, it has the best RMSE and MAE of all the algorithms and configurations of LOCABAL+. This supports the hypothesis that, in this dataset, anonymous trust feedback is a precious source of information to be used in a recommender system, and that correct ratings can be predicted without using personal data about social relations.

Figure 8.1 shows the variation of MAP for all the configurations of LOCABAL+, depending on α and β. By setting α to a constant value (Figure 8.1b), MAP decreases



(a) MAP variation with respect to β

(b) MAP variation with respect to α

Fig. 8.1: MAP variation on Yelp-Hotel. The Y axis represents MAP; the X axis represents β in Figure 8.1a and α in the Figure 8.1b.

	α	β	P	R	F1	MAP	RMSE	MAE	MRR	UCov
LOCABAL+	0.9	0	**0.8253**	**0.7703**	**0.7968**	**0.4818**	**0.8801**	**0.6601**	0.5341	0.6748
SocialMF	-	-	0.8044	0.7687	0.7862	0.4808	0.918	0.701	**0.5399**	0.7055
LOCABAL	0.1	-	0.8041	0.761	0.782	0.4736	0.9369	0.7158	0.5378	0.7014
U2UCF	-	-	0.7906	0.7684	0.7794	0.4723	0.9627	0.7251	0.5239	0.5767
SVD++	-	-	0.7867	0.7607	0.7735	0.4709	0.9832	0.7457	0.5324	**0.7076**
U2USocial	-	-	0.7759	0.7561	0.7658	0.4518	1.0597	0.7975	0.4952	0.2802

Tab. 8.5: Performance@10 on the new data of the Yelp-Hotel dataset.

when β grows. This means that, having blocked the influence of the social component of LOCABAL+, the best results are achieved when the feedback on contributions makes ratings more influent on the Matrix Factorization process. Moreover, by setting β to a constant value (Figure 8.1a), MAP improves when α increases and the best results are achieved with α=0.9, i.e., when the social component, in combination with the other facets of trust, strongly influences preference estimation.

Table 8.5 summarizes the evaluation results on new data (10% of Yelp-Hotel). It can be seen that the results are fairly consistent with those of Table 8.4: LOCABAL+ outperforms the baselines in all performance measures except for User Coverage, as previously, dominated by SVD++ and SocialMF. The main difference in this case is that SocialMF also dominates MRR. We can conclude that LOCABAL+ can be used in dynamic application domains without frequently optimizing the model.

8.7.3.2 Yelp-Food

Table 8.6 shows the evaluation results on the Yelp-Food dataset. The LOCABAL+ configurations outperform the baselines with statistically significant results in all measures except for User Coverage. Moreover, SVD++ has the highest coverage and

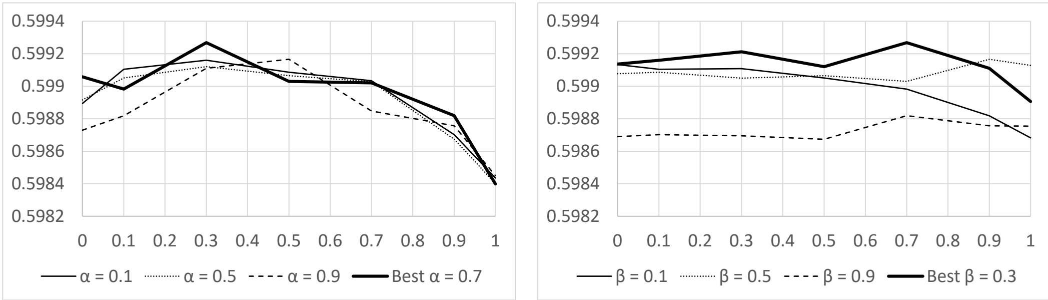
	α	β	P	R	F1	MAP	RMSE	MAE	MRR	UCov
Significance			0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.01
LOCABAL+	0.7	0.3	0.7769	**0.7528**	0.7647	**0.5993**	0.9791	0.7406	**0.7166**	0.8499
LOC+noE	0.9	0.3	0.7769	0.7527	0.7646	0.5993	0.9793	0.7407	0.7165	0.8497
LOC+noS	-	0.3	0.777	0.7527	0.7647	0.5992	0.9789	0.7403	0.7164	0.8497
LOC+noF	0.5	0.7	**0.7783**	0.752	**0.7649**	0.5986	**0.9768**	**0.7389**	0.7163	0.8485
SocialMF	-	-	0.76	0.7312	0.7453	0.5741	1.0358	0.7828	0.7092	0.8642
LOCABAL	0.1	-	0.7586	0.7294	0.7437	0.5724	1.041	0.7866	0.7093	0.8656
U2UCF	-	-	0.7488	0.7337	0.7412	0.5642	1.0796	0.8072	0.6795	0.7025
SVD++	-	-	0.7483	0.7116	0.7295	0.5539	1.0761	0.812	0.7017	**0.8698**
U2USocial	-	-	0.7729	0.7357	0.7539	0.5474	1.0741	0.8178	0.6061	0.2295

Tab. 8.6: Performance@10 on Yelp-Food dataset.

LOCABAL is the second best, followed by SocialMF, but all of them are less accurate and have lower ranking capability than the LOCABAL+ configurations. In this case, LOCABAL+ and LOC+noF are the best performing algorithms but, as LOCABAL+ has the best Recall, MAP and MRR (and LOC+noF the worst ones among LOCABAL+ configurations), we consider LOCABAL+ as the preferable one.

In this dataset, all the LOCABAL+ configurations take multi-dimensional global reputation into account in the computation of multi-faceted trust. Specifically, $\beta = 0.3$ in LOCABAL+, LOC+noE and LOC+noS; moreover, $\beta = 0.7$ in LOC+noF that overlooks the feedback on user contributions. Moreover, social regularization has a medium to high role in the Matrix Factorization process, with the strongest influence in LOCABAL+ ($\alpha = 0.7$) and LOC+noE ($\alpha = 0.9$). In summary, the configurations exploit both multi-faceted trust and social regularization to obtain their best performance but the exclusion of feedback on user contributions raises the importance of the other facets of trust. This can be explained by the fact that users and user contributions receive a low amount of feedback for the evaluation of trustworthiness; e.g., the median number of appreciations is 1 for reviews and 0 for tips and the median number of endorsements to user profiles, including Elite years, compliments and fans, is 6. Thus, the algorithms rely on the joint contribution of all the sources of evidence about trust. Another relevant observation is that the performance of LOC+noE and LOC+noS is similar to that of LOCABAL+. This can be explained by assuming that, in this dataset, the feedback on user profiles and social relations play complementary roles and can replace each other without a major loss of performance.

Figure 8.2 shows the variation of MAP for all the LOCABAL+ configurations depending on α and β. It can be noticed that by setting α to a constant value (Figure 8.2b), MAP first slightly increases but, when $\beta > 0.3$, it quickly decreases. This means that,



(a) MAP variation with respect to β.

(b) MAP variation with respect to α.

Fig. 8.2: MAP variation on Yelp-Food. The Y axis represents MAP; the X axis represents β in Figure 8.2a and α in Figure 8.2b.

	α	β	P	R	F1	MAP	RMSE	MAE	MRR	UCov
LOCABAL+	0.7	0.3	**0.7935**	**0.787**	**0.7903**	**0.5835**	**0.9701**	**0.7376**	**0.6584**	**0.8053**
U2UCF	-	-	0.769	0.7699	0.7694	0.5606	1.0423	0.7807	0.6378	0.7193
U2USocial	-	-	0.7798	0.7708	0.7753	0.5472	1.0568	0.8055	0.5913	0.2595
SocialMF	-	-	0.7818	0.7489	0.765	0.5461	1.0403	0.7829	0.6387	0.8011
LOCABAL	0.1	-	0.7814	0.7478	0.7642	0.5451	1.0449	0.7864	0.6384	0.8025
SVD++	-	-	0.7742	0.7389	0.7561	0.5339	1.0712	0.8056	0.6314	0.8033

Tab. 8.7: Performance@10 on the new data of the Yelp-Food dataset.

regardless of the influence of the social component of LOCABAL+, the algorithm benefits from a moderate support by multi-dimensional global reputation. Notice also that, by setting β to a constant value (Figure 8.2a), results are not particularly affected by the value of α; therefore, social regularization has a generally constant contribution to recommendation performance.

Table 8.7 shows the evaluation of algorithms on new data. It can be seen that LOCABAL+ outperforms the baselines in all performance metrics, including MRR. Thus, we can conclude that also in Yelp-Food LOCABAL+ can be employed in dynamic environments without requiring a frequent optimization of parameters.

8.8 Discussion and Future Work

The experimental results show that, in the Yelp-Hotel and Yelp-Food datasets, LOCABAL+ outperforms all the baseline recommender systems in accuracy, error minimization and ranking capability with a minor loss of User Coverage with respect to SVD++, LOCABAL and SocialMF. Thus, we can say that LOCABAL+ generates better suggestions to marginally fewer people, with a clear positive gain in performance.

Noticeably, LOCABAL+ outperforms the baselines even though MTM is configured to ignore different types of evidence about trust; this finding shows the flexibility of our approach towards the lack of user information. This is generally important because some classes of trust information defined in MTM might not be available in specific recommendation domains. Moreover, it is a key achievement in relation to the management of personal data because, different from the other trust-based recommender systems, LOCABAL+ can work by ignoring data about social links, and by only relying on public anonymous information. This aspect is more and more important given the increasing sensibility of users towards disclosing personal data.

The evaluation results on new data confirm that, on both datasets, LOCABAL+ achieves the best performance in all the measures except for User Coverage (actually, in Yelp-Food LOCABAL+ outperforms the other algorithms in User Coverage as well, but this is not true in Yelp-Hotel). Thus, we conclude that LOCABAL+ can be applied, without losing recommendation capability, to dynamic environments in which it can not be frequently optimized.

The evaluation results are useful to answer our research questions:

$\mathbf{RQ3_1}$: *Can multi-faceted trust be used to improve the performance of a trust-based recommender system with respect to the standard state-of-the-art trust models that only rely on social links and rating similarity among users?*

Our experiments enable us to positively answer $\mathbf{RQ3_1}$ because they provide consistent results on two datasets having different characteristics (e.g., size, distributions of global feedback on users and contributions, etc.); see Section 8.6. Specifically, the superior results of LOCABAL+ with respect to the baselines and, in particular, to LOCABAL, show that by exploiting multiple facets of trust (including global anonymous feedback on users and user contributions), the performance of the recommender system significantly improves. It is worth noticing that, thanks to MTM, trust-based methods can be successfully applied without using personal information about social links. This fact represents a key aspect of our approach compared to existing work on trust-based recommender systems.

Our experiments also show that LOCABAL+ can work without using data about social relations, with a minor loss of performance, if it can use other types of trust feedback. Regarding the other facets of trust, the experiments show that their relative importance depends on the amount and quality of the feedback about users and user contributions available to the recommender system. The two datasets we selected are interesting because the diverse distributions of feedback they provide determine slightly different behavior of the LOCABAL+ configurations. In Yelp-Hotel, which provides a large amount of feedback about user contributions (median = 53 against 27 of Yelp-Food), the algorithm obtains the best performance results by privileging this type of information over multi-dimensional global reputation; moreover, performance results are clearly affected by the omission of the feedback on user contributions. Differently, in the Yelp-Food dataset, which stores scarcer feedback about contributions, the algorithm obtains the best results by balancing the influence of all types of global feedback, including the endorsements to user profiles, and performance increases if feedback on user contributions is ignored.

We thus conclude that global feedback about user trust is a useful type of information in Top-N recommendation. However, in order to decide which types of evidence should be used, it is important to analyze the characteristics of the data to which the recommender system is applied and the features of the various types of feedback that can be used.

Before closing this section it is worth noting that our current work focuses on numerical aspects of the evidence about trust, by measuring the amount of feedback that users and user contributions receive. However, the content of the global feedback provided by users is itself a further source of information about trust that can be analyzed to acquire information about users' behavior. In this perspective, we plan to extend our model to the analysis of the content of reviews (and/or microblogs), which has been largely studied to evaluate their quality (Huang et al., 2015; Mudambi and Schuff, 2010; Chua and Banerjee, 2015; Qazi et al., 2016; Korfiatis et al., 2012; Krishnamoorthy, 2015; Kim et al., 2006), to steer personalized recommendation (Raghavan et al., 2012; Alahmadi and Zeng, 2015; Shen et al., 2019; Hernández-Rubio et al., 2019) and to guide the explanation of recommendations (O'Mahony and Smyth, 2018; Musto et al., 2019).

9 Conclusion and Future Work

In this thesis we analyzed different suggestion models that leverage various types of feedback (i.e. explicit, implicit and anonymous) and that make recommendations at a coarse-grained (i.e. topics or concepts) and fine-grained (i.e. Point of Interests) granularity levels. The common objective of all these works is that of investigating different pieces of information that can be exploited to improve recommendation and/or information filtering, depending on the selected domain. We separately developed the various techniques in order to analyze their impact one by one, and to test their validity in a controlled way. However, the ultimate goal is that of integrating, where possible, multiple techniques to jointly benefit from their value to recommendation. Furthermore, we would like to offer to the users a mixed-initiative interaction model that is able to help them to explore the retrieved and suggested information in order to satisfy their information needs. Thus, besides continuing to develop the models separately and address their related future work that we described above, we plan to use the insights that we collected to build a hybrid recommender system in order to improve the quality of suggestions. We also plan to build and integrate a recommender system in the OnToMap web application in order to offer to the stakeholders a mixed-initiative interaction model that helps them to explore the territory through the navigation of relevant suggestions. In this way we will also be able to carry out online evaluations and to test the model in a real scenario.

In the following, we are going to summarize the models presented in this thesis and describe some future work related to them.

- In Chapter 4, we described a session-based concept suggestion model that supports information search by proposing concepts for query expansion, given the context provided by the observed portions of the search sessions. We focus on presenting types of information that the user might also be interested in, given the data he or she is searching for, at the conceptual level. We aim at supporting the exploration of information spaces including rather different types of data, such as those managed by Geographical Information Systems.

  Our model is based on an analysis of search behavior, that we carried out by studying the AOL query log. Specifically, our model is based on (i) the identification of clusters of concepts that frequently co-occur in search sessions, and (ii) the definition of strategies that select the clusters for concept suggestion by taking the search context that emerges from the first queries of the sessions. The evaluation of our model has provided satisfactory results about its prediction accuracy.

Having performed off-line experiments, we could not test the serendipity of the concept suggestions, as we relied on data about the types of information that users autonomously explored in the search sessions. Our future work includes (i) analyzing users' past behavior for improving concept suggestion, (ii) comparing our model with state-of-the-art query expansion models (Huang et al., 2003; Chen et al., 2008) and (iii) studying the impact of our model on the serendipity of suggestions by testing it in an online experiment with users.

- In Chapter 5, we investigated how semantic granularity in geographical knowledge conceptualization influences concept suggestions in exploratory search support. In our empirical evaluation, we used three ontologies that represent knowledge at different semantic granularity levels. We found that a fine-grained domain conceptualization supports the suggestion of a larger set of concepts, with higher recall. However, the relationship between the formal conceptualization (ontology) and the way people conceptualize and verbally describe geographic space influences precision and, consequently, overall accuracy.

  Even though we carried out this work using a large dataset for the analysis of search behavior, there are limitations in our empirical evaluation that we aim at addressing in our future work. Firstly, we plan to repeat our experiments by exploiting other consolidated ontologies for concept interpretation and suggestion; e.g., ProBase (Wu et al., 2012), or Yago (Suchanek et al., 2008). Secondly, we plan to test accuracy of concept suggestion on other datasets to generalize our results. For instance, we plan to analyze Twitter (`https://twitter.com`) data to extract concept references patterns emerging from social interaction.

- In Chapter 6, we investigated whether the identification of frequently co-occurring interests in information search can be used to improve the performance of KNN collaborative recommender systems. For this purpose, we defined a preference extension model that, applied to a category-based representation of user profiles, infers user preferences by exploiting frequently co-occurring information interests. Then, we implemented the model in the Extended Category-based Collaborative Filtering algorithm (ECCF). This is a variant of User-to-User Collaborative Filtering that works on category-based user profiles, enriched with preferences inferred from general search behavior. For the analysis of user interests, we analyzed the query log of a largely used search engine.

We evaluated ECCF on a large dataset of item ratings, by applying different levels of strictness in the extension of user profiles. The evaluation showed that ECCF outperforms User-to-User Collaborative Filtering in accuracy, MRR, intra-list diversity and user coverage. Interestingly, ECCS also obtains higher accuracy and diversity than the SVD++ recommender system, based on Matrix Factorization; however, ECCS has lower user coverage than SVD++.

In our future work we will focus on the coverage aspect in order to improve the performance of KNN Collaborative Filtering. Moreover, we will carry out further experiments, considering (i) a broader domain than Restaurants and Food, on which we have focused our current work, and (ii) users who have provided few or zero ratings. We will also analyze other datasets to check whether the performance results described in this work can be generalized. Finally, we will compare the performance of ECCF with a larger set of recommendation approaches based on preference extension.

- In Chapter 7, we presented a faceted information exploration approach supporting a flexible visualization of heterogeneous geographic data. Our model provides a multi-category faceted projection of long-lasting geographic maps to answer temporary information needs; this is based on the proposal of efficient facets for data exploration in sparse and noisy datasets. Moreover, the model provides a graphical representation of the search context by means of alternative types of widget that support interactive data visualization, faceted exploration, category-based information hiding and transparency of results at the same time.

  We carried out a user study involving 62 people who have diverse familiarity with technology and with map-based online systems. The results of this study show that, when working on maps populated with multiple data categories, our model outperforms simple category-based map projection and traditional faceted search tools such as checkboxes. Moreover, the layout that uses the sunburst diagram as a graphical widget supports the best user performance and experience, thanks to its clarity and visual compactness. We thus conclude that this implementation is promising for flexible faceted exploration in Geographic Information Search. The described work has limitations that we plan to address:

- Our model only supports the specification of hard visualization constraints on facet values; i.e., the items having a certain value of a facet are either shown, or hidden. However, the user might want to specify preferences. Therefore, similar to what has been done in some related works, we plan to manage soft visualization constraints.
- So far, we present search results in geographic maps and we provide item details in dynamically generated tables showing their properties. In order to enhance data interpretation and sensemaking, we plan to develop additional visualization models supporting visual analytics; e.g., see (Andrienko et al., 2007; Tsai and Brusilovsky, 2019; Cardoso et al., 2019).
- Further experiments are needed to validate the proposed model with a larger set of people and on mobile phones (the OnToMap user interface scales well to the screens of tablets).
- Currently, our model supports a "one size fits all" type of faceted search that exploits general efficiency criteria to guide the user in data exploration. However, some researchers propose to adapt facet suggestions to the user's preferences in order to personalize the navigation of the information space; e.g., see (Tvarožek et al., 2008; Tvarožek and Bieliková, 2010; Koren et al., 2008; Abel et al., 2011). In our future work, we plan to offer multiple data exploration strategies which the user can choose from, including a user-adaptive facet suggestion that depends on her/his preferences and on the search context.
- Depending on their roles, in some scenarios users might need to access different, long-lasting custom views of a shared information space (Rasmussen and Hertzum, 2013). We plan to extend our model by introducing permanent, user-dependent views on map content.

• In Chapter 8, we described the Multi-faceted Trust Model (MTM) and the LOCABAL+ recommender system, which combine social links with global anonymous feedback about users and user contributions to enhance collaborative Top-N recommendation. LOCABAL+ extends the LOCABAL recommender system with multi-faceted trust and with an enhanced regularization of social relations based on both rating similarity and users' multi-dimensional global reputation.

  LOCABAL+ has various advantages with respect to the state-of-the art trust-based recommender systems. In particular, being based on the MTM compositional model, it can be configured to work with different types of evidence about trust, such as anonymous public feedback on user profiles and user contributions, as well as information about social relations. Interestingly, LOCABAL+ can work by ignoring the information about social links, which has been recently found to be problematic in the user acceptance of trust-based recommender systems because it is considered as personal information. Another advantage of LOCABAL+ is that the extension to social regularization makes it possible to select users to steer Matrix Factorization in a more selective way with respect to only considering rating similarity.

Experiments carried out on two public datasets of item reviews show that, with a minor loss of user coverage, LOCABAL+ outperforms state-of-the art trust-based recommender systems and Collaborative Filtering in accuracy, error minimization and ranking capability both when it uses complete information from the datasets and when it ignores social relations (or other types of feedback on users and contributions). It thus represents a flexible approach to trust-based recommendation, suitable to comply with specific data management requirements.

Bibliography

Abdul-Rahman, A. and Hailes, S. (2000). “Supporting Trust in Virtual Communities”. In: *Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 6 - Volume 6*. HICSS ’00. Washington, DC, USA: IEEE Computer Society, pp. 6007– (cit. on p. 29).

Abel, F., Celik, I., Houben, G.-J., and Siehndel, P. (2011). “Leveraging the Semantics of Tweets for Adaptive Faceted Search on Twitter”. In: *The Semantic Web – ISWC 2011*. Ed. by L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, and E. Blomqvist. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–17 (cit. on p. 148).

Ahlberg, C. and Shneiderman, B. (1994). “Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’94. Boston, Massachusetts, USA: ACM, pp. 313–317 (cit. on p. 25).

Akaishi, M., Satoh, K., and Tanaka, Y. (2004). “An Associative Information Retrieval Based on the Dependency of Term Co-occurrence”. In: *Discovery Science: 7th International Conference, DS 2004, Padova, Italy, October 2-5, 2004. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 195–206 (cit. on p. 19).

Alahmadi, D. H. and Zeng, X.-J. (2015). “ISTS: Implicit social trust and sentiment based approach to recommender systems”. In: *Expert Systems with Applications* 42.22, pp. 8840–8849 (cit. on p. 143).

Alotaibi, S. and Vassileva, J. (2016). “Personalized Recommendation of Research Papers by Fusing Recommendations from Explicit and Implicit Social Networks”. In: *Proc. of IFUP 2016: Workshop on Multi-dimensional Information Fusion for User Modeling and Personalization*. Vol. 1618. Halifax, Canada: CEUR, paper 2 (cit. on p. 23).

Amazon.com (2006). “Amazon.com: online shopping for electronics, apparel, etc.” In: `http://www.amazon.com` (cit. on p. 118).

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M., MacEachren, A., and Wrobel, S. (2007). “Geovisual analytics for spatial decision support: Setting the research agenda”. In: *International Journal of Geographical Information Science* 21.8, pp. 839–857 (cit. on pp. 28, 88, 148).

Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., and Voghera, A. (2017a). “Personalised Community Maps”. In: *Int. J. of Electronic Governance* 9.1-2, pp. 156–178 (cit. on p. v).

Ardissono, L., Delsanto, M., Lucenteforte, M., Mauro, N., Savoca, A., and Scanu, D. (2018a). "Map-based Visualization of 2D/3D Spatial Data via Stylization and Tuning of Information Emphasis". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. AVI '18. Castiglione della Pescaia, Grosseto, Italy: ACM, 38:1–38:5 (cit. on pp. vi, 8, 11, 87, 89, 91, 103).

Ardissono, L., Delsanto, M., Lucenteforte, M., Mauro, N., Savoca, A., and Scanu, D. (2018b). "Transparency-based Information Filtering on 2D/3D Geographical Maps". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. AVI '18. Castiglione della Pescaia, Grosseto, Italy: ACM, 56:1–56:3 (cit. on pp. vi, 11).

Ardissono, L., Ferrero, M., Petrone, G., and Segnan, M. (2017b). "Enhancing Collaborative Filtering with Friendship Information". In: *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. UMAP '17. Bratislava, Slovakia: ACM, pp. 353–354 (cit. on pp. 30, 31).

Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., and La Riccia, L. (2016). "Exploration of Cultural Heritage Information via Textual Search Queries". In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '16. Florence, Italy: ACM, pp. 992–1001 (cit. on pp. vi, 11, 12, 20, 27, 39, 50).

Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., and La Riccia, L. (2017c). "OnToMap: Semantic Community Maps for Knowledge Sharing". In: *Proceedings of the 28th ACM Conference on Hypertext and Social Media*. HT '17. Prague, Czech Republic: ACM, pp. 317–318 (cit. on pp. v, 10, 38).

Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., and La Riccia, L. (2017d). "Semantic Interpretation of Search Queries for Personalization". In: *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. UMAP '17. Bratislava, Slovakia: ACM, pp. 101–102 (cit. on pp. vi, 38).

Ardissono, L. and Mauro, N. (2020). "A Compositional Model of Multi-faceted Trust for Personalized Item Recommendation". In: *Expert Systems with Applications*, p. 112880 (cit. on pp. v, 8, 23, 113).

Ardissono, L., Mauro, N., and Savoca, A. (2017e). "Supporting Knowledge Sharing and Learning via Semantic Geographical Maps". In: *Proceedings of the 2017 ACM Workshop on Intelligent Interfaces for Ubiquitous and Smart Learning*. SmartLearn '17. Limassol, Cyprus: ACM, pp. 3–6 (cit. on p. vii).

Ardito, C., Costabile, M., Desolda, G., Lanzilotti, R., Matera, M., Piccinno, A., and Picozzi, M. (2013). "Personal Information Spaces in the Context of Visits to Archaeological Parks". In: *Proc. of the Biannual Conference of the Italian Chapter of SIGCHI*. Trento, Italy: ACM, 5:1–5:4 (cit. on p. 88).

Ardito, C., Costabile, M. F., Desolda, G., and Matera, M. (2016). "Supporting Professional Guides to Create Personalized Visit Experiences". In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '16. Florence, Italy: ACM, pp. 1010–1015 (cit. on p. 88).

Athukorala, L., Medlar, A., Oulasvirta, A., Jacucci, G., and Glowacka, D. (2016). “Beyond relevance: adapting exploration/exploitation in information retrieval”. In: *Proc. of the 21st Int. Conf. on Intelligent User Interfaces (IUI’16)*. Sonoma, California, USA: ACM, pp. 359–369 (cit. on p. 20).

Ballatore, A., Bertolotto, M., and Wilson, D. (2013). “Geographic knowledge extraction and semantic similarity in OpenStreetMap”. In: *Knowledge and Information Systems* 37.1, pp. 68–81 (cit. on pp. 21, 97).

Ballatore, A., Kuhn, W., Hegarty, M., and Parsons, E. (2016). “Special issue introduction: Spatial approaches to information search”. In: *Spatial Cognition & Computation* 16.4, pp. 245–254 (cit. on pp. 13, 21).

Beitzel, S., Jensen, E., Frieder, O., Lewis, D., Chowdhury, A., and Kolcz, A. (2005). “Improving automatic query classification via semi-supervised learning”. In: *Proc. of the Fifth IEEE Int. Conf. on Data Mining*. ICDM ’05. Washington, DC, USA: IEEE Computer Society, pp. 42–49 (cit. on p. 40).

Belkin, N. (1980). “Anomalous states of knowledge as a basis for information retrieval”. In: *Canadian Journal of Information Science* 5, pp. 133–143 (cit. on p. 35).

Bennett, P., White, R., Chu, W., Dumais, S., Bailey, P., Borisyuk, F., and Cui, X. (2012). “Modeling the impact of short- and long-term behavior on search personalization”. In: *Proc. of the 35th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*. SIGIR ’12. Portland, Oregon, USA: ACM, pp. 185–194 (cit. on p. 57).

Booking.com (2019). “Booking.com”. In: `https://www.booking.com` (cit. on pp. 114, 118).

Bradley, K. and Smyth, B. (2001). “Improving Recommendation Diversity”. In: *Proc. of the 12th National Conference in Artificial Intelligence and Cognitive Science*. Ed. by D. O’Donoghue. Maynooth, Ireland, pp. 75–84 (cit. on p. 79).

Brown, G. and Weber, D. (2012). “Measuring change in place values using public participation GIS (PPGIS)”. In: *Applied Geography* 34, pp. 316–324 (cit. on p. 88).

Brusilovsky, P. and He, D. (2018). “Introduction to Social Information Access”. In: *Social Information Access: Systems and Technologies*. Ed. by P. Brusilovsky and D. He. Cham: Springer International Publishing, pp. 1–18 (cit. on p. 14).

Burbach, L., Nakayama, J., Plettenberg, N., Ziefle, M., and Valdez, A. C. (2018). “User Preferences in Recommendation Algorithms: The Influence of User Diversity, Trust, and Product Category on Privacy Perceptions in Recommender Algorithms”. In: *Proceedings of the 12th ACM Conference on Recommender Systems*. RecSys ’18. Vancouver, British Columbia, Canada: ACM, pp. 306–310 (cit. on pp. 16, 114).

Candan, K., Cataldi, M., Caro, L. D., Sapino, M., and Schifanella, C. (2009). “CoSeNa: a context-based search and navigation system”. In: *MEDES ’09 Int. Conf. on Management of Emergent Digital EcoSystems*. Chia Laguna, Italy, Art. 33 (cit. on pp. 20, 24).

Cao, H., Jiang, D., Pei, J., He, Q., Liao, Z., Chen, E., and Li, H. (2008). “Context-aware query suggestion by mining click-through and session data”. In: *Proc. of the 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*. KDD ’08. Las Vegas, Nevada, USA: ACM, pp. 875–883 (cit. on pp. 12, 20).

Cao, N., Sun, J., Lin, Y., Gotz, D., Liu, S., and Qu, H. (2010). “FacetAtlas: multifaceted visualization for rich text corpora”. In: *IEEE Transactions on visualization and Computer Graphics* 16.6, pp. 1172–1181 (cit. on p. 26).

Cao, N., Gotz, D., Sun, J., Lin, Y., and Qu, H. (2011). “SolarMap: multifaceted Visual Analytics for Topic Exploration”. In: *Proc. of the 11th IEEE Int. Conf. on Data Mining*. KDD ’08. Vancouver, BC, Canada: IEEE, pp. 101–110 (cit. on p. 26).

Cardoso, B., Sedrakyan, G., Gutiérrez, F., Parra, D., Brusilovsky, P., and Verbert, K. (2019). “IntersectionExplorer, a multi-perspective approach for exploring recommendations”. In: *International Journal of Human-Computer Studies* 121. Advances in Computer-Human Interaction for Recommender Systems, pp. 73–92 (cit. on pp. 27, 28, 148).

Chaney, A. J., Blei, D. M., and Eliassi-Rad, T. (2015). “A Probabilistic Model for Using Social Networks in Personalized Item Recommendation”. In: *Proceedings of the 9th ACM Conference on Recommender Systems*. RecSys ’15. Vienna, Austria: ACM, pp. 43–50 (cit. on p. 30).

Chang, J. C., Hahn, N., Perer, A., and Kittur, A. (2019). “SearchLens: Composing and Capturing Complex User Interests for Exploratory Search”. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. IUI ’19. Marina del Ray, California: ACM, pp. 498–509 (cit. on p. 25).

Chen, L., Chen, G., and Wang, F. (2015). “Recommender systems based on user reviews: the state of the art”. In: *USER MODELING AND USER-ADAPTED INTERACTION - The Journal of Personalization Research* 25.2, pp. 99–154 (cit. on p. 23).

Chen, Y., Xue, G.-R., and Yu, Y. (2008). “Advertising Keyword Suggestion Based on Concept Hierarchy”. In: *Proc. of the 2008 Int. Conf. on Web Search and Data Mining*. WSDM ’08. Palo Alto, California, USA: ACM, pp. 251–260 (cit. on pp. 12, 20, 146).

Chi, E. H. (2015). “Blurring of the Boundary Between Interactive Search and Recommendation”. In: *Proceedings of the 20th International Conference on Intelligent User Interfaces*. IUI ’15. Atlanta, Georgia, USA: ACM, pp. 2–2 (cit. on p. 2).

Chua, A. and Banerjee, S. (2015). “Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth”. In: *Journal of the Association for Information Science and Technology* 66.2, pp. 354–362 (cit. on p. 143).

Coder-Yu (2019). “Locabal.py”. In: `https://github.com/Coder-Yu/RecQ/tree/master/algorithm/rating/LOCABAL.py` (cit. on p. 136).

Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., and Rau, R. (2011). “OSMonto – An Ontology of OpenStreetMap Tags”. In: *Proc. State of the map Europe (SOTM-EU) 2011*. Vienna, Austria, pp. 55–64 (cit. on pp. 21, 97).

Colby, G. and Sholl, L. (1991). “Transparency and blur as selective cues for complex visual information”. In: *Proceedings of SPIE 1460, Image Handling and Reproduction Systems Integration*. Vol. 1460 (cit. on p. 91).

Coulton, C., Chan, T., and Mikelbank, K. (2011). "Finding Place in Community Change Initiatives: Using GIS to Uncover Resident Perceptions of their Neighborhoods". In: *Journal of Community Practice* 19.1, pp. 10–28. eprint: `https://www.tandfonline.com/doi/pdf/10.1080/10705422.2011.550258` (cit. on p. 88).

Dachselt, R. and Frisch, M. (2007). "Mambo: A Facet-based Zoomable Music Browser". In: *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia*. MUM '07. Oulu, Finland: ACM, pp. 110–117 (cit. on p. 25).

Dachselt, R., Frisch, M., and Weiland, M. (2008). "FacetZoom: A Continuous Multi-scale Widget for Navigating Hierarchical Metadata". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. Florence, Italy: ACM, pp. 1353–1356 (cit. on p. 25).

Danetsoft (2019). "dvd.ciao.co.uk". In: `http://data.danetsoft.com/dvd.ciao.co.uk` (cit. on p. 118).

De Meo, P., Fotia, L., Messina, F., Rosaci, D., and Sarné, G. M. (2018). "Providing recommendations in social networks by integrating local and global reputation". In: *Information Systems* 78, pp. 58–67 (cit. on p. 33).

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). "Indexing by latent semantic analysis". In: *Journal of the American Society for Information Science* 41.6, pp. 391–407. eprint: `https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-4571%28199009%2941%3A6%3C391%3A%3AAID-ASI1%3E3.0.CO%3B2-9` (cit. on p. 4).

Deng, S., Huang, L., and Xu, G. (2014). "Social network-based service recommendation with trust enhancement". In: *Expert Systems with Applications* 41.18, pp. 8075–8084 (cit. on pp. 30, 31).

Desrosiers, C. and Karypis, G. (2011). "A Comprehensive Survey of Neighborhood-based Recommendation Methods". In: *Recommender Systems Handbook*. Ed. by F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. Boston, MA: Springer US, pp. 107–144 (cit. on pp. 22, 71, 113, 117).

Di Noia, T., Ostuni, V. C., Rosati, J., Tomeo, P., Sciascio, E. D., Mirizzi, R., and Bartolini, C. (2016). "Building a relatedness graph from Linked Open Data: A case study in the IT domain". In: *Expert Systems with Applications* 44, pp. 354–366 (cit. on p. 24).

Di Rocco, L. (2016). "Semantic Enhancement of Volunteered Geographic Information". In: *Proc. of Doctoral Consortium of the 15th Edition of AI*IA*. AI*IA-DC 2016. Genova, Italy: CEUR-WS.org, pp. 28–33 (cit. on p. 62).

Dodds, P. S., Muhamad, R., and Watts, D. J. (2003). "An Experimental Study of Search in Global Social Networks". In: *Science* 301.5634, pp. 827–829. eprint: `https://science.sciencemag.org/content/301/5634/827.full.pdf` (cit. on p. 4).

Doychev, D., Lawlor, A., Rafter, R., and Smyth, B. (2014). "An analysis of recommender algorithms for online news". In: *Proc. of CLEF 2014 Conference and Labs of the Evaluation Forum*. Sheffield, UK (cit. on p. 22).

Du, X., Liu, H., and Jing, L. (2017). "Additive Co-Clustering with Social Influence for Recommendation". In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*. RecSys '17. Como, Italy: ACM, pp. 193–200 (cit. on pp. 23, 30, 31).

Epinions (2019). "Epinions datasets". In: `http://www.trustlet.org/epinions.html` (cit. on pp. 30, 118).

Expedia.com (2001). "Expedia". In: `http://www.expedia.com` (cit. on pp. 114, 118).

Fernández-Tobías, I., Tomeo, P., Cantador, I., Di Noia, T., and Di Sciascio, E. (2016). "Accuracy and Diversity in Cross-domain Recommendations for Cold-start Users with Positive-only Feedback". In: *Proc. of the 10th ACM Conference on Recommender Systems*. RecSys '16. Boston, Massachusetts, USA: ACM, pp. 119–122 (cit. on p. 22).

Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., and Motta, E. (2011). "Semantically enhanced information retrieval: an ontology-based approach". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 9.4, pp. 434–452 (cit. on p. 19).

Fonseca, F., Egenhofer, M., Agouris, P., and Câmara, G. (2002a). "Using ontologies for geographic information systems". In: *Transactions in GIS* 3, pp. 231–257 (cit. on p. 21).

Fonseca, F., Egenhofer, M., C.A. Davis, and Câmara, G. (2002b). "Semantic granularity in ontology-driven Geographic Information Systems". In: *Annals of Mathematics and Artificial Intelligence* 36.1-2, pp. 121–151 (cit. on pp. 59, 61).

Gambetta, D. (1988). "Can We Trust Trust?" In: *Trust: Making and Breaking Cooperative Relations*. Ed. by D. Gambetta. Blackwell, pp. 213–237 (cit. on p. 28).

Gauch, S., Chaffee, J., and Pretschner, A. (2003). "Ontology-based personalized search and browsing". In: *Web Intelli. and Agent Sys.* 1.3-4, pp. 219–234 (cit. on p. 57).

Gemmel, J., Schimoler, T., Mobasher, B., and Burke, R. (2012). "Resource recommendation in social annotation systems: a linear-weighted hybrid approach". In: *Journal of computer and system sciences* 78, pp. 1160–1174 (cit. on p. 23).

Giuliano, V. and Jones, P. (1962). *Linear Associative Information Retrieval*. Tech. rep. ESD-TR-62-294. ASTIA (cit. on pp. 20, 36).

Golbeck, J. and Hendler, J. (2004). "Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-Based Social Networks". In: *Engineering Knowledge in the Age of the Semantic Web*. Ed. by E. Motta, N. R. Shadbolt, A. Stutt, and N. Gibbins. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 116–131 (cit. on pp. 28, 30).

Golbeck, J. and Hendler, J. (Jan. 2006). "FilmTrust: movie recommendations using trust in web-based social networks". In: *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.* Vol. 1, pp. 282–286 (cit. on p. 30).

Google (2017). "Knowledge Graph". In: `https://www.google.com/intl/it_it/insidesearch/features/search/knowledge.html` (cit. on pp. 20, 24).

Greenstein-Messica, A., Rokach, L., and Friedman, M. (2017). "Session-based recommendations using item embedding". In: *Proc. of the 22nd Int. Conf. on Intelligent User Interfaces*. IUI '17. Limassol, Cyprus: ACM, pp. 629–633 (cit. on pp. 22, 57).

Gregory, S. (2010). “Finding overlapping communities in networks by label propagation”. In: *New Journal of Physics* 12.103018 (cit. on p. 47).

Groh, G. and Ehmig, C. (2007). “Recommendations in Taste Related Domains: Collaborative Filtering vs. Social Filtering”. In: *Proceedings of the 2007 International ACM Conference on Supporting Group Work*. GROUP ’07. Sanibel Island, Florida, USA: ACM, pp. 127–136 (cit. on p. 30).

Grootjem, F. and T.P. van der Weide (2006). “Conceptual query expansion”. In: *Data & Knowledge Engineering* 56, pp. 174–193 (cit. on p. 19).

Guo, G., Zhang, J., and Yorke-Smith, N. (2015). “TrustSVD: Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings”. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI’15. Austin, Texas: AAAI Press, pp. 123–129 (cit. on pp. 30, 31).

Hearst, M. A. (Aug. 2006). “Design recommendations for hierarchical faceted search interfaces”. In: *Proc. SIGIR 2006, Workshop on Faceted Search*, pp. 26–30 (cit. on pp. 25, 88, 92).

Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., and Yee, K.-P. (Sept. 2002). “Finding the Flow in Web Site Search”. In: *Commun. ACM* 45.9, pp. 42–49 (cit. on p. 25).

Henrich, A. and Lüdecke, V. (2007). “Characteristics of geographic information needs”. In: *Proc. of GIR’07*. Lisboa, Portugal, pp. 1–6 (cit. on p. 21).

Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000). “Explaining Collaborative Filtering Recommendations”. In: *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*. CSCW ’00. Philadelphia, Pennsylvania, USA: Association for Computing Machinery, pp. 241–250 (cit. on p. 22).

Hernández-Rubio, M., Cantador, I., and Bellogín, A. (2019). “A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews”. In: *User Modeling and User-Adapted Interaction* 29.2, pp. 381–441 (cit. on p. 143).

Hildebrand, M., Ossenbruggen, J. van, and Hardman, L. (2006). “/facet: A Browser for Heterogeneous Semantic Web Repositories”. In: *The Semantic Web - ISWC 2006*. Ed. by I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. M. Aroyo. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 272–285 (cit. on p. 26).

Hoeber, O., Yang, X., and Yao, Y. (2005). “Conceptual query expansion”. In: *Advances in Web Intelligence. AWIC 2005, Lecture Notes in Computer Science, vol 3528*. Lodz, Poland, pp. 190–196 (cit. on p. 19).

Hoeber, O. and Yang, X. D. (2006). “A Comparative User Study of Web Search Interfaces: HotMap, Concept Highlighter, and Google”. In: *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*. WI ’06. Washington, DC, USA: IEEE Computer Society, pp. 866–874 (cit. on p. 25).

Hu, G.-N., Dai, X.-Y., Qiu, F.-Y., Xia, R., Li, T., Huang, S.-J., and Chen, J.-J. (Jan. 2018). “Collaborative Filtering with Topic and Social Latent Factors Incorporating Implicit Feedback”. In: *ACM Trans. Knowl. Discov. Data* 12.2, 23:1–23:30 (cit. on p. 31).

Hu, Y., Lv, Z., Wu, J., Janowicz, K., Zhao, X., and Yu, B. (2015). "A multistage collaborative 3D GIS to support public participation". In: *Int. Journal of Digital Earth* 8.3, pp. 212–234 (cit. on p. 88).

Huang, A. H., Chen, K., Yen, D. C., and Tran, T. P. (July 2015). "A Study of Factors That Contribute to Online Review Helpfulness". In: *Comput. Hum. Behav.* 48.C, pp. 17–27 (cit. on p. 143).

Huang, C., Chien, L., and Oyang, Y. (2003). "Relevant term suggestion in interactive web search based on contextual in formation in query session logs". In: *Journal of the American Society for Information Science and Technology* 54.7, pp. 638–649 (cit. on pp. 12, 20, 53, 146).

Hug, N. (2017). "Surprise, a Python library for recommender systems". In: `http://surpriselib.com` (cit. on pp. 79, 136).

J. Van Den Berg and Schuemie, M. (1999). "Information retrieval systems using an associative conceptual space". In: *In Proc. of the 7th European Symposium on Artificial Neural Networks (ESANN'99)*, pp. 351–356 (cit. on p. 19).

Jamali, M. and Ester, M. (2009). "TrustWalker: A Random Walk Model for Combining Trust-based and Item-based Recommendation". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '09. Paris, France: ACM, pp. 397–406 (cit. on pp. 30, 31).

Jamali, M. and Ester, M. (2010). "A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks". In: *Proceedings of the Fourth ACM Conference on Recommender Systems*. RecSys '10. Barcelona, Spain: ACM, pp. 135–142 (cit. on pp. 30, 31, 114, 117).

Jannach, D. and Ludewig, M. (2017). "When Recurrent Neural Networks meet the Neighborhood for Session-Based Recommendation". In: *Proceedings of the 11th ACM Conference on Recommender Systems*. RecSys '17, pp. 306–310 (cit. on p. 22).

Jannach, D., Resnick, P., Tuzhilin, A., and Zanker, M. (Oct. 2016). "Recommender Systems — Beyond Matrix Completion". In: *Commun. ACM* 59.11, pp. 94–102 (cit. on pp. 118, 133).

Janowicz, K., Scheider, S., Pehle, T., and Ha, G. (2012). "LinkedGeoData: A Core for a Web of Spatial Open Data". In: *Semantic Web – Interoperability, Usability, Applicability* 3.4, pp. 333–354 (cit. on p. 21).

Jiang, M., Cui, P., Liu, R., Yang, Q., Wang, F., Zhu, W., and Yang, S. (2012). "Social Contextual Recommendation". In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. CIKM '12. Maui, Hawaii, USA: ACM, pp. 45–54 (cit. on pp. 30, 31).

Jiang, X. and Tan, A. (2009). "Learning and inferencing in user ontology for personalized Semantic Web search". In: *Information Sciences* 179, pp. 2794–2808 (cit. on p. 57).

Jones, C. and Purves, R. (2008). "Geographical Information Retrieval". In: *Int. Journal of Geographical Information Science* 22.3, pp. 219–228 (cit. on p. 13).

Jøsang, A. (2012). “Robustness of Trust and Reputation Systems: Does It Matter?” In: *Trust Management VI*. Ed. by T. Dimitrakos, R. Moona, D. Patel, and D. H. McKnight. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 253–262 (cit. on p. 30).

Jøsang, A. and Goldbeck, J. (2009). “Challenges for robust trust and reputation systems”. In: *5th Int. workshop on Security and trust management (STM 2009)*. Saint Malo, France (cit. on p. 30).

Jøsang, A., Ismail, R., and Boyd, C. (Mar. 2007). “A Survey of Trust and Reputation Systems for Online Service Provision”. In: *Decis. Support Syst.* 43.2, pp. 618–644 (cit. on p. 29).

Joshi, A. and Motwani, R. (2006). “Keyword Generation for Search Engine Advertising”. In: *6th IEEE Int. Conf. on Data Mining - Workshops (ICDMW’06)*. Hong Kong, China: ACM, pp. 490–496 (cit. on p. 19).

Kim, H. and Kim, H.-J. (2014). “A framework for tag-aware recommender systems”. In: *Expert Systems with Applications* 41.8, pp. 4000–4009 (cit. on p. 22).

Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006). “Automatically Assessing Review Helpfulness”. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. EMNLP ’06. Sydney, Australia: Association for Computational Linguistics, pp. 423–430 (cit. on pp. 119, 143).

Koren, J., Zhang, Y., and Liu, X. (2008). “Personalized Interactive Faceted Search”. In: *Proceedings of the 17th International Conference on World Wide Web*. WWW ’08. Beijing, China: ACM, pp. 477–486 (cit. on p. 148).

Koren, Y. and Bell, R. (2011). “Advances in collaborative filtering”. In: *Recommender systems handbook*. Ed. by F. Ricci, L. Rokach, B. Shapira, and P. Kantor. Springer, pp. 145–186 (cit. on pp. 22, 72).

Koren, Y. (2008). “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. Las Vegas, Nevada, USA: ACM, pp. 426–434 (cit. on pp. 31, 72, 84, 117, 123).

Koren, Y., Bell, R., and Volinsky, C. (Aug. 2009). “Matrix Factorization Techniques for Recommender Systems”. In: *Computer* 42.8, pp. 30–37 (cit. on pp. 5, 125).

Korfiatis, N., Garcia-Baríocanal, E., and Sánchez-Alonso, S. (2012). “Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content”. In: *Electronic Commerce Research and Applications* 11.3, pp. 205–217 (cit. on p. 143).

Krishnamoorthy, S. (2015). “Linguistic features for review helpfulness prediction”. In: *Expert Systems with Applications* 42.7, pp. 3751–3759 (cit. on p. 143).

Kuter, U. and Golbeck, J. (2007). “SUNNY: A New Algorithm for Trust Inference in Social Networks Using Probabilistic Confidence Models”. In: *Proceedings of the 22Nd National Conference on Artificial Intelligence - Volume 2*. AAAI’07. Vancouver, British Columbia, Canada: AAAI Press, pp. 1377–1382 (cit. on pp. 23, 30).

Laurini, R. (2015). “Geographic Ontologies, Gazetteers and Multilingualism”. In: *Future Internet* 7, pp. 1–23 (cit. on pp. 13, 21).

Lee, B., Smith, G., Robertson, G. G., Czerwinski, M., and Tan, D. S. (2009). “FacetLens: Exposing Trends and Relationships to Support Sensemaking Within Faceted Datasets”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’09. Boston, MA, USA: ACM, pp. 1293–1302 (cit. on p. 25).

Leung, K. and D.L. Lee, D. (July 2010). “Deriving concept-based user profiles from search engine logs”. In: *IEEE Trans. on Knowl. and Data Eng.* 22.7, pp. 969–982 (cit. on p. 57).

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). “UpSet: Visualization of Intersecting Sets”. In: *IEEE Transactions on visualization and Computer Graphics* 20.12, pp. 1983–1992 (cit. on p. 27).

Li, H., Bhowmick, S. S., and Sun, A. (2011). “CASINO: Towards Conformity-aware Social Influence Analysis in Online Social Networks”. In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. CIKM ’11. Glasgow, Scotland, UK: ACM, pp. 1007–1012 (cit. on p. 32).

Li, H., Wu, D., and Mamoulis, N. (2014). “A Revisit to Social Network-based Recommender Systems”. In: *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*. SIGIR ’14. Gold Coast, Queensland, Australia: ACM, pp. 1239–1242 (cit. on pp. 30, 31).

Li, X., Fang, H., Yang, Q., and Zhang, J. (2018). “Who is Your Best Friend?: Ranking Social Network Friends According to Trust Relationship”. In: *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. UMAP ’18. Singapore, Singapore: ACM, pp. 301–309 (cit. on p. 30).

Li, Y.-M., Wu, C.-T., and Lai, C.-Y. (2013). “A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship”. In: *Decision Support Systems* 55.3, pp. 740–752 (cit. on p. 32).

LibraryThing (2019). “LibraryThing - a home for your books”. In: `https://www.librarything.com` (cit. on pp. 118, 132).

Lionakis, P. and Tzitzikas, Y. (2017). “PFSgeo: Preference-Enriched Faceted Search for Geographical Data”. In: *On the Move to Meaningful Internet Systems. OTM 2017 Conferences*. Ed. by H. Panetto, C. Debruyne, W. Gaaloul, M. Papazoglou, A. Paschke, C. A. Ardagna, and R. Meersman. Cham: Springer International Publishing, pp. 125–143 (cit. on pp. 26, 88).

Liu, F., Yu, C., and Meng, W. (2002). “Personalized web search by mapping user query to categories”. In: *Proc. CIKM’02*. McLean, Virginia, USA (cit. on p. 57).

Liu, F. and Lee, H. J. (July 2010). “Use of Social Network Information to Enhance Collaborative Filtering Performance”. In: *Expert Syst. Appl.* 37.7, pp. 4772–4778 (cit. on pp. 23, 30, 31).

Liu, X. and Aberer, K. (2013). “SoCo: A Social Network Aided Context-aware Recommender System”. In: *Proceedings of the 22Nd International Conference on World Wide Web*. WWW ’13. Rio de Janeiro, Brazil: ACM, pp. 781–802 (cit. on p. 30).

Lobo, M.-J., Pietriga, E., and Appert, C. (2015). “An Evaluation of Interactive Map Comparison Techniques”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: ACM, pp. 3573–3582 (cit. on p. 91).

Loepp, B., Herrmanny, K., and Ziegler, J. (2015). “Blended Recommending: Integrating Interactive Information Filtering and Algorithmic Recommender Techniques”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: ACM, pp. 975–984 (cit. on pp. 27, 28).

Lops, P., de Gemmis, M., and Semeraro, G. (2011). “Content-based recommender systems: state of the art and trends”. In: *Recommender systems handbook*. Ed. by F. Ricci, L. Rokach, B. Shapira, and P. Kantor. Springer, pp. 73–105 (cit. on p. 7).

Lu, Y., Dong, R., and Smyth, B. (2018). “Coevolutionary Recommendation Model: Mutual Learning Between Ratings and Reviews”. In: *Proc. of the 2018 World Wide Web Conference*. WWW ’18. Lyon, France: International World Wide Web Conferences Steering Committee, pp. 773–782 (cit. on p. 23).

Ludewig, M. and Jannach, D. (2018). “Evaluation of Session-based Recommendation Algorithms”. In: *User-Modeling and User-Adapted Interaction* 28.4–5, pp. 331–390 (cit. on p. 22).

Ludewig, M., Mauro, N., Latifi, S., and Jannach, D. (Submitted). “Empirical Analysis of Session-based Recommendation Algorithms”. In: *User Modeling and User-Adapted Interaction* (cit. on pp. v, 18).

Ludewig, M., Mauro, N., Latifi, S., and Jannach, D. (2019). “Performance Comparison of Neural and Non-Neural Approaches to Session-based Recommendation”. In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys ’19, pp. 462–466 (cit. on pp. vi, 18).

Luhn, H. P. (Apr. 1958). “The Automatic Creation of Literature Abstracts”. In: *IBM Journal of Research and Development* 2.2, pp. 159–165 (cit. on p. 3).

Ma, H., King, I., and Lyu, M. R. (May 2011a). “Learning to Recommend with Explicit and Implicit Social Relations”. In: *ACM Trans. Intell. Syst. Technol.* 2.3, 29:1–29:19 (cit. on pp. 30, 31).

Ma, H., Zhou, D., Liu, C., Lyu, M. R., and King, I. (2011b). “Recommender Systems with Social Regularization”. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. WSDM ’11. Hong Kong, China: ACM, pp. 287–296 (cit. on pp. 30, 31).

Ma, H., Zhou, T. C., Lyu, M. R., and King, I. (Apr. 2011c). “Improving Recommender Systems by Incorporating Social Contextual Information”. In: *ACM Trans. Inf. Syst.* 29.2, 9:1–9:23 (cit. on pp. 30, 31).

Magnani, M., Montesi, D., and Rossi, L. (2012). “Conversation retrieval for microblogging sites”. In: *Information retrieval* 15.3-4, pp. 354–372 (cit. on p. 5).

Mandala, R., Tokunaga, T., and Tanaka, H. (1999). "Combining multiple evidence from different types of thesaurus for query expansion". In: *Proc. of the 22nd Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR '99)*. New York, NY, USA: ACM, pp. 191–197 (cit. on p. 19).

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2014). "The Stanford CoreNLP Natural Language Processing Toolkit". In: *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60 (cit. on p. 64).

Marchiorini, G. (2006). "Exploratory search: from finding to understanding". In: *Communications of the ACM* 49.4, pp. 41–46 (cit. on pp. 25, 35, 88).

Marsden, P. and Friedkin, N. (1993). "Network Studies of Social Influence". In: *Sociological Methods & Research* 22.1, pp. 127–151. eprint: `https://doi.org/10.1177/0049124193022001006` (cit. on p. 30).

Massa, P. and Avesani, P. (2007). "Trust-aware Recommender Systems". In: *Proceedings of the 2007 ACM Conference on Recommender Systems*. RecSys '07. Minneapolis, MN, USA: ACM, pp. 17–24 (cit. on pp. 30, 31).

Mauro, N. (2017). "Intelligent and Personalized Community Maps". In: *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion*. IUI '17 Companion. Limassol, Cyprus: ACM, pp. 181–184 (cit. on p. vii).

Mauro, N. (2018). "Suggestion Models in Geographic Exploratory Search". In: *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. IUI '18. Tokyo, Japan: ACM, pp. 669–670 (cit. on p. vii).

Mauro, N. (2019). "Supporting the Exploration of Cultural Heritage Information via Search Behavior Analysis". In: *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. UMAP'19 Adjunct. Larnaca, Cyprus: ACM, pp. 371–376 (cit. on p. vii).

Mauro, N. and Ardissono, L. (2017a). "Concept-aware Geographic Information Retrieval". In: *Proceedings of 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. Leipzig, Germany: ACM, pp. 34–41 (cit. on pp. v, 11, 12, 20, 27, 39, 50).

Mauro, N. and Ardissono, L. (2017b). "Thematic Maps for Geographical Information Search". In: *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. UMAP '17. Bratislava, Slovakia: ACM, pp. 337–342 (cit. on p. vii).

Mauro, N. and Ardissono, L. (2018). "Session-based Suggestion of Topics for Geographic Exploratory Search". In: *Proceedings of 23rd International Conference on Intelligent User Interfaces*. IUI '18. Tokyo, Japan: ACM, pp. 341–352 (cit. on pp. v, 8, 35).

Mauro, N. and Ardissono, L. (2019). "Extending a Tag-based Collaborative Recommender with Co-occurring Information Interests". In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '19. Larnaca, Cyprus: ACM, pp. 181–190 (cit. on pp. vi, 8, 71).

Mauro, N., Ardissono, L., Di Rocco, L., Bertolotto, M., and Guerrini, G. (2019a). "Impact of Semantic Granularity on Geographic Information Search Support". In: *Proceedings of 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. Santiago, Chile: IEEE, pp. 323–328 (cit. on pp. vi, 8, 59).

Mauro, N., Ardissono, L., and Hu, Z. F. (2019b). "Multi-faceted Trust-based Collaborative Filtering". In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP '19. Larnaca, Cyprus: ACM, pp. 216–224 (cit. on pp. vi, 8, 23, 113, 116).

Mauro, N., Ardissono, L., and Lucenteforte, M. (Submitted). "Faceted Search of Heterogeneous Geographic Information for Dynamic Map Projection". In: *Information Processing & Management* (cit. on pp. v, 8, 87).

Mcnally, K., O'Mahony, M. P., and Smyth, B. (Aug. 2014). "A Comparative Study of Collaboration-based Reputation Models for Social Recommender Systems". In: *User Modeling and User-Adapted Interaction* 24.3, pp. 219–260 (cit. on pp. 23, 29).

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* 27.1, pp. 415–444 (cit. on p. 30).

Medlar, A., Pyykkö, J., and Glowacka, D. (2017). "Towards fine-grained adaptation of exploration/exploitation in Information Retrieval". In: *Proc. of the 22nd Int. Conf. on Intelligent User Interfaces*. IUI '17. Limassol, Cyprus: ACM, pp. 623–627 (cit. on p. 20).

Misztal, B. (1996). *Trust in modern societies*. Cambridge, MA: Polity Press (cit. on p. 29).

Molina, M. and Bayarri, S. (2011). "A multinational SDI-based system to facilitate disaster risk management in the Andean community". In: *Computers & Geosciences* 18, pp. 1501–1510 (cit. on p. 19).

Moradi, P. and Ahmadian, S. (2015). "A reliability-based recommendation method to improve trust-aware recommender systems". In: *Expert Systems with Applications* 42.21, pp. 7386–7398 (cit. on pp. 30, 31).

Moro, A., Raganato, A., and Navigli, R. (2014). "Entity Linking meets Word Sense Disambiguation: a Unified Approach". In: *Transactions of the Association for Computational Linguistics (TACL)* 2, pp. 231–244 (cit. on pp. 11, 64).

Mudambi, S. and Schuff, D. (2010). "What makes a helpful online review? A study of customer reviews on Amazon.com". In: *MIS Quarterly* 34.1, pp. 185–200 (cit. on p. 143).

Muhammad, K., Lawlor, A., Rafter, E., and Smyth, B. (2015). "Great Explanations: Opinionated Explanations for Recommendations". In: *Case-Based Reasoning Research and Development*. Ed. by E. Hüllermeier and M. Minor. Cham: Springer International Publishing, pp. 244–258 (cit. on p. 23).

Mui, L., Mohtashemi, M., and Halberstadt, A. (Jan. 2002). "A computational model of trust and reputation". In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pp. 2431–2439 (cit. on p. 29).

Musat, C. and Faltings, B. (2015). "Personalizing product rankings using collaborative filtering on opinion-derived topic profiles". In: *Proc. 24th IJCAI*. Buenos Aires, Argentina, pp. 830–836 (cit. on p. 23).

Musto, C., Basile, P., Lops, P., de Gemmis, M., and Semeraro, G. (2017a). “Introducing Linked Open Data in Graph-based Recommender Systems”. In: *Inf. Process. Manage.* 53.2, pp. 405–435 (cit. on p. 24).

Musto, C., Franza, T., Semeraro, G., de Gemmis, M., and Lops, P. (2018a). “Deep Content-based Recommender Systems Exploiting Recurrent Neural Networks and Linked Open Data”. In: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. UMAP ’18. Singapore, Singapore: ACM, pp. 239–244 (cit. on p. 24).

Musto, C., Lops, P., de Gemmis, M., and Semeraro, G. (2017b). “Semantics-aware Recommender Systems exploiting Linked Open Data and graph-based features”. In: *Knowledge-Based Systems* 136, pp. 1–14 (cit. on p. 24).

Musto, C., Lops, P., Gemmis, M. de, and Semeraro, G. (2019). “Justifying Recommendations Through Aspect-based Sentiment Analysis of Users Reviews”. In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. UMAP ’19. Larnaca, Cyprus: ACM, pp. 4–12 (cit. on p. 143).

Musto, C., Semeraro, G., Lovascio, C., Gemmis, M. de, and Lops, P. (2018b). “A Framework for Holistic User Modeling Merging Heterogeneous Digital Footprints”. In: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. UMAP ’18. Singapore, Singapore: ACM, pp. 97–101 (cit. on p. 22).

Nakamoto, R., Nakajima, S., Miyazaki, J., and Uemura, S. (2007). “Tag-based contextual collaborative filtering.” In: *IAENG Int. Journal of Computer Science* 34.2 (cit. on p. 23).

O’Donovan, J. and Smyth, B. (2005). “Trust in Recommender Systems”. In: *Proceedings of the 10th International Conference on Intelligent User Interfaces*. IUI ’05. San Diego, California, USA: ACM, pp. 167–174 (cit. on pp. 30, 32).

O’Mahony, M. P. and Smyth, B. (2018). “From Opinions to Recommendations”. In: *Social Information Access: Systems and Technologies*. Ed. by P. Brusilovsky and D. He. Cham: Springer International Publishing, pp. 480–509 (cit. on pp. 119, 143).

OCG (2017). “GeoSPARQL - A Geographic Query Language for RDF Data”. In: `http://www.opengeospatial.org/standards/geosparql` (cit. on p. 26).

OpenStreetMap Contributors (2017). “OpenStreetMap”. In: `https://www.openstreetmap.org` (cit. on p. 10).

Oramas, S., Ostuni, V., Di Noia, T., Serra, X., and Di Sciascio, E. (2015). “Sound and music recommendation with knowledge graphs”. In: *ACM Transactions on Intelligent Systems and Technology* 8.2, Art. 21 (cit. on pp. 20, 24).

Oren, E., Delbru, R., and Decker, S. (2006). “Extending Faceted Navigation for RDF Data”. In: *The Semantic Web - ISWC 2006*. Ed. by I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. M. Aroyo. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 559–572 (cit. on pp. 27, 92, 95, 99, 100).

OSM Contributors (2019). “taginfo”. In: `https://taginfo.openstreetmap.org/` (cit. on p. 96).

Page, L., Brin, S., Motwani, R., and Winograd, T. (Nov. 1999). *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report 1999-66. Previous number = SIDL-WP-1999-0120. Stanford InfoLab (cit. on pp. 4, 32, 120).

Palacio, D., Derungs, C., and Purves, R. (2015). "Development and evaluation of a geographic information retrieval system using fine grained toponyms". In: *Journal of Spatial Information Science JoSIS* 11, pp. 1–29 (cit. on pp. 21, 61).

Palumbo, E., Rizzo, G., and Troncy, R. (2017). "Entity2Rec: Learning User-Item Relatedness from Knowledge Graphs for Top-N Item Recommendation". In: *Proc. of the Eleventh ACM Conference on Recommender Systems*. RecSys '17. Como, Italy: ACM, pp. 32–36 (cit. on p. 24).

Papadakos, P. and Tzitzikas, Y. (2014). "Hippalus: preference-enriched faceted exploration". In: *Proc. of the workshops of EDBT/ICDT Joint Conference*. Vol. 1133. Athens, Greece: CEUR, pp. 167–172 (cit. on p. 26).

Parvin, H., Moradi, P., and Esmaeili, S. (2019). "TCFACO: Trust-aware collaborative filtering method based on ant colony optimization". In: *Expert Systems with Applications* 118, pp. 152–168 (cit. on pp. 30, 31).

Peltonen, J., Belorustceva, K., and Ruotsalo, T. (2017). "Topic-relevance map: visualization for improving search result comprehension". In: *Proc. of the 22nd Int. Conf. on Intelligent User Interfaces*. IUI '17. Limassol, Cyprus: ACM, pp. 611–622 (cit. on p. 26).

Petrelli, D., Mazumdar, S., Dadzie, A.-S., and Ciravegna, F. (2009). "Multi Visualization and Dynamic Query for Effective Exploration of Semantic Data". In: *The Semantic Web - ISWC 2009*. Ed. by A. Bernstein, D. R. Karger, T. Heath, L. Feigenbaum, D. Maynard, E. Motta, and K. Thirunarayan. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 505–520 (cit. on pp. 26, 27).

Porrini, R., Palmonari, M., and Vizzari, G. (2014). "Composite match autocompletion (COMMA): a semantic result-oriented autocompletion technique for e-marketplaces". In: *Web Intelligence and Agent Systems: an international journal* 12, pp. 35–49 (cit. on p. 20).

Qazi, A., Syed, K. B. S., Raj, R. G., Cambria, E., Tahir, M., and Alghazzawi, D. (2016). "A concept-level approach to the analysis of online review helpfulness". In: *Computers in Human Behavior* 58, pp. 75–81 (cit. on p. 143).

Qian, F., Zhao, S., Tang, J., and Zhang, Y. (2016). "SoRS: Social recommendation using global rating reputation and local rating similarity". In: *Physica A: Statistical Mechanics and its Applications* 461, pp. 61–72 (cit. on pp. 31, 32).

Qiu, Y. and Frei, H. (1993). "Concept based query expansion". In: *Proc. of the 16th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*. SIGIR '93. New York, NY, USA: ACM, pp. 160–169 (cit. on p. 19).

Quercia, D., Schifanella, R., and Aiello, L. (2014). "The Shortest Path to Happiness: Recommending Beautiful, Quiet, and Happy Routes in the City". In: *Proc. of the 25th ACM Conference on Hypertext and Social Media*. HT '14. Santiago, Chile: ACM, pp. 116–125 (cit. on p. 88).

Rafailidis, D. and Crestani, F. (2017). “Learning to Rank with Trust and Distrust in Recommender Systems”. In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*. RecSys ’17. Como, Italy: ACM, pp. 5–13 (cit. on p. 32).

Raghavan, S., Gunasekar, S., and Ghosh, J. (2012). “Review Quality Aware Collaborative Filtering”. In: *Proceedings of the Sixth ACM Conference on Recommender Systems*. RecSys ’12. Dublin, Ireland: ACM, pp. 123–130 (cit. on pp. 119, 143).

Ragone, A., Tomeo, P., Magarelli, C., Di Noia, T., Palmonari, M., Maurino, A., and Di Sciascio, E. (2017). “Schema-summarization in Linked-data-based Feature Selection for Recommender Systems”. In: *Proc. of the Symposium on Applied Computing*. SAC ’17. Marrakech, Morocco: ACM, pp. 330–335 (cit. on p. 24).

Rasmussen, R. and Hertzum, M. (2013). “Visualizing the application of filters: A comparison of blocking, blurring, and colour-coding whiteboard information”. In: *International Journal of Human-Computer Studies* 71.10, pp. 946–957 (cit. on p. 148).

Ratzan, L. (2004). *Understanding information systems: what they do and why we need them*. Chicago, USA: American Library Association (cit. on p. 35).

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). “GROUPLENS: an open architecture for collaborative filtering of netnews”. In: *Proc. of the 1994 ACM conference on Computer Supported Cooperative Work (CSCW 1994)*. Chapel Hill. North Carolina: ACM, pp. 175–186 (cit. on p. 5).

Resnick, P. and Zeckhauser, R. (2002). “Trust among strangers in Internet transactions: empirical analysis of eBay’s reputation system”. In: *The Economics of the Internet and E-commerce (Advances in Applied Microeconomics*. Vol. 11. Emerald Group Publishing Limited, pp. 127–157 (cit. on p. 29).

Ricci, F., Rokach, L., and Shapira, B. (2011). “Introduction to Recommender Systems Handbook”. In: *Recommender Systems Handbook*. Ed. by F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. Boston, MA: Springer US, pp. 1–35 (cit. on pp. 5, 27, 74, 126).

Richthammer, C., Weber, M., and Pernul, G. (2017). “Reputation-Enhanced Recommender Systems”. In: *Trust Management XI*. Ed. by J.-P. Steghöfer and B. Esfandiari. Cham: Springer International Publishing, pp. 163–179 (cit. on p. 30).

Rieh, S. Y. and Xie, H. ((2006). “Analysis of multiple query reformulations on the web: The interactive information retrieval context”. In: *Information Processing & Management* 42.3, pp. 751–768 (cit. on p. 44).

Rocchio, J. J. (1971). “Relevance Feedback in Information Retrieval”. In: Harvard Univ., Cambridge, MA: ISR-9 (cit. on p. 4).

Ronen, R., Yom-Tov, E., and Lavee, G. (2016). “Recommendations meet web browsing: enhancing collaborative filtering using internet browsing logs”. In: *Proc.of 2016 IEEE 32nd International Conference on Data Engineering (ICDE)*. Kelsinki, Finland, pp. 1230–1238 (cit. on p. 23).

Sacco, G. M. (2000). “Dynamic taxonomies: a model for large information bases”. In: *IEEE Trans. on Knowledge and Data Engineering* 12.3, pp. 468–479 (cit. on p. 25).

Salton, G. (1968). *Automatic Information Organization and Retrieval.* McGraw Hill Text (cit. on p. 4).

Sanderson, M. and Croft, W. B. (May 2012). "The History of Information Retrieval Research". In: *Proceedings of the IEEE* 100.Special Centennial Issue, pp. 1444–1451 (cit. on p. 3).

Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). "Collaborative Filtering Recommender Systems". In: *The Adaptive Web: Methods and Strategies of Web Personalization*. Ed. by P. Brusilovsky, A. Kobsa, and W. Nejdl. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 291–324 (cit. on p. 6).

Sen, S., Vig, J., and Riedl, J. (2009). "Tagommenders: Connecting Users to Items Through Tags". In: *Proc. of the 18th International Conference on World Wide Web*. WWW '09. Madrid, Spain: ACM, pp. 671–680 (cit. on p. 23).

Shen, R.-P., Zhang, H.-R., Yu, H., and Min, F. (2019). "Sentiment based matrix factorization with reliability for recommendation". In: *Expert Systems with Applications* (cit. on p. 143).

Sieg, A., Mobasher, B., and Burke, R. (2007). "Web Search Personalization with Ontological User Profiles". In: *Proc. of the 16th ACM Conf. on Information and Knowledge Management*. CIKM '07. Lisbon, Portugal: ACM, pp. 525–534 (cit. on pp. 22, 72, 80).

Sieg, A., Mobasher, B., and Burke, R. (2010). "Ontology-based collaborative recommendation". In: *Computing* (cit. on pp. 22, 57, 72).

Smyth, B., Balfe, E., Freine, J., Briggs, P., Coyle, M., and Boydell, O. (2005). "Exploiting query repetition and regularity in an adaptive community-based web search engine". In: *User Modeling and User-adapted Interaction - The Journal of Personalization Research* 14.5, pp. 383–423 (cit. on p. 57).

Sontag, D., Collins-Thompson, K., Bennett, P., White, R., Dumais, S., and Billerbeck, B. (2012). "Probabilistic Models for Personalizing Web Search". In: *Proc. of the 5th ACM Int. Conf. on Web Search and Data Mining*. WSDM '12. Seattle, Washington, USA: ACM, pp. 433–442 (cit. on p. 57).

Sparck Jones, K. (1988). "Document Retrieval Systems". In: ed. by P. Willett. London, UK, UK: Taylor Graham Publishing. Chap. A Statistical Interpretation of Term Specificity and Its Application in Retrieval, pp. 132–142 (cit. on p. 4).

Stadler, C., Martin, M., and Auer, S. (2014). "Exploring the Web of Spatial Data with Facete". In: *Proc. of the 23rd International Conference on World Wide Web*. WWW '14 Companion. Seoul, Korea: ACM, pp. 175–178 (cit. on p. 26).

Su, K., Xiao, B., Liu, B., Zhang, H., and Zhang, Z. (2017). "TAP: A personalized trust-aware QoS prediction approach for web service recommendation". In: *Knowledge-Based Systems* 115, pp. 55–65 (cit. on p. 32).

Suchanek, F., Kasneci, G., and Weikum, G. (2008). "YAGO: A Large Ontology from Wikipedia and WordNet". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 6.3, pp. 203–217 (cit. on p. 146).

Sutcliffe, A., Ennis, M., and Hu, J. (Nov. 2000). "Evaluating the Effectiveness of Visual User Interfaces for Information Retrieval". In: *Int. J. Hum.-Comput. Stud.* 53.5, pp. 741–763 (cit. on p. 25).

Tang, J., Gao, H., and Liu, H. (2012). "MTrust: Discerning multi-faceted trust in a connected world". English (US). In: *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. WSDM 2012, pp. 93–102 (cit. on p. 124).

Tang, J., Hu, X., Gao, H., and Liu, H. (2013). "Exploiting Local and Global Social Context for Recommendation". In: *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. IJCAI '13. Beijing, China: AAAI Press, pp. 2712–2718 (cit. on pp. 16, 23, 30, 31, 72, 113, 114, 117, 120, 124).

Taube, M., Gull, C. D., and Wachtel, I. S. (1952). "Unit terms in coordinate indexing". In: *American Documentation* 3.4, pp. 213–218. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.5090030404` (cit. on p. 3).

Teevan, J., Dumais, S., and Horvitz, E. (2005). "Personalizing Search via Automated Analysis of Interests and Activities". In: *Proc. of the 28th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*. SIGIR '05. Salvador, Brazil: ACM, pp. 449–456 (cit. on pp. 22, 57).

Tsai, C.-H. and Brusilovsky, P. (Aug. 2019). "Exploring Social Recommendations with Visual Diversity-Promoting Interfaces". In: *ACM Trans. Interact. Intell. Syst.* 10.1, 5:1–5:34 (cit. on pp. 27, 148).

Tso-Sutter, K. H. L., Marinho, L. B., and Schmidt-Thieme, L. (2008). "Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms". In: *Proceedings of the 2008 ACM Symposium on Applied Computing*. SAC '08. Fortaleza, Ceara, Brazil: ACM, pp. 1995–1999 (cit. on p. 23).

Tvarožek, M., Barla, M., Frivolt, G., Tomša, M., and Bieliková, M. (2008). "Improving Semantic Search via Integrated Personalized Faceted and Visual Graph Navigation". In: *Proceedings of the 34th Conference on Current Trends in Theory and Practice of Computer Science*. SOFSEM'08. Nový Smokovec, Slovakia: Springer-Verlag, pp. 778–789 (cit. on p. 148).

Tvarožek, M. and Bieliková, M. (2010). "Generating Exploratory Search Interfaces for the Semantic Web". In: *Human-Computer Interaction*. Ed. by P. Forbrig, F. Paternó, and A. Mark Pejtersen. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 175–186 (cit. on p. 148).

Tzitzikas, Y., Manolis, N., and Papadakos, P. (Apr. 2017). "Faceted Exploration of RDF/S Datasets: A Survey". In: *J. Intell. Inf. Syst.* 48.2, pp. 329–364 (cit. on p. 26).

Vahedian, F., Burke, R., and Mobasher, B. (2017). "Multirelational Recommendation in Heterogeneous Networks". In: *ACM Trans. Web* 11.3, pp. 1–34 (cit. on p. 24).

Victor, P., Cornelis, C., Cock, M. D., and Teredesai, A. M. (Jan. 2011). "Trust- and Distrust-Based Recommendations for Controversial Reviews". In: *IEEE Intelligent Systems* 26.1, pp. 48–55 (cit. on p. 32).

Voghera, A., Crivello, R., Ardissono, L., Lucenteforte, M., Savoca, A., and L. La Riccia (2016). "Production of spatial representations through collaborative mapping. An experiment". In: *Proc. of 9th Int. Conf. on Innovation in Urban and Regional Planning (INPUT 2016)*, pp. 356–361 (cit. on pp. 12, 39, 61, 90).

Voghera, A., La Riccia, L., Lucenteforte, M., and Savoca, A. (2018). “Rappresentare il territorio attraverso il collaborative mapping. Il progetto OnToMap”. In: *Smartness e healthiness per la transizione verso la resilienza. Orizzonti di ricerca interdisciplinare sulla città e il territorio*. Vol. 86.1. BETHA - Built Environment Technologies and Healthy Architectures 4. Milano, Italy: Franco Angeli, pp. 239–252 (cit. on p. 90).

W3C (2017). “SPARQL Query Language for RDF”. In: `https://www.w3.org/TR/rdf-sparql-query/` (cit. on p. 26).

W3C (2018). “Linked Data”. In: `https://www.w3.org/standards/semanticweb/data` (cit. on p. 25).

Wang, Y., Huang, H., and Feng, C. (2017). “Query expansion based on a feedback concept model for microblog retrieval”. In: *Proc. of the 26th Int. Conf. on World Wide Web*. WWW ’17. Perth, Australia: International World Wide Web Conferences Steering Committee, pp. 559–568 (cit. on pp. 19, 59).

Wang, Y., Li, H., Wang, H., and Zhu, K. (2012). “Concept-based web search”. In: *Conceptual Modeling. ER 2012. Lecture Notes in Computer Science, vol 7532*. Ed. by P. Atzeni and D. Cheung. Berlin, Heidelberg: Springer, pp. 449–462 (cit. on pp. 19, 39).

White, R., Bennett, P., and Dumais, S. (2010). “Predicting short-term interests using activity-based search context”. In: *Proc. of the 19th ACM Int. Conf. on Information and Knowledge Management*. CIKM ’10. Toronto, ON, Canada: ACM, pp. 1009–1018 (cit. on p. 57).

White, R., Bilenko, M., and Cucerzan, S. (2007). “Studying the use of popular destinations to enhance web search interaction”. In: *Proc. of the 30th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*. SIGIR ’07. Amsterdam, The Netherlands: ACM, pp. 159–166 (cit. on p. 38).

White, R. and Roth, R. (2006). “Supporting exploratory search”. In: *Communications of the ACM* 49.4, pp. 36–39 (cit. on pp. 25, 35).

WordNet (2017). “WordNet - a lexical database for English”. In: `https://wordnet.princeton.edu/` (cit. on pp. 39, 78).

Wu, W., Li, H., Wang, H., and Zhu, K. (2012). “Probase: A Probabilistic Taxonomy for Text Understanding”. In: *Proc. of the 2012 ACM SIGMOD Int. Conf. on Management of Data*. Scottsdale, Arizona, USA: ACM, pp. 481–492 (cit. on pp. 39, 78, 146).

Xiang, R., Neville, J., and Rogati, M. (2010). “Modeling Relationship Strength in Online Social Networks”. In: *Proceedings of the 19th International Conference on World Wide Web*. WWW ’10. Raleigh, North Carolina, USA: Association for Computing Machinery, pp. 981–990 (cit. on p. 125).

Yang, B., Lei, Y., Liu, J., and Li, W. (Aug. 2017). “Social Collaborative Filtering by Trust”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.8, pp. 1633–1647 (cit. on pp. 23, 30, 31, 72).

Yang, X., Steck, H., and Liu, Y. (2012). “Circle-based Recommendation in Online Social Networks”. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’12. Beijing, China: ACM, pp. 1267–1275 (cit. on pp. 30, 31).

Yee, K.-P., Swearingen, K., Li, K., and Hearst, M. (2003). “Faceted Metadata for Image Search and Browsing”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’03. Ft. Lauderdale, Florida, USA: ACM, pp. 401–408 (cit. on p. 25).

Yelp (2019a). “Yelp”. In: `https://www.yelp.com` (cit. on pp. 114, 118).

Yelp (2019b). “Yelp Dataset Challenge”. In: `https://www.yelp.com/dataset_challenge` (cit. on pp. 73, 77, 116, 118, 127).

Yuan, Q., Chen, L., and Zhao, S. (2011). “Factorization vs. Regularization: Fusing Heterogeneous Social Relationships in Top-n Recommendation”. In: *Proceedings of the Fifth ACM Conference on Recommender Systems*. RecSys ’11. Chicago, Illinois, USA: ACM, pp. 245–252 (cit. on p. 31).

Zhang, Z.-K., Zhou, T., and Zhang, Y.-C. (2011). “Tag-aware recommender systems: a state-of-the-art survey”. In: *Journal of computer science and technology* 26.5, p. 767 (cit. on p. 22).

Zhao, T., McAuley, J., and King, I. (2019). “LibraryThing Dataset”. In: `http://deepx.ucsd.edu/datasets/librarything/lthing_data.tar.gz` (cit. on p. 116).

Zhen, Y., Li, W.-J., and Yeung, D.-Y. (2009). “TagiCoFi: Tag Informed Collaborative Filtering”. In: *Proceedings of the Third ACM Conference on Recommender Systems*. RecSys ’09. New York, New York, USA: ACM, pp. 69–76 (cit. on p. 23).

List of Figures

List of Tables