AN ONTOLOGY-BASED RECOMMENDER SYSTEM USING SCHOLAR'S BACKGROUND KNOWLEDGE

BAHRAM AMINI VALASHANI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2016

AN ONTOLOGY-BASED RECOMMENDER SYSTEM USING SCHOLAR'S BACKGROUND KNOWLEDGE

BAHRAM AMINI VALASHANI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2016

DEDICATION

To our prophet, ***Mohammad***, the last messenger of truth, fraternization and kindness.

To my family, ***my wife Maryam, and my sons Mohammad Javad and Amir Ali.***

To my supervisors, ***Dr. Roliana Ibrahim, Dr. Mohd Shahizan Othman, and Associate Professor Dr. Mohammad Ali Nematbakhsh.***

And to all who supported me in this study.

ACKNOWLEDGEMENT

In the name of Allah, Most Gracious, Most Merciful. All praise and thanks are due to Allah S.W.T, and peace and blessings be upon his messenger, Mohammed (peace be upon him). I thank Allah for granting me perseverance and strength I needed to complete this thesis. Here, I would like to express my heartfelt gratitude to my supervisors Dr. Roliana Ibrahim and Dr. Mohd Shahizan Othman, for their encouragement, guidance, critics, and supports. They have been incredibly understanding, helpful, and generous throughout the process, and have truly been great mentors, and I owe here my deepest thanks. In addition, I wish to express my sincere appreciation to my external supervisor, Associated Professor Dr. Mohammad Ali Nematbakhsh from the University of Isfahan, Iran, for encouraging, assisting, and supervising me through this research. Without his guidance and advices, this study would not have been possible.

Also, in conducting this thesis, I was in contact with many researchers and academicians who have contributed towards my understanding and thoughts. I would like to thank Professor Dr. Bob Colomb, Professor Dr. Nguyen, Dr. Mohammad Nazir Ahmad, and the members of Software Engineering Group (SERG) for their comments and advice. Besides, I would like to thank the authority of Universiti Teknologi Malaysia (UTM) for providing me with a good environment and facilities which I needed during the process. My sincere appreciation also extends to all of my friends and others who have provided assistance at various occasions. Their tips and views were useful indeed.

ABSTRACT

Scholar's recommender systems recommend scientific articles based on the similarity of articles to scholars' profiles, which are a collection of keywords that scholars are interested in. Recent profiling approaches extract keywords from the scholars' information such as publications, searching keywords, and homepages, and train a reference ontology, which is often a general-purpose ontology, in order to profile the scholars' interests. However, such approaches do not consider the scholars' knowledge because the recommender system only recommends articles which are syntactically similar to articles that scholars have already visited, while scholars are interested in articles which contain comparatively new knowledge. In addition, the systems do not support multi-area property of scholars' knowledge as researchers usually do research in multiple topics simultaneously and are expected to receive focused-topic articles in each recommendation. To address these problems, this study develops a domain-specific reference ontology by merging six Web taxonomies and exploits Wikipedia as a conflict resolver of ontologies. Then, the knowledge items from the scholars' information are extracted, transformed by DBpedia, and clustered into relevant topics in order to model the multi-area property of scholars' knowledge. Finally, the clustered knowledge items are mapped to the reference ontology by using DBpedia to create clustered profiles. In addition a semantic similarity algorithm is adapted to the clustered profiles, which enables recommendation of focused-topic articles that contain new knowledge. To evaluate performance of the proposed approach, three different data sets from scholars' information in Computer Science domain are created, and the precisions in different cases are measured. The proposed method, in comparison with the baseline methods, improves the average precision by 6% when the new reference ontology along with the full scholars' knowledge is utilized, by an extra 7.2% when scholars' knowledge is transformed by DBpedia, and further 8.9% when clustered profile is applied. Experimental results certify that using knowledge items instead of keywords for profiling as well as transforming the knowledge items by DBpedia can significantly improve the recommendation performance. Besides, the domain-specific reference ontology can effectively capture the full scholars' knowledge which results to more accurate profiling.

ABSTRAK

Sistem-sistem pengesyor bagi sarjana mencadangkan artikel-artikel saintifik berdasarkan kesamaan artikel dengan profil sarjana iaitu satu koleksi kata kunci yang diminati oleh para sarjana. Pendekatan pemprofilan kebelakangan ini mengekstrak kata kunci daripada maklumat sarjana seperti penerbitan, pencarian kata kunci, dan laman utama, dan melatih sebuah ontologi rujukan yang pada kebiasaannya adalah satu ontologi kegunaan umum bagi memprofil minat para sarjana. Walau bagaimanapun, pendekatan sedemikian tidak mempertimbangkan pengetahuan para sarjana kerana sistem pengesyor hanya mengesyor artikel-artikel yang secara sintetiknya serupa dengan artikel yang telah sarjana lawati, manakala mereka berminat dengan artikel-artikel yang mungkin mengandungi pengetahuan baru berbanding pengetahuan sedia ada. Sebagai tambahan, sistem-sistem tersebut tidak menyokong penyelidikan pelbagai bidang pengetahuan sarjana sedangkan penyelidik biasanya membuat kajian dalam pelbagai topik pada masa yang sama, dan menjangkakan untuk menerima artikel-artikel berfokuskan topik dalam setiap cadangan. Bagi menangani masalah-masalah itu, kajian ini membangunkan satu rujukan ontologi domain spesifik dengan menggabungkan enam taksonomi pada web dan mengeksploitasi Wikipedia sebagai satu penyelesai konflik ontologi. Kemudian, butir-butir pengetahuan daripada maklumat sarjana diekstrak, dipindahkan melalui DBpedia dan dikelompokkan kepada topik-topik yang relevan untuk memodelkan pelbagai bidang pengetahuan sarjana. Akhirnya, butir-butir kelompok pengetahuan itu dipetakan pada ontologi rujukan dengan menggunakan DBpedia bagi mencipta profil-profil berkelompok. Seterusnya, satu algoritma persamaan semantik diadaptasikan kepada profil berkelompok yang membolehkan cadangan kepada artikel-artikel yang berfokus topik yang mengandungi pengetahuan baru. Bagi menilai prestasi kaedah yang dicadangkan, tiga set data berbeza daripada maklumat para sarjana dalam domain Sains Komputer telah dibangunkan dan ketepatan dalam kes-kes berlainan diukur. Berbanding dengan kaedah garis dasar, kaedah yang dicadangkan meningkatkan purata ketepatan sebanyak 6% apabila ontologi rujukan beserta pengetahuan penuh sarjana digunakan, melebihi sebanyak 7.2% apabila pengetahuan sarjana dipindahkan melalui DBpedia, dan seterusnya sebanyak 8.9% apabila profil berkelompok diaplikasikan. Hasil uji kaji mengesahkan bahawa penggunaan butir-butir pengetahuan berbanding kata kunci bagi pemprofilan dan juga pemindahan butir-butir pengetahuan melalui DBpedia dapat meningkatkan prestasi cadangan dengan signifikan. Di samping itu, ontologi rujukan domain spesifik boleh merangkumi pengetahuan penuh sarjana dengan lebih berkesan serta seterusnya membawa kepada pemprofilan yang lebih tepat.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACL ARC	-	ACL Anthology Reference Corpus
AP	-	Average Precision
BOTW	-	Best Of The Web Taxonomy
CB	-	Content Based
CBR	-	Case-Based Reasoning
CBRS	-	Constraint Based Recommender System
CF	-	Collaborative Filtering
CL	-	Computational Linguistic
CRF	-	Conditional Random Fields
CS	-	Computer Science domain
CV	-	Curriculum Vitae
DBLP	-	DBLP dataset
DF	-	Document Frequency
DL	-	Digital Library
DRS	-	Demographic Recommender System
En	-	Entropy-based feature selection
FE	-	Feature Extraction
FS	-	Feature Selection
HTML	-	Hyper Text Markup Language
IPL	-	The Internet Public Library
ID	-	Identifier
IS	-	Information System
IR	-	Information Retrieval
JSD	-	Jensen-Shannon Divergence
K-NN	-	k-Nearest Neighbor classifier
LCC	-	Library of Congress Classification
LSI	-	Latent Semantic Indexing

KBRS	-	Knowledge Based Recommender System
MAS	-	Microsoft Academic Search
ME	-	Margin Error
MEP	-	Mediated Profiles data set
MI	-	Mutual Information
NLP	-	Natural Language Processing
ODP	-	Open Directory Project
OWL	-	Web Ontology Language
POS	-	Part-Of-Speech
RDF	-	Resource Description Locator
RS	-	Recommender System
SA	-	Spreading Activation
SNA	-	Social Networks Analysis
SPD	-	Starting Point Directory
TF	-	Term Frequency
TM	-	Text Mining
TS	-	Term Strength
URI	-	Unified Resource Identifier
URL	-	Uniform Resource Locator
VLIB	-	The Web Virtual Library
VLR	-	Volunteer Researchers data set
VSM	-	Vector Space Model
WSD	-	Word Sense Disambiguation
XDL	-	arXiv Data set

LIST OF SYMBOLS

tf-idf	-	Term frequency-inverse document frequency
R	-	Real number
Chi^2, χ^2	-	Chi-Square statistic
K	-	Number of clusters
Z*	-	Score for level of confidence
σ	-	Standard Deviation
$\bar{x}$	-	Mean

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter describes an overview of scholars' recommender systems, the characteristics of scholar's domain, research problem and objectives, research scope, and importance of the study. Moreover, the contribution of the research and the structure of the thesis are explained.

1.1 Overview

Typically, a Digital Library (DL) contains electronic collections of various information and services including scientific articles and search services, which supplies to the scholar community (Smeaton and Callan, 2005). Digital libraries join scholars, information, and technologies to provide an infrastructure for convenient information search and retrieval. As the flood of information in digital libraries is explosive, Information Retrieval (IR) from digital libraries becomes a complex and challenging task (Weiss et al., 2010). Unfortunately, traditional digital libraries retrieve a large number of irrelevant information in response of researchers' queries, raising "information overload", which overwhelm them with retrieving enormous useless articles (Uchyigit, 2009; Sugiyama and Kan, 2010).

To address this problem, personalization approaches have been recently proposed, which filter out irrelevant articles and recommend unseen articles as similar as possible to the scholar's interests (Uchyigit and Ma, 2008). In fact, personalized approaches explore users' interests and conforms the result of queries by filtering out unrelated results and recommend new items (articles) that scholars might be interested in (Castellano et al. 2009). Hence, recommendation technologies qualify digital libraries by filtering irrelevant information and re-ranking the retrieved information to improve information retrieval, and ultimately, increasing users' satisfaction.

There are two broad categories of recommender systems including content-based and collaborative filtering (Ricci et al., 2011). The content-based (CB) approaches (Pazzani and Billsus, 2007; Lops et al., 2011) identify the common characteristics of items that have been received favorable ratings from users (scholars), and then recommend new items that have the similar characteristics. In CB, it is assumed that rich information about the items is available and the items are represented in the form of a feature vector. For example, for text documents such as news, articles, or Web page content, vectors often contain the term frequency/inverse document frequency (tf-idf) weights of the most informative keywords (Salton et al., 1975).

However, content-based systems suffer from two major problems: over-specialization and limited content analysis (Ricci et al., 2011). Over-specialization (Anand and Mobasher, 2005) may occur when the prediction of a user for an item is high and the item is too similar to the items liked previously by the user. For example, in scholar domain, the system may recommend an article which has the same topic or the set of keywords as those articles which have been previously read by the scholar. Therefore, the system may fail to recommend articles that are different but still interesting to the scholar. The problem of limited content analysis (Lops et al., 2011) emerges from the fact that little information about the content of items or users' behavior is available. Typically, the lack of insufficient information is due to privacy, low quality (noisy data), access issue, or imprecise of content. For instance, a scholar might avoid providing demographic information (privacy issue), or ambiguous keywords in articles may result in imprecise content.

The collaborative filtering (CF) (Schafer et al., 2007; Roza et al., 2010) are pure usage-based approaches which rely on the ratings of users on items as well as rating of other users in the same community on items. The main idea is that the rating of a user for an unseen item is likely to be similar to that of other users that have rated the item in a similar manner. In other words, a user is likely to rate an item A similar to rating of similar users that have given rating to A. However, the usage-based approach exposes four important drawbacks because the recommendation process depends highly on the existing user transaction data. First, the rating of articles should be available before suggesting an article to scholar. It is referred to as new item problem (Dai and Mobasher, 2009). Second, new scholars to the system requires to rate a number of articles before obtaining appropriate recommendations, referred to new user problem (Anyanwu and Sheth, 2003). Third, when the proportion of scholars to articles is significantly small, only a few numbers of articles will be recommended. Lastly, similar scholars who rated similar set of articles might possess different level of knowledge, resulting in distinct preferences, and in turn, imprecise recommendation.

To point out the source of issues, "user profile" is focused: Recommender systems model users' interests in "profiles", which involve the features of most important facts that users are particularly interested in (Schiaffino and Amandi, 2009). A user profile is a structured representation of characteristics and features of users' past experiences such as searched keywords, knowledge, or history of feedback (Kadima and Malek, 2010; Snasel et al., 2010). User profiles are created manually by user's information (explicit method), or created automatically during the course of user's interactions with the system using the user's contextual information and intelligent techniques (Wei and Lei, 2009). A user profile is the core component of a recommender system, and is being treated as a network of concepts that is updated based on the user's feedback (Lopes, Martins Souto, et al., 2007). There are a number of profiling methods including ontology-based, which engage ontologies for profiling.

In ontology-based profiling, concepts are usually extracted from a pre-existed taxonomy or so-called reference ontology (Mohammed et al. 2010). Reference ontology is a hierarchy of topics, where the topics are used to classify items being recommended. Thus, an ontology-based user profile is a set of concepts (or nodes) in a hierarchical structure, each node annotated with an interest score, which represent the degree of interest that the user has interested in that concept (Sieg et al., 2010). Many research has proven that ontology-based approaches are effective for profiling, which successively improves the recommendation precision (Yujie and Licai, 2010; Sieg et al., 2010).

Today, recommender systems are widely used in a variety of information systems including digital libraries (Lopes, Martins, et al., 2007), educational systems (Satyanarayana and Rajagopalan, 2007), and treated as a tool to deliver personalized information (Abbar et al., 2007). Personalization approaches are applied in many areas of scholar's domain such as ranking Web sites (Zhuhadar and Nasraoui, 2010), e-learning (Tsatsou et al., 2009), and scientific articles (Yang, 2010; Sugiyama and Kan, 2010). Good examples of successful recommender system for digital libraries are ACM, PubMed, Elsevier, Google Scholar, and CiteSeerX, which rely on the content of articles as well as collaborative information of scholars (i.e., the rating of scholars on articles). Moreover, emerging technologies such as agents (Godoy and Amandi, 2007), mobile (Ricci, 2010), and the social Web (Groh et al., 2012) are incorporated into basic tasks of recommender systems such as resource discovery, distributed information gathering, inferring user preferences (user modeling), and item filtering to provide more sophisticated recommendations.

1.2 Characteristics of Scholar’s Domain

As mentioned above, many researches focused on technical properties of recommender systems, whereas this study aims to interconnect cognitive characteristics of scholars to the recommender system. Thus, the characteristics of scholars are reviewed here. Scholar’s domain is associated with particular properties that make it distinctive (Drachsler et al., 2007). In general, research task is knowledge-intensive because the navigation through research space is associated with many decision choices (Eppler, 2006). Scholars study articles to capture “new knowledge” and get into deeper understanding of subjects that they are interested in. Relevant articles provide researcher with opportunities to make “connections” between their prior knowledge and the “new knowledge” being captured (Strangman and Hall, 2009). From a scholar's viewpoint, an article is interesting if it has been grounded on his background knowledge, i.e., the article contains relevant concepts of some topics that have been known to the scholar (Strangman and Hall, 2009). Scholars' prior knowledge is the starting point of conducting a research task (Berkovsky et al. 2007). A typical research demands frequent analysis and reading of different articles (e.g., digital papers, books, and similar) to update research path. Individual scholar has their own tendency and attitude in selecting an article for reading because of differences in their proficiency level and background knowledge (Bitonto et al., 2010).

Moreover, scholars acquire knowledge upon a "cumulative learning" process (Gagné, 1968), where current knowledge helps to capture relevant new knowledge. Gagne (1968) stresses the cumulative nature of learning in which mastery of higher-level knowledge depends primarily upon the prior mastery of lower-order knowledge or concepts. Accordingly, the knowledge is arranged in a hierarchical structure, so that successful learning begins with studying lower-order (or basic) knowledge and progresses upwards. Figure 1.1 depicts the cumulative learning process in general. It shows that scholars capture new knowledge in a pyramid style (or hierarchical structure), where starts from learning basic knowledge and grows towards more advanced knowledge. In each progress, new knowledge is captured based upon the known knowledge. Accordingly, a scholar can effectively learn, for example, about the "Text Mining", if he has already basic knowledge about "Concept Extraction", "Text Clustering", and "Text Categorization".

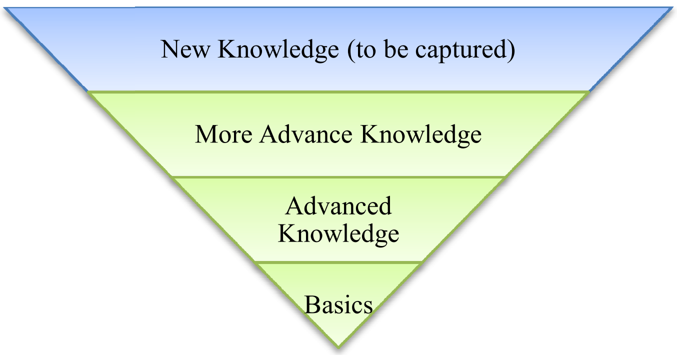

Figure 1.1: Hierarchy of learning, the relationship between acquisition of new knowledge (blue part) and the known knowledge (green parts)

In addition, full researchers (i.e., professors) often do search in several topics simultaneously, as they work in different research projects and capture knowledge in multiple topics. For instance, a professor does research in "Database Systems" topic while conducting another research in "Semantic Web". This is recalled as "multi-area" property of scholar domain. Thus, the scholars' profiles should model multiple research topics, and hence, the recommender system should suggest focused-topic articles (Adomavicius et al., 2011). Traditional recommender systems, on the other hand, conclude that the researcher has sufficient knowledge in several topics, though the fact is true, but the result of filtering is divergence to multiple topics with low precision, i.e., the user receive articles which encompass keywords of several topics.

1.3 Problem Statement

Developing a recommender system which recommends scientific articles relevant to research topics and fit into the research path is a challenging problem (Devedžić 2006). Such system essentially requires an extensive analysis of research context as well as scholars' prior knowledge (Berkovsky et al. 2007). In traditional approaches, it is assumed that the new choices of a scholar are "very similar" to the choices made by him in the nearby past (Yang et al., 2010; Sugiyama and Kan, 2010; Jomsri et al., 2012), while it is in sharp contrast with the characteristics of scholars in which continually seeking for new knowledge and following up learning task (Bogers and Bosch, 2008). In fact, scholars are interested in articles which encompass new knowledge in line with their knowledge scope- the articles that are correspond to the search keywords while encompass relatively new knowledge (Strangman and Hall, 2009). Such new knowledge is called "complementary knowledge" as it fulfills the scholars' needs and is relevant to the research topic.

In practice, such systems (Ricci et al., 2011) represents the scholars' documents (publications, Web page information, citation information) as well as scholars' profiles by a statistical model such as Vector Space Model (Salton et al. 1975), and calculate the similarity of documents with the profiles using a statistical method such as Cosine Similarity (Manning et al., 2009), Jaccard, or Sorenson-Dice measures. Such statistical methods rely on surface overlap of participant vectors. Accordingly, two vectors are similar if they share the same vocabularies (Huang 2008). Though, in statistical methods, the precision is pretty acceptable, but the role of words in the vectors are neglected (Khan et al. 2010), and thus, the presence of new knowledge items (in the sense of words, terms, etc) in articles results in low precision, and likely rejection of valuable articles.

Moreover, as mentioned in previous section, full researchers do search in several topics simultaneously. Thus, the scholar's profiles should model multiple topics of scholars' interests, and in turn, underlying recommender system can recommend focused-topic articles. For this reason, traditional recommender systems such as (Liang et al., 2008; Duong et al., 2009; Yang et al., 2010) conclude that the researcher has adequate knowledge in multiple topics, so the result of filtering process is divergence to several topics with low precision. For example, if a scholar is interested in multiple topics A, B, and C, then traditional recommender system tries to suggest articles that simultaneously encompass all three topics, and an article which only contains a single topic failed to be recommended.

Such recommender systems utilize ontologies for modeling scholars' interests (He and Fang, 2008) using a prebuilt reference ontology. Reference ontology acts as an initial model of scholar's preferences (Liao et al., 2010; Mohammed et al., 2010). However, such reference ontologies are disqualified, as they lack sufficient ontological concepts in representing scholars' knowledge (low coverage), the volume of knowledge covered in ontologies are limited, the knowledge is not adequately domain specific (Tao et al., 2008), and worse, unable in modeling the multi-area property of researchers. They rather encompass general-purpose topics of scholars' domain and are incomplete in representing the concepts of scholars' knowledge (Pierrakos and Paliouras, 2010; Lops et al., 2011). Hence, it is crucial to profile scholars' knowledge with a qualified ontology, which provides an extensive and domain-specific collection of topics.

As a result, it is necessary to investigate ontology-based profiling methods, similarity computing algorithms, and related methods, to adapt them into the scholar's domain in order to address the aforementioned problems. In other words, the main problem is how to deploy recommender systems into the research scenario in which tackles three dimensions of scholar domain: 1) Collects data as much as possible from all areas of scholars' background knowledge and vectorization them to represent multi-area property of scholar knowledge, 2) Construct a special-purpose ontology which represents scholars' knowledge in a hierarchical structure and assists in filtering of articles which contain new knowledge, and 3) Profiles scholars' knowledge in which filters out articles as relevant as possible to the scholars' background knowledge (Riedl 2009), and recommends focused-topics articles.

1.4 Research Question

Achieving higher precision is the ultimate goal of most recommender systems (Gunawardana and Shani, 2009). Considering the characteristics of scholar's domain and the problems described in Section 1.3, this study enhances profiling of scholars' knowledge and enables filtering of focused-topic articles which contain complementary knowledge. Following this goal, the main research question can be formulated as follows:

> ***How to enhance scholars' background knowledge in an ontology-based profiling approach that enables filtering of focused-topic articles which contain complementary knowledge?***

To narrow down the main question, the following sub-questions are set:

1. How can scholar's knowledge be sufficiently extracted from the scholars' context, and be represented in multiple area of knowledge?

2. How can a qualified reference ontology be constructed from domain ontologies which represents the real structure of scholars' knowledge and supports the profiling requirements?

3. How can scholars' knowledge be profiled to enable filtering of articles which supply complementary knowledge?

4. How can similarity computing algorithms be adapted to the profiling method to support filtering of focused-topic articles?

Question 1 deals with context modeling techniques including feature extraction, feature selection, and feature transformation in scholar's domain, while question 2 investigates an approach for constructing a reference ontology which improves the domain coverage for scholars' knowledge. Question 3 deals with profiling approach using the reference ontology to enable filtering of articles, containing complementary knowledge. Question 4 investigates the adaptation of similarity algorithms with profiling approach in order to enable the filtering of focused-topic articles.

1.5 Research Objectives

The main goal of this study is to improve the recommendation precision by profiling scholars' knowledge using a new structure of reference ontology, and filtering focused-topic articles which offer complementary knowledge to scholars. To achieve this goal, the following objectives have been carried out:

1- To enhance capturing scholars' knowledge for profiling by incorporating new knowledge resources and representing scholars' knowledge in multi-areas.

2- To develop a qualified and special-purpose reference ontology based on the structure of scholars' knowledge in order to improve the profiling method.

3- To propose a profiling method using the new reference ontology in order to improve recommendation performance by filtering focused-topic articles which contain complementary knowledge.

1.6 Importance of the Study

The explosion growth of Internet, online digital libraries, and social networks are generating incredibly large amounts of useful data, and the tremendous growth in computational power is increasing the desire of users to personalize information access. Thus, the overwhelming amount of data necessitates appropriate mechanisms for efficient information filtering and recommendation (Uchyigit, 2009; Sugiyama and Kan, 2010). The potential value of personalization has become clear both as a means for benefit of end-users, and as an enabler of better information services. An exciting characteristic of recommender systems for scholars is that they draw attention of the community while posing very interesting research challenges as well (Riedl, 2009).

In spite of significant progress in the field, and the community efforts to bringing the benefits of new techniques to end-users, there are still important gaps that make personalization and adaptation inadequate to the scholars (Sharma and Gera, 2013). Research activities still often focus on narrow problems such as incremental accuracy improvements of current techniques, or tend to overspecialize on a few problems (typically collaborative recommenders, sometimes often because of the availability of data sets) (Middleton et al., 2009; Yang and Hsu, 2010; Zuva et al., 2012). Thus, there is a good point to take one step forward to seek a new perspective which has been rarely investigated in recommender systems - a new approach for profiling where cognitive information of scholars, i.e., the scholars' background knowledge, with a special focus on Semantic Web technologies in a working application is investigated.

Furthermore, many works have recently been published in the field of recommender system for scholars (Adomavicius and Tuzhilin, 2005; Hong et al., 2009), but more research should be conducted to advance the state-of-the-art in scholar domain (Yujie and Licai, 2010). This is because existing algorithms and techniques in other domains such as e-commerce (He and Fang, 2008), online news (IJntema et al., 2010), and movies (Christakou and Stafylopatis, 2005) cannot be directly applied to scholar domain. As explained in Section 1.2 this domain is associated with particular characteristics that makes it distinctive (Devedžić, 2006; Drachsler et al., 2007).

1.7 Research Scope

To measure the impact of applying scholar's background knowledge into the recommendation process, the following features are considered in advance:

1. Scholar refers to a researcher in any sub fields of Computer Science, including master student, graduate or Ph.D. student, or full researcher.

2. The source of scholar knowledge is bounded to text-based academic resources.

3. The core technology for modeling scholars' knowledge is Semantic Web, particularly ontology-based approaches.

4. This study mainly focuses on unobtrusive knowledge extraction from context, i.e., scholars do not fully interfere with knowledge extraction.

1.8 Research Contributions

Overall, this research introduces a new approach of recommender system for scholars' domain, and makes the following particular contributions to the field:

- To extend scholars' context, new resources including scholars' formal education and mediated profiles on the Web are examined in the profiling process. Such resources have not yet been used for profiling.

- In order to extract feature vectors, the C-Value/NC-Value which have not been used in recommender systems are employed that properly recognize the semantic meaning of associated words and increase the discriminative power of feature vectors.

- To improve the feature selection process, term transformation using DBpedia has been introduced.

- To support the multi-area property of scholars' knowledge, knowledge items (feature vectors) are clustered into semantically related terms using a semantic similarity method.

- To improve the coverage and richness of reference ontology, a new ontology based on the hierarchical structure of scholars' knowledge has been developed.

- To measure the similarity between the scholars' profiles and candidate articles (in particular, to filtering focused-topic articles), unlike traditional approaches, a semantic-based similarity has been employed.

- A new semantic-based profiling method has been employed in profiling, which improves the overall recommendation precision and recommends articles which contain complementary knowledge.

1.9 Structure of the Thesis

In Chapter 2, the state-of-the-art of recommender systems and ontology-based profiling approaches are discussed. Besides, issues with existing user modeling, particularly ontology-based profiling, and the research gap are highlighted. Also, feature extraction, feature reduction, and clustering approaches in which used in recommender systems are analyzed. In Chapter 3, the overall methodology for achieving research objectives has been described. The focus of methodology is to incorporating scholars' background knowledge into the profiling process. In particular, the theoretical framework, research map, and system framework for capturing scholars' knowledge, developing reference ontology, and profiling scholars' knowledge using semantic-based methods are discussed.

In Chapter 4, a method for capturing scholars' background knowledge from scholars' context is introduced. The method extends the contextual information by incorporating "formal education" and "mediated profiles" which had not been yet employed for profiling. In addition, feature transformation has been introduced which improves the quality of extracted key terms. Term clustering is also applied which addresses multi-area property of scholars' knowledge. Chapter 5 discusses a method for construction reference ontology using several Web taxonomies. Our method investigates existing Web taxonomies which have been exploited either for profiling scholars' interests or provide sufficient ontological concepts in the domain.

Chapter 6 also discusses the implementation of third objectives and describes the experiments and achieving precision using scholars' background knowledge. It compares the proposed method with benchmark recommender systems, which have been discussed in Chapter 2. Chapter 7 finalizes the study by making conclusion about the results, highlights the contribution in detail, discusses new challenges, and suggests future works and extensions.

1.10 Summary

This chapter outlines the essential parts of the study for incorporating scholars' background knowledge into an ontology-based recommender system. The background of research problem and research goal as well as problem statement along with the research questions and objectives are described. Scope of the research and a brief of achieving contribution of the study are also mentioned. This chapter serves as an introductory part of the thesis.

REFERENCES

Aamodt, A. and Plaza, E., (1994). Case-Based Reasoning : Foundational Issues , Methodological Variations , and System Approaches. In *AI Communications*. IOS Press, pp. 39–59.

Abbar, S., Bouzeghoub, M. and Lopez, S., (2007). Context-Aware Recommender Systems: A Service-Oriented Approach. In *VLDB' 2009*. Lyon, France: ACM, pp. 1–6.

Adomavicius, G., Manouselis, N. and Kwon, Y., (2011). Multi-Criteria Recommender Systems. In *Recommender Systems Handbook*. Springer, pp. 769–803.

Adomavicius, G. and Tuzhilin, A., (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), pp.734–749.

Agirre, E. and Edmonds, P., (2006). *Word Sense Disambiguation Text , Speech and Language Technology*, Springer.

Aïmeur, E. et al., (2006). Privacy-preserving demographic filtering. In *ACM symposium on Applied computing - SAC '06*. Dijon, France: ACM Press, pp. 872–878.

Akthar, F. and Hahne, C., (2013). *RapidMiner Studio 6 User Manual*, RapidMiner GmbH.

Alelyani, S., Tang, J. and Liu, H., (2013). *Feature Selection for Clustering : A Review* C. Aggarwal & C. Reddy, eds., CRC Press.

Altman, D.G. et al., (2000). *Statistics with confidence: Confidence with intervals and statistical guidelines* 2nd ed., Boston: Blackwell BMJ.

Amatriain, X. et al., (2011). Data Mining Methods for Recommender Systems. In *Recommender Systems Handbook*. pp. 1–549.

Anand, S.S. and Mobasher, B., (2005). Intelligent Techniques for Web Personalization. In *Intelligent Techniques for Web Personalization Lecture Notes in Computer Science Volume 3169*. Springer, pp. 1–36.

Anyanwu, K. and Sheth, A., (2003). ρ-Queries: Enabling Querying for Semantic Associations on the Semantic Web. In *Proceedings of the Twelfth International World-Wide Web Conference*.

Apidianaki, M., Neumann, R.J. Von and Cedex, O., (2013). LIMSI : Cross-lingual Word Sense Disambiguation using Translation Sense Clustering. In *Second Joint Conference on Lexical and Computational Semantics (SEM)*. pp. 178–182.

Aquin, M. and Schlicht, A., (2009). Criteria and Evaluation for Ontology Modularization Techniques. In *Modular Ontologies*. Springer, pp. 67–89.

De Araujo, F.F., Lopes, F.L.R. and Loscio, B.F., (2010). MeMO: A Clustering-based Approach for Merging Multiple Ontologies. In *2010 Workshops on Database and Expert Systems Applications*. IEEE, pp. 176–180.

Belei, D.M., (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), pp.77–84.

Berkovsky, S., Kuflik, T. and Ricci, F., (2007). Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction*, (18), pp.245–286.

Biemans, H.A. and Simons, P.R.J., (1996). Contact-2 : a computer-assisted instructional strategy for promoting conceptual change. *Instructional Science*, 24, pp.157–176.

Bird, S. et al., (2008). The ACL Anthology Reference Corpus : A Reference Dataset for Bibliographic Research in Computational Linguistics. In *Sixth International Conference on Language Resources and Evaluation*. European Language Resources Association, Paris.

Bitonto, P. Di, Laterza, M. and Roselli, T., (2010). A recommendation method for e-learning environments : the rule-based technique. *Journal of e-Learning and Knowledge Society*, 6(3), pp.31–40.

Bizer, C. et al., (2009). DBpedia - A Crystallization Point for the Web of Data. *Elsevier, Web Semantics*, 7(3), pp.154–165.

Bogers, T. and Bosch, A., (2008). Recommending Scientific Articles Using CiteULike. In *RecSys'08*. ACM, pp. 8–11.

Borges, H.L. and Lorena, A.C., (2010). A Survey on Recommender Systems for News Data. *Knowledge Creation Diffusion Utilization*, pp.129–151.

Bransford, J., Brown, A. and Cocking, R., (2004). *How People Learn- Brain,Mind, Experience, and School*, Washington DC: National Academy Express.

Brewster, C. et al., (2004). Data Driven Ontology Evaluation. In *In Int. Conf. on Language Resources and Evaluation*.

Bridge, D., Goker, M., et al., (2006). Case-based recommender systems. *The Knowledge Engineering Review*, 20(3), pp.315–320.

Bridge, D., Goker, M.H. and McGinty, L., (2006). Case-based recommender systems. *The Knowledge Engineering Review*, 20(3), pp.315–320.

Budanitsky, A. and Hirst, G., (2001). Semantic distance in WordNet: An experimental , application-oriented evaluation of five measures. In *Workshop on WordNet and Other Lexical Resources at the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics*. pp. 29–34.

Burke, R., (2002). Hybrid Recommender Systems : Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4), pp.331–370.

Burke, R., (2007). Hybrid Web Recommender Systems. In *The Adaptive Web, LNCS 4321*. pp. 377–408.

Burke, R., (2000). Knowledge-based recommender systems. *Encyclopedia of library and information systems*, 69, pp.175–186.

Butts, C.T., (2008). Social network analysis : A methodological introduction. *Asian Journal of Social Psychology*, 11, pp.13–41.

Cantador, I., Bellogín, A. and Castells, P., (2008). A Multilayer Ontology-based Hybrid Recommendation Model. *AI Communications - Recommender Systems*.

Cassel, L., Clements, A. and Davies, G., 2008, *Computer Science Curriculum 2008: An Interim Revision of CS 2001*,

Castellano, G. et al., (2009). Innovations in Web Personalization. In *Web Personalization in Intelligent Environments Studies in Computational Intelligence, SCI 229*. pp. 1–26.

Chakrabarti, S., Van Den Berg, M. and Dom, B., (1999). Focused crawling: A new approach to topic-specific Web resource discovery. *Computer Networks*, 31(11), pp.1623–1640.

Chen, J. et al., (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, 36(3), pp.5432–5435.

Chen, R., Bau, C. and Yeh, C., (2011). Merging domain ontologies based on the WordNet system and Fuzzy Formal Concept Analysis techniques., 11, pp. 1908–1923.

Christakou, C. and Stafylopatis, A., (2005). A hybrid movie recommender system based on neural networks. In *5th International Conference on Intelligent Systems Design and Applications 2005, ISDA '05*. pp. 500–505.

Chua, S. and Kulathuramaiyer, N., (2004). Semantic Feature Selection Using WordNet. In *M International Conference on Web Intelligence (WI'04)*. IEEE Compter Society, pp. 166–172.

Cover, T.M. and Hart, P.E., (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), pp.21–27.

Craven, M., Dipasquo, D. and Freitag, D., (1998). Learning to Extract Symbolic Knowledge from the World Wide Web. In *National Conference on Artificial Intelligence - AAAI*. pp. 509–516.

Dai, H. and Mobasher, B., (2009). Integrating Semantic Knowledge with Web Usage Mining for Personalization. In *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications*.

Damljanovic, D., Stankovic, M. and Laublet, P., (2012). Linked data-based concept recommendation: Comparison of different methods in open innovation scenario. In *ESWC 2012, LNCS 7295*. pp. 24–38.

Das, S. et al., (2011). On Identifying Academic Homepages for Digital Libraries. In *JCDL'11 Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*.

Davis, R., Shrobe, H. and Szolovits, P., (1993). What Is a Knowledge Representation? *AI Magazine*, 14(1), pp.17–33.

Davis, S.J. and Winek, J., (1989). Improing expository writing by increasing background knowledge. *International Reading Association*, 33(3), pp.178–181.

Deerwester, S. et al., (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41(6), pp.391–407.

Degemmis, M. and Lops, P., (2007). *A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation*, Springer.

Degemmis, M., Lops, P. and Semeraro, G., (2006). Learning Semantic User Profiles from Text. In *ADMA'06, LNAI 4093*. Springer Berlin / Heidelberg, pp. 661–672.

Deng, Z. et al., (2004). A Comparative Study on Feature Weight in Text. In *APWeb 2004*. Springer Berlin / Heidelberg, pp. 588–597.

Devedžić, V., (2007). Personalization Issues. In *Integrated Series in Information Systems*.

Dey, A.K. and Abowd, G.D., (1999). Towards a Better Understanding of Context and. In *HUC '99 Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*. pp. 304–307.

Dolog, P. and Nejdl, W., (2007). Semantic Web Technologies for the AdaptiveWeb. In *The AdaptiveWeb, LNCS 4321*. Springer Berlin Heidelberg, pp. 697–719.

Dong, X., Halevy, A. and Madhavan, J., (2005). Reference Reconciliation in Complex Information Spaces. In *SIGACM-SIGMOD'05 international conference on Management of data*. Maryland, USA: ACM, pp. 85–96.

Drachsler, H., Hummel, H.G.K. and Koper, R., (2007). Recommendations for learners are different: Applying memory-based recommender system techniques to lifelong learning. In *Proceeding of the 1st Workshop on Social Information Retrieval for Technology-Enhanced Learning & Exchange*. pp. 18–26.

Duong, T.H. et al., (2009). A Collaborative Ontology-Based User Profiles System. In *ICCCI 2009, LNAI 5796*. Springer Berlin / Heidelberg, pp. 540–552.

Eirinaki, M., Mavroeidis, D. and Tsatsaronis, G., (2006). Introducing Semantics in Web Personalization : The Role of Ontologies. , pp.147–162.

Eppler, M.J., (2006). A Framework for Information Quality Management. In *Managing Information Quality*. Springer Berlin Heidelberg, pp. 57–210.

Eric, B., (1992). A Simple Rule-Based Part of Speech Tagger. In *Applied Natural Language Processing*. pp. 152–155.

Eyharabide, V. and Amandi, A., (2011). Ontology-based user profile learning. *Applied Intelligence*, 36(4), p.2012.

Fareh, M., Boussaid, O. and Chalal, R., (2013). Merging ontology by semantic enrichment and combining similarity measures. *Int. J. Metadata, Semantics and Ontologies,*, 8(1), pp.65–74.

Feldman, R. and Sanger, J., (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*,

Felfernig, A. and Burke, R., (2008). Constraint-based Recommender Systems : Technologies and Research Issues. In *Int. conf. on Electronic Commerce (ICEC)'08*. Innsburuck, Austria: ACM, pp. 19–22.

Forman, G., (2003). An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *The Journal of Machine Learning Research*, 3, pp.1289–1305.

Frantzi, K., Ananiadou, S. and Mima, H., (2000). Automatic Recognition of Multi-Word Terms : the C-value/NC-value Method. *International Journal on Digital Libraries, Springer*, 3(2), pp.115–130.

Freitas, A., Curry, E. and Oliveira, J.G., (2012). Querying Heterogeneous Datasets on the Linked Data Web. *Internet Computing, IEEE*, pp.24 – 33.

Gagné, R.M., (1988). *Conditions of Learning*, Holt, Rinehart and Winston.

Gagné, R.M., (1968). Learning Hierarchies. In D. Ely P. & T. Plomp, eds. *Classic Writings on Instructional Technology*. Libraries Unlimited, pp. 63–84.

Gal, A. and Shvaiko, P., (2009). Advances in Ontology Matching. In *Advances in Web Semantics I, LNCS 4891*. Springer Berlin Heidelberg, pp. 176–198.

Gan, G., Ma, C. and Wu, J., (2007). *Data Clustering Theory, Algorithm, and Application*, American Statistical Association.

Gangemi, A. et al., (2005). A theoretical framework for ontology evaluation and validation. In *the 2nd Italian Semantic Web Workshop, CEUR Workshop Proceedings*. Trento, Italy.

Gangemi, A., Steve, G. and Giacomelli, F., (1999). ONIONS : An Ontological Methodology for Taxonomic Knowledge Integration. In *ECAI-96 Workshop on Ontological Engineering*. pp. 183–220.

García, P. et al., (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education*, 49(3), pp.794–808.

Gauch, S., Speretta, M., Chandramouli, A., et al., (2007). User Profiles for Personalized Information Access. In *The Adaptive Web, LNCS 4321*. pp. 54–89.

Gauch, S., Speretta, M. and Pretschner, A., (2007). Ontology-Based User Profiles for Personalized Search. In *Ontologies, Integrated Series in Information Systems*. Springer US, pp. 665–694.

Gaudioso, E., (2008). Evaluation of recommender systems : A new approach. *Expert Systems with Applications*, 35, pp.790–804.

Gemmis, M. De et al., (2009). Preference Learning in Recommender Systems. In *ECML/PKDD Workshop on Preference Learning*.

Godoy, D. and Amandi, A., (2007). An Agent-Based Recommender System to Support Collaborative Web Search Based on Shared User. In *CRIWG '07, LNCS 4715*. pp. 303–318.

Godoy, D. and Schiaffino, S., (2004). Interface agents personalizing Web-based tasks. *Cognitive Systems Research*, 5, pp.207–222.

Gómez-pérez, A. and Manzano-macho, D., (2003). A survey of ontology learning methods and techniques OntoWeb Consortium. In *OntoWeb: Ontology-based Information Exchange for Knowledge Management and Electronic Commerce*. pp. 10–86.

Grigonyte, G., (2010). *Building and Evaluating Domain Ontologies. NLP Contributions*, Logos Verlag Berlin GmbH.

Groh, G., Birnkammerer, S. and Köllhofer, V., (2012). Social Recommender Systems Georg. In *Recommender Systems for the Social Web*. pp. 3–42.

Guarino, N., (1998). Formal Ontology and Information Systems. In *Proceedings of FOIS'98,*. Trento, Italy: IOS Press, pp. 3–15.

Guarino, N. and Giaretta, P., (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarificatio. *Knowledge Acquisition*, pp.25–32.

Gunawardana, A. and Shani, G., (2009). A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. *Journal of Machine Learning Research*, 10, pp.2935–2962.

Haghighi, P.D., Krishnaswamy, S. and Zaslavsky, A., (2008). Reasoning about Context in Uncertain Pervasive. In *EuroSSC 2008, LNCS 5279*. Springer Berlin / Heidelberg, pp. 112–125.

Haldar, R. and Mukhopadhyay, D., (2011). Levenshtein Distance Technique in Dictionary Lookup Methods : An Improved Approach. In *Web Intelligence & Distributed Computing Research Lab*.

Hanushek, E.A., Jackson, J.E. and Rossi, P.H., (1977). *Statistical Methods for Social Scientists*, Elsevier Ltd.

Haque, R., Penkale, S. and Way, A., (2013). Bilingual Termbank Creation via Log-Likelihood Comparison and Phrase-Based Statistical Machine Translation. In *4th International Workshop on Computational Terminology (Computerm)*. Dublin, Ireland.

Hausenblas, M., (2011). On Entities in theWeb of Data. In *REST: From Research to Practice*. Springer Science+Business, pp. 425–440.

He, S. and Fang, M., (2008). Ontological User Profiling on Personalized Recommendation in e-Commerce. In *IEEE International Conference on e-Business Engineering, ICEBE '08.* IEEE Compter Society, pp. 585–589.

Henderson, L., (2009). *Automated Text Classification in the DMOZ Hierarchy*, Available at: http://cs.anu.edu.au/students/projects/09S2/Reports/Lachlan Henderson.pdf.

Herlocker, J.L. et al., (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, 22(1), pp.5–53.

Hong, J., Suh, E. and Kim, S., (2009). Context-aware systems: A literature review and classification. *Expert Systems With Applications*, 36(4), pp.8509–8522.

Hotho, A., Staab, S. and Stumme, G., (2003). Wordnet improves Text Document Clustering. In *Proc. of the SIGIR 2003 Semantic Web Workshop*.

Huang, A., (2008). Similarity Measures for Text Document Clustering. In *NZCSRSC 2008*. pp. 49–56.

Huang, J., Ertekin, S. and Giles, C.L., (2006). Efficient Name Disambiguation for Large-Scale Databases. In *PKDD 2006, LNAI 4213*. Springer-Verlag Berlin Heidelberg, pp. 536–544.

Huang, L. et al., (2012). Learning a Concept-based Document Similarity Measure. *Journal of the American Society for Information Science and Technology*, 63(8), pp.1593–1608.

Hunt P., D., (2003). The concept of knowledge and how to measure it. *Journal of Intellectual Capital*, 4(1), pp.100–113.

IJntema, W. et al., (2010). Ontology-based news recommendation. In *1st International Workshop on Data Semantics - EDBT '10*. Lausanne, Switzerland: ACM, pp. 11–20.

Isaksson, A. et al., (2008). Cross-validation and bootstrapping are unreliable in small sample classification. *Pattern Recognition Letters*, 29, pp.1960–1965.

Jack, K. et al., (2012). Mendeley ' s open data for science and learning : a reply to the DataTEL challenge. *Int. J. Technology Enhanced Learning*, 4(1/2), pp.31–46.

Jonassen, D., Mayes, T. and McAleese, R., (1993). A Manifesto for a Constructivist Approach to Uses of Technology in Higher Education. In *Designing Environments for Constructive Learning*. Springer New York, pp. 231–247.

Kadima, H. and Malek, M., (2010). Toward ontology-based personalization of a recommender system in social network. In *Proceedings of the International Conference of Soft Computing and Pattern Recognition, SoCPaR 2010*. pp. 119–122.

Kalyanpur, A. et al., (2006). Swoop: A Web Ontology Editing Browser. *Web Semantics: Science, Services and Agents on the Web*, 4(2), pp.144–153.

Keet, C.M. (Marijke), (2004). *Aspects of Ontology Integration*. School of Computing, Napier University.

Khan, A., Baharudin, B. and Khan, K., (2010). Semantic Based Features Selection and Weighting Method for Text Classification. *IEEE*, 2, pp.850–855.

Kim, S.N. et al., (2010). SemEval-2010 Task 5 : Automatic Keyphrase Extraction from Scientific Articles. *Computational Linguistics*, (July), pp.21–26.

Kodakateri, A. et al., (2009). Conceptual Recommender System for CiteSeerX. In *RecSys'09*. NY, USA: ACM, pp. 241–244.

Kotis, K., Vouros, G. and Stergiou, K., (2006). Towards automatic merging of domain ontologies: The HCONE-merge approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1), pp.60–79.

Kozakov, L. et al., (2004). Glossary extraction and utilization in the information search and delivery system for IBM Technical Support. *IBM SYSTEMS JOURNAL*, 43(3), pp.546–563.

Kujawa, S. and Huske, L., (1995). *Strategic teaching and reading project guidebook*, Oak Brook, IL: NCREL North Central Regional Educational Laboratory.

Kumar, A., (2010). Collaborative Web Recommendation Systems-A Survey Approach. *Global Journal of Computer Science and Technology*, 9(5), pp.30–35.

Lee, L., (2000). Measures of Distributional Similarity. In *ACL'99 Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. PA, USA: ACM, pp. 25–32.

Lesk, M., (1987). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In *SIGDOC'86 Proceedings of the 5th annual international conference on Systems documentation*. pp. 24–26.

Li, S., Wu, G. and Hu, X., (2011). Hierarchical User Interest Modeling for Chinese Web Pages. In *3rd International Conference on Internet Multimedia Computing and Service (ICIMCS'11)*. pp. 164–169.

Liang, T. et al., (2008). A semantic-expansion approach to personalized knowledge recommendation. *Elevier, Decision Support Systems*, 45(3), pp.401–412.

Liao, I.-E. et al., (2010). A library recommender system based on a personal ontology model and collaborative filtering technique for English collections. *Emeral Group Publishing*, 28(3), pp.386–400.

Liao, S. et al., (2009). PORE: a personal ontology recommender system for digital libraries. *Emeral Group Publishing*, 27(3), pp.496–508.

Liu, T. et al., (2003). An Evaluation on Feature Selection for Text Clustering. In *Proceedings of the 20th Int. Conf. on Machine Learning (ICML-2003)*. Washington DC, pp. 488–495.

Lopes, G.R., Martins Souto, M.A., et al., (2007). A Personalized Recommender System for Digital Libraries. In *WebMedia'08 Proceedings of the 14th Brazilian Symposium on Multimedia and the Web*. pp. 59–66.

Lopes, G.R., Martins, M.A. and Wives, L.K., (2007). A Personalized Recommender System for Digital Libraries. In *Proceedings of the 14th Brazilian Symposium on Multimedia and the Web*. NY, USA: ACM Press, pp. 59–66.

Lopes, L. and Vieira, R., (2012). Improving Portuguese Term Extraction. In *PROPOR 2012, LNAI 7243*. Springer-Verlag Berlin Heidelberg, pp. 85–92.

Lops, P., Gemmis, M. De and Semeraro, G., (2011). Content-based Recommender Systems : State of the Art and Trends. In *Recommender Systems Handbook*. Springer Science+Business, pp. 73–105.

Lorenzi, F. and Ricci, F., (2005). Case-Based Recommender Systems : a Unifying View. In *Intelligent Techniques for Web Personalization-Lecture Notes in Computer Science*. Springer-Verlag Berlin Heidelberg, pp. 89–113.

Manning, C., Raghavan, P. and Schütze, H., (2009). *An Introduction to Information Retrieval*, Cambridge University Press.

Manouselis, N. et al., (2010). TEL as a Recommendation Context. In N. M. et Al., ed. *Recommender Systems for Learning*. New York, NY: Springer New York, pp. 21–37.

Manouselis, N., Kosmopoulos, T. and Kastrantas, K., (2009). Developing a Recommendation Web Service for a Federation of Learning Repositories. In *International Conference on Intelligent Networking and Collaborative Systems, INCOS '09*. pp. 208–211.

Marcus, M. et al., (2013). The Penn TREEBANK : Annotating predicate argument structure. *ARPA Human Language Technology Workshop (HLT'94)*, pp.114–119.

Mascardi, V., Locoro, A. and Rosso, P., (2010). Automatic ontology matching via upper ontologies: A systematic evaluation. *IEEE Transactions on Knowledge and Data Engineering*, 22(5), pp.609–623.

McDonald, S. and Stevenson, R.J., (1998). Effects of Text Structure and Prior Knowledge of the Learner on Navigation in Hypertext. *Human Factors*, pp.18–27.

Medelyan, O. et al., (2009). Mining meaning from Wikipedia. *Journal of Human Computer Studies*, 67(9), pp.716–754.

Medelyan, O., Witten, I.H. and Milne, D., (2008). Topic Indexing with Wikipedia. In *Proceedings of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy*. pp. 19–24.

Middleton, S.E., Roure, D. De and Shadbolt, N.R., (2009). Ontology-Based Recommender Systems. In *Handbook on Ontologies, International Handbooks on Information Systems*. pp. 779–796.

Middleton, S.E., Shadbolt, N.R. and Roure, D.C. DE, (2004). Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), pp.54–88.

Mihalcea, R., (2007). Using Wikipedia for Automatic Word Sense Disambiguation. In *Proceedings of the Conference of NAACL'07*. pp. 196–203.

Miraoui, M., Tadj, C. and Soukra, R. De, (2008). Architectural Survey of Context-Aware Systems in Pervasive Computing Environment. *Ubiquitous Computing and Communication Journal (UBICC)*, 3(3), pp.68–76.

Mobasher, B., (2007a). Data Mining for Web Personalization. In *The Adaptive Web, LNCS 4321*. Springer Berlin / Heidelberg, pp. 90–135.

Mobasher, B., (2007b). Recommender Systems. *Kuenstliche Intelligenz*, 21(3), pp.41–43.

Mohammed, N., Duong, T.H. and Jo, G.S., (2010). Contextual Information Search Based on Ontological User Profile. In *ICCCI 2010, LNAI 6422*. pp. 490–500.

Mohsenzadeh, M., Shams, F. and Teshnehlab, M., (2005). A New Approach for Merging Ontologies. *World Academy of Science, Engineering and Technology*, pp.153–159.

Navigli, R., (2009). Word sense disambiguation- A Survey. *ACM Computing Surveys*, 41(2), pp.1–69.

Nguyen, N.T., (2006). A Consensus-based Approach for Ontology Integration. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW06)*. pp. 514–517.

Noy, N.F. and Musen, M.A., (1999). An Algorithm for Merging and Aligning Ontologies: Automation and Tool Support. In *Workshop on Ontology Management at the 16th National Conference on Artificial Intelligence (AAAI-99)*. AAI Press, pp. 17–27.

Noy, N.F., Musen, M.A. and Informatics, S.M., (2000). PROMPT : Algorithm and Tool for Automated Ontology Merging and Alignment. In *17th National Conference on Artificial Intelligence (AAAI-2000)*. Austin, Texas.

Park, Y., (2002). Identification of Probable Real Words: An Entropy-based Approach. In *ULA '02 Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition*. Stroudsburg, PA, USA: ACM, pp. 1–8.

Pazzani, M.J., (1999). A Framework for Collaborative , Content-Based and Demographic Filtering. *Artificial Intelligence Review - AIR*, 13(5-6), pp.393–408.

Pazzani, M.J. and Billsus, D., (2007). Content-Based Recommendation Systems. In *The Adaptive Web, LNCS 4321*. Springer Berlin Heidelberg, pp. 325–341.

Pedersen, T. et al., (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40, pp.288–299.

Pedersen, T., (2007). Unsupervised Corpus-Based Methods for WSD. In *Word Sense Disambiguation: Algorithms and Applications*. pp. 133–166.

Piao, S. et al., (2010). Evaluating Tools for Automatic Concept Extraction: a Case Study from the Musicology Domain. In *Proceedings of The Digital Economy All Hands Meeting - Digital Futures 2010 Conference*.

Pierrakos, D. and Paliouras, G., (2010). Personalizing Web Directories with the Aid of Web Usage Data. *IEEE Transactions on Knowledge and Data Engineering*, 22(9), pp.1331–1344.

Popescu, A.-M. and Etzioni, O., (2005). Extracting Product Features and Opinions from Reviews. In *HLT '05 Proceedings of Human Language Technology and Empirical Methods in Natural Language Processing*. pp. 339–346.

Pu, P., Chen, L. and Hu, R., (2012). Evaluating recommender systems from the user's perspective : survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22, pp.317–355.

Raunich, S. and Rahm, E., (2011). ATOM: Automatic Target-driven Ontology Merging. In *27th International Conference on Data Engineering (ICDE)*. IEEE, pp. 1276 – 1279.

Resnik, P., Laboratories, S.M. and Drive, T.E., (1995). Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In *Proc. of IJCAI-95*. pp. 448–453.

Ricci, F., (2010). Mobile Recommender Systems. *Information Technology & Tourism*, 12(3), pp.205–231.

Ricci, F. et al., (2011). *Recommender Systems Handbook*, Springer Science+Business.

Riedl, J., (2009). Research Challenges in Recommender Systems. *ACM RECsys'9*, pp.1–4.

Rocha, L.M., (2002). TalkMine: a soft computing approach to adaptive knowledge recommendation. In *Soft computing agents*. Germany, Germany: Physica-Verlag GmbH Heidelberg, pp. 89–116.

Rubin, D.L., Noy, N.F. and Musen, M.A., (2007). Protégé: A Tool for Managing and Using Terminology in Radiology Applications. *Journal of Digital Imaging*, 20(1), pp.34–46.

Saerens, M., (2008). Evaluating performance of recommender systems : An experimental comparison. , pp.8–11.

Saini, P.S., Sona, D. and Ronchetti, M., (2004). *A Baseline Approach for the Automatic Hierarchical Organization of Learning Resources*,

Salton, G. and Buckley, C., (1988). On the Use of Spreading Activation Methods in Automatic Information Retrieval Information Retrieval. In *SIGIR '88 Proc. of the 11th annual international ACM SIGIR conference*. pp. 147–160.

Salton, G., Wong, A. and Yang, C.S., (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11), pp.613–620.

Sánchez, D. et al., (2012). Ontology-based semantic similarity : A new feature-based approach. *Expert Systems with Applications*, 39, pp.7718–7728.

Satyanarayana, K. and Rajagopalan, S. p., (2007). Recommender System for Educational Institues. *Asian Journal of Information Technology*, 6(9), pp.964–969.

Schafer, J. Ben et al., (2007). Collaborative Filtering Recommender Systems. In *The Adaptive Web, LNCS 4321*. Springer Berlin Heidelberg, pp. 291–324.

Schiaffino, S. and Amandi, A., (2009). Intelligent User Profiling. In *Artificial Intelligence, LNAI 5640*. IFIP International Federation for Information Processing, pp. 193–216.

Schiaffino, S., Amandi, A. and Nacional, U., (2006). Polite Personal Agents. *Intelligent Systems, IEEE*, 21(1), pp.12–19.

Semeraro, G. et al., (2001). Combining Learning and Word Sense Disambiguation for Intelligent User Profiling. *Knowledge Creation Diffusion Utilization*, pp.2856–2861.

Sharma, L. and Gera, A., (2013). A Survey of Recommendation System : Research Challenges. *International Journal of Engineering Trends and Technology*, 4(5), pp.1989–1992.

Sharman, R., Kishore, R. and Ramesh, R., (2007). *ONTOLOGIES: A Handbook of Principles, Concepts and Applications in Information Systems*,

Shvaiko, P., (2013). Ontology Matching : state of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, 25(1), pp.158–176.

Sieg, A. and Burke, R., (2007). Web Search Personalization with Ontological User Profiles. In *CIKM'07*. Lisboa, Portugal: ACM, pp. 525–534.

Sieg, A., Mobasher, B. and Burke, R., (2010). Improving the Effectiveness of Collaborative Recommendation with Ontology-Based User Profiles. In *HetRec'10*. Barcelona, Spain: ACM.

Sieg, A., Mobasher, B. and Burke, R., (2007a). Learning Ontology-Based User Profiles : A Semantic Approach to Personalized Web Search. *IEEE Intelligent Informatics Bulletin*, 8(1), pp.7–18.

Sieg, A., Mobasher, B. and Burke, R., (2007b). Ontological User Profiles for Personalized Web Search. *AAAI Worksop, Intelligent Techniques for Web Personalization (ITWP07)*, pp.84–91.

Smeaton, A.F. and Callan, J., (2005). Personalisation and Recommender Systems in Digital Libraries. *International Journal on Digital Libraries*, 5(4), pp.299–308.

Smith, B., (2004). Beyond Concepts: Ontology as Reality Representation. In *Proceedings of FOIS 2004, International Conference on Formal Ontology and Information Systems*. Turin, pp. 71–83.

Smyth, B., (2007). Case-Based Recommendation. In *The AdaptiveWeb, LNCS 4321*. Springer-Verlag Berlin Heidelberg, pp. 342–376.

Stamper, J. et al., (2010). PSLC DataShop: A Data Analysis Service for the Learning Science Community. In *Intelligent Tutoring Systems (ITS 2010) LNCS 6095*. pp. 455–455.

Steinbach, M., (2000). A Comparison of Document Clustering Techniques. In *KDD workshop on text mining (2000)*. pp. 1–20.

Strangman, N. and Hall, T., (2009). *Background Knowledge*, Wakefield, MA.

Strobbe, M. et al., (2011). Hybrid reasoning technique for improving context-aware applications. *Knowledge and Information Systems*, 31(3), pp.581–616.

Strube, M. and Ponzetto, S.P., (2006). WikiRelate! Computing Semantic Relatedness Using Wikipedia. *American Association for Artificial Intelligence (AAAI)*, pp.1419–1424.

Styperek, A., Ciesielczyk, M. and Szwabe, A., (2014). Semantic search engine with an intuitive user interface. In *ACIIDS 2014, Part I, LNAI 8397*. pp. 383–384.

Sugimoto, C.R. and Mccain, K.W., (2009). Ontology-Based Information Extraction: An Introduction and a Survey of Current Approaches. *Journal of Information Science*, 4, pp.1–15.

Sugiyama, K. and Kan, M., (2010). Scholarly Paper Recommendation via User' s Recent Research Interests. In *JCDL'10*. pp. 29–38.

Syafrullah, M. and Salim, N., (2010). Improving Term Extraction Using Particle Swarm Optimization Techniques. *J. of Computer Science*, 6(3), pp.323–329.

Tamma, V., (2010). Semantic Web Support for Intelligent Search and Retrieval of Business Knowledge. *Intelligent Systems, IEEE*.

Tang, J. et al., (2008). ArnetMiner : Extraction and Mining of Academic Social Networks. In *KDD'08*. pp. 990–998.

Tao, X. et al., (2008). An Ontology-based Framework for Knowledge Retrieval. In *WI-IAT '08. IEEE/WIC/ACM International*.

Tartir, S., Arpinar, I.B. and Sheth, A.P., (2010). Ontological Evaluation and Validation. In R. Poli, ed. *Theory and Applications of Ontology :Computer Applications*. Springer Science+Business Media, pp. 115–130.

Trajkova, J. and Gauch, S., (2004). Improving Ontology-Based User Profiles. In *Recherche d'Information Assistee par Ordinateur (RIAO '04)*. pp. 380–390.

Tran, T., Cimiano, P. and Ankolekar, A., (2008). A Rule-Based Adaption Model for Ontology-Based Personalization. , 135, pp.117–135.

Tsatsou, D., Menemenis, F. and Davis, P.C., (2009). A Semantic Framework for Personalized Ad Recommendation based on Advanced Textual Analysis. *Text*, pp.217–220.

Uchyigit, G., (2009). Semantically Enhanced Web Personalization. In *Web Mining Appl. in E-Commerce & E-Services, SCI 172*. Springer-Verlag Berlin Heidelberg, pp. 25–43.

Uchyigit, G. and Ma, M.Y., (2008). *Personalization Techniques and Recommender Systems*, World Scientific Publishing Co.

Verbert, K. et al., (2011). Dataset-driven Research for Improving Recommender Systems for Learning. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge, LAK'11*. pp. 44–53.

Vu, T., Aw, A.T. and Zhang, M., (2004). Term Extraction Through Unithood And Termhood Unification. , pp.631–636.

Wang, X. and Peristeras, V., (2008). oWSD: A Tool for Word Sense Disambiguation in Its Ontology Context. In *7th International Semantic Web Conference (ISWC2008)*. Karlsruhe, Germany, pp. 2–3.

Wartena, C. and Brussee, R., (2008). Topic Detection by Clustering Keywords. In *Database and Expert Systems Applications - DEXA'08*. pp. 54–58.

Wei, G., Bao, M. and Wu, S., (2010). Research on Ontology-Based Text Representation of Vector Space Model. *IEEE*, pp.1–4.

Wei, L. and Lei, S., (2009). Integrated Recommender Systems Based on Ontology and Usage Mining. , pp.114–125.

Weidong, Z. et al., (2009). Role-activity diagrams modeling based on workflow mining. In *2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009*. IEEE Compter Society, pp. 301–305.

Weiss, S.M. et al., (2010). Information Retrieval and Text Mining. In *Fundamentals of Predictive Text Mining, Texts in Computer Science*. pp. 75–90.

Werner, D., Cruz, C. and Nicolle, C., (2012). Ontology-based Recommender System of Economic Articles. In *8th International Conference on Web Information Systems and Technologies*. Porto, Portugal, pp. 725–728.

White, R.W., Bailey, P. and Chen, L., (2009). Predicting User Interests from Contextual Information. In *SIGIR'09*. ACM.

Witten, I.H., (2005). Text mining. In *Practical handbook of internet computing*. Boca Raton, Florida.: Chapman & Hall.

Wu, Y., Siy, H. and Fan, L., (2008). Discovering Meaningful Clusters from Mining Software Engineering Literature. In *Software Engineering and Knowledge Engineering - SEKE*. pp. 613–618.

Xinglin, L. et al., (2012). Text Similarity Computing Based on Thematic Term Set. *International Journal of Advancements in Computing Technology (IJACT)*, 4(April), pp.338–345.

Yang, S.-Y., Hsu, C.-L. and Lu, S.-H., (2010). Developing an Ontology-Supported Information Recommending System for Scholars. *IEEE*, 37(10), pp.223–228.

Yang, Y. and Pedersen, J.O., (1997). A Comparative Study on Feature Selection in Text Categorization. In *ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning*. Morgan Kaufmann, pp. 412–420.

Yao, L., Tang, J. and Li, J., (2007). A Unified Approach to Researcher Profiling. In *IEEE Computer Society*. pp. 359–365.

Yujie, Z. and Licai, W., (2010). Some Challenges for Context-aware Recommender Systems. In *5th International Conference on Computer Science and Education (ICCSE)*. pp. 362–365.

Zhang, Z., Brewster, C. and Ciravegna, F., (2008). A Comparative Evaluation of Term Recognition Algorithms. In *The sixth international conference on Language Resources and Evaluation, (LREC 2008)*. Marrakech, Morocco.

Zhiqiang, L. and Port, R., (2009). Measuring Semantic Similarity between Words Using Wikipedia. In *International Conference on Web Information Systems and Mining*. pp. 251–255.

Zhou, L., (2007). Ontology learning : state of the art and open issues. *Information Technology and Management*, 8(3), pp.241–252.

Zhou, X. et al., (2012). The State-of-the-Art in Personalized Recommender Systems for Social Networking. In *Artificial Intelligence Review*. Springer, pp. 119–132.

Zhuhadar, L. and Nasraoui, O., (2010). A Hybrid Recommender System Guided by Semantic User Profiles for Search in the E-learning Domain. *JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE*, 2(4), pp.272–281.

Zimmermann, A. et al., (2007)., An Operational Definition of Context. In *CONTEXT'07 Proceedings of the 6th international and interdisciplinary conference on Modeling and using context*. pp. 558–571.