TKK Dissertations 222
Espoo 2010

METHODS AND APPLICATIONS FOR ONTOLOGY-BASED RECOMMENDER SYSTEMS

Doctoral Dissertation

Tuukka Ruotsalo

Aalto University
School of Science and Technology
Faculty of Information and Natural Sciences
Department of Media Technology

TKK Dissertations 222
Espoo 2010

METHODS AND APPLICATIONS FOR ONTOLOGY-BASED RECOMMENDER SYSTEMS

Doctoral Dissertation

Tuukka Ruotsalo

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Information and Natural Sciences for public examination and debate in Auditorium AS1 at the Aalto University School of Science and Technology (Espoo, Finland) on the 7th of June 2010 at 12 noon.

Aalto University
School of Science and Technology
Faculty of Information and Natural Sciences
Department of Media Technology

Aalto-yliopisto
Teknillinen korkeakoulu
Informaatio- ja luonnontieteiden tiedekunta
Mediatekniikan laitos

Distribution:
Aalto University
School of Science and Technology
Faculty of Information and Natural Sciences
Department of Media Technology
P.O. Box 15500
FI - 00076 Aalto
FINLAND
URL: http://media.tkk.fi/
Tel. +358-9-470 22870
Fax +358-9-470 25014
E-mail: tuukka.ruotsalo@tkk.fi

ISBN 978-952-60-3150-7
ISBN 978-952-60-3151-4 (PDF)
ISSN 1795-2239
ISSN 1795-4584 (PDF)
URL: http://lib.tkk.fi/Diss/2010/isbn9789526031514/

TKK-DISS-2753

Multiprint Oy
Espoo 2010

Aalto University

ABSTRACT OF DOCTORAL DISSERTATION	AALTO UNIVERSITY SCHOOL OF SCIENCE AND TECHNOLOGY P.O. BOX 11000, FI-00076 AALTO http://www.aalto.fi
Author Tuukka Ruotsalo	
Name of the dissertation Methods and applications for ontology-based recommender systems	
Manuscript submitted 19th of January, 2010	Manuscript revised 14th of April, 2010
Date of the defence 7th of June, 2010	
☐ Monograph	☒ Article dissertation (summary + original articles)

Faculty	Faculty of Information and Natural Sciences
Department	Department of Media Technology
Field of research	Media Technology
Opponent(s)	Professor Ray Larson
Supervisor	Professor Eero Hyvönen
Instructor	Professor Eero Hyvönen

Abstract

Recommender systems are a specific type of information filtering systems used to identify a set of objects that are relevant to a user. Instead of a user actively searching for information, recommender systems provide advice to users about objects they might wish to examine. Content-based recommender systems deal with problems related to analyzing the content, making heterogeneous content interoperable, and retrieving relevant content for the user.

This thesis explores ontology-based methods to reduce these problems and to evaluate the applicability of the methods in recommender systems. First, the content analysis is improved by developing an automatic annotation method that produces structured ontology-based annotations from text. Second, an event-based method is developed to enable interoperability of heterogeneous content representations. Third, methods for semantic content retrieval are developed to determine relevant objects for the user.

The methods are implemented as part of recommender systems in two cultural heritage information systems: CULTURESAMPO and SMARTMUSEUM. The performance of the methods were evaluated through user studies. The results can be divided into five parts. First, the results show improvement in automatic content analysis compared to state of the art methods and achieve performance close to human annotators. Second, the results show that the event-based method developed is suitable for bridging heterogeneous content representations. Third, the retrieval methods show accurate performance compared to user opinions. Fourth, semantic distance measures are compared to study the best query expansion strategy. Finally, practical solutions are developed to enable user profiling and result clustering.

The results show that ontology-based methods enable interoperability of heterogeneous knowledge representations and result in accurate recommendations. The deployment of the methods to practical recommender systems show applicability of the results in real life settings.

Keywords Ontology-based recommender systems, Information storage and retrieval, Content analysis

ISBN (printed) 978-952-60-3150-7	ISSN (printed) 1795-2239
ISBN (pdf) 978-952-60-3151-4	ISSN (pdf) 1795-4584
Language English	Number of pages 96 + 80
Publisher Department of Media Technology	
Print distribution Department of Media Technology	

☒ The dissertation can be read at http://lib.tkk.fi/Diss/

Aalto-yliopisto

VÄITÖSKIRJAN TIIVISTELMÄ	AALTO-YLIOPISTO TEKNILLINEN KORKEAKOULU PL 11000, 00076 AALTO http://www.aalto.fi

Tekijä Tuukka Ruotsalo

Väitöskirjan nimi

Menetelmiä ja sovelluksia ontologiaperustaisiin suosittelujärjestelmiin

Käsikirjoituksen päivämäärä 19.01.2010	Korjatun käsikirjoituksen päivämäärä 14.04.2010
Väitöstilaisuuden ajankohta 07.06.2010	
☐ Monografia	☒ Yhdistelmäväitöskirja (yhteenveto + erillisartikkelit)

Tiedekunta	Informaatio- ja luonnontieteiden tiedekunta
Laitos	Mediatekniikan laitos
Tutkimusala	Mediatekniikka
Vastaväittäjä(t)	Professori Ray Larson
Työn valvoja	Professori Eero Hyvönen
Työn ohjaaja	Professori Eero Hyvönen

Tiivistelmä

Suosittelujärjestelmät ovat informaationsuodatusjärjestelmiä, joiden tavoitteena on tunnistaa tietylle käyttäjälle relevantit informaatiokohteet. Sen sijaan, että käyttäjä aktiivisesti etsisi informaatiota, suosittelujärjestelmä voi tiedottaa käyttäjää informaatiokohteista, joihin käyttäjä mahdollisesti haluaisi tutustua.

Sisältöperustaiset suosittelujärjestelmät tunnistavat relevantit informaatiokohteet niiden sisällön perusteella. Alueen tutkimusongelmia ovat automaattinen sisällön analysointi, heterogeenisen sisällön yhteentoimivuus ja tiedonhaun menetelmät, sekä niiden käyttäminen relevanttien informaatiokohteiden tunnistamiseen.

Tässä työssä keskityttiin ontologiaperustaisten menetelmien kehittämiseen näiden ongelmien vähentämiseksi. Ensiksi kehitettiin automaattisen annotoinnin menetelmä, jolla rakenteista ontologiaperustaista annotaatiota voidaan tuottaa tekstistä. Toiseksi kehitettiin tapahtumaperustainen tietämyksen esittämismalli, jolla mahdollistetaan heterogeenisten sisällönkuvailujen yhteentoimivuus. Kolmanneksi kehitettiin menetelmiä semanttiseen tiedonhakuun, joilla rakenteisesta tietämyksestä voidaan tunnistaa käyttäjälle relevantit informaatiokohteet.

Menetelmät on toteutettu osina KULTTUURISAMPO- ja SMARTMUSEUM -järjestelmiä kulttuuriperintöalueella ja niiden toimintaa on arvioitu käyttäjäkokein. Tutkimuksessa syntyi viidenlaisia tuloksia. Ensiksi sisällönanalyysimenetelmiä parannettiin käytössä oleviin menetelmiin nähden. Toiseksi tapahtumaperustaisella tietämyksen esittämismallilla mahdollistettiin heterogeenisen sisällön yhteentoimivuus. Kolmanneksi tiedonhaussa saavutettiin lähes yhtä hyvä tarkkuus, kuin mihin ihmiset pystyivät samassa tehtävässä. Neljänneksi, semanttisen etäisyyden arviointiin kehitettyjä menetelmiä vertailtiin parhaan kyselynlaajennusstrategian löytämiseksi. Viidenneksi, käytännöllisiä ratkaisuja kehitettiin käyttäjäprofilointiin ja tulosten klusterointiin.

Tulokset osoittavat, että ontologiaperustaisilla menetelmillä voidaan parantaa heterogeenisten sisältöjen yhteentoimivuuttta ja niiden avulla käyttäjille voidaan tarjota täsmällisiä suosituksia. Menetelmät ovat osoittautuneet toimiviksi myös osana reaalimaailman järjestelmiä.

Asiasanat ontologiaperustaiset suosittelujärjestelmät, informaation tallennus ja haku, sisällön analyysi

ISBN (painettu) 978-952-60-3150-7	ISSN (painettu) 1795-2239
ISBN (pdf) 978-952-60-3151-4	ISSN (pdf) 1795-4584
Kieli Englanti	Sivumäärä 96 + 80

Julkaisija Mediatekniikan laitos

Painetun väitöskirjan jakelu Mediatekniikan laitos

☒ Luettavissa verkossa osoitteessa http://lib.tkk.fi/Diss/

Preface

This research was carried out during the years 2005–2010 in the Semantic Computing Research Group (SeCo) at the Department of Media Technology, the Helsinki University of Technology (now Aalto University) and the Department of Computer Science, University of Helsinki, Finland. During the spring semester of the year 2008, I was a visiting scholar in the Web and Media group at the Vrije Universiteit Amsterdam (VU) in the Netherlands.

This work would not have been possible without the help, encouragement and support from several people. Hereby I wish to thank all the people who were involved in the work and made this thesis possible.

I want to thank my supervisor professor Eero Hyvönen for all his time, support and feedback. He has been a professional support for guiding the research and responsible for creating an exuberant and inspiring research environment. This thesis would not have been complete without a proficient guidance of professor Guus Schreiber, who was my supervisor at the VU. I also appreciate the comments and suggestions for improvement by professors Frank van Harmelen and Airi Salminen, the pre-examiners of this thesis.

I'm grateful to my colleagues and other researchers, who have contributed to this thesis, indirectly or directly. I warmly thank my colleagues, especially doctoral students in the SeCo group, for their help, support and patience. Tomi, Eetu, Kim, Osma, Katri, Reetta, Jussi, Matias, Jouni, thanks for all of you. Also all other present and past members of the SeCo group deserve thanks. I am also thankful for professor Pirkko Oittinen and doctoral students from the visual media group. Especially Stina, Mari and Jan have devoted time for splendid discussions at the coffee room.

I am also grateful for people from other research groups who I have had an opportunity to work with. I want to thank the SMARTMUSEUM team for making me part of a great and successful EU project. The whole VU team deserves special thanks for the support in research, but also for the great spring and summer in Amsterdam. Special thanks go to Mark, Laura and Alistair, Antoine and Cécile, Michiel and Andra, Anna, Willem and Vera, Ryan (UCB), Véronique, Borys, Bob, Lora, Jacco, Jan, Alia and all other Web and Media, and AI group members.

Funding for the research was provided by the Finnish Funding Agency for Technology and Innovation (Tekes) (FinnONTO and Semantic Web 2.0), the European Union's 7th Framework program (Smartmuseum), and the Research Foundation of TKK. In different stages the projects have had more than forty additional funding organizations, both public and private. I wish to thank all of the organizations for their support.

Although conducting research has been a big part of my life during the last years, it is not worth much without having fantastic friends and family. I want to thank all of my friends for proving me that there are more important things in life than research and doctoral thesis. Special thanks go to Aki and Mirkka, Antti and Laura, Iivari, Antto and Anne, Tiina and Jorge, Outi and Joel, Saijaleena and AJ, and Thomas and Milka. Among being friends, Saara, Jaan-Olle, Maija and Jani have also been like-minded in praising, but also in moaning about research and scientific work in general.

I want to thank my wife Anna-Kaisa for her love and support throughout the years. My life is full of love and happiness because of you! Finally, I want to dedicate this thesis to my dear parents, my mother Hannele and my step-father Petri. Thank you for always giving me the support, freedom and understanding in all my decisions and choices.

Helsinki, 26th of April, 2010.

Tuukka Ruotsalo

Contents

List of Publications

This thesis consists of an overview and the following publications which are referred to in the text by their Roman numerals.

I Tuukka Ruotsalo and Eero Hyvönen: An Event-Based Approach for Semantic Metadata Interoperability. 2007. In *K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux (Eds.): Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference (ISWC/ASWC 2007),* LNCS 4825, pages 409–422. Springer-Verlag, Berlin, Heidelberg.

II Tuukka Ruotsalo and Eero Hyvönen: A Method for Determining Ontology-Based Semantic Relevance. 2007. In *R. Wagner, N. Revell, and G. Pernul (Eds.): Proceedings of the 18th International Conference on Database and Expert Systems Applications (DEXA 2007),* LNCS 4653, pages 680–688. Springer-Verlag, Berlin, Heidelberg.

III Tuukka Ruotsalo, Lora Aroyo, and Guus Schreiber: Knowledge-Based Linguistic Annotation of Digital Cultural Heritage Collections. IEEE Intelligent Systems, vol. 24, no. 2, pages 64–75, IEEE Computer Society, 2009.

IV Tuukka Ruotsalo, Eetu Mäkelä, Tomi Kauppinen, Eero Hyvönen, Krister Haav, Ville Rantala, Matias Frosterus, Nima Dokoohaki, and Mihhail Matskin: Smartmuseum: Personalized Context-aware Access to Digital Cultural Heritage. 2009. In Proceedings of the International Conference on Digital Libraries and the Semantic Web 2009 (ICSD 2009), September, pages 178–192. Università di Trento, Trento, Italy.

V Tuukka Ruotsalo and Eetu Mäkelä: A Comparison of Corpus-Based and Structural Methods on Approximation of Semantic Relatedness in Ontologies. International Journal On Semantic Web and Information Systems, vol. 5, no. 4, pages 39–56, IGI Global, 2009.

Author's Contribution

In the conference article I, a method for making metadata conforming to heterogeneous schemas semantically interoperable is presented. A case study of transforming three different schemas and datasets was conducted. The resulting event-based knowledge representation was deployed to an ontology-based recommender system for the CULTURESAMPO portal. As the first author of the paper, Tuukka Ruotsalo conducted the study, designed and developed the method, constructed the recommender system, and was responsible for writing the article.

In the conference article II, a method for determining ontology-based semantic relevance is presented. A user study was conducted to evaluate the performance of the method. The method was implemented as part of the recommender system of the CULTURESAMPO portal. As the first author of the paper, Tuukka Ruotsalo conducted the user study, designed and developed the method, constructed the recommender system, and was responsible for writing the article.

In the journal article III, a method for automatic annotation of objects in digital cultural heritage collections is presented. Given a set of objects, each accompanied by a text description, a set of structured vocabularies, a metadata schema, and a training set of annotations of the text descriptions, the method produces annotations for the objects. The method was evaluated through a user study. As the first author of the paper, Tuukka Ruotsalo conducted the study, designed and developed the method, implemented the system, and was responsible for writing the article.

In the conference article IV, the SMARTMUSEUM recommender system is presented. The main contribution of the article is a recommendation method that extends the method reported in II by introducing simplified scoring and indexing. The recommendation method was also extended by a user-profile -based adaptation method and a result-clustering method. As the first author of the paper, Tuukka Ruotsalo designed and developed the method. Other authors participated in the implementation and the design of the system. As the first author of the paper, Tuukka Ruotsalo was responsible for writing the article.

In the journal article V, the performance of two distinct approaches to determine semantic relatedness, corpus-based and structural methods, are compared. As the first author of the paper, Tuukka Ruotsalo conducted the study, and mainly implemented the methods. The first author was responsible for writing the article. The second author participated in the implementation, the reporting and the design of the study.

In addition, the author has participated in developing the CULTURESAMPO portal and the SMARTMUSEUM system. The implementation of the methods in the scope of the CULTURESAMPO portal is further discussed in article [62], and in scope of the SMARTMUSEUM system in the article [76]. The overall description of the CULTURESAMPO system is reported in the articles [60, 59]. The author has also contributed to related research on the areas of semantic information access [84], user interface technologies [85], and ontology infrastructure [64, 63].

List of Abbreviations

AAT	Art and Architecture Thesaurus
CBIR	Content-based Image Retrieval
CF	Collaborative Filtering
CIDOC-CRM	International Committee for Documentation Conceptual Reference Model
DC	Dublin Core
ICA	Independent Component Analysis
idf	Inverse Document Frequency
IE	Information Extraction
IF	Information Filtering
IR	Information Retrieval
kNN	k-Nearest Neighbor
LSA	Latent Semantic Analysis
NER	Named Entity Recognition
OWL	Web Ontology Language
PoS	Part-of-Speech
RDF(S)	Resource Description Framework (Schema)
SRL	Semantic Role Labeling
SUS	System Usability Scale
SVD	Singular Value Decomposition
tf	Term Frequency
TGN	Getty Thesaurus of Geographic Names
ULAN	Union List of Artist Names
URI	Uniform Resource Identifier
VRA	Visual Resources Association
VSM	Vector Space Model
VSM-TS	Vector Space Model for Triple Space
YSO	General Finnish Ontology

1 Introduction

1.1 Background

While storing digital information has become possible, retrieving and accessing resources in the growing collections is far from trivial. We are facing a mixture of information originating from professionally managed collections such as image or text databases to individually or collaboratively created content such as personal image collections, online encyclopedias or even the World Wide Web itself.

While the explosion of on-line information has enabled accessing digital information, it has also brought to the forefront the problem of finding useful information and making sense of large multi-dimensional information spaces. One of the main challenges that information systems confront is the retrieval of information to satisfy users' information needs [90].

Digital information is mainly accessed using *information retrieval* (IR) systems. IR systems assume that the users are able to express their information need in the form of a query [6]. In its most common form, a user enters a set of keywords which summarize the user's information need. Given the query, the goal of an IR system is to retrieve information which is relevant to the information need of the user.

Recommender systems form a specific type of *information filtering* (IF) technique that attempts to present information objects that are likely of interest to the user. Instead of users actively searching for information, recommender systems provide advice to users about objects they might wish to examine [18]. Recommendations can be based on the content of the objects or observations of user behavior. In [104], the shift from active search to discovery is characterized as follows:

"The Web, they say, is leaving the era of search and entering one of discovery. What's the difference? Search is what you do when you're looking for something. Discovery is when something wonderful that you didn't know existed, or didn't know how to ask for, finds you."

Recommender systems have been an active area of research, but also a source for abundance of practical applications. Recommender systems have been used in a number of different applications such as recommending books, music [78], movies [94], videos [54], other products [122, 121], news [71], identifying web pages that will be of interest for the user, or suggesting alternate ways of searching for information [9].

In its most common formulation, the recommendation problem is reduced to the problem of estimating ratings for objects that have not been seen by a user [2]. To achieve this, recommender systems use several distinct techniques and can be categorized into three main categories [2]. In the case of *collaborative filtering* (CF) [44] the user will be recommended objects that people with similar tastes and preferences liked in the past. In the case of *content-based recommender systems* [99, 105] the user will be recommended relevant objects based on the content of the objects the user is examining or has examined in the past. *Hybrid systems* combine collaborative filtering and content-based approaches [19, 136, 107, 8].

Recommender systems that are based on collaborative filtering have been successful [112], but they are not suitable for all use cases. For example, in cases where the number of users is small relative to the number of content objects in the system, coverage of the ratings can be sparse [8]. In cases where the population of users or the objects are varying the problem becomes even more crucial. Another problem is that if a user's tastes are unusual compared to the rest of the users, there will not be other users with similar tastes and predictions can not be drawn. A central problem affecting collaborative filtering systems is the availability of user preferences. In many applications, users are not willing to identify themselves and the tracking of the user behavior may be difficult.

The problems that collaborative filtering systems suffer from can be avoided in case it is possible to derive recommendations based on the content of objects. The content-based approach to recommendation has its roots in information retrieval research. The retrieval of objects is based on data structures that are created using features present in or extracted from the content descriptions of the objects [8]. In case *user profiling* is possible, the features of the content that the user has preferred in the past can be stored in the user's profile and used in the retrieval.

1.2 Scope

This thesis focuses on improvements in content-based recommender systems. Specifically, this thesis concentrates on methods that make use of *ontologies*. Such systems are called *ontology-based recommender systems*. In computer science, an ontology can be defined as a specification of a representational vocabulary consisting of definitions of classes, relations, functions, and other objects for a shared domain of discourse [46].

Ontology-based methods can be used to reduce problems that content-based recommender systems are known to suffer from. These problems concern the way the systems analyze the content they recommend, the way they retrieve the content, and the way they treat heterogeneously represented content [125, 2, 52]:

- *Content Analysis*: The features used to represent the objects need to be automatically extracted or manually associated with the objects.
- *Content Heterogeneity*: The representations of the objects can be mutually incompatible.
- *Content Retrieval*: The retrieval of the objects is limited to the features that are explicitly associated with the objects.

Associating features with the objects manually can be a cumbersome task. Therefore *automatic content analysis* is often used. In its simplest form, words in the textual descriptions of the objects are directly used as features [8, 105]. *Information extraction* techniques can be used to distill structured data or knowledge from text by identifying references to concepts and named entities as well as stated relationships between them [98, 29]. The resulting structured data can then be used as features to represent the objects. Techniques used in content-based image retrieval (CBIR) [75] can be used to extract features from images or videos. However, CBIR techniques suffer from the *semantic gap*, which is the discrepancy between the information that can be derived from the low-level image data and the interpretation that users have about the content [55].

In this thesis, the problem of content analysis is limited to content descriptions that are textual. In particular, the focus is on automated techniques that are able to analyze text and produce structured data.

Even if structured data were available for the recommender system to use, it may not always be sufficient and can suffer from content heterogeneity. Content heterogeneity means the mismatch between different data representations and conceptualizations used to describe the objects. *Syntactic heterogeneity* refers to differences among local data formats. Syntactic heterogeneity problems can be solved by modifying data to enforce homogeneity, or they can be dealt with in the applications [133].

Semantic heterogeneity occurs when the data describing the same or related real-world entities is represented in different ways [126, 24]. Semantic heterogeneity can refer to, for example, naming conflicts, when different databases use different names to represent the same concepts, or domain conflicts, when different databases use different values to represent the same or similar concepts. In addition, it can refer to structural conflicts, when different databases use different data organization to represent the same concepts [24].

Many recommender systems rely on syntactic content and similarity measures that operate on syntactic content [13, 8, 105]. In scope of content-based recommender systems, the handling of semantic heterogeneity is manifested as the ability of the recommender system to provide content that is similar at the semantic level, but can be represented with different names, values and structures. *Semantic similarity measures* function as mechanisms for comparing objects that can be retrieved or integrated across heterogeneous repositories [114]. In the case of recommender systems these measures can be used to assess how similar the objects are to the features stored in the user profile. For example, if a user has only visited an object annotated as manufactured in Paris and no other objects annotated as manufactured in Paris are available, objects annotated as manufactured in Montmartre could be recommended because Paris and Montmartre are related on the semantic level.

Another problem often faced by retrieval methods in content-based recommender systems is *over-specialization*. Over-specialization means a situation, where the system only recommends objects that score high against a user's profile and the user is limited to being recommended objects that are very similar to those already rated [2]. The problem with over-specialization is not only that a content-based system cannot recommend objects that are different from anything the user has seen before. In some cases, objects should not be recommended if they are too similar to something the user has already seen, such as a different news article describing the same event or a different photo of the same artifact [2]. The diversity of recommendations is often a desirable feature in recommender systems. Ideally, the user should be presented with a range of options instead of only the objects with highest similarity.

Ontologies have been applied to a variety of recommender systems to reduce content heterogeneity and improve content retrieval. For example, in [91, 21, 93, 92, 143, 83, 23] good results to cope with content heterogeneity have been obtained by using subsumption hierarchies to generalize user profiles. In [74, 100, 144], ontologies are used on a specific domain of product descriptions, and a hand crafted ontology is built just for this purpose.

In [101], similar approach is adopted for television program domain, and in [96] for e-tourism domain complemented with mining the user behavior. In [134, 61] ontology-based recommender systems are applied to a museum domain. A number of methods to determine semantic similarity for objects described using ontology-based knowledge-representation have been proposed (see [35] for review).

Despite all these studies, the benefits that ontologies can provide for recommender systems still remain incomplete in many ways. In the previous studies limited knowledge representation, content analysis, and content retrieval methods are used.

1.3 Contributions

This thesis describes work that has been carried out to develop ontology-based recommender systems for the cultural heritage information systems CULTURESAMPO [I,II] [62, 115], and SMARTMUSEUM [IV][76], and proposes methods to improve content analysis [III], deal with semantic heterogeneity [I], and enable accurate content retrieval [II]. In particular, this thesis concentrates on methods that are based on ontologies. Contributions are made on four areas:

- *Content analysis*: A method that produces structured ontology-based annotation using information extraction was developed [III]. The developed method was found to improve information extraction from text descriptions compared to a state of the art method, and achieved performance close to human annotators.

- *Content Heterogeneity*: A method that enables interoperability between heterogeneous structured ontology-based annotations was developed [I]. The method results to event-based knowledge representation that was used in a recommender system in the CULTURESAMPO portal.

- *Content Retrieval*: Three methods that utilize the improved content analysis and structured knowledge-representation were developed. A method to determine semantic relevance was first presented in [II]. In addition, methods that can be used to determine semantic relatedness of concepts in ontologies were compared [V]. Further, a method that simplifies the method presented in [II] and combines it with user profiling and clustering to avoid over-specialization, was presented [IV]. The methods were found to perform accurately in a user study.

- *Applications*: Two applications on cultural heritage domain were developed: a recommender system for the CULTURESAMPO portal [62, 115] and a recommender system for the SMARTMUSEUM mobile system [IV]. The methods were implemented and deployed in the applications, and found satisfying in user trials.

1.4 Structure of this Thesis

The structure of this thesis is as follows. Section 2 reviews the related work. Section 3 gives an overview of the research approach; the materials, the research methods used, and presents the results in the four contribution areas: content analysis, content heterogeneity, content retrieval, and applications. In section 4, external validity and the limitations of the research are discussed, and future research directions suggested.

2 Related Research

Research on content-based recommender systems is multi-disciplinary and requires combining methods from a number of areas. In this section, related research is presented on seven areas: content-based recommender systems, information retrieval methods, knowledge representation, semantic relatedness approximation, information extraction, user profiling, and finally evaluation of recommender systems.

2.1 Content-based Recommender Systems

Content-based recommender systems analyze the content of the objects to identify the ones that are of interest to the user [8]. Objects are recommended based on a comparison between their content and a user profile. In case the user can not be identified, the user profile may consist of only the object the user is examining at the time the recommendations are retrieved. The more detailed information about the user is available, the more complete user profile may be built.

The recommendation problem can be formulated as follows [2]. A recommender system maps each user profile - object pair to a particular rating value by estimating the rating function R:

$$R : UserProfiles \times Objects \rightarrow Ratings \tag{2.1}$$

The rating function can be estimated in a way that the highest rated object O' (or a number of highest rated objects) are selected:

$$O' = \text{arg max}_{U \in UserProfiles, O \in Objects} R(U, O) \tag{2.2}$$

In content-based recommendation methods, the rating $R(U, O)$ of object O for the user profile U is typically estimated based on the ratings assigned in the user profile U to other objects that are relevant to object O in terms of their content [1].

For example, in an art domain, a content-based recommender system tries to understand user preferences by analyzing commonalities among the content of the artifacts. These commonalities could be based on features such as the style, the creator, and the place of manufacturing. The artifacts that have a high degree of relevance to the user's profile are recommended [1].

The definition of the rating function requires to measure the similarity between the user profile and the objects. The content of the objects O are characterized using a set of features, here defined as $Content(O)$. In addition, profile of a user U needs to be defined. The user profiles are also defined in terms of features that characterize the objects, here $ContentBasedProfile(U)$. The rating function can now be written as a score function of the content-based profile and a content object:

$$R(U, O) = score(ContentBasedProfile(U), Content(O)). \tag{2.3}$$

In the case of content-based recommender systems, where the scoring is based on the content descriptions available in text or structured annotation, the score function can be implemented using methods developed in IR research.

2.2 Information Retrieval Methods

The main IR approaches are based on the Boolean model, the vector space model (VSM), and probabilistic models [6]. The simplest retrieval approach is the Boolean model that considers the features to be present or absent in an object and assigns a binary value for each feature in each object [6]. The Boolean model has disadvantages, such as that it returns too few or too many objects and is unable to rank the objects. VSM allows relevance rankings and partial matches of objects. Probabilistic models treat the process of document retrieval as probabilistic inference. Similarities are computed as probabilities that a document is relevant or not relevant for a given query. Probabilistic models have shown good retrieval performance, but do not exceed the performance of VSM [87]. However, they allow relevance feedback and prior information to be easily incorporated in the model. IR systems that use language models build a probabilistic model from the document and the query based on an n-gram language model [87].

The difference between VSM and probabilistic IR systems is not remarkable. According to [87], it is possible to change an existing vector-space IR system into a probabilistic system simply by adopting term weighting formulas from probabilistic models. The language model approach has been successful in terms of retrieval performance, but does not significantly improve the retrieval performance of the VSM. Therefore, the VSM remains the most successful IR approach [6].

Ontology-based Information Retrieval Methods

Light-weight ontologies provide controlled vocabularies that can be used in annotation of objects. This approach has brought improvements over classic keyword-based search through e.g. query expansion based on class hierarchies and other relationships [39, 26], or multifaceted searching and browsing [141, 61, 53].

Ontology-based information retrieval systems developed so far typically use a logic-based search model that is based on an ideal view of the information space as consisting of non-ambiguous, non-redundant, formal pieces of ontological knowledge [26, 138, 61, 82]. In this view, the information retrieval problem is reduced to a data retrieval task [6]. For example, in the MUSEUMFINLAND system a faceted search system and a rule-based recommendation system were proposed to access digital museum collections [61]. Such an approach can be satisfying for users when the users can interact with the system and refine the queries. However, such a system is not able to rank the objects and it can be difficult for users to digest different viewpoints in the result list returned by the system.

In the case of content-based recommender systems the profiles can be large and it is not likely that all of the features that appear in an individual profile appear in an individual object. This emphasizes the importance of ranking. Recently, ranking of the ontology-based search has been enabled by extending VSM to combine text-based features and ontology-based features [23]. However, this approach considers only individual concepts and does not enable retrieval based on a more complex annotation structure.

In this thesis, the vector space model (VSM) [117] is utilized to enable retrieval of objects annotated with a complex annotation structure. The VSM enables straightforward representation of the objects and fast computation of the score function. Furthermore, this thesis extends the retrieval model by retrieval result clustering, where the initially highest ranked objects are clustered based on the annotation structure to avoid over-specialization.

Vector Space Model

VSM is a straightforward numeric representation of the features of the objects in an i dimensional Euclidean space, where each dimension corresponds to a feature in the possible feature space. In the VSM, the features in both O and U are represented as a vector of weights $W = (w_1, ..., w_k)$, where each weight w_i denotes the importance of the feature

i for an object. A weight for a feature i for an object j is therefore indexed as $w_{i,j}$. In case of a content-based profile, the weight represents the importance of feature i for the user, and in case of an object, it represents the weight of feature i for the object.

The features (often called index terms in IR) are usually assumed to be mutually independent [6]. This clearly is a simplification because often occurrences of the features are not uncorrelated [87]. However, the independence assumption allows fast indexing and computation.

Feature Weighting

It is well known that weighting of the features can lead to improvement in the retrieval performance of the system [6]. It is intuitive that some features can be more important in scoring than others. For example, consider music albums that are characterized by the features of a recording company and a music genre. Having a very specific genre, such as *sister funk*, could relate records fairly close to each other, while a more generic genre such as *African-American music* could be less important. On the other hand, the recording company information could relate objects. In case the user tends to like a music released on Warner Bros. Records, the importance of the feature could be relatively low because Warner Bros. Records has published records of hundreds of artists that compose very different kinds of music. On the other hand, if the record company is very small, and thus concentrated on releasing only very specific kind of music, such as Warp records, the importance of this feature could be relatively high.

These aspects can be captured using a weighting scheme. A well known weighting scheme for the vector space model is term frequency–inverse document frequency (tf-idf) [118]. It is based on the idea that features that are common in the object set under interest affect the scoring less than features that are rare in the object set under interest. This can be motivated by the fact that common features are not very good at distinguishing the relevant objects from the non-relevant objects. On the other hand, the more often a feature is present in the scope of a certain object, the more relevant it can be assumed. Term frequency (tf) is the number of times a certain term, or in our case feature, appears in a object. In normalized form tf is:

$$tf_{i,j} = \frac{N_{i,j}}{\sum_k N_{k,j}}, \tag{2.4}$$

where $N_{i,j}$ is the number of times a feature i is mentioned in the object j and $\sum_k N_{k,j}$ is the sum of the number of occurrences of all features in the object j. Inverse document frequency (idf) is defined as:

$$idf_i = log\frac{N}{n_i}, \tag{2.5}$$

where n_i is the number of objects, where the feature i appears and N is the total number of objects in the system. The weight of an individual feature is given by:

$$w_{i,j} = tf_{i,j} \cdot idf_i. \tag{2.6}$$

The importance increases proportionally to the number of times a feature appears in the object description, but is offset by the frequency of the feature in the object collection. High tf-idf value is determined for features that are rare in the object collection and appear many times in the object under interest.

Scoring

In the vector model the feature vectors can be used to compute the degree of similarity between each object O stored in the system and the profile of the user U. The vector model evaluates the similarity between the vector representing an individual object W_{O_j} and the user profile W_U. The similarity between the vectors can be quantified, for example, using the cosine of the angle between the user profile vector and the object vector:

$$score(O_j, U) = sim(W_{O_j}, W_U) = \frac{W_{O_j} \cdot W_U}{|W_{O_j}||W_U|} = \frac{\sum_{i=1}^{k} w_{i,j} \cdot w_{i,u}}{\sqrt{\sum_{i=1}^{k} w_{i,j}^2} \cdot \sqrt{\sum_{i=1}^{k} w_{i,u}^2}}, \quad (2.7)$$

where k is the total number of features in the system, j is an index for an object, u is an index for a profile, and i is an index for a feature.

The dot product of the vectors is normalized using the Euclidean distance between the vectors. Thus, the vector model ranks the objects according to their similarity to the profile.

Alternative Scoring

Many variations of the weighting scheme and the scoring function exist [87]. Many practical search engine implementations treat the cosine similarity in a slightly modified manner. In the SMARTMUSEUM system, the open source search engine Apache Lucene[1] was used. It computes the cosine similarity using the following scoring function[2]:

[1] http://lucene.apache.org/

[2] The factors not affecting computing in the methods reported in this thesis are omitted. Full documentation can be found at: http://lucene.apache.org/java/2_9_0/api/all/org/apache/lucene/search/Similarity.html.

$$score(U, O_j) = cf(U, O_j) \cdot qb(U) \cdot \frac{W_{O_j} \cdot W_U}{|W_U|} \cdot dln(O_j), \tag{2.8}$$

where cf is a coord-factor, qb is a query boost, and dln is a document length normalizer.

The major modification that Lucene does is that it removes the normalization with respect to W_{O_j}. This is because the normalization of the object information O_j can be problematic in that it removes all object length information (number of features present in the object).

In Lucene, the normalization effect is encapsulated in dln, which ensures that objects with less present features contribute more to the score. In fact, the normalization term $|W_U|$ now only contributes to keeping scores between different queries or profiles comparable to each other. The other modifications are the possibility to boost the value of the features at retrieval time using the query boost $qb(U)$. This can be useful especially in cases where the weight for each feature can be determined with some other technique, such as a user interface control or a feature expansion strategy that adds additional features to the profile with some weights. The cf boosts the similarity of the profile to an object based on the fraction of the features present in the object compared to all of the features in the profile. In other words, the profile score is up-weighted with respect to which share of its features are found in the object.

$$cf(U, O_j) = \frac{mf}{k}, \tag{2.9}$$

where mf is the number of matching features and k is the total number of features in the profile.

This score function is implemented in Lucene as the Practical Scoring Function, formally

$$score(U, O_j) = cf(U, O_j) \cdot qn(U) \cdot \sum_{i=1}^{k} (tf_{W_{O_j i}} \cdot idf_i^2 \cdot qb(W_{U_i}) \cdot ln(W_{O_j})), \quad (2.10)$$

where the query norm qn makes scores between queries comparable, qb is a function returning a boost for a single profile feature, and i is the feature index ranging from the first feature with index 1 up to k, which is the number of features, and ln is a length norm that up-weights documents with less present features. The qn factor does not affect ranking (since all ranked objects are multiplied by the same factor), but rather just attempts to make scores from different profiles comparable. It is computed as:

$$qn(U) = \frac{1}{(\sum_i^k (idf(i) \cdot qb(i))^2)^{\frac{1}{2}}}. \quad (2.11)$$

The length norm is computed as:

$$ln(W_{O_j}) = \frac{1}{\sqrt{nf}}, \quad (2.12)$$

where nf is the number of features present used to index the object.

Lucene calculates the tf-idf in a modified way. For tf it uses:

$$tf = freq_{i,j}^{\frac{1}{2}}, \quad (2.13)$$

and for idf:

$$idf = 1 + log(\frac{N}{n_i + 1}). \quad (2.14)$$

Despite its simplicity, the VSM with tf-idf weighting is currently the most common way to represent objects in any information retrieval system [87]. The popularity of VSM can be explained by the speed of vector operations that it allows. Methods exist for performing dimension reduction [12, 72], and VSM has also performed well in retrieval quality [87]. It has been shown that it is difficult to improve the performance of the VSM approach in IR without query expansion or relevance feedback [6].

2.3 Knowledge Representation

Content-based recommender systems are designed mostly to recommend text-based objects and employ techniques to represent object features that are directly acquired from the textual descriptions of the objects [8, 105]. In the case of VSM, the objects in the systems are described with feature vectors W_o that are constructed based on the occurrences of the words in the text descriptions of the objects. For example, the content-based component of the Fab system [8], which recommends Web pages to users, represents Web page content with the 100 most important words. Similarly, the Syskill & Webert system [105] represents documents with the 128 most informative words. Each word is seen as a separate feature that characterizes the object [2].

Such a representation, where the content expressed in natural language is directly indexed using words as the features, has limitations. The problem with indexing directly with unstructured text is that the syntactic and lexical realizations of the sentences may vary. For example, consider the following sentences: ”The work was created in France in 1888 by Van Gogh.” and ”In Arles, Van Gogh painted the still life in the late 19th century.”. Both of the sentences express the same semantic content but have very different syntactic and lexical realizations.

If the relevance of the second sentence would be rated based on features from the first sentence, the rating could only be made through the words ”Van” and ”Gogh” that occur

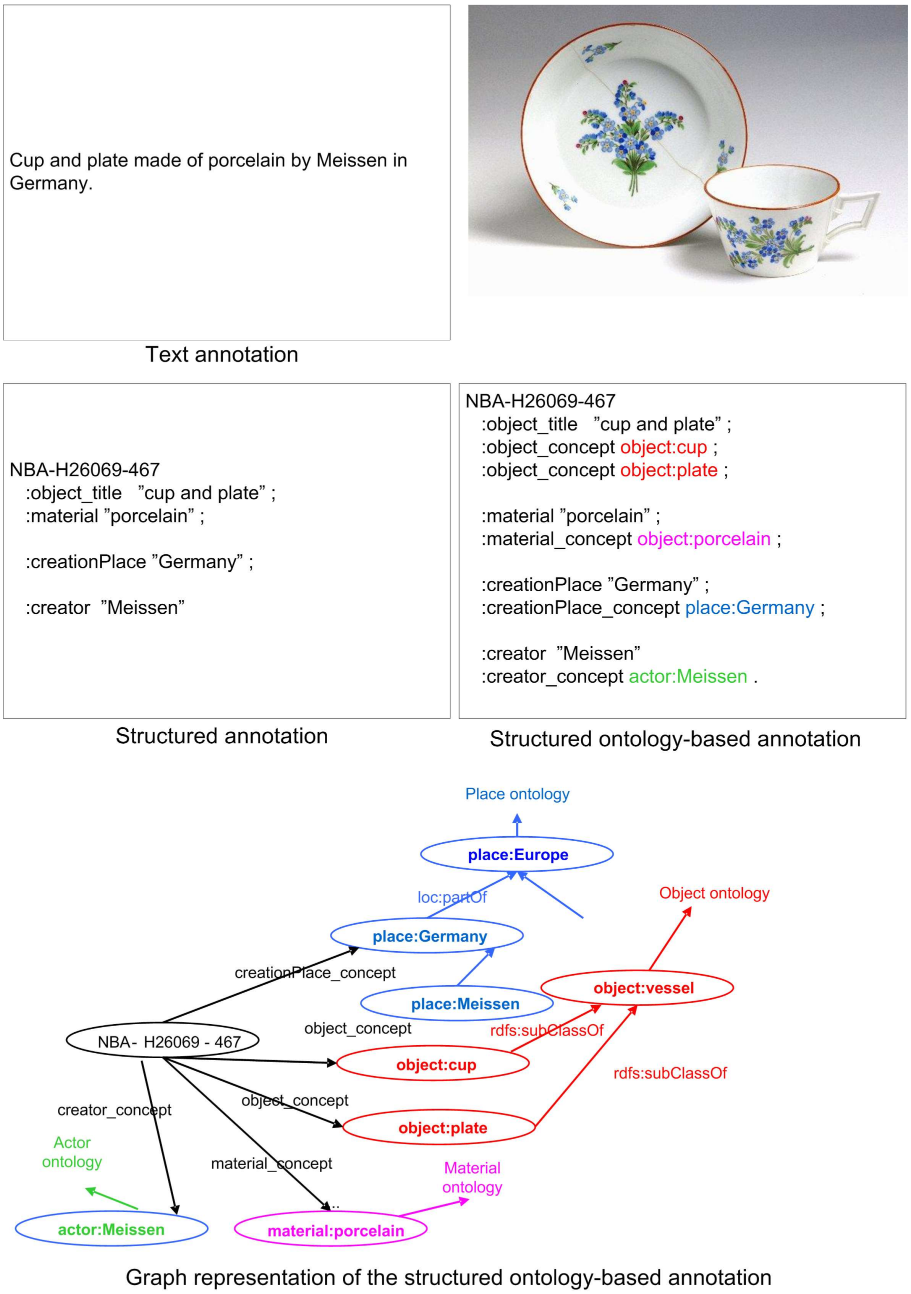

Figure 2.1: Annotation of a "cup and plate" from National Museum of Finland. Annotation is presented on three levels: text, structured, and ontology-based structured. A graph representation of the ontology-based structured annotation is also illustrated. The Figure is modified from [58].

in both of the sentences. The rating could be improved by modeling the concepts and their relations in the sentences and using the resulting structures as features. For example, ”Van Gogh” could be identified as a person and as the creator of the artwork, and ”France” and ”Arles” could be identified as place names. In case the background knowledge that ”Arles” is part of ”France” would be available the connection on the semantic level between these place names could be possible. Such features can be represented using ontology-based structured annotation, which is discussed in the following sections.

Structured Annotation

According to [55], annotation is defined as: ”information that is explicitly related to an object with the purpose of describing the object for future reference and retrieval”. Another definition is given in [3], where annotations are defined as: ”metadata, that is, additional data which relate to an existing content and clarify the properties and semantics of the annotated content”.

The latter definition refers to what in this thesis is called structured annotation. Structured annotation means annotation that corresponds to a knowledge representation that clarifies the semantics of the annotated content according to some schema. Structured annotation enables more carefully defined features to be used in the feature vectors. These features correspond to properties in a schema or a standard element set. When a controlled set of concepts that are defined in ontology are used as the values of the properties of the schema, the annotation is called ontology-based structured annotation. Textual annotation refers to content descriptions that are expressed in natural language.

Ontologies

In computer science, an ontology can be defined as a specification of a representational vocabulary consisting of definitions of classes, relations, functions, and other objects for a shared domain of discourse [46]. Thesauri, ontologies and lexical databases are not clearly distinguishable from each other [135], but rather they define similar vocabularies with different levels of formal semantics. In this thesis, formal semantics refer to structures that can be used by automated reasoning procedures to give additional statements about the structure in some logic. Consensus on ontological definitions among members of a community is an important difference between ontologies and conceptual models [34]. Conceptual models are application-dependent, but ontologies are only based on people's understanding of the domain [49].

When referring to ontology in this thesis, a lightweight ontology is meant. Lightweight ontologies embed limited knowledge, but semantics are expressed explicitly. Typically lightweight ontologies document the different meanings of lexical entries (for example, bank as a financial institution and bank as a river bank), ensure the correctness of the transitive subsumption relations (bank is a kind of financial institution), explicate hierarchical relations, such as meronymy (bank is a part-of the financing sector), and document related concepts (bank is related to financing).

Metadata Schemas

Metadata schemas can be used to increase the structure of the annotation [55]. Metadata schemas consist of elements or properties that indicate the way the concepts in the ontologies are linked to the objects that are being annotated [55]. With a metadata schema one can, for example, distinguish the creator of an object from the creation place of an object. It is important to note that schema definitions can be based on ontology definitions and vice versa [50].

The differences between textual, structured and ontology-based structured annotation are illustrated in Figure 2.1, where a partial annotation of an artifact ”cup and plate” from the National Museum of Finland is shown. The possible annotation is presented on the three levels respectively. A graph representation of the ontology-based structured annotation is also illustrated.

Using ontologies one can, for example, define that there are two entities named "Meissen", one that is a city and is a part of Germany, and another that is a factory located in the city of Meissen. Using a metadata schema one can, for example, define that "Meissen (factory)" is the creator of an object.

Annotation Heterogeneity

Annotations in real life collections can be anything from text to structured ontology-based annotations. The question of how annotation should be represented is non-trivial. For some purposes, only textual annotation, where natural language is used to give descriptions for resources, can be suitable. Other scenarios could require ontology-based structured annotation.

Matching textual annotations and structured ontology-based annotations require extracting the necessary concepts and relations from text. However, semantic heterogeneity in annotation can also occur in the case of ontology-based structured annotation, where different structures and concepts can be used to describe similar objects. If the objects originate from different collections, alternative representations of the objects are difficult to avoid [106].

Annotation Interoperability

Explicit and formal definition of semantics of the concepts has recently guided researchers to apply formal ontologies as a solution to reduce semantic heterogeneity in annotation [47]. Concepts from ontologies can be used to define the meaning of the values used in the structured annotation.

The heterogeneity can occur also at the metadata schema level. Different approaches have been proposed to enable automatic matching and mapping between the metadata schemas (see [111] for review). The automatic methods rarely find all of the correspondences and therefore rule-based approaches are often used [116]. Schema mapping can be done by finding correspondences between schemas pairwise or by using a global commonly agreed schema [32]. The commonly agreed schema approach presents an abstract global schema that can model the local schemas, or can be extended to model the local schemas. Examples of commonly agreed schemas are standard element sets, such as the Dublin Core metadata element set [67] or event-based approaches, such as CIDOC-CRM [31].

The Dublin Core (DC) metadata element set is a widely used standard element set. It defines 15 main elements to describe objects, such as creator, date and type. One can define local schemas according to DC element set. For example, VRA [5] metadata schema extends the DC element set with elements tailored to the needs of visual objects.

CIDOC-CRM is an example of the event-based approach proposed for cultural heritage domain [31]. CIDOC-CRM defines an ontology that consists of a set of classes and properties. Examples of classes are *Event*, *Visual Item*, and *Person*. For example, an object representing a particular person, such as Napoleon, could be an instance of the class Person and an object representing the Battle of Waterloo an instance of the class Event. Further, the objects and events can have properties that characterize them. For example, the object representing Napoleon could have the property *participated in* having the object representing *the Battle of Waterloo* as the value.

Ontology Languages

The languages developed by the Semantic Web community have been adopted to support the conceptual representation of ontologies and annotation schemas. The Resource Description Framework (Schema) (RDF(S)) [15] and OWL [89], can be used to formally describe concepts and properties between them. The constructs in these languages have predefined semantics. Resources (concepts and properties) can be defined and described using these constructs. The RDF(S) language, and the semantics defined for it, can be used to describe, for example, the subsumption and the type relations. One can, for example, express that there are classes named human and mammal, a class named human is a subclass of class named mammal and that there is a resource named Napoleon that has a type human and that through subsumption also has a type mammal. The resources can be identified with Uniform Resource Identifiers (URI's). RDF has been defined as a general model for describing web resources, using a variety of syntax formats. The RDF data model is similar to conceptual modeling approaches such as Entity-Relationship diagrams [25], as it is based upon the idea of making statements about resources in the form of subject-predicate-object statements, also known as triples in RDF terminology. An annotation A_O for an object O consist of a set of triples $\{t\}$.

$$A_O = \{t\}, \tag{2.15}$$

A triple t can be written as

$$t =< s, p, o >, \tag{2.16}$$

where s is called the subject, p is called the property and o is called the object. The subject, the property and the object in the triple are resources, and the object can also be a literal. For example, the object type of ”cup and plate” shown in Figure 2.1 can be written as the triple:

`<A-H26069-467, object_concept, object:cup>.`

In case the following ontological background knowledge exists

`<object:cup, rdfs:subClassOf, object:vessel>,`

a triple

`<NBA-H26069-467, object_concept, object:vessel>,`

can be inferred.

The deductive closure and the set of triples describing the reasoned annotation are denoted as

$$A'_O = \{t'\}, \tag{2.17}$$

where the set $\{t'\}$ now includes all the triples $t_{1...k}$ present in the deductive closure of the triples for the object O.

2.4 Semantic Relatedness Approximation

While the adoption of ontologies in recommender systems has been found to be useful [93], ontologies are not necessarily directly suitable to be used across different applications and domains [127]. This is because all of the concepts and relations important for the domain are not necessarily defined in the ontologies. In VSM, this can cause a sparsity problem, i.e. the features in the profile vectors may not match the features in the object vectors. However, the features can still be semantically related and thus should be matched. This has raised the question if the missing relations could be acquired automatically to fit the needs of a specific sub-domain [102].

The acquisition of the relations can be seen as a semantic relatedness approximation problem. Two approaches to approximate semantic relatedness can be identified: measures that make use of the structure of the ontology, and corpus-based methods that make use of an external document collection.

Structural Measures

The backbone of the ontology graph is the subsumption hierarchy which, for example in the case of WordNet [95], accounts for close to 80% of the relations [17]. Therefore, the structural measures are mainly based on metrics that make use of the subsumption hierarchies.

A simple way to compute semantic relatedness in a subsumption hierarchy is to identify relatedness with the path length between the concepts. This approach is taken, for example, by Rada and colleagues [110] and by Leacock and Chodorow [73], where the path length is normalized with the maximum depth of the subsumption hierarchy.

Despite its apparent simplicity, an acknowledged problem with the edge-counting approach is that it typically relies on uniform distances. Some subsumption hierarchies are much denser than others and therefore the depth of the taxonomy should be taken into account [113]. This feature is considered in the measure proposed by Wu and Palmer [139]. It takes into account the fact that two classes near the root of a hierarchy are close to each other in terms of edges but can be very different conceptually, while two classes deeper in the hierarchy can be separated by a larger number of edges and can still be closer conceptually.

Other techniques include Resnik's Information-based Approach [113] and Jiang and Conrath's Combined Approach [68]. The key idea underlying Resnik's approach is the intuition that one criterion of similarity between two concepts is the extent to which they share information in common. Jiang and Conrath's Combined Approach is based on combining the hierarchical measures and corpus statistics.

Corpus-based Methods

Information retrieval research has proposed a number of unsupervised methods, typically based on dimensionality reduction or clustering techniques that can be used to find semantic relations between terms based on a document collection or a text corpus. A well known method of this type is Latent Semantic Analysis (LSA) [72].

LSA utilizes the idea that relationships between terms within a document collection can be deduced from their occurrence patterns across the documents. Singular value decomposition (SVD) is applied to a term-document matrix to obtain a projection of both documents and terms into a lower dimensional space. Relatedness calculations between terms (or documents) can then be performed in the lower dimensional space.

Methods to determine specific kinds of labeled relations have also been proposed [45, 11]. The main research direction has been to mine taxonomic relations to form subsumption hierarchies for the backbones of ontologies [51, 69, 81, 124]. There are also several studies that propose learning non-taxonomic relationships from text. Approaches have been developed for learning part-of relations [10, 132], qualia relations [27], causation relations [41], and other non-taxonomic relations [119].

2.5 Information Extraction

Producing structured ontology-based annotations is a major bottleneck of many real world systems and, if done manually, can lead to low utilization of the systems [131]. Many objects are, however, accompanied by a textual description. Such information is frequently available, for example, in newspaper and journal articles, descriptions in music or art databases, on-line encyclopedias, and many other portals and web sources.

Information Extraction (IE) is any process which selectively structures and combines data which is found in one or more text documents or textual annotations [29]. The results of IE have been used in finding good indexing features for IR [87]. By features, such as named entities, or structured representation of the data, one effectively extends the simple bag-of-words model of IR [86].

The components of a typical IE system (based on [29]) are depicted in Figure 2.2[3]. Each of the components is discussed below.

[3] The original list of components includes a filtering component that selects the most important pieces of text for more detailed analysis. The filtering component has been omitted, because in this thesis the focus of information extraction is in extracting structures from textual annotations that can be considered relevant for the object and filtering is not required.

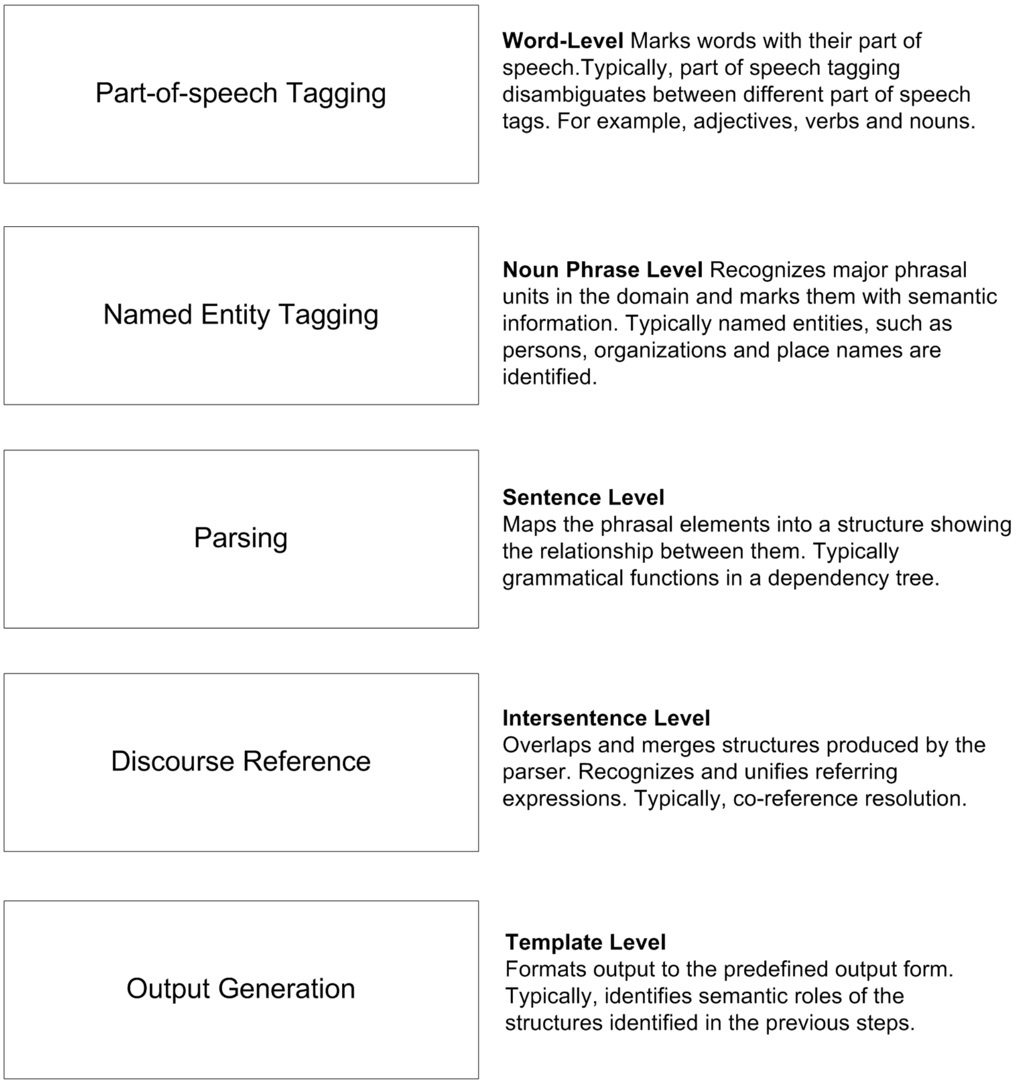

Figure 2.2: Typical components of an IE system. Adapted from [29].

Part-of-Speech Tagging

Part-of-speech (PoS) tagging marks words with their part of speech. For example, in the sentence : "Books are made of ink, paper, and glue.", the word "books" is a plural form of a noun, while in the sentence : "Mr Y books the tickets.", "books" is a verb. PoS tagging helps to identify the meaning of a word on a word class level. For the sentence "Barack

Obama gave a victory speech in Chicago", a PoS tagger can give the following output [4]:

```
Barack/NNP
Obama/NNP
gave/VBD
a/DT
victory/NN
speech/NN
in/IN
Chicago/NNP
```

The word ”gave” is recognized as a past tense verb, ”a” as a determiner, ”victory” and ”speech” as a nouns (common, singular or mass), ”Chicago” as noun (proper, singular), ”in” as a preposition or conjunction (subordinating), and ”Barack” and ”Obama” are tagged as a noun (proper, singular).

Named Entity Tagging

A named entity recognition (NER) system is able to identify a word or a sequence of words that form a proper name like ”Barack Obama”, ”Chicago”, or ”Nokia” and tag it with semantic class information. These classes include names of people, organizations and places. For example, an organization "Nokia" and a place "Nokia" can be disambiguated and tagged. For the example sentence, a NER system tags the phrase: ”Barack Obama” as a person and ”Chicago” as a place.

[4] The examples are produced using Stanford NLP tools (available at: http://nlp.stanford.edu/) that use Penn Treebank II tags.

Parsing

Parsing operates on a sentence level and maps the phrasal elements of a sentence into a structure showing the relationships between them. For the example sentence, a parser may produce the following parse tree [5]:

```
(ROOT
  (S
    (NP (NNP Barack_Obama))
    (VP (VBD gave)
      (NP (DT a) (NN victory) (NN speech))
      (PP (IN in)
        (NP (NNP Chicago))))))
```

The parse tree determines the relations between the phrasal elements of the sentence. Even more detailed description is obtained using a dependency parser [30]:

```
nsubj(gave-2, Barack_Obama-1)
det(speech-5, a-3)
nn(speech-5, victory-4)
dobj(gave-2, speech-5)
prep_in(gave-2, Chicago-7)
```

The dependency parser is able to determine the grammatical functions between the words in the sentence. For example, ”Barack_Obama” is tagged as a nominal subject of the sentence and the governor for the nominal subject here is the verb ”gave”. The word ”speech” is the direct object of the verb.

[5] Note that the named entity ”Barack Obama” recognized using a NER system is now fed for the system as ”Barack_Obama”

Discourse Reference

Dependency parsing and named entity recognition reveal the syntactic and simple semantic structures of the sentence. However, these techniques do not seize the problems related to discourse references. Well known problems here are anaphora and co-reference resolution [97]. Anaphora is an instance of an expression referring to another. For example, in sentences "Obama was in Chicago." and "He gave a speech.", the pronoun "He" refers to the named entity "Obama". The resolution may also take place in the form of co-reference. For example, in the sentences "Obama was in Chicago." and "The president gave a speech.", the noun "president" refers to the named entity "Obama".

Output Generation

Output generation of IE means classifying words or word chunks, such as named entities, into values of properties of a pre-defined template, such as a metadata schema. While the referred techniques can be used to comprise more accurate indexing terms, such as proper names or temporal expressions, they do not reveal the semantics of the sentences. The latest research direction in determining such roles automatically in text is called Semantic Role Labeling (SRL) [40, 42].

For example, in the sentence "Barack Obama gave a victory speech in Chicago", using a dependency parser, it is possible to determine that ”Barack_Obama” is the nominal subject of the sentence, but this does not determine that ”Barack_Obama” is the *agent* of the sentence. In other words, that it was ”Barack_Obama” *who* gave the speech and not, for example, ”speech” *who* gave ”Barack_Obama”.

SRL is based on the assumption that syntactic features of a sentence acquired using PoS tagging, NER tagging, and parsing can be used to predict the semantic roles of the word chunks in the sentence. In the example sentence, the information that ”Barack Obama” is a named entity, it appears before the verb, is a nominal subject of the verb and the sentence has an active voice, could be used to predict that ”Barack Obama” is the agent of giving the speech.

2.6 User Profiling

Personalization in the context of recommender systems can be defined as the process of customizing the content to the specific and individual needs of each user [33]. The process of the creation of an information base that contains the preferences, characteristics, and activities of users is called user profiling [33].

User profiling can be knowledge-based or behavior-based [91]. Knowledge-based approaches engineer static models of users, for example, based on demographic categories, and match users to the closest model. The user profiling approach used by most recommender systems is behavior-based, which uses the user’s behavior as a model and behavioral logging or explicit user ratings are employed to obtain the necessary data [93]. A behavior-based approach that takes advantage of content descriptions of objects is called a content-based approach.

A content-based approach assumes the existence of content descriptions for each object and builds a model of user preferences using these content descriptions. The profiling can also be based on the rating data obtained from the user [4]. If the content of the object contains information, for example, about the target audience of the objects, the approach can be extended to knowledge-based approach. This thesis concentrates on methods that are content-based. This means that the recommendation methods do not make use of information about other users, which is the case in collaborative filtering systems. The content-based techniques can be categorized into three main categories: vector space approach, classification approach, and ontology-based approach.

Vector Space Approaches

In the vector space approach, both objects and user profiles are represented as vectors of weighted features according to the vector space model. Based on what the user has found relevant in the past, the profile vector can be modified and the recommendation task can still be based on comparing the similarity between the vectors. A well known technique to perform this operation is Rocchio [6], where the features appearing in the objects indicated relevant by the user during the retrieval process are up-weighted in the profile vector, and the features not appearing in the objects indicated relevant are down-weighted.

Classification Approaches

If a user has determined some relevant and non-relevant objects, it is possible to build a classifier, rather than re-weighting and expanding the query or the user profile. Here, the problem is turned into a classification problem where objects can be classified as relevant or as non-relevant.

Classification of objects according to user preferences can be done using a variety of machine learning techniques, such as k-Nearest Neighbor (kNN) [93, 105], decision trees [105], support vector machines [37], or naive Bayes classifiers [99, 105]. Probabilistic models are also suitable for more complex scenarios than just predicting correct objects based on user relevance feedback. This is because the probabilistic framework provides a convenient and principled way to include various kinds of prior information into the model [87].

Ontology-based Personalization

Ontology-based user profiling approaches are designed to reduce the semantic gap between the low-level features extracted from documents, such as bag of words, and the more abstract, conceptual views of user interests [43]. For example, in the Foxtrot and Quickstep systems [93] interest profiles are represented using concepts from the ontologies, allowing other interests to be inferred that go beyond those directly available in the content descriptions. The profiles are represented using concept vectors, and a kNN classifier is used to determine the relevant objects.

The current systems utilize ontologies by extending the bag of words model [38, 93]. This can reduce the gap between the concepts used in the ontology-based structured annotation and the concepts in the user profiles. Domain ontologies are used to bridge the concepts in the profiles and the objects by using subsumption hierarchies to generalize the concepts [93]. In [38], such user profiles are generated by analyzing the behavior of the user, specifically the content, length, and time spent on each Web page they visit.

The current methods utilize ontologies to improve performance of user profiling, but they do not consider more complex annotation structures than simple concept sets and hierarchies. In this thesis, the user profiles are modeled as vectors of triples. This allows the user model to represent features that occur in ontology-based structured annotations.

2.7 Evaluation of Recommender Systems

Recommender systems can be implemented using various techniques and methods that can together approximate relevant objects and present them for the user. But how do we know whether these systems are beneficial for the user?

Relevance

The key utility measure in evaluating information retrieval or filtering systems, that content-based recommender systems are, is user satisfaction [66]. In this context, user satisfaction can be measured using relevance. Traditionally, relevance is defined as system relevance, that is the relation between a query and information objects retrieved, or failed to be retrieved, by a given method.

However, relevance can also cover topical, cognitive, situational, or motivational factors [120]. For example, topical relevance measures the relation between the subject or the topic expressed in a query, and a topic or a subject covered by the retrieved objects. Cognitive relevance takes into account the state of knowledge and cognitive information need of a user. Situational relevance considers also the task, or problem at hand, and the motivational relevance the intents and goals of the user.

Evaluation Settings

Reliable evaluation, that would take into account all types of relevance, can be expensive and difficult to conduct. Therefore, it seems plausible to assume that system relevance and topical relevance are the most important factors affecting the recommender system quality [87].

Evaluating system relevance and topical relevance is referred as retrieval performance evaluation [6], where a relevance rating determined by the method is compared to a relevance assessment by human annotators. In many cases, information retrieval systems are evaluated with laboratory experiments, where retrieval performance evaluation is carried out with standard benchmark datasets [52]. It has been suggested that the actual evaluation of recommender systems should be based on a so called ”find good objects” task [52]. This task focuses on suggesting specific objects to their users, providing users with a ranked list of the recommended objects, along with a rating that predicts how much the users would like them. This is the core recommendation task. In many systems, a fixed amount of the highest rated recommendations are shown [52].

The ”find good objects” task captures an important aspect of topical and motivational relevance in real life systems. It has been noticed that many of the users using real life recommender systems find it pleasant to just browse [52]. Whether one models this activity as learning or as entertainment, it seems that recommender systems are also widely used in other tasks than searching for pre-known objects.

In such cases, determining retrieval performance using accuracy measures against a benchmark dataset may be misleading. This is because the user opinions of good objects in benchmark datasets can be based on, for example, purchase decisions. Therefore, interesting objects that the users receive through browsing, but are not willing to buy, are not judged relevant in the benchmark dataset [52]. If laboratory experiments are used, it is important that the tasks the method is designed to support are similar to the tasks supported by the system from which the relevance assessments of humans are collected [52, 93]. The system should not be benchmarked with a dataset collected for another intended use case or from another domain.

Tasks based on real world systems also touch another important aspect of recommender systems related to relevance: the user interface and visualization. For example, a recommender system can generate explanations that are important especially on complex domains, or enable user interaction to improve the usability of a system. These functionalities can strongly influence user satisfaction, but are not measured in the basic retrieval performance paradigm [87]. These aspects can be evaluated in task-based experiments, where users use the system in the intended usage context [66].

Retrieval Performance Evaluation

If a benchmark dataset for the domain under interest is not available, user testing needs to be carried out to ensure a valid evaluation setting [52]. This ensures that the dataset is representative for the domain and for the intended use case. In case the evaluation is performed using the ”find good objects” task, the retrieval performance evaluation of the systems and its components can be based on the accuracy metrics of IR.

Accuracy metrics measure how close the relevance ratings predicted by a method are to the relevance assessments by users. The relevance assessments by users is also called a gold standard. Commonly used accuracy metrics are recall, precision and accuracy [87]. Recall RE is the fraction of the relevant objects which has been retrieved and precision P is the fraction of the retrieved objects which is relevant.

Table 2.1: Contingency table for retrieval performance measures.

	Relevant	Non-relevant
Retrieved	true positives (tp)	false positives (fp)
Not retrieved	false negatives (fn)	true negatives (tn)

These measures can be defined using the contingency table 2.1. Now one can write:

$$RE = \frac{tp}{(tp + fn)} \tag{2.18}$$

$$P = \frac{tp}{(tp + fp)} \tag{2.19}$$

Accuracy A can be defined as:

$$A = \frac{(tp + tn)}{(tp + fp + fn + tn)}. \tag{2.20}$$

Precision and recall are vulnerable measures because often when precision increases, recall decreases and vice versa. Therefore, a single measure that can be used to estimate a balanced performance in terms of precision and recall can be useful. A single measure that trades off precision versus recall is the F-measure. The traditional F-measure or balanced F-score (F_1 score) is the harmonic mean of precision and recall:

$$F_1 = \frac{(2PRE)}{(P + RE)}. \tag{2.21}$$

As can be observed, precision and recall operate on a binary relevance assessment scale. Generalized precision and generalized recall, originally defined by Kekäläinen and Järvelin [70], are measures that take into account graded relevance assessments. Due to graded relevance assessments the distance between the relevance assessment by human annotators and the relevance rating given by the method are not necessarily on a binary scale, but are measured as an interval.

Ehring and Euzenat have defined the measure in more general manner in the scope of ontology matching [32], where the generalized precision and recall are calculated based on an overlap function between a gold standard and the result given by the method. In [32], generalized precision gP and generalized recall gR are defined as:

$$gP(A, G) = \frac{overlap(A, G)}{|A|}, \tag{2.22}$$

$$gR(A, G) = \frac{overlap(A, G)}{|G|}, \tag{2.23}$$

where G is the set of objects in the gold standard and A is the set of objects given by the method.

The overlap function should return the value 1 if the score in the gold standard and the score given by the method are the same [32]. In this way, the precision is 1 as long as there is no difference in the score in the gold standard and in the score given by the method. The overlap function can now be defined as the difference between the grade given by the gold standard $G(O_i)$ and the grade given by the method $A(O_i)$ for each object O_i as: $1 - |G(O_i) - A(O_i)|$. Intuitively, the generalized precision measures the proportion of error between the gold standard and the method with respect to the number of objects retrieved, and the generalized recall measures the proportion of error between the gold standard and the method with respect to the all objects in the gold standard. If all and only all of the objects are retrieved or judged by the method and the gold standard, the generalized precision and generalized recall becomes equal and can be called generalized accuracy gA. This is typical for a classification task, where a classifier is used to predict the relevance rating for objects in the gold standard.

These measures require a relevance assessment that can be assessed by human annotators. Several human annotators can be used to ensure an unbiased assessment. The agreement among annotators, called inter-annotator agreement, can then be measured. This can be done using Kappa statistics [28, 22]. The statistical significance of the retrieval performance can be ensured using significance tests (see [56] for an overview).

Evaluation of Subtasks

The different components of a recommender system put together can lead to a working system with a good overall retrieval performance. However, this does not tell much about the performance of the individual components, such as information extraction or retrieval components. For example, what was the role of information extraction method, did it perform better than a simpler one, or did the query expansion strategy chosen perform better than another technique. This suggests that ultimately, the performance of the system should be evaluated as a whole, but also with emphasis on the individual components.

The evaluation of all components is possible in one run, but in practice can be tricky and complex [123]. Therefore, the components are typically evaluated individually, and their performance compared to alternative methods and human relevance assessment.

Beyond Retrieval Performance

A recommender system should avoid over-specialization, offer relevant objects, and satisfy the information need of the user in the intended use case. However, no systematic attempts to measure all these aspects in a laboratory experiment exist [52].

A task-based evaluation setting can be used to measure the performance of the system in the intended use case. In such a setting users perform tasks that are assumed in the intended use cases. These can vary from known item searching to muddled topic or content searching [66]. In a known item task users try to find a specific object based on known features, such as the creator of the object. The other extreme is the muddled topic or content searching task, where users explore contents or subject matters in novel information environments to solve vaguely defined work tasks [66]. The former can be evaluated using formal studies, where accuracy measures can be used. The latter requires observing the users, conducting interviews, or using post-test questionnaires.

In addition to retrieval performance and task-based evaluation, the system performance should be kept in mind. If a system performs with high accuracy and the user finds the system satisfying according to a task, but the computational cost of obtaining this is beyond the level that the user or the system provider is ready to accept in a real life setting, the system performance is low and may lead to low acceptance of the system. Therefore, a constructive approach needs to be taken and real world systems have to be created to ensure that the complete systems actually can be built based on the suggested components, and that they serve the users with acceptable system performance.

3 Overview of Research

Despite all of the advances on methods supporting recommender systems and success of practical applications, the current generation of content-based recommender systems still requires improvements to make recommending methods more effective and applicable to a broader range of domains and applications. This thesis tackles the problems of recommender systems related to automatic content analysis, bridging heterogeneous content, content retrieval, and the performance of the methods as part of real life recommender systems. In the following sections, main research questions are defined and the research approach is discussed. The research reported in this thesis builds on top of a work carried out in earlier projects. Therefore, the research context is also discussed. Further, an overview of the developed methods is given, and the results discussed.

3.1 Research Questions

First, the problem of content analysis in the scope of information extraction is investigated. The focus is in automatic content analysis that aims to automatically produce ontology-based structured annotation from textual annotation. The first research question is:

1. *How can structured ontology-based annotations be produced automatically from textual annotations?*

Ontology-based structured annotation typically originates from different sources and corresponds to different kinds of metadata schemas. This causes semantic heterogeneity and sparsity in the vector space model. The second research question is:

2. *How can semantic interoperability between heterogeneous structured ontology-based annotations be obtained?*

Semantic interoperability enables integration of heterogeneous structured annotations, and ontologies provide background knowledge that can be used to further derive information about the annotations. However, developing methods that are able to determine ratings for structured ontology-based annotations is an open problem. The third research question is:

3. *How can content retrieval in ontology-based recommender systems be enabled?*

The first three research questions deal with two issues. First, enabling to build the necessary components and systems, and second, enabling acceptable system performance of the components. Even if the methods would perform well in terms of retrieval performance and would enable to build components supporting content analysis, enabling interoperability, and content retrieval, the methods should also be applicable to recommender applications that users find satisfactory. The fourth research question is:

4. *Do users find the recommender systems utilizing the developed methods satisfying?*

3.2 Approach

Four different kinds of research areas were studied: content analysis, content heterogeneity, content retrieval, and applicability and performance of the developed methods in real life systems. Suitable benchmark datasets were not available neither for the domain, nor for studying the research problems defined in this thesis. Therefore appropriate research methods and datasets were selected separately for each study. In particular, three different methodological approaches were used: user study, case study, and constructive approach. The focus of the thesis is on the accuracy of the components of the recommender systems. However, solutions are also sought to ensure relevance of the methods as part of real life systems and their intended use-cases.

The research questions are studied in the digital cultural heritage domain. The analyzes are limited to descriptions available in digital format and accessible through knowledge systems.

The annotations of digital cultural heritage objects often concentrate on the manufacturing and preservation of the objects, such as who created the object, where and when it was created, and in case of tangible objects, where it is currently located. The annotations also document the subject matter of the objects, such as what is the style or genre that the object represents and what the object depicts.

The research on semantic relatedness approximation methods was performed in the news domain. The news domain is in many ways similar to cultural heritage. It involves descriptions of people, places and objects and real world events where the objects, people, places, and other entities participate.

There are restrictions and possibilities that these domains entail. Cultural heritage is a knowledge-rich domain, in which large bodies of structured background knowledge are available in form of vocabularies and ontologies, and experts agree on the main concepts and relations. The news domain is broader, and specific background-knowledge does not necessarily exist. This is due to the fact that the news domain documents current events, for which the participating entities and their relations are not necessarily documented in ontologies or vocabularies.

3.3 Research Context

The research reported in this thesis has been conducted as a part of three large research projects. The FinnONTO projects[6] have been creating a basis for a national metadata, ontology, and ontology service framework in Finland, and demonstrating its usefulness in practical applications. The research in the project has been carried out by more than 30 people in different stages and it has resulted in both large knowledge bases and software that have been utilized in the research reported in this thesis. First, the YSO ontology, its extensions, and content annotated using these ontologies have enabled access to a unique knowledge-base of cultural heritage data. Development and experimentation of ontology-based methods, reported in articles I,II,V would not have been possible without such knowledge base. The project also produced a software framework to index and process RDF(S) data. The software framework was used and further developed in the research reported in this thesis.

The research reported in article III was conducted under the MultimediaN e-Culture project[7]. Ontologies, content and APIs developed in the MultimediaN e-Culture project were used in the research reported in this thesis. The SMARTMUSEUM project[8] was a EU FP7 funded project with partners from a number of European countries. The recommender system back-end was developed based on the FinnONTO software framework. In research reported in article IV, annotated data was provided by the Heritage Malta and the Institute and Museum of the History of Science in Florence, Italy. The user interface development was conducted by INRIA in France, and WebGate JSC in Bulgaria, and the user profiling server was implemented by Apprise Ltd. in Estonia.

[6]https://www.seco.tkk.fi/projects/finnonto/
[7]http://e-culture.multimedian.nl/
[8]http://www.smartmuseum.eu/

3.4 Content Analysis

In this thesis, an automatic annotation method was developed [III]. The method is able to automatically produce structured ontology-based annotation from textual annotation. The state of the art research is able to determine semantic roles for word chunks of a natural language sentences in a benchmark corpus using SRL [40]. The focus of the method developed is in semantic role labeling of real life texts with a goal to produce ontology-based structured annotation, where the target template conforms to a metadata schema. The state of the art methods were extended with ontology-based features and compared to the state of the art techniques and human performance in the same task. In addition, the effect of using ontologies as background knowledge for the method was measured.

Approach

The developed method is based on semantic role labeling [40], where the syntactic features of a sentence are used to predict the role of each of the constituents of the sentence. The developed method extends the current state of the art by using ontologies as background knowledge and considering metadata schemas as the target templates.

The overall architecture of the approach is presented in Figure 3.1. It consists of three phases: (1) linguistic analysis, (2) concept identification, and (3) role identification. The linguistic analysis is first performed for a sentence in the textual description. The resulting syntactic features are then used to perform the concept identification. Finally, the role identification is performed based on both the linguistic analysis and the concept identification. The purpose of the concept identification phase is to determine the concepts that have correspondences in the ontologies and are therefore candidates for annotation. The purpose of the role identification phase is to determine the semantic role, if any, that these concepts play in the annotation. The exact description of the developed method and the target metadata schema are reported in the article III.

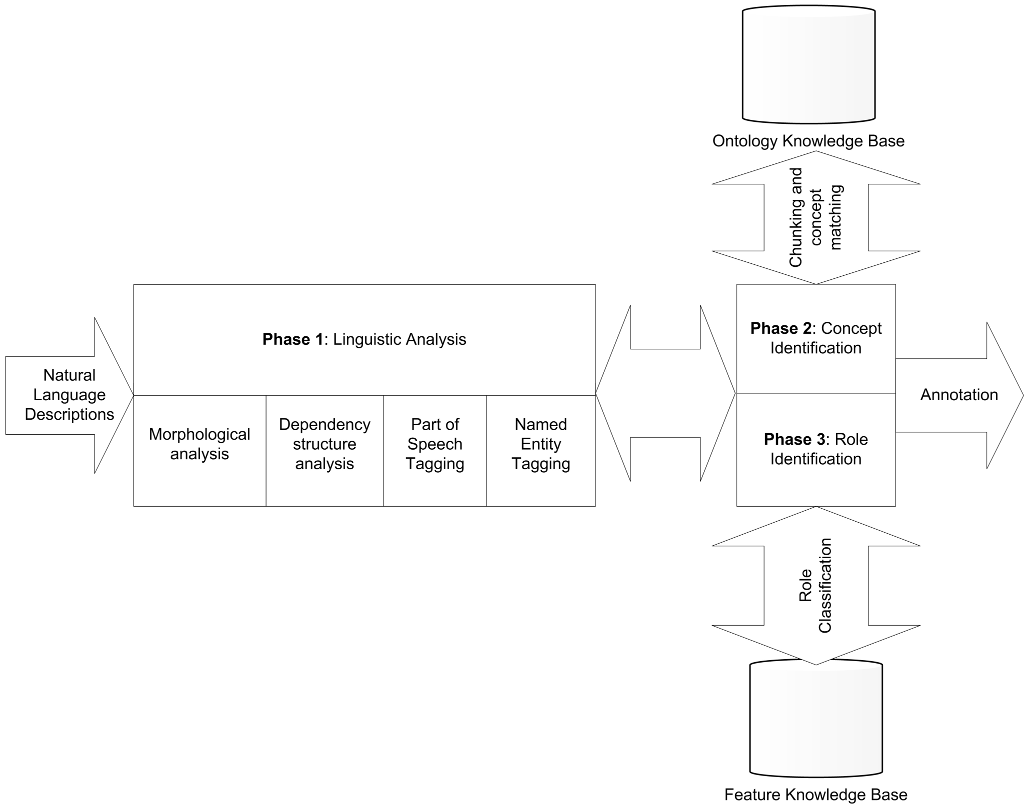

Figure 3.1: Architecture of the content analysis system.

Research Methods and Dataset

The developed method was evaluated through a user study. A gold standard dataset was acquired in a user study, in which fourteen human annotators participated. Retrieval performance was measured using precision, recall, accuracy and F_1 measure. The developed method was compared to a baseline method and human performance in the same task. Inter-annotator agreement was measured using Cohen's Kappa.

The dataset consists of textual annotations of 750 masterpieces of the Rijksmuseum Amsterdam. The structured ontology-based annotation of the textual descriptions of 250 objects in the dataset was done in the user study using four ontologies: AAT, TGN, ULAN and WordNet. The annotation has been performed using a VRA (Visual Resources Association)[9] specialization of the DC metadata schema tailored to the needs of artwork annotation. The Art and Architecture Thesaurus (AAT) is a structured vocabulary of around 34,000 concepts, including 131,000 terms, descriptions, and other information relating to fine art, architecture, decorative arts, archival materials, and material culture. The Getty Thesaurus of Geographic Names (TGN) is a structured vocabulary containing around 912,000 records, including 1,1 million names, place types, coordinates, and descriptive notes, focusing on places important for the study of art and architecture. The Union List of Artist Names (ULAN) is a structured vocabulary containing around 120,000 records, including 293,000 names and biographical and bibliographic information about artists and architects, including a wealth of variant names, pseudonyms, and language variants. WordNet is a general lexical database in which nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. WordNet also provides relations for hyponymy, meronymy and troponymy.

Evaluation

The developed method with the ontology-based features achieved an accuracy of 0.61 (Cohen's Kappa 0.54) and the baseline method, that used only syntactic and lexical features, achieved an accuracy of 0.58 (Cohen's Kappa 0.49). The difference between the developed method and the baseline is statistically significant $(p < 0.01)$. The human annotators achieved an accuracy of 0.65 (Cohen's Kappa 0.58). The overall F_1 measure of the developed method compared to the F_1 measure of the baseline method was statistically significant $(p < 0.05)$. Cohen's Kappa shows moderate to substantial agreement of human annotators. The details of the experiments and results achieved for each metadata schema role are available in the article III.

[9] http://www.vraweb.org/resources/datastandards/vracore3/

3.5 Content Heterogeneity

In this thesis, an event-based method was developed to enable interoperability of heterogeneous annotations [I]. An event-based knowledge representation has been argued to be suitable to describe cultural heritage content [31]. Cultural heritage content is often described as narratives that consist of events where different objects participate; who did, what, where, and when? In [137], users are argued to use the systems in an event-centric way meaning that users organize their memories as events that they have experienced and use such patterns when accessing information. The representation of events in a way that interoperability between data would be achieved and retrieval of the content in applications would benefit from the representations is a central topic of this thesis.

The method developed in this thesis is based on an idea to reduce content heterogeneity by making the knowledge embedded in the metadata schema structures interoperable and explicit by transforming the schemas into a shared, event-based representation of knowledge. The method explicates the knowledge by using a set of thematic roles [128] and domain ontology. In this way, the ontology that is used to describe the conceptualization of the domain can be reused and only the schemas that are typically specific for different content types need to be transformed.

Approach

Our event-based schema introduces relations enabling representation of the original metadata as events with associated thematic roles and quality roles, an idea proposed in the fields of knowledge representation, natural language processing, and discourse modeling [7, 128].

A distinction is made between *metadata schema*, *domain ontology* and *event-based metadata* conforming to an *event-based knowledge representation schema*. The event-based knowledge representation schema specifies a way to represent heterogeneous metadata schemas using domain ontologies. The metadata is represented by instantiating domain ontology concepts and by assigning relations between the instances in terms of the event-based knowledge representation schema.

The method for mapping the metadata schemas to the event-based knowledge representation is based on a classification of the relations of the metadata schema according to meta-properties [48, 88]. Based on the meta-properties a set of rules can be written using a logic programming language. The annotations that are instantiations of metadata schemas are then transformed to the event-based knowledge representation according to the rules. Definitions of the meta-properties, the classification and the mapping principles, the set of thematic roles, and the resulting rules used are reported in the article I.

Research Methods and Dataset

The method was evaluated as a case study, where metadata schemas were analyzed. Three different metadata schemas and the corresponding metadata were used in the study: descriptions of artifacts conforming to the DC like metadata schema of the MUSEUMFINLAND system [61], descriptions of paintings conforming to the CIDOC-CRM used in the Finnish National Gallery, and descriptions of artists conforming to the ULAN schema [130]. The General Finnish Ontology YSO [65] was used as the domain ontology.

Evaluation

The event-based knowledge representation was found to lead to the following benefits. First, semantic interoperability of syntactically different schemas can be obtained by defining the meaning of metadata schemas in terms of the underlying domain ontology concepts. This enables the usage of the transitive subsumption hierarchies of the domain ontology in reasoning. Second, it is possible to exploit additional semantic reasoning by explicating the hidden implicit semantics of metadata schemas. This is achieved by more explicit descriptions of the relational roles in terms of domain ontologies. Third, the event-based knowledge representation reduces the number of different properties to be dealt with in the reasoning phase. Fourth, the problem of aligning different metadata schemas onto each other becomes easier by using a a single knowledge representation model. The number of pairwise mappings between n schemas is $O(n \cdot (n-1)/2)$, but there are only $O(n)$ mappings between the schemas and the event-based knowledge representation model.

The event-based knowledge representation schema was able to represent all of the needed implicit metadata. However, some difficulties were encountered when using the method. Some of the relations referred to local domain ontology resources that had to be mapped separately. For example, the *ulan:gender* relation in the ULAN dataset referred to *ulan:female* or *ulan:male*, that were mapped to the corresponding concepts in the domain ontology.

Another problem was how to enrich the metadata with new thematic roles. For example, the content descriptions for the subject matter of the objects contained values such as *yso:horse*, *yso:ride*, and *yso:man* without relations to each other. Thematic roles can easily be resolved by a human annotator, e.g. that a man rides a horse, and not that a horse rides a man. However, selecting the fillers of the roles often requires tacit human knowledge and is difficult for fully automated methods. The detailed description of the study is available in article I.

3.6 Content Retrieval

While Boolean retrieval models may be manageable and sufficient for small knowledge bases, they do not scale properly for large object repositories where searches typically return hundreds or thousands results [23]. Boolean search does not provide clear ranking criteria, without which the search system may become useless if the search space is large.

In this thesis, VSM was utilized to enable the ranking of ontology-based structured annotations in content retrieval. Two variations of a retrieval method were developed using VSM. First, a method for the CULTURESAMPO portal [II], and second, a method for the SMARTMUSEUM system [IV]. In the SMARTMUSEUM system a user profiling system was built [IV]. This brought up an over-specialization problem. To tackle the problem, a result clustering method was incorporated in the VSM retrieval model. This enabled fast ranking computation to find the highest scored objects and enabled non-over-specialized view of the data for the users. The clustering could also be performed on-line, because the objects were ranked and only the highest scored objects were required to be clustered. Finally, the ontology-based retrieval methods were found to lead to accurate recommendations. However, it is debatable whether only the subsumption reasoning is enough for the query expansion. Therefore, the content retrieval method was extended with semantic relatedness approximation [paper:Paper5] and different approximation techniques were compared to find the best query expansion strategy [V]. This section will present the VSM adapted to ontology-based structured annotations and the extensions developed to improve the performance of the method.

Vector Space Model for Triple Space

Assuming that the annotations are represented as structured ontology-based annotations that are described using triples, it is possible to define a Vector Space Model for triple space (VSM-TS). In VSM-TS each object O is now represented using the annotation A_O, i.e. a set of triples. Based on the definition, it is possible to define both of the vectors W_U and W_{O_j} as vectors in a triple space $W = (t_1, ..., t_k)$. The features in the vectors are the triples that are present in the user profile and in the annotation of the object respectively.

Two different approaches were explored to determine a relevance rating in VSM-TS. It is possible to use reasoning when computing the scoring function. In this case indexing is required only for the original triples in the annotations A_O, and the computation is performed in the retrieval phase. Another option is to perform the reasoning in the indexing phase and store the reasoned annotations A'_O into the index. In this case, standard scoring functions, such as cosine similarity, and computation can be used in the retrieval phase.

In this thesis, both of the approaches were explored. The former was implemented as part of the recommender system in CULTURESAMPO portal [II], where a scoring function and tf-idf based weighting scheme that weighted RDF(S) triples were developed. A disadvantage of this approach was found to be that measuring the similarity of the vectors becomes more computationally complex, because the relevance between the triples must be computed for each feature in the retrieval phase.

However, because the mapping from the original triples to the deductive closure is linear, it is possible to directly perform the mapping of the triples in the indexing phase and to instantiate the feature vectors in the VSM-TS directly using the reasoned triples A'_O. In this case it is possible to use a standard weighting scheme and scoring methods. In the SMARTMUSEUM system [IV], these were implemented using Apache Lucene's practical score function (Equation 2.10).

In [II], a weighting scheme was used to weight the triples by adapting tf-idf weighting for triple space. Classes and instances were weighted based on their occurrence in an individual annotation and in the whole knowledge base to adapt the idf. The rarer the triples that match are, the less they contribute to the total score. For example, objects that originate from the same large collection are matched based on this information, but because there is a large number of other objects from the same collection, the weight for this particular feature is low. On the other hand, triples describing a rare subject matter receive higher weight and lead to a higher score. The tf effect can be achieved through reasoning. For example, in case an object is annotated with triples describing a subject matter with concepts of different animals, say elephants, lions and tigers, it receives higher tf for a concept animals because all of the three triples match to the concept animals through subsumption reasoning. On the other hand, the idf for the concept animals is lower than for the concept elephant, lion or tiger. The tf can also be observed directly from the annotation. For example, in the case of annotation resulting from information extraction process, some triples may have many occurrences in an individual annotation.

In [IV], the objects are directly indexed with reasoned annotations. In this case, separate weighting scheme is not needed, because each triple represents a feature in the vector space and standard tf-idf and cosine similarity can be used to compute the score function.

Research Methods and Dataset

The method described in [II] was evaluated through a user study in which seven voluntary media technology students and faculty members from the Helsinki University of Technology participated. The participants had experience with recommender systems, but they were not experts in the cultural heritage domain. The dataset of the CULTURESAMPO portal[10] was used in the experiments. It contained structured ontology-based annotations of three types of objects: images of museum objects, images of photographs and images of paintings. The objects had been annotated by domain experts in Finnish museums. A transformation to RDF was performed and the values of the metadata schema elements of the annotations were mapped to YSO. Ambiguous references were manually disambiguated to refer to correct concepts and the annotations were transformed to the event-based knowledge representation.

[10]The version of the dataset in year 2007.

Seven objects were randomly selected as source objects. No weights were available for the triples in the profile vector, that consisted of the triples from a source object. The triples in the profile were weighted on a binary scale. The recommendations were then computed for each source object. The computing was performed against a knowledge base that contained structured ontology-based annotations of nearly 10.000 objects. The five top-ranked recommendations for each source object given by the method were considered the higher relevance group. The other five, the lower relevance group, were a sample of the lower half of the objects based on the median rating. This resulted to a sample of 70 objects.

The task of the users was to classify the objects in the sample to belong under a certain source object based on what they would like to be recommended when examining the source object. The users also had a possibility to classify objects non-relevant to any of the source objects. After the initial classification, the users were asked to further classify the objects under each of the source objects to higher and lower relevance group.

The purpose of the study was to test the hypothesis that there is a difference in the retrieval performance between higher and lower relevance groups. In other words, if the objects rated high by the method were more often relevant than the objects rated low. This is intuitive, because typically recommender systems only show the k-top objects for the user. In addition, an accuracy that the method achieved for the higher and lower relevance group was measured using the user assessments as the gold standard.

In the article II, precision, recall, inter-annotator agreement, and statistical significance of the difference between the lower relevance and higher relevance groups, were not reported. These were calculated later and are reported here.

Evaluation

For the higher relevance group the precision was 0.91 and the recall was 0.82. For the lower relevance group the precision was 0.64 and the recall was 0.72. The χ^2 test showed that the difference between the groups was statistically significant ($p < 0.05$). Cohen's Kappa showed substantial agreement between the annotators (Kappa = 0.67).

It is notable that the lower relevance group was sampled below the median of the rating given by the score function and still received relatively high precision and recall. This indicates that some objects with lower scores were also found relevant. The users were interviewed after the user study. Five of the seven users mentioned the difficulty of deciding which were the most important dimensions to which the classification should be based on. This advocates the need of methods that can avoid over-specialization and allow different viewpoints to the data.

Semantic Relatedness Approximation

Subsumption inference can be used to deduce additional statements about the objects. For example, in case an object is manufactured in Montmartre, it can be inferred as been manufactured in Paris, France, Europe, and Earth. However, also expansion to other concepts than to the ones explicitly stated or that can be inferred through logical reasoning could be useful. For example, if a user is interested in objects related to schools, the user might also be interested in objects related to teaching. Such a relation is not necessarily explicitly stated in the ontology and therefore should be acquired by other means. Acquiring relations can be performed by approximating relatedness of the concepts in the ontologies by using methods that make use of the ontology structures or external information sources.

In this thesis, the performance of three well known semantic relatedness approximation methods, the Wu-Palmer measure, the Leacock-Chodorow measure and LSA, were compared to find out the best performing methods. The detailed definitions and comparison of semantic relatedness approximation methods can be found in the article V.

The Wu-Palmer measure was found to perform best with an appropriate cut-off value and was implemented in the SMARTMUSEUM system. The approximation was utilized in content retrieval by expanding the user profile vector W_U with additional triples. Each triple W_{U_i} can be expanded into new triples based on the relatedness value determined by the relatedness approximation method. The relatedness value is determined for each resource of the triple, and all triple combinations of resources that have a relatedness value over a threshold value are constructed and added to the profile vector. The threshold for relatedness can be obtained, for example, from a user interface control [IV].

Research Methods and Dataset

A user study was conducted to measure the performance of three different semantic relatedness approximation methods: LSA, Leacock-Chodorow and Wu-Palmer. Fifteen users participated (Kappa = 0.68).

The Helsingin Sanomat News Corpus was used as the dataset for the study. The dataset consists of 883 randomly selected articles from the Finnish newspaper ”Helsingin Sanomat”. Each article consists of the heading and the article body. YSO was used as the domain ontology. A sample of 3168 concept pairs appearing in the intersection of YSO and the articles was annotated by the users as relevant or non-relevant. This set of concept pairs was used as a gold standard to evaluate the performance of the methods.

Evaluation

The overall performance of the corpus-based method Latent Semantic Analysis (generalized accuracy = 0.84) was found more accurate than the structural measures proposed by Wu and Palmer (generalized accuracy = 0.74), and Leacock and Chodorow (generalized accuracy = 0.51). However, both of the structural measures had substantially better performance than LSA when cut-off values were used. The concept pairs approximated by the best performing structural measure Wu-Palmer and latent semantic analysis show a low level of overlap. LSA is superior in filtering out the non-relevant relations, and is able to find relations in which the structural measures fail. Structural measures show good overall performance even with a low cut-off value. LSA finds relations specific to the corpus, but only a limited number of the relations that are present in the ontology [V]. Such a low level of overlap of LSA and Wu-Palmer measure indicates that the structural measures and corpus-based measures are complementary and a combination of methods should be used to achieve good performance. The results are statistically significant $(p < 0.000001)$.

Result Clustering

The scoring functions are used to determine the rating of the objects given a profile. However, scoring alone is not necessarily the best way to determine the objects to be recommended.

Objects are returned as a ranked list based on the rating given by the score function. While the ranking of the objects is important, to avoid over-specialization, users may also want to receive recommendations from the different viewpoints specified in their user profiles. For example, consider a user profile with three triples each defining an material of an object, say ”brass”, ”copper”, and ”copper alloy”, and one triple defining a type of the object, say ”vase”. In case the recommendations would be obtained directly using the score function in VSM-TS, and assuming that all materials would have approximately the same tf-idf value, all objects that have two of the materials mentioned, say copper and copper alloy, would be ranked higher than any of the objects having a material brass or having a type vase alone. This easily leads to a situation where the top ranked objects appearing in the user interface only consist of very similar objects that are ranked high based on a subset of features in the profile; in the example case, objects with materials copper and copper alloy. However, from the perspective of the user, it could be more interesting to obtain objects also based on other sets of features that are less important based on the scoring function, but still score high based on different features.

In this thesis, the over-specialization problem was approached by using clustering of the objects that were rated high by the scoring function [IV]. The clustering is based on the matching triples collected for each of the top k objects given by the scoring function. The FastICA algorithm was used to perform independent component analysis (ICA) [57][11] for the retrieved objects to find clusters. The clusters were labeled by including the labels of the five most common triples occurring in the cluster excluding triples that occur in all of the clusters. The details of the implementation can be found in the article IV.

Research Methods and Dataset

The clustering method was tested and implemented using the SMARTMUSEUM dataset. The dataset consists of structured ontology-based annotations of 500 museum objects and points of interest from the collections of the Institute and Museum of the History of Science, Florence and Heritage Malta. The objects are structurally annotated using AAT and TGN ontologies. A metadata schema corresponding to DC was used in the annotation. In addition, the metadata schema included properties enabling descriptions of the target group, age group, suggested education, and other demographic properties that can help personalizing the content for a specific user.

[11] The Java implementation of FastICA (http://sourceforge.net/projects/fastica/) is used.

Figure 3.2: Clusters determined for a test user profile.

Demonstrating Example

A formal evaluation to determine the quality of the clustering was not conducted within the scope of this thesis, but the method was initially tested with ten test user profiles defined by domain experts from the Institute and Museum of the History of Science, Florence. Figure 3.2 shows a test web interface of the content-based recommender system of the SMARTMUSEUM system. One of the test user profiles is inserted to the system and two clusters determined for the profile using the clustering method. The test user profile consists of user preferences expressed with four triples. First, objects that are annotated to have a type of instruments. Second, objects are annotated to have a material copper. Third, objects that are annotated to have a material copper alloy, and fourth, objects that are annotated to have the subject astronomy. The method finds two separate clusters. The first cluster marked with dashed line in the Figure 3.2 consists of objects that are instruments and have materials copper and copper alloy. The second cluster consists of objects that are instruments, have the material copper and the subject astronomy. The objects in the second cluster receive lower ranks based on the score function, but can still be relevant for the user in addition to the objects in the former cluster.

The clustering method seems to generalize for the test user profiles. However, the labeling and content of some of the clusters were found to have weaknesses. First, in some cases depending on the query expansion level, the labeling of the clusters generates too general or too few labels. Second, typically the clustering method generates one cluster that contains objects that have very low rank and is mainly based on triples that occur in other clusters, but not in this particular cluster. A revised labeling method and a cutoff for rank values, instead of a fixed number of objects, in a retrieval phase could improve the system performance by gathering only the relevant labels and filtering out objects that are have low rank. In addition, a formal retrieval performance evaluation should be conducted to determine the retrieval performance and identify other possible weaknesses of the method.

User Profiling

So far relevance rating of objects based on two feature vectors, the object vector W_{O_j} and the user profile vector W_U, has been discussed. It has been shown that the object vector can be constructed from text, and harmonized using event-based knowledge representation. In addition, it has been shown that the user profile vector can be expanded based on semantic relatedness approximation, and that over-specialization can be avoided using clustering. However, it has been assumed that the initial features in the user profile are available. In this section, a method to construct the initial user profile vector W_U is discussed.

In the SMARTMUSEUM system, a user profile can be constructed manually by the user. This is performed by inserting triples to the profile directly through a user interface or by user profiling that can be done based on the behavior of a user [IV]. The user can tag an interesting object with an "I like" tag or an "I dislike" tag. The triples in the annotation of the object are added to the user profile with count 1. For the triples that already are in the profile, the count is increased by the number of the times an object where the triple appears is tagged. This results in a list of triples that the user has marked relevant or non-relevant and the count of each triple. The relevant and non-relevant ratings are averaged. Thus, the triple in a profile is a pair of average vote and the number of times the triple has been tagged. It is now possible to calculate the likelihood for the triple in a user profile [IV] and use the most likely triples as a query in the vector space model.

A rationale behind choosing a probabilistically motivated approach over a straightforward vector space approach was that a context aware version of the profile was also developed. In the context-aware version each triple can be conditioned using a context. The proposed method is explained in detail in article IV.

The user profiling was implemented as a practical solution part of the SMARTMUSEUM system, for which user trials were conducted in two museums. The results of these user trials will be discussed later. A retrieval performance evaluation of the approach has not been conducted.

3.7 Applications

The presented methods have been implemented in two recommender systems. The first version of the content retrieval method operating on an event-based knowledge representation was implemented in the CULTURESAMPO portal [I,II] and later extended with query expansion based on semantic relatedness approximation and result clustering, and implemented in the SMARTMUSEUM system [IV].

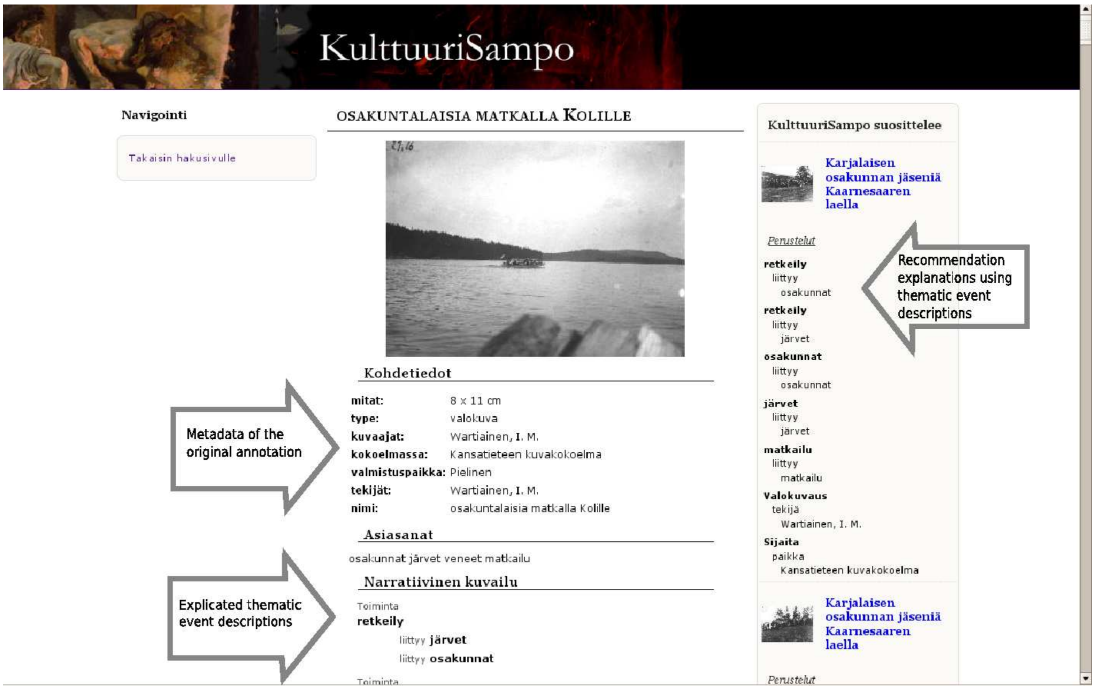

Figure 3.3: A screen capture of an object page of the CULTURESAMPO portal. The results of the recommender system are on the right side of the screen capture.

Recommender System for the Culturesampo portal

CULTURESAMPO is a demonstration application for publishing and accessing cultural heritage contents on the Web. It is based on a metadata infrastructure that relies on the use of ontologies [62, 60]. The system presents new solutions to interoperability problems of dealing with multiple ontologies of different domains, and to problems of integrating multiple metadata schemas and cross-domain content. The system provides search and recommendation functionalities. In addition, the content can be accessed through nine thematic perspectives including map views, a time view, and a story and narrative based access [59]. The methods described in this thesis, were implemented in the first version of the CULTURESAMPO portal[12].

[12]The first version of the CULTURESAMPO portal was accessible on the Web between 2007-2008.

The CULTURESAMPO portal does not store user profiles and the recommendations were determined based on the content of the object being examined by the user. In other words, the user profile consists of only the features in the object that the user is examining.

Figure 3.3 illustrates the user interface of an object page of the CULTURESAMPO portal. It shows a page about a photograph concerning a student union traveling by boat to the Koli mountain in Karelia. On the right side the system gives recommendation links to other content objects with explanations such as “hiking related to a student association” and “traveling related to a student association”. The event-based system gives these links because the image describes a “hiking” event with a “student association” and “lake” in participant roles. The method also gives links to content objects that are “stored” in the same collection, “photographed” by the same person, and so on.

Recommender System for the Smartmuseum System

SMARTMUSEUM is a demonstration system for mobile on-site personalized access to digital cultural heritage. It supports two principal scenarios: inside and outside scenario. In the outside scenario, the system enables recommending points of interest, such as museums and sights while the user is mobile. In the inside scenario, the user indicates visiting a museum and the system recommends objects inside the museum. The system also has other functionalities, such as on-site video streaming, speech synthesis, collaborative filtering based recommendations, and Radio Frequency Identification and Global Positioning System -based object and location identification.

The content-based recommender system of SMARTMUSEUM was implemented by using the presented methods for content retrieval, user profiling, semantic relatedness approximation, and clustering [IV]. In addition, for the outside scenario, bounding-box based geographical search functionalities were implemented to restrict the recommendations to objects located near the user [IV]. Figure 3.4 shows two screen captures from the user

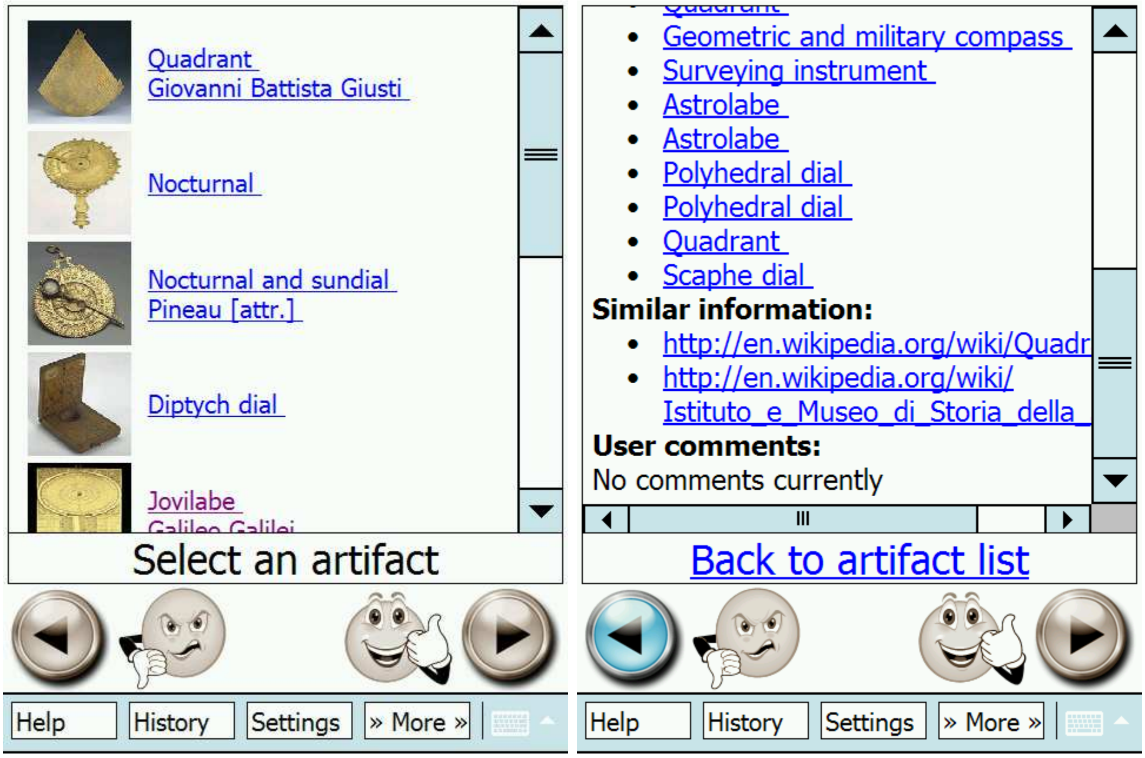

Figure 3.4: A screen capture of two screens of the SMARTMUSEUM system. The screen captures show the user interface of a recommendation list (left) and a page showing links to relevant content based on a selection of an object from the initial list (right).

interface, where a recommendation list (left) and a page showing links to other relevant content (right) are provided for the user in an inside scenario. The recommendations are presented as a flattened list, but are based on scoring, semantic relatedness approximation, and clustering. The user is able to construct and update the user profile by voting on each object. The recommendation method is also used to construct the related objects list by using the features in the user profile and the features in the present object as a query vector. In this case, the features in the current object are temporarily added to the user profile or up-weighted based on the original user profile. This enables recommendations related to the object examined by the user, that can be scored based on the user's profile.

The SMARTMUSEUM system was evaluated in two user trials. The user trials were designed and conducted by the museum staff in two museums: the Fine Arts Museum in Malta, and Institute and Museum of the History of Science in Florence. The user trials and the obtained results are shortly presented here, because the user trials were not reported in the articles that are part of this thesis.

The user trials were organized by the museum staff who assisted the participants in using the system. The participants were first given a 30 minute presentation about the system including instructions how to create a user profile, log into the SMARTMUSEUM system, use the recommender system, and to use the system in one's own mobile phone. Eight Personal Digital Assistant devices were made available for users to try out the system on their own time. A post-test questionnaire based on System Usability Scale (SUS) [16] was handed for the users after the user trial. SUS is a low-cost usability scale that can be used for assessments of systems usability. It does not provide detailed insight to usability because it only has limited number of questions. However, it is a method that is suitable to conduct usability studies in on-site user trials, when controlled experiments are not possible or suitable. All together 24 responses were gathered. Eleven responses in the case of Fine Arts Museum in Malta[13], and thirteen in the case of Institute and Museum of the History of Science in Florence.

Table 3.7 shows the questions and results of the post-test questionnaire[14]. The system was found easy to use and to improve museum experience (95.8%). The system was found easy to learn to use (95.8%) and users believed that they would not need technical support to use the system (75%). A majority of the users thought that the functions were well integrated (58.4%) or were neutral on the subject (33.3%) and very few of the users thought that there were inconsistencies in the system (4.1%). In addition, users believed

[13] Fine Arts Museum in Malta is a part of Heritage Malta

[14] The original questionnaire also had strongly agree and strongly disagree categories as defined in SUS. These were combined into agree and disagree categories.

Table 3.1: Results of the SUS post-test questionnaire of the user trials. A = Agree, N = Neutral, D = Disagree

Question	A(%)	N (%)	D (%)
I thought the system was easy to use and improved my museum experience.	95.8	0	4.2
I think that I would like to use such a system in other museums.	91.7	8.3	0
I found the system unnecessarily complex.	12.5	8.3	79.2
I think that I would need the support of a technical person to be able to use this system.	12.5	12.5	75
I found the various functions in this system were well integrated.	58.4	33.3	8.3
I thought there was too much inconsistency in this system.	4.1	29.2	66.7
I would imagine that most people would learn to use this system very quickly.	58.4	33.3	8.3
I found the system very cumbersome to use.	4.2	12.5	83.3
I felt very confident using the system.	54.2	33.3	12.5
I needed to learn a lot of things before I could get going with this system.	0	4.2	95.8

that they would use the system again in other museums (91.7%). The agreement between the users was moderate to substantial (Kappa = 0.59).

In addition, the users were asked if they had any problems using the system, suggestions for improvements, or general comments on the system. A major suggestion that rose was map support to navigate inside the museums. Another suggestion was that the related objects list (shown on the right side of Figure 3.4) could also show objects from other museums and collections. Users also wanted to see explanations of why the objects were related to the one they were examining, and why they were related to the user profiles of the users. The users also thought that it could be easier if they could plan a tour beforehand using a web interface and retrieve the tour when accessing the museum. Users that were domain experts or representatives of some museum suggested that the system should provide pre-defined stereotypic user profiles that a user could choose from when entering the museum.

4 Conclusions and Discussion

The focus of this thesis is in improvement and development of methods for ontology-based recommender systems and testing them in practical applications. Research has been done in four areas: content analysis, content heterogeneity, content retrieval, and applications. The results are next discussed in the scope of the research questions.

4.1 Research Questions Revisited

The first research question reflects the area of content analysis:

1. *How can structured ontology-based annotations be produced automatically from textual annotations?*

The thesis contributed a method that produces ontology-based structured annotation using information extraction techniques, especially semantic role labeling extended with ontological features [III]. The method was found to perform close to the accuracy that humans achieved in the same task and outperformed the baseline method to which it was compared.

The performance of the method differed in the case of some roles [III]. A possible explanation could be that the sentence context was not enough to make a distinction between the depicted and the factual information. In addition, the ontologies used often did not contain corresponding concepts for some specific roles. For example, person names were often not present in the ontologies.

The experiment was carried out with non-expert annotators in a rather specialized domain. The results show that the concordance of the annotators is relatively low in the case of some roles [III]. This suggests that future research could be carried out to compare the concordance of expert annotators, and subsequently measure the performance of the method when more consistent training data is available.

Recent research in natural language processing and information extraction, such as statistical syntactic parsers and NER systems [36], has enabled advances in computational natural language understanding [40]. However, as shown in this study, our hybrid approach, with both statistical methods and ontologies, results in higher performance. It is important to note that this approach is restricted to domains for which ontologies are available. Previous research in SRL has achieved high accuracy in role identification when using hand-corrected parse trees on artificial datasets [40, 108]. Nevertheless, it has been shown that these techniques generalize to other stricter domains only when appropriate training data is available [109]. This suggests that the performance of both statistical tools used for the linguistic analysis and ontologies are dependent on the domain in which they are applied. Yet, the annotation method proposed in this thesis is based on a feature set that could be applied to other domains similar to cultural heritage, such as the news domain.

The method presented in [III] concentrated mainly on the role identification task and therefore a relatively simple method was used for concept identification. Although high accuracy was obtained in empirical evaluation for role identification, in this study the bias in concept identification was not measured.

The second research question reflects the area of content heterogeneity:

2. *How can semantic interoperability between heterogeneous structured ontology-based annotations be obtained?*

Content heterogeneity was studied and a method that is able to bridge heterogeneous structured data was developed [I]. The proposed method utilized event-based knowledge representation to reduce semantic heterogeneity. The performance of the method was studied as a case study and it was successfully used to harmonize three metadata schemas. Further, the resulting knowledge representation was utilized in the VSM-TS based recommending method [I].

While the case study presented in this thesis confirmed that the event-based knowledge representation schema was able to represent the implicit metadata in the three schemas that were studied, some difficulties were encountered when using the method. Some of the relations referred to local domain ontology resources that had to be mapped separately onto YSO concepts. The ontology matching community has tackled this problem (see [35] for an overview of the state of the art).

Another problem was how to enrich the metadata with new thematic roles. Thematic roles can be resolved by a human annotator because selecting the fillers of the roles often requires tacit human knowledge, but can be difficult for automated methods.

The third research question reflects the area of content retrieval:

3. *How can content retrieval in ontology-based recommender systems be enabled?*

Three methods that utilize the structured knowledge-representation were developed. A method to determine semantic relevance was first introduced [II]. The method achieved high retrieval performance in the user study. Further, a method that simplifies the method was proposed in [II]. Combining it with semantic relatedness approximation and clustering was proposed in [IV]. In addition, measures to determine semantic relatedness in ontologies were compared in [V]. LSA was found to be the most accurate method in general, but the Wu-Palmer measure had superior performance when cut-off values were applied. In closer analysis, the methods were found complementary to each other.

A fair state of the art baseline method does not exist to compare the performance with the developed retrieval method. Because the purpose of the study was not to improve recommendations for objects with text descriptions, but rather it was assumed that heterogeneous annotations that originate from diverse sources should interoperate, a comparison to a method that uses only text-based features was not performed. However, the effect of automatically acquired ontology-based structured annotation to the recommendation of text-based descriptions could be an interesting research prospect. A comparison to semantic vector space models [79, 80, 103], where only simple linguistic features rather than full ontology-based structured annotation is used, could supplement the results obtained in this thesis. The same applies to other classification methods operating on a semantic feature space [142]

While highly accurate retrieval performance was obtained using pre-defined test profiles, the user profiling methods were not formally evaluated. This would require real life usage statistics from the actual systems that were not available. The clustering of retrieval results was not formally evaluated in the scope of this thesis, but could be compared with other clustering methods in terms of both retrieval performance and system performance.

The semantic relatedness approximation methods were compared in a controlled user study, but their effect on the retrieval performance of a scoring function was not measured. A comparison of a larger set of structural measures [113, 77, 68] and corpus-based methods such as latent Dirichlet allocation [14] could supplement the study performed in the scope of this thesis. Although good results have been obtained in ontology-based recommender systems [93], ontology-based information retrieval [23], and the use of ontologies in topic detection and tracking [86], the effect of the ontologies on the performance of recommender systems still requires further evidence.

The fourth research question reflects the area of applications:

4. *Do users find the recommender systems utilizing the developed methods satisfying?*

The methods were implemented in two recommender systems in the cultural heritage domain. A recommender system for the web portal CULTURESAMPO [I,II] was first developed. The method was modified and additional features to support user profiling, result clustering and semantic relatedness approximation were developed for the SMARTMUSEUM system [IV]. The system was evaluated in two user trials in two museums. Users found the system easy to use and indicated that the system improved their museum experience. In addition, the users expressed that they would like to use the system again in other museums. The main suggestions for improvements were navigation support, possibility to relate content to objects in other museums and collections, and explanation support for the user interface.

4.2 External Validity

Content-based recommender systems have known limitations. Specifically, content-based recommender systems have only limited content analysis capabilities and therefore they are most useful in domains where metadata can be extracted automatically or where it has been provided manually. It would be much more difficult to use the systems to recommend, for example, un-annotated audio and video streams. Furthermore, ontology-based recommender systems assume the existence of formal ontologies for the particular application domain. Content-based systems are also not able to determine recommendations based on latent features that are not part of the annotations, but affect the human opinion about the objects. For example, in case of a movie, the general opinion about the movie can be difficult, if not impossible, to obtain based on the content or even the annotation of the object alone.

Despite these limitations, the methods proposed in this thesis were found useful both as part of practical recommender systems and in terms of retrieval performance. However, the studies were done using separate datasets and the retrieval experiments lacked fair baseline methods. This implies that the results are valid in the research context that they were performed in, but the effect of all proposed methods, in terms of the performance of the whole system, have not been verified.

The formal evaluations of the methods were performed as user studies where the accuracy of each of the methods was compared to a gold standard. Questionnaires were used in the user trials. Extensive field studies could have revealed what users actually do in their own contexts, showing common uses and usage patterns, problems and unmet needs. On the other hand, the methods were implemented in recommender systems and their performance was demonstrated in a real life context. Also the user studies were conducted using data and tasks from the domain under interest.

It was not possible to collect data to evaluate user profiling methods, because we were only able to conduct short-term studies. The retrieval methods developed in this thesis are based on VSM. Though the methods were found to perform with high accuracy, they were not compared to alternative retrieval methods. Studies have suggested that some information retrieval tasks can be performed just as successfully with less accurate methods [52]. However, it is pointed out that if subjects continually had to put more attention to the quality of the offered recommendations, perhaps they would grow dissatisfied and eventually stop using the system [52]. Such a comparison was not possible in the scope of this thesis.

In this thesis, system performance, such as the amount of time used to determine the ratings, was not measured. Intuitively, the system performance was acceptable in both applications. However, in the case of the CULTURESAMPO recommender system, the recommendations were computed as a batch process. This was done because the computing of the vector operations was too slow to achieve acceptable system performance. The computation of the indexes of the SMARTMUSEUM recommender system for a test set of 100,000 objects required approximately 6 hours. Formal experiments to evaluate the system performance were not conducted.

The main question related to the external validity of the results considers the domains under interest. The performance of the methods was measured only in the cultural heritage and news domains. The methods could be applicable also to other knowledge-rich domains, where ontology-based structured annotation can be useful and for which ontologies are available. For example, health information [129] is a potentially applicable domain. The automatic annotation method [III] is limited to domains where natural language is used to describe the content. The method classifies the content according to a set of ontologies and the VRA metadata schema. The method is based on supervised machine learning and can be trained using other ontologies and metadata schemas. Therefore, the method is limited to domains for which the ontologies and training data are available or can be constructed. The retrieval methods presented in [II] and in [IV] are limited to cases where ontological concepts can be used to express the annotation. The retrieval method, clustering and user profiling can be adopted to broad domains in a sense that they operate in numerical space and no customization is required.

On the other hand, the developed methods are highly dependent on the ontologies used to capture the knowledge of the domain. In case of the retrieval methods, the correctness of the subsumption hierarchies and the conceptual coverage of the domain under interest is important, while the information extraction method is also dependent on the coverage and quality of the lexical information available in the ontologies. The ontologies that were used in the studies are professionally curated and carefully designed by domain experts. The lexical and conceptual coverage of YSO ontology is not as detailed as, for example, AAT in the cultural heritage domain or WordNet in the general lexical domain. However, YSO is based on YSA, the General Finnish Thesaurus which is the most extensively used thesaurus in Finland. Furthermore, most of the data available in the used collections was already annotated with YSA.

The information extraction method is limited to the English language, but could be adapted to other languages for which Penn Treebank based parsing is applicable. The method to reduce content heterogeneity is based on the usage of rules. This requires manual work, but enables accurate transformation into the event-based knowledge representation.

Another notable feature of the developed systems is the way the content is mainly consumed. In our case the systems are intended for use cases where they assist collection browsing and museum visits. The recommender systems do not support users looking for known objects, but rather offer a variety of options for the user. This can affect the willingness of the users to give relevance feedback and emphasizes the importance of content analysis and content-based techniques that are able to determine recommendations even without extensive user profiling.

The usage context also raises another question related to the users and the assumed usage scenarios. Layman users participated in the user studies which may cause bias compared to professional users. On the other hand, three user studies that ensured the relevance of the systems in the actual tasks were conducted. In fact, the systems developed are meant to be used mostly by layman users and therefore the results may even give insight to the performance of the methods in the actual usage context. In addition, the user studies were extensive enough that statistical significance was ensured and inter-annotator agreement validated.

4.3 Future Work

Content analysis methods could benefit from the following supplements. Information about a dynamic context [108], addressing how other constituents in the sentence were classified, was not used in the reported study. In fact, only features extracted from a single sentence and paths to the main verb of the sentence were used. Adding features that would consider more extensive context and discourse reference, rather than a single sentence, could lead to improved performance [140]. Advanced classification strategies could also result in a gain of the method performance [108]. For example, using separate classifiers to distinguish between the depiction information and the factual information. Improvement with respect to the named entities could be achieved by using anaphora or co-reference resolution. Additional ontology-based features could also be explored. Unsupervised or semi-supervised methods to perform semantic role labeling are a challenging but important future research area.

The content heterogeneity research could benefit from ontology mapping techniques and automatic semantic role labeling of structured ontology-based annotation, where the exact semantic roles have not been specified. This problem is a topic of ongoing research especially in the SRL field [40, 109], where natural language is used as a source for the structure, but requires further development in the heterogeneous schema integration field. Methods that make use of minimal supervision and are able to produce event-based knowledge representation are an important future research direction.

The intent of the research was not to determine the improvements in performance compared to methods operating on text-based annotations, but the starting point rather assumed that heterogeneous annotations can originate from many sources, structured and non-structured. However, because many objects remain without structured ontology-based annotation, an overall performance against methods that operate on text-based annotations could be potentially interesting. Also the role of user interfaces, for example, systems that are able to give explanations are an important future research area.

The study on semantic relatedness approximation could be extended to learning rules and more specific relations, such as subsumption hierarchies. It is also notable, that in our approach, the semantic relatedness is measured between concepts. This means that the independence is assumed not only at the feature level (triples), but on a concept level (resources). Therefore, the semantic relatedness approximation could be investigated in the triple space. Probabilistic user profiling and retrieval methods could enable incorporating more appropriate priors and model dependencies between the features.

Finally, hybrid systems that capture the advantages of both collaborative filtering techniques and ontology-based techniques are an emerging research area [20, 19]. Such approach could lead to substantially better recommender systems that are able to capture the common sense knowledge available in ontologies and the wisdom of the crowds.

References

[1] Gediminas Adomavicius, Ramesh Sankaranarayanan, Shahana Sen, and Alexander Tuzhilin. 2005. Incorporating contextual information in recommender systems using a multidimensional approach. ACM Transactions on Information Systems 23, no. 1, pages 103–145.

[2] Gediminas Adomavicius and Alexander Tuzhilin. 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, no. 6, pages 734–749.

[3] Maristella Agosti and Nicola Ferro. 2007. A formal model of annotations of digital content. ACM Transactions on Information Systems 26, no. 1. Article 3.

[4] Sarabjot Singh Anand, Patricia Kearney, and Mary Shapcott. 2007. Generating semantically enriched user profiles for Web personalization. ACM Transactions on Internet Technology 7, no. 4. Article 22.

[5] Visual Resources Association. 2009. The visual resources association core categories, version 3.0. Referenced 5.1.2010. Available online at: http://www.vraweb.org/projects/vracore3/.

[6] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 1999. Modern information retrieval. Addison-Wesley, ACM Press, New York.

[7] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In: Proceedings of the 17th International Conference on Computational Linguistics, pages 86–90. Association for Computational Linguistics, Morristown, NJ, USA.

[8] Marko Balabanović and Yoav Shoham. 1997. Fab: content-based, collaborative recommendation. Communications of the ACM 40, no. 3, pages 66–72.

[9] Doug Beeferman and Adam Berger. 2000. Agglomerative clustering of a search engine query log. In: KDD '00: Proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 407–416. ACM, New York, NY, USA.

[10] Matthew Berland and Eugene Charniak. 1999. Finding parts in very large corpora. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pages 57–64. Association for Computational Linguistics, College Park, Maryland, USA.

[11] Chris Biemann. 2005. Ontology learning from text: a survey of methods. LDV-Forum 20, no. 2, pages 75–93.

[12] Ella Bingham and Heikki Mannila. 2001. Random projection in dimensionality reduction: applications to image and text data. In: KDD '01: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining, pages 245–250. ACM, New York, NY, USA.

[13] Yolanda Blanco-Fernández, José J. Pazos-Arias, Alberto Gil-Solla, Manuel Ramos-Cabrer, Martín López-Nores, Jorge García-Duque, Ana Fernández-Vilas, Rebeca P. Díaz-Redondo, and Jesús Bermejo-Muńoz. 2008. A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems. Knowlegde-Based Systems 21, no. 4, pages 305–320.

[14] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. Journal of Machine Learning Research 3, pages 993–1022.

[15] D. Brickley and R. V. Guha. 2004. RDF vocabulary description language 1.0: RDF Schema W3C recommendation 10 February 2004. Recommendation, World Wide Web Consortium.

[16] J. Brooke. 1996. SUS: A Quick and Dirty Usability Scale. In: P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland (editors), Usability Evaluation in Industry. Taylor & Francis, London.

[17] Alexander Budanitsky and Graeme Hirst. 2006. Evaluating WordNet-based measures of lexical semantic relatedness. Computational Linguistics 32, no. 1, pages 13–47.

[18] Robin Burke. Knowledge-based Recommender Systems. In: A. Kent (editor), Encyclopedia of Library and Information Systems, volume 69. Supplement 32.

[19] Robin Burke. 2002. Hybrid recommender systems: survey and experiments. User Modeling and User-Adapted Interaction 12, no. 4, pages 331–370.

[20] Iván Cantador, Alejandro Bellogín, and Pablo Castells. 2008. A multilayer ontology-based hybrid recommendation model. AI Communications 21, no. 2-3, pages 203–210.

[21] Iván Cantador, Alejandro Bellogín, and Pablo Castells. 2008. Ontology-based personalised and context-aware recommendations of news items. In: WI-IAT '08: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pages 562–565. IEEE Computer Society, Washington, DC, USA.

[22] Jean Carletta. 1996. Assessing agreement on classification tasks: the Kappa statistic. Computational Linguistics 22, no. 2, pages 249–254.

[23] Pablo Castells, Miriam Fernandez, and David Vallet. 2007. An adaptation of the vector-space model for ontology-based information retrieval. IEEE Transactions on Knowledge and Data Engineering 19, no. 2, pages 261–272.

[24] Stefano Ceri and Jennifer Widom. 1993. Managing semantic heterogeneity with production rules and persistent queues. In: VLDB '93: Proceedings of the 19th International Conference on Very Large Data Bases, pages 108–119. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN 1-55860-152-X.

[25] Peter Pin-Shan Chen. 1976. The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems 1, no. 1, pages 9–36.

[26] V. Christophides, G. Karvounarakis, D. Plexousakis, Michel Scholl, and Sotirios Tourtounis. 2004. Optimizing taxonomic semantic web queries using labeling schemes. Web Semantics: Science, Services and Agents on the World Wide Web 1, no. 2, pages 207 – 228. 2003 World Wide Web Conference.

[27] Philipp Cimiano and Johanna Wenderoth. 2005. Automatically learning qualia structures from the Web. In: DeepLA '05: Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition, pages 28–37. Association for Computational Linguistics, Morristown, NJ, USA.

[28] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, no. 1, pages 37–46.

[29] Jim Cowie and Wendy Lehnert. 1996. Information extraction. Communications of the ACM 39, no. 1, pages 80–91.

[30] Marie-Catherine de Marneffe and Christopher D. Manning. 2008. Stanford typed dependencies manual. Referenced 30.11.2009. Available online at: http://nlp.stanford.edu/software/dependencies_manual.pdf.

[31] Martin Doerr. 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI Magazine 24, no. 3, pages 75–92.

[32] Marc Ehrig and Jérôme Euzenat. 2005. Relaxed precision and recall for ontology matching. In: Integrating Ontologies '05, Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies, volume 156 of *CEUR Workshop Proceedings*. CEUR-WS.org.

[33] Magdalini Eirinaki and Michalis Vazirgiannis. 2003. Web mining for Web personalization. ACM Transactions on Internet Technology 3, no. 1, pages 1–27.

[34] Ramez Elmasri and Shamkant B. Navathe. 2006. Fundamentals of database systems (5th Edition). Addison Wesley.

[35] Jérôme Euzenat and Pavel Shvaiko. 2007. Ontology matching. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[36] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by Gibbs sampling. In: ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pages 363–370. Association for Computational Linguistics, Morristown, NJ, USA.

[37] Angel Garcia-Crespo, Juan Miguel Gomez-Berbis, Ricardo Colomo-Palacios, and Francisco Garcia-Sanchez. 2009. Using support vector machines for feature-oriented profile-based recommendations. International Journal of Advanced Intelligence Paradigms 1, no. 4, pages 418–431.

[38] Susan Gauch, Jason Chaffee, and Alaxander Pretschner. 2003. Ontology-based personalized search and browsing. Web Intelligence and Agent Systems 1, no. 3-4, pages 219–234.

[39] Susan Gauch, Jeason Chaffee, and Alaxander Pretschner. 2003. Ontology-based personalized search and browsing. Web Intelligence and Agent Systems 1, no. 3–4, pages 219–234.

[40] Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. Computational Linguistics 28, no. 3, pages 245–288.

[41] Roxana Girju and Dan I. Moldovan. 2002. Text mining for causal relations. In: Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference, pages 360–364. AAAI Press.

[42] Ana-Maria Giuglea and Alessandro Moschitti. 2006. Semantic role labeling via FrameNet, VerbNet and PropBank. In: ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, pages 929–936. Association for Computational Linguistics, Morristown, NJ, USA.

[43] Daniela Godoy and Analia Amandi. 2005. User profiling for Web page filtering. IEEE Internet Computing 9, no. 4, pages 56–64.

[44] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. 1992. Using collaborative filtering to weave an information tapestry. Communications of the ACM 35, no. 12, pages 61–70.

[45] Asunción Gómez-Pérez and David Manzano-Macho. 2004. An overview of methods and tools for ontology learning from texts. Knowledge Engineering Review 19, no. 3, pages 187–212.

[46] T. R. Gruber. 1993. A translation approach to portable ontology specifications. Knowledge Acquisition 5, no. 2, pages 199–220.

[47] N. Guarino. 1998. Formal ontology and information systems. In: N. Guarino (editor), Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998., pages 3–15. IOS Press, Amsterdam.

[48] Nicola Guarino. 1992. Concepts, attributes and arbitrary relations: some linguistic and ontological criteria for structuring knowledge bases. Data and Knowledge Engineering 8, no. 3, pages 249–261.

[49] Nicola Guarino. 1995. Formal ontology, conceptual analysis and knowledge representation. International Journal of Human-Computer Studies 43, no. 5-6, pages 625 – 640.

[50] Farshad Hakimpour and Andreas Geppert. 2001. Resolving semantic heterogeneity in schema integration. In: FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems, pages 297–308. ACM, New York, NY, USA.

[51] Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th Conference on Computational Linguistics, pages 539–545. Association for Computational Linguistics, Morristown, NJ, USA.

[52] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 22, no. 1, pages 5–53.

[53] Michiel Hildebrand, Jacco van Ossenbruggen, and Lynda Hardman. 2006. /facet: A Browser for Heterogeneous Semantic Web Repositories. In: The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, pages 272–285. Springer.

[54] Will Hill, Larry Stead, Mark Rosenstein, and George Furnas. 1995. Recommending and evaluating choices in a virtual community of use. In: CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 194–201. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA.

[55] Laura Hollink. 2006. Semantic annotation for retrieval of visual resources. SIKS Dissertation Series. Vrije Universiteit Amsterdam.

[56] David Hull. 1993. Using statistical testing in the evaluation of retrieval experiments. In: SIGIR '93: Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval, pages 329–338. ACM, New York, NY, USA.

[57] Aapo Hyvärinen and Erkki Oja. 1997. A fast fixed-point algorithm for independent component analysis. Neural Computing 9, no. 7, pages 1483–1492.

[58] Eero Hyvönen. 2009. Presentation given at the 35th anniversary seminar of Finnish terminology center. Referenced 5.1.2010.

[59] Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkarinen, Joonas Laitio, and Katariina Nyberg. 2009. CultureSampo – Finnish culture on the Semantic Web 2.0. thematic perspectives for the end-user. In: Pro-

ceedings, Museums and the Web 2009, Indianapolis, USA. Archives & Museum Informatics, Toronto, Canada.

[60] Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputtia, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkarinen, Joonas Laitio, and Katariina Nyberg. 2009. CultureSampo – A national publication system of cultural heritage on the Semantic Web 2.0. In: Proceedings of the 6th European Semantic Web Conference (ESWC2009), Heraklion, Greece, pages 851–856. Springer-Verlag.

[61] Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila, and Suvi Kettula. 2005. MuseumFinland–Finnish museums on the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web 3, no. 2-3, pages 224 – 241. Selcted Papers from the International Semantic Web Conference, 2004 - ISWC, 2004.

[62] Eero Hyvönen, Tuukka Ruotsalo, Thomas Häggström, Mirva Salminen, Miikka Junnila, Mikko Virkkilä, Mikko Haaramo, Eetu Mäkelä, Tomi Kauppinen, and Kim Viljanen. 2007. CultureSampo – Finnish culture on the Semantic Web: the vision and first results. In: K. Robering (editor), Information Technology for the Virtual Museum, pages 33–58. LIT Verlag, Berlin.

[63] Eero Hyvönen, Arttu Valo, Ville Komulainen, Katri Seppälä, Tomi Kauppinen, Tuukka Ruotsalo, Mirva Salminen, and Anu Ylisalmi. 2005. Finnish national Ootologies for the Semantic Web – towards a content and service infrastructure. In: Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005). Universidad Carlos III de Madrid, Leganés (Madrid), Spain.

[64] Eero Hyvönen, Kim Viljanen, Eetu Mäkelä, Tomi Kauppinen, Tuukka Ruotsalo, Onni Valkeapää, Katri Seppälä, Osma Suominen, Olli Alm, Robin Lindroos, Teppo Känsälä, Riikka Henriksson, Matias Frosterus, Jouni Tuominen, Reetta

Sinkkilä, and Jussi Kurki. 2007. Elements of a national Semantic Web infrastructure - case study Finland on the Semantic Web. In: Proceedings of the First International Semantic Computing Conference (IEEE ICSC 2007). IEEE Computer Society, Irvine, California.

[65] Eero Hyvönen, Kim Viljanen, Jouni Tuominen, and Katri Seppälä. 2008. Building a national Semantic Web ontology and ontology service infrastructure—the FinnONTO approach. In: Proceedings of the 5th European Semantic Web Conference (ESWC 2008), pages 95–109. Springer.

[66] Peter Ingwersen and Kalervo Järvelin. 2005. The turn: integration of information seeking and retrieval in context. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[67] Dublin Core Metadata Initiative. 2008. Dublin core metadata element set, version 1.1. Referenced 6.11.2009. Available online at: http://dublincore.org/documents/dces/.

[68] J. J. Jiang and D. W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In: Proceedings of the International Conference Research on Computational Linguistics (ROCLING X), pages 19–33.

[69] Xing Jiang and Ah-Hwee Tan. 2005. Mining ontological knowledge from domain-specific text documents. In: ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining, pages 665–668. IEEE Computer Society, Washington, DC, USA.

[70] Jaana Kekäläinen and Kalervo Järvelin. 2002. Using graded relevance assessments in IR evaluation. Journal of American Society for Information Science and Technology 53, no. 13, pages 1120–1129.

[71] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. 1997. GroupLens: applying collaborative filtering to Usenet news. Communications of the ACM 40, no. 3, pages 77–87.

[72] Thomas Landauer, P. W. Foltz, and D. Laham. 1998. Introduction to latent semantic analysis. Discourse Processes 25, no. 1, pages 259–284.

[73] Claudia Leacock and Martin Chodorow. 1998. Combining local context and WordNet similarity for word sense identification. In: Christiane Fellbaum (editor), WordNet: An Electronic Lexical Database, pages 265–283. The MIT Press, Cambridge, MA.

[74] Taehee Lee, Jonghoon Chun, Junho Shim, and Sang-Goo Lee. 2006-2007. An ontology-based product recommender system for B2B marketplaces. International Journal of Electronic Commerce 11, no. 2, pages 125–155.

[75] Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain. 2006. Content-based multimedia information retrieval: state of the art and challenges. ACM Transactions on Multimedia Computing, Communications and Applications 2, no. 1, pages 1–19.

[76] Innar Liiv, Tanel Tammet, Tuukka Ruotsalo, and Alar Kuusik. 2009. Personalized context-aware recommendations in SMARTMUSEUM: combining semantics with statistics. In: Proceedings of the The Third International Conference on Advances in Semantic Processing (SEMAPRO 2009). IEEE Computer Society, Sliema, Malta.

[77] Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In: Proceedings of the 17th International Conference on Computational Linguistics, pages 768–774. Association for Computational Linguistics, Morristown, NJ, USA.

[78] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 7, no. 1, pages 76–80.

[79] G. Liu. 1994. The semantic vector space model (SVSM): a text representation and searching technique. In: Proceedings of the Twenty-Seventh Hawaii International

Conference on System Sciences., volume 4, pages 928–937. Vol.IV: Information Systems: Collaboration Technology Organizational Systems and Technology.

[80] Geoffrey Z Liu. 1997. Semantic vector space model: implementation and evaluation. Journal of the American Society for Information Science 48, no. 5, pages 395–417.

[81] Alexander Maedche and Steffen Staab. 2001. Ontology learning for the Semantic Web. IEEE Intelligent Systems 16, no. 2, pages 72 – 79.

[82] Alexander Maedche, Steffen Staab, Nenad Stojanovic, Rudi Studer, and York Sure. 2003. SEmantic portAL: The SEAL Approach. In: Dieter Fensel, James A. Hendler, Henry Lieberman, and Wolfgang Wahlster (editors), Spinning the Semantic Web, pages 317–359. MIT Press.

[83] Veronica Maidel, Peretz Shoval, Bracha Shapira, and Meirav Taieb-Maimon. 2008. Evaluation of an ontology-content based filtering method for a personalized newspaper. In: RecSys '08: Proceedings of the 2008 ACM Conference on Recommender Systems, pages 91–98. ACM, New York, NY, USA.

[84] Eetu Mäkelä, Tuukka Ruotsalo, and Eero Hyvönen. 2007. Automatic exhibition generation based on semantic cultural content. In: Poster Proceedings of the 6th International Semantic Web Conference. Busan, Korea.

[85] Eetu Mäkelä, Kim Viljanen, Olli Alm, Jouni Tuominen, Onni Valkeapää, Tomi Kauppinen, Jussi Kurki, Reetta Sinkkilä, Teppo Känsälä, Robin Lindroos, Osma Suominen, Tuukka Ruotsalo, and Eero Hyvönen. 2007. Enabling the Semantic Web with ready-to-use Web widgets. In: Proceedings of the First Industrial Results of Semantic Technologies Workshop. CEUR Vol-293, Busan, Korea.

[86] Juha Makkonen, Helena Ahonen-Myka, and Marko Salmenkivi. 2004. Simple semantics in topic detection and tracking. Information Retrieval 7, no. 3-4, pages 347–368.

[87] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. Introduction to information retrieval. Cambridge University Press, 1st edition.

[88] C. Masolo, G. Guizzardi, L. Vieu, E. Bottazzi, and R. Ferrario. 2005. Relational roles and qua-individuals. In: Proceedings of the AAAI Fall Symposium on Roles, an Interdisciplinary Perspective. Hyatt Crystal City, Arlington, Virginia.

[89] Deborah L. McGuinness and Frank van Harmelen. 2004. OWL Web ontology language overview. Referenced 5.1.2010. Available online at: http://www.w3.org/TR/owl-features/.

[90] Herbert Menzel. 1966. Information needs and uses in science and technology. Annual Review of Information Science and Technology 1, pages 41 – 69.

[91] Stuart E. Middleton, David C. De Roure, and Nigel R. Shadbolt. 2001. Capturing knowledge of user preferences: ontologies in recommender systems. In: K-CAP '01: Proceedings of the 1st International Conference on Knowledge Capture, pages 100–107. ACM, New York, NY, USA.

[92] Stuart E. Middleton, Nigel R. Shadbolt, and David C. De Roure. 2003. Capturing interest through inference and visualization: ontological user profiling in recommender systems. In: K-CAP '03: Proceedings of the 2nd International Conference on Knowledge Capture, pages 62–69. ACM, New York, NY, USA.

[93] Stuart E. Middleton, Nigel R. Shadbolt, and David C. De Roure. 2004. Ontological user profiling in recommender systems. ACM Transactions on Information Systems 22, no. 1, pages 54–88.

[94] Bradley N. Miller, Istvan Albert, Shyong K. Lam, Joseph A. Konstan, and John Riedl. 2003. MovieLens unplugged: experiences with an occasionally connected recommender system. In: IUI '03: Proceedings of the 8th International Conference on Intelligent User Interfaces, pages 263–266. ACM, New York, NY, USA.

[95] George A. Miller. 1995. WordNet: a lexical database for English. Communications of the ACM 38, no. 11, pages 39–41.

[96] Rokia Missaoui, Petko Valtchev, Chabane Djeraba, and Mehdi Adda. 2007. Toward recommendation based on ontology-powered Web-usage mining. IEEE Internet Computing 11, no. 4, pages 45–52.

[97] Ruslan Mitkov. 2002. Anaphora resolution. Studies in Language and Linguistics. Longman.

[98] Raymond J. Mooney and Razvan Bunescu. 2005. Mining knowledge from text using information extraction. SIGKDD Explorations Newsletter 7, no. 1, pages 3–10.

[99] Raymond J. Mooney and Loriene Roy. 2000. Content-based book recommending using learning for text categorization. In: DL '00: Proceedings of the fifth ACM conference on Digital libraries, pages 195–204. ACM, New York, NY, USA.

[100] Sharmin Moosavi, Mohammadali Nematbakhsh, and Hadi Khosravi Farsani. 2009. A semantic complement to enhance electronic market. Expert Systems with Applications 36, no. 3, Part 2, pages 5768 – 5774.

[101] Yannick Naudet, Armen Aghasaryan, Yann Toms, and Christophe Senot. 2008. An ontology-based profiling and recommending system for mobile TV. In: SMAP '08: Proceedings of the 2008 Third International Workshop on Semantic Media Adaptation and Personalization, pages 94–99. IEEE Computer Society, Washington, DC, USA.

[102] Roberto Navigli and Paola Velardi. 2004. Learning domain ontologies from document warehouses and dedicated Web sites. Computational Linguistics 30, no. 2, pages 151–179.

[103] O. Nouali and P. Blache. 2004. A semantic vector space and features-based approach for automatic information filtering. Expert Systems with Applications 26, no. 2, pages 171 – 179.

[104] Jeffrey M. O'Brien. 2006. The race to create a 'smart' Google. Referenced 12.8.2009. Available online at: http://money.cnn.com/magazines/fortune/fortune_archive/2006/11/27/8394347/.

[105] Michael Pazzani and Daniel Billsus. 1997. Learning and revising user profiles: the identification of interesting Web sites. Machine Learning 27, no. 3, pages 313–331.

[106] Michael Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In: The Adaptive Web, pages 325–341. Springer.

[107] Michael J. Pazzani. 1999. A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review 13, no. 5-6, pages 393–408.

[108] Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2005. Support vector learning for semantic argument classification. Machine Learning 60, no. 1-3, pages 11–39.

[109] Sameer S. Pradhan, Wayne Ward, and James H. Martin. 2008. Towards robust semantic role labeling. Computational Linguistics 34, no. 2, pages 289–310.

[110] R. Rada, H. Mili, E. Bicknell, and M. Blettner. 1989. Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man and Cybernetics 19, no. 1, pages 17–30.

[111] Erhard Rahm and Philip A. Bernstein. 2001. A survey of approaches to automatic schema matching. The VLDB Journal 10, no. 4, pages 334–350.

[112] Paul Resnick and Hal R. Varian. 1997. Recommender systems. Communications of the ACM 40, no. 3, pages 56–58.

[113] Philip Resnik. 1995. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pages 448–453.

[114] M. Andrea Rodríguez and Max J. Egenhofer. 2003. Determining semantic similarity among entity classes from different ontologies. IEEE Transactions on Knowledge and Data Engineering 15, no. 2, pages 442–456.

[115] Tuukka Ruotsalo, Katri Seppälä, Kim Viljanen, Eetu Mäkelä, Jussi Kurki, Olli Alm, Tomi Kauppinen, Jouni Tuominen, Matias Frosterus, Reetta Sinkkilä, and Eero Hyvönen. 2008. Ontology-based interoperability of digital collections. Signum , no. 5.

[116] Gunter Saake, Kai-Uwe Sattler, and Stefan Conrad. 2005. Rule-based schema matching for ontology-based mediators. Journal of Applied Logic 3, no. 2, pages 253 – 270. Logic-based Methods for Information Integration.

[117] G. Salton, A. Wong, and C. S. Yang. 1975. A vector space model for automatic indexing. Communications of the ACM 18, no. 11, pages 613–620.

[118] Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. Information Processing and Management 24, no. 5, pages 513–523.

[119] David Sánchez and Antonio Moreno. 2008. Learning non-taxonomic relationships from Web documents for domain ontology construction. Data and Knowledge Engineering 64, no. 3, pages 600 – 623.

[120] Tefko Saracevic. 1975. Relevance: A review of and a framework for the thinking on the notion of information science. Journal of American Society for Information Science 26, no. 6, pages 321–334.

[121] J. Ben Schafer, Joseph Konstan, and John Riedi. 1999. Recommender systems in e-commerce. In: EC ’99: Proceedings of the 1st ACM Conference on Electronic Commerce, pages 158–166. ACM, New York, NY, USA.

[122] J. Ben Schafer, Joseph A. Konstan, and John Riedl. 2001. E-commerce recommendation applications. Data Mining and Knowledge Discovery 5, no. 1-2, pages 115–153.

[123] Nigel Shadbolt, Kieron O'Hara, and Louise Crow. 1999. The experimental evaluation of knowledge acquisition techniques and methods: history, problems, and new directions. International Journal of Human-Computer Studies 51, no. 4, pages 729–755.

[124] Barforoush A. A. Shamsfard, M. 2004. Learning ontologies from natural language texts. International Journal of Human-Computer Studies 60, no. 1, pages 17–63.

[125] Upendra Shardanand and Pattie Maes. 1995. Social information filtering: algorithms for automating "word of mouth". In: CHI '95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 210–217. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA.

[126] Amit P. Sheth and James A. Larson. 1990. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys 22, no. 3, pages 183–236.

[127] Elena Simperl. 2009. Reusing ontologies on the Semantic Web: A feasibility study. Data and Knowledge Engineering 68, no. 10, pages 905 – 925.

[128] J. F. Sowa. 2000. Knowledge representation: logical, philosophical and computational foundations. Brooks/Cole, Pacific Grove, CA.

[129] Osma Suominen, Eero Hyvönen, Kim Viljanen, and Eija Hukka. 2009. HealthFinland – a national semantic publishing network and portal for health information. Web Semantics: Science, Services and Agents on the World Wide Web 7, no. 4, pages 287 – 297. Semantic Web challenge 2008.

[130] The J. Paul Getty Trust. 2009. Union list of artist names. Referenced 5.1.2010. Available online at: http://www.getty.edu/research/conducting_research/vocabularies/ulan/.

[131] Victoria Uren, Philipp Cimiano, José Iria, Siegfried Handschuh, Maria Vargas-Vera, Enrico Motta, and Fabio Ciravegna. 2006. Semantic annotation for knowl-

edge management: requirements and a survey of the state of the art. Web Semantics: Science, Services and Agents on the World Wide Web 4, no. 1, pages 14 – 28.

[132] Willem Robert van Hage, Hap Kolb, and Guus Schreiber. 2006. A method for learning part-whole relations. In: Proceedings of the 5th International Semantic Web Conference, pages 723–735. Springer.

[133] Vincent Ventrone. 1991. Semantic heterogeneity as a result of domain evolution. SIGMOD Record 20, no. 4, pages 16–20.

[134] Yiwen Wang, Natalia Stash, Lora Aroyo, Peter Gorgels, Lloyd Rutledge, and Guus Schreiber. 2008. Recommendations based on semantically enriched museum collections. Web Semantics: Science, Services and Agents on the World Wide Web 6, no. 4, pages 283 – 290. Semantic Web Challenge 2006/2007.

[135] Christopher Welty and Nicola Guarino. 2001. Supporting ontological analysis of taxonomic relationships. Data and Knowledge Engineering 39, no. 1, pages 51 – 74.

[136] Sung-Shun Weng, Binshan Lin, and Wen-Tien Chen. 2009. Using contextual information and multidimensional approach for recommendation. Expert Systems with Applications 36, no. 2, pages 1268–1279.

[137] Utz Westermann and Ramesh Jain. 2007. Toward a Common Event Model for Multimedia Applications. IEEE Multimedia 14, no. 1, pages 19–29.

[138] Jan Wielemaker, Michiel Hildebr, Jacco Van Ossenbruggen, and Guus Schreiber. 2008. Thesaurusbased search in large heterogenous collections. In: Proceedings of The Semantic Web - ISWC 2008 7th International Semantic Web Conference. Springer-Verlag.

[139] Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In: Proceedings of the 32nd annual meeting on Association for Computational Lin-

guistics, pages 133–138. Association for Computational Linguistics, Morristown, NJ, USA.

[140] Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 88–94.

[141] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. 2003. Faceted metadata for image search and browsing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 401–408. ACM, New York, NY, USA.

[142] Yun Fei Yi, Cheng Hua Li, and Wei Song. 2008. Email Classification Using Semantic Feature Space. In: International Conference on Advanced Language Processing and Web Information Technology, 2008. ALPIT ’08., pages 32–37.

[143] Cai-Nicolas Ziegler, Georg Lausen, and Joseph A. Konstan. 2008. On exploiting classification taxonomies in recommender systems. AI Communications 21, no. 2-3, pages 97–125.

[144] Cai-Nicolas Ziegler, Georg Lausen, and Lars Schmidt-Thieme. 2004. Taxonomy-driven computation of product recommendations. In: CIKM ’04: Proceedings of the 13th ACM Conference on Information and Knowledge Management, pages 406–415. ACM, New York, NY, USA.

ISBN 978-952-60-3150-7
ISBN 978-952-60-3151-4 (PDF)
ISSN 1795-2239
ISSN 1795-4584 (PDF)