Using Process Ontologies to Contextualize Recommender Systems in Engineering Projects for Knowledge Access Improvement

Mahsa Mehrpoor[1], Jon Atle Gulla[2], Dirk Ahlers[2], Kjetil Kristensen[1], Ole Ivar Sivertsen[1], Soroush Ghodrat

[1] Department of Engineering Design and Materials, Faculty of Engineering Science and Technology, NTNU, Trondheim, Norway
[2] Department of Computer and Information Science, Faculty of Information Technology, Mathematics and Electrical Engineering, NTNU, Trondheim, Norway
mahsa.mehrpoor@ntnu.no
jon.atle.gulla@idi.ntnu.no
dirk.ahlers@idi.ntnu.no
kjetil.kristensen@kristensenconsulting.com
ole.ivar.sivertsen@ntnu.no
s.ghodrat@gmail.com

Abstract: Knowledge and information are highly important resources in today's knowledge economy and vital in achieving organization's goals. Particularly in engineering projects, users' tasks are highly dependent on accessing, using, and reusing these resources. Users already spend a lot of time searching for relevant knowledge. As the total volume of documents across different sources and repositories increases, users face additional overhead related to search and retrieval. Knowledge workers across multiple disciplines experience fierce competition and a persistent pressure to deliver value-added contributions in a competitive global business environment with complex, multidisciplinary problems. Simple search engines are often not sufficient since they are not designed to retrieve those relevant documents that match the user's current work situation. Therefore, the need for a semantics-based solution has been identified. This paper describes the early stages of a PhD project that proposes a tailored recommender system for improving knowledge accessibility in an engineering setting. The recommender system will be developed for and validated in a multidisciplinary engineering project as a case study. We take the advantage of content-based filtering and collaborative filtering along with semantic technologies to provide relevant and accurate recommendations. In order to contextualize user's work situations during project development process, recommender system utilizes a tailored process ontology to be able to explore different dimensions of user's situations. By merging the concepts derived from the ontology, the current work situation of the given user is identified and varied fine-grained user profiles will be created at real time called dynamic user profiles. Therefore, recommender system is able to set the scope of user's interest to the exact level that the user desires in his current situation. To classify the identified relevant documents, we propose creating concept profiles that are originated from process ontology concepts for further recommendation according to collaborative approach. This paper describes the recommender system components and proposes a framework of the target recommender system and discusses how its components are integrated and interact in order to improve information access in the engineering project.

Keywords: Work situation; Recommender Systems; Semantic recommendation; Process ontology; Dynamic User profile; Information Access

1. Introduction

Knowledge and information are highly important resources in today's knowledge economy and are vital in achieving organization's goals. One of the issues that organizations are concerned about is how to manage these important resources in order to be able to align them with their objectives and goal achievement strategies. The more complicated the corpus of organization and knowledge resources is, the more effort they need to spend to come up with a tailored solution of knowledge management. Users' tasks are highly dependent on accessing, using, and reusing the knowledge resources. There are many factors that need to be considered about the user in a particular work situation (Mehrpoor et al. 2014) since the way the users interact with the system for exploring their required information influences the management of knowledge resources (Freund 2008). Common search engines have been used for information retrieval but their results are based on the search keywords entered by the user. Search engines are not designed to filter search results according to user's work situation. Finding semantic relations between user's interest and knowledge resource content might be helpful to retrieve more relevant results. Semantic search engines (Jayavel et al. 2013) consider contextual meaning of the search query and provide more relevant results. However, the search string might lack the essential concepts that lead to a proper context identification. So, the approach that we have chosen for relevant information retrieval is recommender systems (Ricci et al. 2011) to improve context identification (Ahlers and Mehrpoor 2014) of semantic search engines. Recommender systems can assist knowledge managers to provide right information for the right users and also reduce search costs.

We propose a tailored recommender system for knowledge management improvement in an engineering setting. This paper describes the early stages of work and is organized as follows: Section 2 provides an overview of related work in the area of recommender systems and using ontologies and user profiles for recommendation improvements. Section 3 elaborates the case study and formulates the challenges of knowledge access in the engineering settings. Section 4 describes the proposed recommender system to address these challenges. Section 5 is about the future steps of the PhD research and Section 6 concludes the paper.

2. Background

This section discusses the background for the proposed solution of knowledge access improvement in the engineering projects, namely recommender systems, ontologies as semantic technologies, vector models, information extraction and indexing tools, and user profiling in recommender systems.

The two fundamental approaches of recommendation are content-based filtering and collaborative filtering. The content-based approach focuses on the similarities between the items and combines it with users' preferences (Lops et al. 2011). The collaborative filtering approach focuses on the similarity between the preferences of the target user with other users (Schafer et al. 2007). The preferences of the user are derived from different explicit and implicit methods such as the history of user's activities in the system as user behavior or by explicitly asking user's interests and storing them in user profiles (Pazzani and Billsus 2007). User profiles might be created statically or dynamically (Hong et al. 2013). Static user profiles contain the information manually added by user; versus, dynamic user profiles are derived from users' behavior, history of their activities and so on. To improve performance, both approaches can be combined into hybrid recommender systems (Garcin et al. 2012). The level of relevance for documents is most commonly calculated using the vector space model in content-based filtering approaches (Werner and Cruz 2013). In addition, the context-aware approach focuses on the contextual information (Dey 2001) and how the main contextual information of the user play a role in identifying the relevant information which fits the user's interests (Adomavicius and Tuzhilin 2011, White et al. 2009, Bouneffouf 2013). Especially when there are only a few ratings available from the users, contextual information becomes more important to be taken into account and analyzed to know the users in more detail (Ma et al. 2011).

In order to improve the recommendation solutions, semantic technologies are applied. Ontologies are one the main semantic tools. They have been used for knowledge management and constructing a semantic model for the concepts of the domain in order to provide more semantic relations between the concepts and avoid the limitations in text-mining techniques (He 2013). Ontologies have been used in different recommender systems. (Paiva et al. 2013) proposes a common hierarchical architecture for ontology-based recommender systems that consists of four layers of context, discovery, recommendation and ontology to provide the relevant recommendation for the given user. For presenting semantic description of both items and user profile, (Werner et al. 2013) proposed an ontology-based recommender system for recommending economic articles.

In addition to the usage of ontology for knowledge modeling, they have used ontology as a fundamental tool for indexing and annotating articles, which makes the system less dependent to a specific area. (Yu et al. 2007) used ontology to contextualize the user, content and domain in three individual ontologies. Then, a recommendation method with four steps of semantic relevance calculation, recommendation refining, learning path generation, and recommendation augmentation, is proposed for providing relevant recommendations. Recommender systems should be able to extract the information in order to analyze it and identify similarities between the objects. There are other available libraries for information extraction and retrieval like Elastic-search (Banon 2012) which is used in this work. It supports different document formats and will enable the recommender system to search and analyze in real time.

3. Case study: Knowledge access in a multidisciplinary engineering project and its challenges

Shell Eco-Marathon is an annual competition that challenges student teams to design and build ultra energy-efficient vehicles. Every year a team of master students at NTNU participates in this competition and designs the DNV GL fuel fighter (Bøvre et al. 2014). The team consists of different engineering disciplines such as mechanics, electronics, materials, cybernetics, aero-dynamics and other sub-disciplines. They are working together in a multidisciplinary work environment over one year.

One of their challenges is knowledge access and sharing. During the project development, engineers would often benefit from reusing archived knowledge and information from previous competitions, but having loads of unstructured documents in different formats makes it difficult for them to search and to get access to all the relevant information for reuse, and it often causes them to start from scratch, failing to make full use of available and potentially valuable resources. The current solution of the team for knowledge access is to communicate with other students of past years' competition to ask for the required knowledge and how to explore the relevant archived knowledge resources. They organize meetings with the past team members or send emails, which can be time-consuming and inefficient. This mirrors a standard situation in other companies, where new people always come into the company or new project teams are formed, who then have to learn their way around knowledge management and document storage systems.

An appropriate solution for knowledge access improvement in the engineering project requires analysis of both users and knowledge resources. We need to explore more information to figure out how well it is possible to tailor the documents to the users' assigned tasks. These issues are addressed in the following.

3.1 User information needs

Engineers with different proficiencies and different levels of expertise are involved in the project. Particular responsibilities are assigned to engineers in different phases of the project development from requirements analysis to design, implementation, test and evaluation. Some of the engineers are more involved to the early stages such as mechanical people, some of them contribute more in the late stages such as electrical people, and some people like project managers have constant responsibilities in the whole project development process. Their responsibilities contain certain role(s) and task(s) and they work with specific machine component(s). During their task performance, some the engineers need to interact with each other. For instance, the design engineer needs to interact with aero-dynamics engineer to come up with an aero-dynamics form of the machine. Also, electronics and cybernetics people work together in some phases to make a component of the machine. In addition, there are some inter-relations between machine components and they themselves break into some sub-components, which makes them dependent for design concerns.

As an example, the mechanical engineer may have one or more roles such as team leader with some assigned tasks such as requirements analysis and designing of specific components of the machine such as steering system, brake and wheels. In order to find the relevant information, the mechanical engineer looks for those directories of document storage that look related to his assigned tasks but since the document storage is not well-structured and not all the information that he needs is stored in the place that he expects, it makes knowledge exploration challenging and he has to spend more time on knowledge and information seeking rather than on efficient task performance. Therefore, a tool that helps him in finding his desired knowledge and information is missing.

3.2 Document storage analysis

Available archived documents are stored in a shared file system. According to the analysis of the document storages of the last three years of the competition, the type of stored documents are quite varied. Many different unstructured documents have been identified in different formats such as text-based, multimedia, modeling, programming formats, html files and the other types that are created along with the output of the specified applications. In the case of the mechanical engineer, he is interested in modeling documents, a group of textual documents and a sort of modeling images. A proper solution for the identification of relevant documents for both textual and non-textual document types is required. Figure 1 depicts engineers in two disciplines of mechanics and cybernetics and their particular contexts.

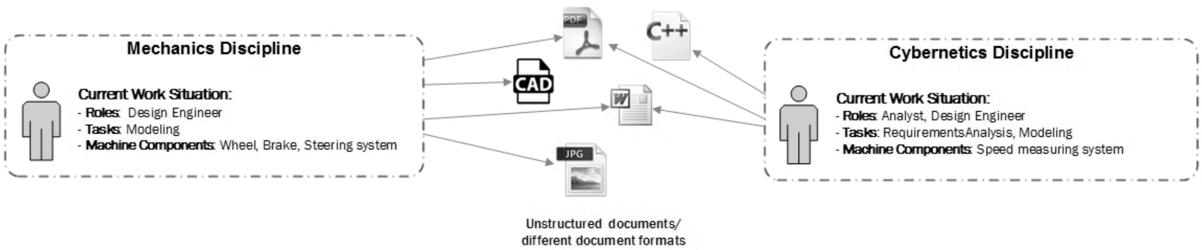


Figure 1. Engineers in particular work situations and their required documents

4. Tailored recommender system for knowledge access improvement

This section describes the essentials for the target recommender system. Each sub-section introduces a component of the recommender system and discusses why it is used and how it will be employed. The proposed framework consisting of all the introduced components is presented and the integration and interaction of the components is described.

4.1 Recommendation approaches

Recommendation techniques follow varied approaches according to the conducted literature review. Depending on the features and specifications of the use case, one or more of these approaches are used to provide the potentially right knowledge and information for the right users. In our use case, from one side there are users in certain work situations and from the other side there are documents that are classified in two groups of textual and non-textual documents.

Textual document formats are mostly office documents and PDFs. Also, non-textual document formats are mostly binary files such as multimedia documents, CAD files and so on. In order to identify the similar documents to the given user's preferences, the content of the documents needs to be analyzed. Therefore, we follow the content-based recommendation approach. To explore the content of documents precisely, one of the available techniques is to index the documents and make them searchable by using textual search engine libraries (Hatcher and Gospodnetic 2004). In addition, meaningful meta-data included in the document itself might be helpful in document's content identification. Information extraction libraries are used to annotate and index the content of documents. According to the background of our work, we use Elastic-search (Banon 2012) for information extraction and indexing the documents. In our example of the mechanical engineer, mainly those documents that have the most information pieces about mechanics, designing and modeling of wheels, brake and steering system are likely to be relevant for him. However, it should be taken into consideration that the work situation of the user is not constant during the project development and at each stage his required information narrows down to a specific scope; for instance, in the early phases, he may work on modeling the wheels. So, information about brake is not his interest in this phase.

In the two document groups, the results of content analysis would be different. In textual documents group, since they are text-based, the probability of exploring similar information pieces that fit user's interests are higher than non-textual documents. Since there is not enough meaningful text-based content embedded in non-textual documents, it limits information extraction and indexing process for annotating these types of documents. So, another recommendation approach should also be applied. For this part, we need to investigate the users' side more and focus on the information that we can gain from their behavior. Therefore, we follow collaborative filtering approach along with content-based approach to study user's side more and come up with better recommendation results. If the mechanical engineer uses an image and rates it, specifies it as relevant, then it is inferred that this image is included in such a situation. Consequently, for another user in a similar situation, the identified document is more likely to be the right candidate for recommending to this user. However, it remains uncertain how accurate the explicit feedback of the first user is in general. Explicit feedback is the feedback given directly by the user like scoring the document. The users do not always tend to rate the documents and not all the time their rating results are reliable. Implicit feedback is also considered which is inferred from the user's behavior indirectly like the amount of time that users spend for studying a document or whether users open the document at all. Another challenge is the number of available users, new users and new documents which causes the cold-start problem where the relevance of new or unrated items in unknown (Gunawardana and Meek 2009). Since at the early phases of this approach very few documents are viewed by the users, the numbers of identified relevant documents are fewer and not sufficient for a collaborative filtering recommendation process.

Identifying the documents that are similar to user's interest only provides potentially relevant documents. However, the question still remains how we can identify the level of relevancy of each potentially relevant document or in other words, how we can rank the documents. We need to identify how close each document is to the given user's situation. To do this, we use Elastic-search capabilities for scoring the documents which partly includes technique like VSM, Vector Space Model (Werner and Cruz 2013). In addition to this method, we also rank the documents using the explicit and implicit feedback. Not all the documents can be scored properly using Elastic-search especially non-textual documents.

To take more advantage of the two discussed recommendation approaches, the documents that are viewed by users can be classified according to the situation of their viewers. For classification of the documents, we propose creating profiles that match users' contextual features, which we elaborate on section 4.3.

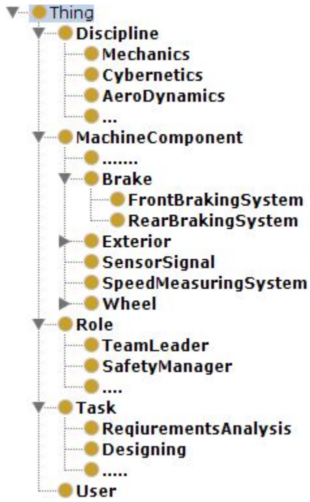
4.2 Process ontology and factors of user's context identification

In this research, we aim to apply ontology as one of the knowledge management tools with the inspiration from OBIE systems, Ontology-Based Information Extraction systems (Wimalasuriya and Dou 2010). In our recommender system, ontology is used as the fundamental resource of contextual knowledge to help us in identifying the situation of the user. For building the ontology, a scenario-based methodology called NeOn methodology is used. The NeOn methodology supports a knowledge reuse approach (Suárez-Figueroa et al.

2012).

Figure 2. Main concepts of the process ontology

Among 9 scenarios defined by this methodology, the first scenario is selected which is "From specification to implementation" since the development of the ontology is from scratch. In ontology requirements specification activity, ORSA, the requirements for the ontology have been specified. The ontology environment has been studied by interviewing the project team. In the conceptualization phase, the extracted terms of ORSA is conceptualized. Afterwards, the conceptual model has been transformed into a formal model and implemented using Protégé (Protege 2015). The main concepts are shown in Figure 2 that consists of 5 dimensions of user, discipline, role, task, and machine component. Every dimension describes a part of the work situation of the user and by combining their leaf concepts, the concepts defined in the last level of the ontology graph, different work situations are described. Figure 3 illustrates some of the leaf concepts of the



ontology.

Figure 3. Parts of the hierarchical structure of the concepts of the ontology

4.3 User profiling and Concept profiles

Recommending the relevant documents to the user requires collecting some information about user's preferences in order to compare documents with users' preferences to identify similarities. The scope of user's preferences includes all the essential entities that characterize the situation of the user in his work environment. These entities might be different according to different duties of the user and it causes some modification in user's preferences scope.

We use process ontology as the generic static user profile that contains different aspects of users' situations. Utilizing the process ontology also enables the system to create many different fine-grained user profiles which we call them dynamic user profiles. Creating dynamic user profiles help the recommender system to set the scope of users' interest to the exact level that the user desires. Therefore, the recommended items are much closer to the users' information needs and not too many relevant and irrelevant documents are explored from the document storage. Without considering the dynamic user profiles, recommender system covers all the preferences of the user in all the project stages and many documents will be recommended that not all of them are relevant at this stage of the project. Dynamic user profile is the combination of the individual leaf concept profiles of the ontology and we call them dynamic since they are mutable and are only alive until the end of the life-time of the current situation of the given user.

As discussed earlier, identified relevant documents will be classified to some profiles. These profiles match the leaf concepts of the ontology; so we have a profile for mechanics, a profile for electronics, a profile for reporting, and so on. The identified relevant documents will be appended to these profiles for further recommendations to the users who have similar situations.

In our example, different combinations of the involved aspects to engineer's work situations might cause varied user profiles. Figure 4 (a) illustrates the maximum number of possible user profiles for a particular user. Also, in (b), the maximum number of possible user profiles for the whole system, which can be created by leaf ontology concepts, is depicted. The more roles and tasks the user has, the more situations are identified and consequently the more user profiles could be created in the whole project development process. However, it should also be considered that whether there are any constraints on the relations between disciplines, roles, tasks and components. For instance, wheels are only related to steering system and are not related to brakes, technically. So the number of possible combination is reduced. These relations are inferred from the built process ontology. This knowledge can also be used to improve the indexing of documents.

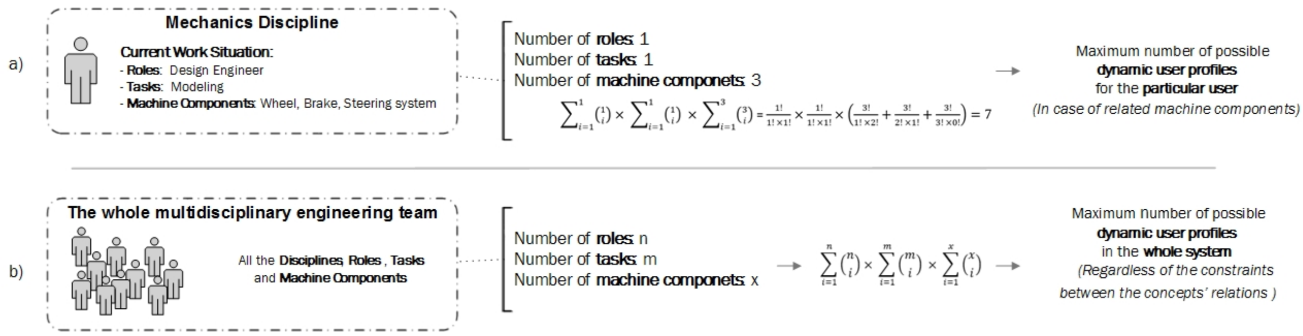


Figure 4. Users' work situations and possible dynamic user profiles

4.4 Framework of the proposed recommender system

After introducing the essential components of our recommender system, we now describe how these components are integrated and interact with each other as the framework of the recommender system as represented in Figure 5.

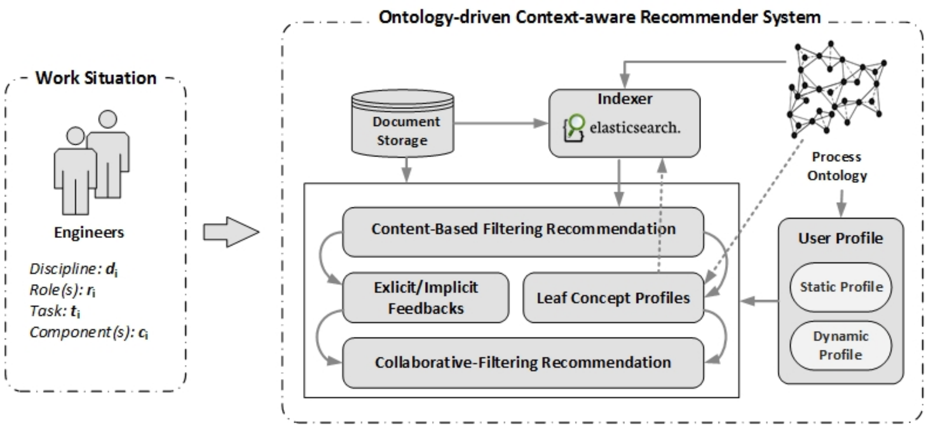


Figure 5. The framework of the ontology-driven context-aware recommender system

The process ontology as the context-based domain knowledge used in multiple parts of the recommender system. The stored documents in the document storage are posted to Elastic-search to be indexed and become searchable. Notice that not all the documents may index properly and have only limited metadata available. When one of the users e.g. the mechanical engineer logins to the system, he selects his current work situation derived from the process ontology. At this stage, we consider the ontology as the static user profile. The current identified user's situation is the dynamic user profile. The indexed documents that are similar to the identified user's situation are explored by Elastic-search according to their level of relevance. Ranked documents are represented to the mechanical engineer and here we use explicit/implicit feedback such as asking the user to score the selected document or considering the time spent for studying the document. So, we capture how well the document matches user's context. At this stage, we need to classify the identified relevant document in concept profiles for later recommendations to all those users' situation that have similarities with the features of this document according to collaborative approach. Here again we take the advantage of the synergy of collaborative filtering and content-based approach. For those documents that are not indexed properly, we aim to use collaborative filtering approach by using some feedback from the users' side and follow the similar approach for textual documents; however, as mentioned earlier, the cold-start problem is still a challenge. Also notice that, the ontology will be developed and improved based on the results that we achieve in the process of documents exploration to improve the ontology concepts and the matching by adapting it to documents inputs.

5. Future work

In our future work, we intend to expand the development of the process ontology. We will investigate how matchable the indexed documents are with the concepts defined in process ontology. We will investigate possible ways to explore all the relevant resources from the document storage. We should be able to match the documents with the ontology and if required involve more aspects to the ontology such as how the documents are classified in the file system and what document formats are more specified in each classification group particularly for those documents that are not indexed properly.

To take the further advantage of the ontology, we will utilize the relations among concepts of the process ontology to identify more possibilities for recommendation. While the users' contexts are being compared, these defined semantic relations help the recommender system to logically infer more items for recommendation since ontology supports inference and reasoning. These inferred recommendable items might not be identified if the approach is only limited to users' behavior and feedback. The task of the recommender system is not only to recommend more, but to recommend more accurate and relevant.

6. Conclusion

The advantages of using the recommender systems in the professional workplaces are discussed in this paper and a structured framework is proposed to create a tailored recommender system. To the best of our knowledge, this is a novel approach in the area of recommender systems which contextualizes users' work situation by using ontologies. We take the advantage of the two fundamental approaches of recommendation along with studying user's context, inspired from context-aware approaches. Monitoring the users from different aspects helps us to provide more relevant recommendations for the target users.

We propose the solution of creating varied fine-grained user profiles on the fly by utilizing the ontology. These dynamic user profiles help the recommender system to focus on the current scope of user's interest and retrieve those items that are closer to this particular area and do not involve all the user's preferences. This leads to spending less time in the information retrieval process and more accurate results. Any user is able to define a new situation during the project development process at real time and utilize the recommender system for relevant knowledge access.

7. References

Adomavicius, G. and Tuzhilin, A. (2011) Context-aware recommender systems. Recommender Systems Handbook. Springer.

Ahlers, D. and Mehrpoor, M. (2014) Semantic Social Recommendations in Knowledge-Based Engineering. CEUR Workshop Proceedings. vol. 1210.

Banon, S. (2012) "Elasticsearch", [online], http://www.elastic.co/products/elasticsearch

Bouneffouf, D. (2013) Situation-aware approach to improve context-based recommender system. arXiv preprint arXiv:1303.0481.
Bøvre, M., Kyllo, M., Aalberg, J., Nordal, I. O. S. and Imaz Abal, L. (2014) DNV GL Fuel Fighter: Development and Construction of Vehicles for Participation in Shell Eco-marathon 2014. Master Thesis, Norwegian University of Science and Technology.
Dey, A. K. (2001) Understanding and using context. Personal and ubiquitous computing, 5**,** 4-7.
Freund, L. S. (2008) Exploiting Task-Document Relations in Support of Information Retrieval in the Workplace. PhD Thesis, University of Toronto.
Garcin, F., Zhou, K., Faltings, B. and Schickel, V. (2012) Personalized news recommendation based on collaborative filtering. Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01, IEEE, 437-441.
Gunawardana, A. and Meek, C. (2009) A unified approach to building hybrid recommender systems. Proceedings of the third ACM conference on Recommender systems, ACM, 117-124.
Hatcher, E. and Gospodnetic, O. (2004) Lucene in action, Manning Publications.
He, W. (2013) Examining students' online interaction in a live video streaming environment using data mining and text mining. Computers in Human Behavior, 29**,** 90-102.
Hong, W., Zheng, S. and Wang, H. (2013) Dynamic user profile-based job recommender system. Computer Science and Education (ICCSE), 8th International Conference on, IEEE, 1499-1503.
Jayavel, S., Anouncia, M. and Kapoor, A. (2013) Semantic Search Engine. International Journal of Recent Contributions from Engineering, Science and IT (iJES), 1**,** pp. 19-21.
Lops, P., De Gemmis, M. and Semeraro, G. (2011) Content-based recommender systems: State of the art and trends. Recommender systems handbook. Springer.
Ma, H., Zhou, T. C., Lyu, M. R. and King, I. (2011) Improving recommender systems by incorporating social contextual information. ACM Transactions on Information Systems (TOIS), 29**,** 9.
Mehrpoor, M., Gjaerde, A. and Sivertsen, O. I. (2014) Intelligent services: A semantic recommender system for knowledge representation in industry. Engineering, Technology and Innovation (ICE), IEEE.
Paiva, F. A. P. D., Costa, J. A. F. and Silva, C. R. M. (2013) A Hierarchical Architecture for Ontology-based Recommender Systems. Computational Intelligence and 11th Brazilian Congress on Computational Intelligence (BRICS-CCI and CBIC), IEEE, 362-367.
Pazzani, M. J. and Billsus, D. (2007) Content-based recommendation systems. The Adaptive Web. Springer.
Protégé. (2015) "Protégé, open-source ontology editor", [online], http://protege.stanford.edu/
Ricci, F., Rokach, L., Shapira, B. and Kantor, P. B. (2011) Recommender systems handbook, Springer.
Schafer, J. B., Frankowski, D., Herlocker, J. and Sen, S. (2007) Collaborative filtering recommender systems. The adaptive web. Springer.
Suárez-Figueroa, M., Gómez-Pérez, A. and Fernández-López, M. (2012) The NeOn Methodology for Ontology Engineering. In: Suárez-Figueroa, M. C., Gómez-Pérez, A., Motta, E. & Gangemi, A. (eds.) Ontology Engineering in a Networked World. Springer Berlin Heidelberg.
Werner, D. and Cruz, C. (2013) A method to manage the precision difference between items and profiles: In a context of content-based recommender system and vector space model. Signal-Image Technology and Internet-Based Systems (SITIS), IEEE, 337-344.
Werner, D., Cruz, C. and Nicolle, C. (2013) Ontology-based recommender system of economic articles. arXiv preprint arXiv:1301.4781.
White, R. W., Bailey, P. and Chen, L. (2009) Predicting user interests from contextual information. Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. Boston, MA, USA: ACM.
Wimalasuriya, D. C. and Dou, D. (2010) Ontology-based information extraction: An introduction and a survey of current approaches. Journal of Information Science, 39: 211-224.
Yu, Z., Nakamura, Y., Jang, S., Kajita, S. and Mase, K. (2007) Ontology-based semantic recommendation for context-aware e-learning. Ubiquitous Intelligence and Computing. Springer.