IMPROVING THE MAPPING PROCESS IN ONTOLOGY-BASED USER PROFILES FOR WEB PERSONALIZATION SYSTEMS

Ahmad Hawalah and Maria Fasli
Department of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K.

Keywords: User profile, Ontology, Mapping, Concept clustering, Web personalisation.

Abstract: Web personalization systems that have emerged in recent years enhance the retrieval process based on each user's interests and preferences. A key feature in developing an effective web personalization system is to build and model user profiles accurately. In this paper, we propose an approach that implicitly tracks users' browsing behaviour in order to build an ontology-based user profile. The main goal of this paper is to investigate techniques to improve the accuracy of this user profile. We focus in particular on the mapping process which involves mapping the collected web pages the user has visited to concepts in a reference ontology. For this purpose, we introduce two techniques to enhance the mapping process: one that maintains the user's general and specific interests without the user's involvement, and one that exploits browsing and search contexts. We evaluate the factors that impact the overall performance of both techniques and show that our techniques improve the overall accuracy of the user profile.

1 INTRODUCTION

Although the Internet and the WWW have enhanced access to information, their rapid expansion have also caused information overload to such an extent that the process of finding a specific piece of information or a suitable product may often become frustrating and time-consuming for users. One way to deal with this problem is through adaptive or personalization web systems (Pignotti, Edwards and Grimnes, 2004, Challam, Gauch and Chandramouli, 2007, Sieg, Mobasher and Burke, 2007 and Pan, Wang and Gu, 2007). The ultimate objective of these systems is to provide personalized services or products with respect to each user's requirements. Today, the use of personalization systems is widespread in many application domains. For example, in the domain of e-learning, personalization has been used to provide each user with specific information that meets his or her needs and knowledge (Azouaou and Desmoulins, 2007). In the e-commerce domain, a personalization system plays an important role in recommending products or services based on the user's needs and interests; for instance, when a user navigates through a specific section in a retail book store, the system can recommend books that suit his or her characteristics and preferences (Gorgoglione, Palmisano & Tuzhilin, 2006).

Of course, all of these systems require some information about users in order to learn and respond to their interests and needs. Each system independently models and builds a user's profile, which is a representation of known information about that individual, including demographic data, interests, preferences, goals and previous history. However, one of the main challenges in current personalization systems is that they rely mostly on static or low-level dynamic user profiles (Felden and Linden, 2007 and Trajkova and Gaunch, 2004) which constrain the personalization process because they use the same user information over time, often leading to recommendations of irrelevant services as the user's needs and interests change.

One way to overcome this challenge is by building an ontological user profile that dynamically captures and learns the user's interests and preferences (Challam, Gauch and Chandramouli, 2007, Daoud, Tamine and Boughanem, 2008, Middleton et al. 2004 and Felden and Linden, 2007). The process of building such a profile is complex and requires multiple tasks. These tasks can be divided into three main phases: the information retrieval (IR) phase, which consists of preparing the reference ontology, collecting user navigation

Hawalah A. and Fasli M..
IMPROVING THE MAPPING PROCESS IN ONTOLOGY-BASED USER PROFILES FOR WEB PERSONALIZATION SYSTEMS.
DOI: 10.5220/0003179003210328
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 321-328
ISBN: 978-989-8425-40-9

behaviour, and mapping URLs to the reference ontology; the profile adaptation and learning phase, which forms the dynamic user profile; and the personalization engine phase, which provides recommendations based on the dynamic user profile. In this paper, we investigate the first phase, mapping URLs to a reference ontology, which is essential for the subsequent phases. Indeed, capturing inaccurate user interests in the first phase would directly affect the personalization performance. Therefore, this paper gives much needed attention to the first phase by introducing two novel algorithms that aim to improve the mapping process. These two algorithms each have characteristics that enhance particular aspects of the mapping process. The first algorithm, called Gradual Extra weight (GEW), is applied to an ontology to maintain a balance between a user's general and specific interests. The second algorithm, called the Contextual Concept Clustering (3C), is designed to exploit the user's context and thereby improve the mapping accuracy.

The paper is structured as follows. Next we discuss related work and following that we discuss the process of modelling the user profile. Section four presents the details of the two techniques and the next section presents a set of experiments that have been conducted along with the results. The paper ends with the conclusions and pointers to future work.

2 PREVIOUS WORK

A number of approaches have been presented to improve the overall accuracy of the mapping process. One such approach is to use a reference ontology with a limited number of levels. Liu et al. (2002), for example, mapped users' interests to a level-two ontology, while other approaches utilized a three-level ontology to map users' interests (Chen, Chen and Sun, 2002 and Trajkova & Gauch, 2004). As to retrieval precision, using a limited number of levels has been reported to improve overall accuracy, but a great limitation of this approach is that levels two or three of the ontology may still be too general to represent a user's actual interests.

Another approach that has been applied to map interests to an ontology is adding a pre-defined percentage of each sub-concept's weight to its super-concept. The idea behind this approach is that if a user is interested in a particular concept, then he or she also has some interest in its more general super-concept. Middleton et al. (2004) and Kim et al. (2007) applied this approach by adding an extra 50% for each concept's weight to its super-concept's weight, and then repeating the process until the root. Although this method offered an improvement over the original cosine similarity approach, its accumulation behaviour led to more emphasis on the top-level concepts, which are too general to represent a user's actual interests, while the middle and low-level concepts receive less attention. Daoud, Tamine and Boughanem (2008), on the other hand, assumed that representing interests with two levels of the ontology is too general, while leaf-node representation is too detailed, and that the most relevant concept is the one that has the greatest number of dependencies. Based on these assumptions, they proposed a sub-concept aggregation scheme, the main goal of which was to represent all user interests with three levels in the ontology. The weight of a level-three concept in this system is calculated by adding the weights of all its sub-concepts and then associating each user's interests to one level-three concept.

All of the approaches that have been proposed to improve the mapping process have limitations. Users tend to have general and specific interests on a wide range of topics. Therefore, assuming a two or three-level representation of all users' interests cannot be accurate or particularly effective. For instance, in the Open Directory Project (ODP) ontology, level-two Computers/Programming and level-three Computers/Programming/Languages are both too general to represent, for example, an interest in Java or C# programming languages. On the other hand, approaches that rely on adding extra weight to a super-concept based on its sub-concepts also suffer from a serious limitation since the accumulation behaviour leads to more emphasis on top-level concepts, which are too general to represent actual user interests. Therefore, we need a new approach that is capable of maintaining both general and specific interests. The focus of this paper is on introducing new techniques that can maintain a balance between general and specific interests.

3 PROFILE MODELLING

In this section, we present our approach for modelling ontological user profiles. The process of modelling user profiles involves three aspects: (i) tracking the user behaviour; (ii) using a reference ontology; (iii) mapping concepts to the ontology. These are explained in the subsequent sections.

3.1 Tracking User Behaviour

In order to learn and discover user interests, some information about users is required. Since collecting user data explicitly adds more burden on users (Kim and Chan, 2003), in this system we aim at collecting user browsing behaviour implicitly. The main data that we need to observe in this system is the visited websites and timestamp which denotes the date/time at which a website is visited. For each session, the user navigation behaviour is recorded and stored in a log file. After each session, the contents of each visited website are extracted. It is essential at this point to remove all the noise by applying text analysis techniques. Various algorithms are applied like tokenization, sentence splitting and stemming (Porter, 1980). After performing the text analysis, the processed data are stored in the processed log file (P-log file).

3.2 Using a Reference Ontology

Ontology representation is a rich knowledge representation which has been proven to provide a significant improvement in the performance of the personalization systems (Trajkova and Gauch, 2004, Challam, Gauch and Chandramouli, 2007, Daoud, Tamine and Boughanem, 2008, Middleton et al., 2004). In this paper, an ontology plays a key role in modelling the user profile. A reference (or domain) ontology provides a clear illustration of contents of a particular domain of application (Trajkova and Gauch, 2004). The ontology is modelled in a hierarchal way in which super-concepts are linked to sub-concepts. Another potential feature of using a reference ontology is that it could be agreed and shared between different systems, and therefore, user interests and preferences which mapped to the ontology can be easily shared between different systems. Unlike flat representations, a reference ontology provides a richer representation of information in that semantic and structural relationships are defined explicitly. In this paper, user interests are generated from the reference ontology based on the user browsing behaviour. After each browsing session, the visited websites are mapped to the reference ontology in order to classify each webpage to the right concept. A vector space mechanism is used in this paper as the main classifier (see equation 1).

$$\text{Term weight} = (tf_{ij} * idf_i) \qquad (1)$$

Despite the fact that the term frequency or vector space classifier is one of the simplest classification methods, it has a few drawbacks. One important drawback is that this classifier distinguishes between terms or vectors that have the same root. Words such as “play”, “plays” and “played” are processed as different words. This makes the classifier less effective in that the dimensionality of the terms increases. In order to reduce the dimensionality of the terms, we use the Porter stemming algorithm to remove term suffix and return each term to its stem (Porter, 1980). Stop words also can be removed from the reference ontology. Words such as “and”, “he” and “but” add more noise to the classifier and consequently lead the classifier to be less effective.

3.3 Mapping Concept to an Ontology

Once the term weights are calculated for each term in the ontology, any vector similarity method can be used to map visited web pages to appropriate concepts (or classes) in the reference ontology. In this paper, a cosine similarity algorithm (Baeza & Ribeiro, 1999) which is a well known algorithm is applied to classify websites to the right concepts.

4 IMPROVING THE MAPPING PROCESS

In this section we introduce two novel approaches that are capable of maintaining the user’s general and specific interests without the user’s participation.

4.1 Gradual Extra Weight (GEW)

The idea behind GEW is that if a user is interested in a particular concept, then he also has some interest in its super-concept. Unlike other approaches that were discussed in section 2, in this approach we make no assumption about the number of levels in an ontology as the specification of each ontology varies. Additionally, we do not assign a specific percentage of a sub-concept to be added to its super-concept. Instead, we propose an auto-tuning mechanism in which the percentage value of each sub-class that is added to its super-class is tailored to different levels on the ontology (see equation 2).

$$\text{Extra percentage } (\boldsymbol{EP}) = (\mathbf{CL}/2)*10 \qquad (2)$$

Where:
CL: the current sub-class's level

In this approach, we assume that the concepts deep in any ontology are more closely related than those in higher levels. Therefore, the Extra Percentage in our approach is calculated by dividing the level of a sub-concept by two and then the result is multiplied by 10 to give the extra percentage to be added from the sub-concept to its super-concept. As we move up towards the root, the percentage is reduced. In this case, we keep a balance between the general and the specific interests. Algorithm 1 describes the procedure used to map and calculate the EP.

Input: reference ontology and web pages that need to be mapped
Output: URLs with top **α** concepts from the ontology with updated similarity weights
RO= reference ontology
RO= $\{c_1, \dots, c_n\}$, concepts with associated documents.
c_{Level} = *level of a concept c*
LOG= log file that contains user's browsing behaviour
LOG =$\{URL_1, \dots, URL_n\}$ *visited web sites.*
*EP= Extra percentage= (CL /2)*10)*
SR= Sim Results between URL and concepts after applying GEW.
// apply original Cosine similarity for each URL

Foreach $URL_i \in LOG$ **do**
  Extract contents;
  Apply dimensionality reduction techniques;
  Calculate TFIDF;
  Foreach $c_i \in$ RO **do**
    Calculate sim($URL_{contents}$, c_{doc}); // using cosine algorithm
    SR.Add(URL_i,c_i,sim);// add URL, concepts and sim weight.
  End
  // Select top **α** concepts to apply GEW on.
  SR.sort by weight;
  Foreach $c_i \in$ RS and c_i.count $\geq$ **α do**
    If sim.weight > 0 **then**
      Calculate **EP= (**c_{Level}/2)*10**;**
      Extra_weight= EP× $c_i.sim\_weight$;
      c_i.superclass-weight += Extra_weight;
    End
  End
  SR.sort by weight; // re-order SR after applying GEW.
End

Algorithm 1: Gradual Extra Weight.

4.2 Contextual Concept Clustering (3C)

Though the GEW approach may improve the process of mapping web pages to concepts, correct mapping cannot be guaranteed as not all the visited web pages usually have good representative contents. Therefore, we further improve the mapping process by taking advantage of having a log file that stores the entire user's browsing behaviour. Usually, when users browse the Internet, they tend to visit several web pages that represent one interest. We call our mechanism Contextual Concept Clustering (3C) because the context of the user behaviour is considered. To illustrate, a visited web page could be clustered to one concept in one session, but in another session, the same web page could be grouped to a different cluster. This behaviour could be significant in the process of finding the right user interests. Therefore, we introduce the 3C mechanism that aims at grouping related web pages with the same concept into one cluster. For each browsing session, we first apply the GEW approach on each concept for each URL. After applying the GEW to all concepts, the top $\boldsymbol{\beta}$ similarities are used to represent each web page. We select the top $\boldsymbol{\beta}$ results because in some cases the concept with the highest similarity does not give a correct view of a web page. This could be due to poor or irrelevant information in a web page, or it could be simply due to a high level of noise. As a result, we avoid such a scenario by considering all the top $\boldsymbol{\beta}$ concepts and treat them as possible candidates. The context is then exploited by selecting the common concept that is associated with different web pages. This common concept eventually is selected to represent a web page. If there is no common concept, the concept with the highest similarity weight is selected. The full 3C algorithm is described in Algorithm 2.

Input: Similarity results (SR) after applying GEW.
Output: URLs mapped to concepts.
RO= reference ontology.
RO=$\{c_1, \dots, c_n\}$, concepts with associated documents.
LOG= log file that contains user's browsing behaviour
LOG =$\{URL_1, \dots, URL_n\}$ *visited web sites*
CLU_CON= { *TC*1, ...,*TC*n},*concepts with total sum of weights*
FIN_CLU= final result after applying 3C algorithm
// Select the highest $\boldsymbol{\beta}$ concepts similarity for each URL.
For each $URL_j \in$ SR **do**
  Select top $\boldsymbol{\beta}$ concepts;
End
// find all concepts that appear in different URLs
For each distinct $C_j \in SR$ **do**
  Total_weight= 0;
  CLU_CON.Add(C_j, Total_weight);
End
For each $TC_j \in CLU\_CON$ **do**
  For each $URL_j \in$ SR **do**
    If $URL_j.Contains(TC_j)$ **then**
      Total_weight += $TC_j.weight$;
    End If
  End
  CLU_CON.update(TC_j, Total_weight);
End
CLU_CON.sort by Total_weight DESC;
// assign correct concepts to URLs
For each $TC_j \in CLU\_CON$ **do**
  For each $URL_j \in$ SR **do**
    If $URL_j.contains(TC_j)$ **then**
      If *FIN_CLU.Does_not_contain(*URL_j*)***then**
        FIN_CLU.Add(URL_j, TC_j*;)*
      End if
    End if
  End
End

Algorithm 2: Contextual Concept Clustering (3C).

5 EVALUATING GEW AND 3C

In order to evaluate the two proposed approaches we first create a reference ontology using the Open Directory Project (ODP). Then, we build the classifier using the TF-IDF classifier. In the next stage, we create a set of tasks and invite 5 users to perform these tasks. Finally, we evaluate different characteristics that impact on the overall performance of both algorithms, and then we employ four different mapping approaches to test and compare them individually. The following sections describe all these stages in detail.

5.1 Creating a Reference Ontology

For evaluation purposes, we use the ODP concept hierarchy as a reference ontology (ODP, 2010), and more specifically the ***computer*** category. The computer directory contains more than 110,000 websites categorized in more than 7000 categories. In order to train the classifier for each category, all the websites under each category were fetched. Furthermore, all the contents of all websites were extracted and combined in one document. That is, each category is represented by one document. All the non-semantical classes (e.g. alphabetical order) were removed to keep only those classes that are related to each other semantically. This resulted in a total of 4116 categories and about 100,000 training websites whose contents were extracted and combined in 4116 documents in our reference ontology. Dimensionality reduction techniques such as Porter stemming and stop words removal are also applied to all the 4116 documents. The TF-IDF classifier (equation 1) is then used to give each term $\boldsymbol{t}$ in each document $\boldsymbol{d}$ a weight from 0 to 1.

5.2 Collecting Real Usage Data

In order to collect user browsing behaviour, a Firefox browser is used with a modified add-on component called Meetimer (Meetimer, 2010). A SQLite database is used to store all the user's sessions. For the purpose of the evaluation process, 35 different concepts from the computer ontology were selected, and a set of tasks were created. Tasks took the form of finding a specific piece of information, or writing a short paragraph. Five users were invited to perform a total of 90 tasks in a one month period. For each session throughout the month, users were asked to select 3 tasks and try to answer these tasks by browsing and searching for related web pages. The sequence of the tasks was not fixed but users were given freedom. After each session, users were asked to write down what tasks they performed. These data represent users' actual interests which will be matched against the mapped concepts generated by different mapping approaches.

After one month, five log files from five different users were collected. These five users together surfed 1,899 web pages. We started processing the collected data to create processed log files (P-log files) by fetching all the visited web pages, and extracting all their contents. We then applied the GEW and 3C algorithms and compared the accuracy results against the users' actual interests. Next we describe what aspects we have analyzed and what experiments have been conducted.

5.3 Experiments

Three experiments are proposed to analyze different aspects that impact on the overall performance of GEW and 3C.

5.3.1 Pruning Non-relevant Concepts

In this experiment, we want to determine a threshold value ($\boldsymbol{\alpha}$ in GEW algorithm) that could remove non-relevant concepts to create a more accurate user profile. For this reason, we apply the GEW algorithm to: all the retrieved concepts from the original cosine similarity, top 50, top 20, top 10 and top 5 results. We use the following accuracy measure (equation 3) to compute the accuracy.

$$Accuracy = \frac{\#\ of\ positive\ mapped\ web\ pages}{total\ number\ of\ web\ pages} \qquad (3)$$

Table 1 shows the accuracy percentages for all the five thresholds after comparing all concepts with the users' actual interests. In Table 1, it can be clearly seen that that the accuracy of applying the GEW algorithm to all the concepts is relatively low (30%). While applying GEW to the top 50, 20, 10 and 5 concepts achieved a considerable increase in the accuracy (71%, 74.90%, 76% and 75.35% respectively). This shows that applying the GEW to all concepts could cause inflation in the weight of the non-relevant concepts. However, applying GEW on the top 10 results provided the highest accuracy. This is because the top 10 results could hold the most important concepts that are likely to be related to a web page. Based this results, we assign $\boldsymbol{\alpha}$ in the GEW algorithm to be 10 in the next experiment.

Table 1: Accuracy of all web pages that visited by all users considering different threshold values.

Threshold	Top5	Top10	TOP20	TOP50	All
Precision	75.3%	76.8%	74.9%	71%	30%

In the following experiment, we try to identify the value of $\boldsymbol{\beta}$ which is used in the 3C algorithm as a threshold. In the next experiment, we apply the 3C algorithm to the top 30, top 20, top 10 and top 5 URLs. Table 2 shows the accuracy percentages for all the four thresholds after comparing all concepts with users' actual interests.

Table 2: Accuracy of using different threshold values for the 3C algorithm.

Threshold	top 5	top 10	top 20	top 30
Accuracy	76.8%	76.1%	75.2%	73.2%

It can be clearly seen from the table 2 that the accuracy of all the thresholds have achieved close accuracy results. However, the top 5 threshold achieved the highest accuracy result, while top 10 and top 20 achieved slightly less accuracy results. Based on the above results, we assign $\boldsymbol{\beta}$ in 3C algorithm to be 5 in all the following experiments.

5.3.2 Rank Ordering

In this experiment, we analyzed the performance of the 3C algorithm when the concepts are clustered based on ordering concepts by number of web pages and by the total similarity weight for each concept. For both techniques, we calculated the precision for each user's profile. Figure 1 shows the ordering accuracy results for both techniques.

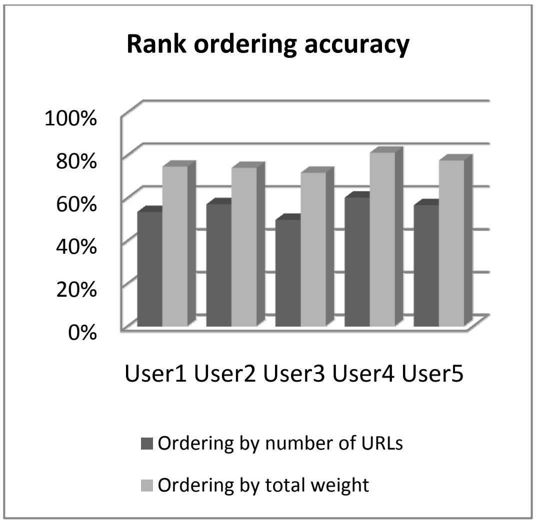

Figure 1: Rank ordering accuracy for each user.

It can be clearly seen that there is a considerable difference between ordering concepts by number of web pages and by total similarity weight for each concept. This could be due to the fact that many concepts in the log file could share the same super-concept. As a result, when clustering those concepts by the number of URLs, the common super-concept which is likely to be too general is selected. Consequently, most of the concepts in the user profile will be too general to represent users' actual interests. On the other hand, ordering concepts by the accumulated weight rather than the number of URLs, achieved a high average accuracy of about 75.68%. This result demonstrates that clustering and ordering concepts by the accumulated similarity weights provides more effective representation of users' interests and preferences.

5.3.3 Comparing Mapping Techniques

In this experiment, we aimed at comparing our mapping techniques (GEW and 3C) to three different mapping techniques in the literature. The first technique is the original cosine similarity which computes the similarity between each URL and all documents in the ontology. The second technique which was suggested by Middleton et al. (2004) and Kim et al. (2007), is adding 50% of each sub-concept's weight to its super-concept, and repeats this process until the root. Finally, the last approach is the Sub-class Aggregation Scheme that was proposed by Daoud, Tamine and Boughanem (2008). For this experiment, each visited web page for each user was mapped by applying all four techniques. Figure 2 illustrates the overall accuracy for each mapping technique for each user.

According to Figure 2, it is clear that the original cosine similarity and sub-class aggregation schemes performed poorly for all users (46.17% and 45% respectively). The main reason that the original cosine similarity showed the lowest precision is that the most inaccurate mapped concepts are too specific and detailed. Similarly, the sub-class aggregation scheme showed a poor precision of 45%. This is because all the visited web pages were mapped to only level three classes. The accumulation behaviour of adding the weights of all the sub-classes under the level three super-classes causes inflation on the weights of level three super-classes. Consequently, no level two classes were ever mapped to any web page. On the other hand, adding 50% of each sub-class to its super-class shows a considerable improvement in the accuracy average of 60%. This improvement could be attributed to the fact that if a user is interested in a concept, then he/she also has some interest in its super-concept. Although this method improved the overall mapping precision, the percentage which is added to the super-classes is very high (50%). As a result, 80% of all the incorrectly mapped web pages were mapped to level 1 and 2 super-concepts that are too general to represent user interests. For our proposed GEW and 3C algorithms, the reported results were interesting. The overall precision shows a noteworthy improvement of average of 75%. This major improvement demonstrates that GEW and 3C can overcome some of the drawbacks of other approaches. Furthermore, the GEW and 3C methods have shown to keep a balance between the general and the specific interests. Nevertheless, although the GEW and 3C achieved great results, they have one limitation. That is, the 3C approach does not take into the account the semantic relationships between concepts. In order to improve the performance further and as part of our future work these relationships need to be taken into account.

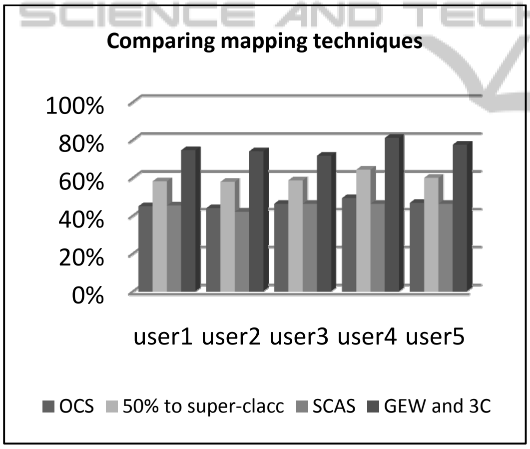

Figure 2: A comparison of 4 different mapping techniques: OCS: original cosine similarity, 50% from sub-class to its super-class, sub-class aggregation scheme and GEW & 3C.

6 CONCLUSIONS

Web personalization systems enable users to search for and retrieve information which is tailor-made to their interests and preferences. However, creating an accurate user profile unobtrusively and adapting it dynamically is a complex problem. In this paper, we presented two novel mapping algorithms (GEW and 3C) that were used to improve the overall accuracy of the ontological user profile. Our paper revolves around discovering user interests by mapping visited web pages to an ontology based on the user browsing behaviour. Our evaluation results demonstrate that applying the GEW and 3C mapping algorithms for modelling user profiles can effectively improve the overall performance. The experimental results show that the process of mapping user interests can be significantly improved by 28% when utilizing the GEW and 3C algorithms. As part of further work, we will try to enhance the mapping process further by exploiting the semantic relationships between concepts in the ontology.

REFERENCES

Azouaou, F. and Desmoulins, C., 2007. Using and modeling context with ontology in e-learning. *International Workshop on Applications of Semantic Web Technologies for E-Learning*, Dublin, Ireland.

Baeza, R. and Ribeiro, B., 1999. *Modern Information Retrieval.* Addison-Wesley.

Challam, V., Gauch, S., and Chandramouli, A., 2007. Contextual Search Using Ontology-Based User Profiles, *Proceedings of RIAO 2007*,Pittsburgh, USA.

Chen, C., Chen, M., Sun, Y., 2002. PVA: A Self-Adaptive Personal View Agent System. *Journal of Intelligent Information Systems,* vol. 18, no. 2, pp. 173–194.

Daoud, M., Tamine, L., Boughanem, M., 2008. Learning user interests for session-based personalized search. *In ACM Information Interaction in context (IIiX).*

Felden, C. and Linden, M., 2007. Ontology-Based User Profiling. *Lecture Notes in Computer Science*. vol. 4439, pp. 314-327.

Gorgoglione, M., Palmisano, C. and Tuzhilin, A., 2006. *Personalization in Context: Does Context Matter When Building Personalized Customer Models?* ICDM 2006. pp. 222-230.

Kim H., Chan P.,2003. Learning Implicit User Interest Hierarchy for Context in Personalization*. In Proceedings of the 2003 International Conference on Intelligent user interfaces 2003*, Miami, Florida.

Kim, J., Kim, J., Kim, C., 2007. Ontology-Based User Preference Modeling for Enhancing Interoperability in Personalized Services. *Universal Access in HCI,* Part III, HCII 2007, LNCS 4556. pp. 903–912.

Lin, R., Kraus, S. and Tew, J., 2004. Osgs - a personalized online store for e-commerce environments. *Information Retrieval journal*, vol.7(3–4).

Liu, F., Yu, C., Meng, W., 2002. Personalized web search by mapping user queries to categories. *In Proceedings of CIKM.* Virginia, vol.02, pp.558-565.

Meetimer, 2010. https://addons.mozilla.org/en-US/firefox/addon/5168/, April 2010.

Middleton, S., Shadbolt, N., Roure, D., 2004. Ontological user profiling in recommender systems. *ACM Trans.*

Open Directory Project, 2010. http://dmoze.org, April 2010.

Pan, X., Wang, Z. and Gu, X., 2007, Context-Based Adaptive Personalized Web Search for Improving

Information Retrieval Effectiveness, *International Conference on Wireless Communications, Networking and Mobile Computing*, 2007. pp.5427-5430.

Pignotti E., Edwards, P. and Grimnes, G., 2004. Context-Aware Personalised Service Delivery. *European Conference on Artificial Intelligence, ECAI 2004*. pp.1077–1078.

Porter, M., 1980. An algorithm for suffix stripping. Program 14,3,pp.130-137.

Sieg, A., Mobasher, B. and Burke, R., 2007. Representing context in web search with ontological user profiles. *In Proceedings of the Sixth International and Interdisciplinary Conference on Modeling and Using Context*. Roskilde, Denmark.

Trajkova, J., Gauch, S., 2004. Improving ontology-based user profiles. *In Proceedings of RIAO.*

Van, C., Robertson, S., Porter, M., 1980. New models in probabilistic information retrieval. British Library Research and Development Report, no. 5587. London. UK.