RECOMMENDATION OF PERSONALIZED RSS FEEDS BASED ON ONTOLOGY APPROACH AND MULTI-AGENT SYSTEM IN WEB 2.0

YASYN ELYUSUF1, MOHAMMED AMINE ALIMAM2, HAMID SEGHIOUER3

1, 2, 3 Laboratory, Operational Research and Applied Statistics LIROSA
Faculty of Sciences, University Abdelmalek Essadi, Tetuan, Morocco.
E-mail: 1 elyusufyasyn@gmail.com, 2 amine.alimam@gmail.com, 3 hamid.seghiouer@gmail.com

ABSTRACT

Nowadays, multi-agent systems (MAS) are used in many fields such as industry, education, finance, etc. MAS counts among the most promising technological paradigms in the development of Web applications. They can contribute significantly to improving the quality of the use of these applications. In this paper, we propose a new recommendation approach for personalized RSS (Really Simple Syndication) feeds in Web2.0 applications (Wiki, Blogs, Social Networks…). This approach is based on a process of coupling ontologies and multi-agent systems and benefits from the synergy between the two concepts. In this context, we propose a model which allows human agents (scientists, experts …) to cooperate with software agents to help the system to recommend personalized RSS feeds for users’ interests. In the first section we present personalization techniques using the ontological approach. In the second section we present the new concept of coupling between Multi Agent Systems and ontologies. In the third section, we present our approach to identifying ontology-based user profiles and multi agent system for personalized RSS feeds. The conclusion comes in the last section.

Keywords: Recommendation, User Profile, RSS, Ontology, MAS, Web2.0.

1. INTRODUCTION

Actually the number of people using Web2.0 applications is increasing exponentially. In addition, the functions of these applications are becoming far richer. The traditional approach followed to enable the users to control the complexity and richness of these applications consists in making the user interaction with these simple applications easier, through meeting the expectations of all users [1]. Several methods have been used for personalizing information and search in the web. M. Daoud and L. Tamine - Lechani [2] present an approach for building ontology-based user profiles to obtain personalized search for information. In the same way, M.A. Alimam, Y. Elyusufi, H. Seghiouer present in [3] an automatic user profiling for personalized access to information based on ontological approach. I.F Moawad and H. Talha present in [4] another personalization method based on a multi-agent system to personalize the web search results. The major limitation of the classical profiling is that it is based on a general approach that consistently evaluates user requests and deliver results without considering the context of research. However as stated in [5], different personalization techniques are combined to overcome the disadvantages of single personalization techniques in recommendation systems, this is called a hybrid recommender system. In this very context, we propose a new recommendation method for personalized (Really Simple Syndication RSS) feeds in Web 2.0 applications (Wiki, Blogs, Social Networks,…), based on a coupling process of ontologies and multi-agent systems to take advantage of the synergy between the two approaches. In this paper we take the ontological approach and Multi-Agents-System as the basis to identify user profiles to be used to recommend personalized RSS feeds for each profile. The use of a standard publication language, like RSS, will help to simplify our architectural design and to solve the interoperability problems. This choice will simplify the personalization of the web content by searching only for Web repositories in relation with the interest centers of users. To implement this method we will adopt software agents that interact with human agents (experts, scientists, …) in order to categorize internet users in terms of profiles. The proposed system is based firstly on MAS to manage the analysis of the text stream generated by the
users, and secondly on the reference ontology to categorize users according to suitable profiles. This categorization is mainly based on three process: analysis of messages, profile identification and finally the recommendation process of personalized RSS feeds for each profile.

2. USING ONTOLOGY IN PERSONALIZATION

P. Hagen, H. Manning, and R. Souza define personalization in [6] as follows " the ability to provide content and services adapted to users based on their preferences and behaviors ". Several areas of application are using users profiles, for reasons related to personalization, with different needs. Depending on the area, personalization consists of one or more of the following tasks: filtering a flow of information, guide the search in an wide information space, recommend a set of information to the user, adjust the results of a request to the profile, adapt the interaction to the user situation (interface, interaction). In this context M. Daoud and L. Tamine-Lechani present in [2] an approach for building ontology-based user profiles to personalize search information. On the other hand, B. R. Suteja, S. Guritno, R. Wardoyo and A. Ashari present in [8] an ontology- based approach for online learning personalization in order to guide the learning process. Also, M.A. Alimam and H. Seghiouer [9] present another approach based on ontology for personalized E-Learning environment. In the same way K.M. Oliveira, F. Bacha, H. Mnasser and Mourad present in [10] a new application of a domain ontology to generate personalized user interfaces. Y. Elyusufi, H. Seghiouer and A. Benkaddour present in the same direction [17] a technique for personalizing web interfaces based on ontological approach. However, the utilization of ontologies in user profiling techniques has gained much attention since it allows inference to be employed, enabling interests to be discovered that were not directly observed in the user’s behavior. The method we propose in this article will enable the analysis of the users’ messages by using Multi-agent approach. This new concept of coupling between MAS and ontologies will firstly help the system in user profiling process, and secondly facilitate the communication between the ontology designers (Scientists, experts . . .).

3. COUPLING MAS AND ONTOLOGY

Both approaches ‘ multi-agent systems ’ and ‘ ontologies ’ are often mentioned together in the literature [11][12][13]. We can only assume that they are two related areas. More specifically, they constitute two complementary approaches. Multi-agent systems are used to answer questions related to communication. However, ontologies enable to solve common problems of language and representation of knowledge domains. The main link between multi-agent systems and ontologies is the use of ontologies to make multi-agent systems function [11], particularly in the case of agents in Semantic Web [12]. There are also multi-agent systems that support processing on ontologies. In this sense K. Lister, L. Sterling and K. Taveter used MAS to align the ontology resources [13]. Moreover the communication between agents is of paramount importance for multi-agent systems (MAS) [14] [15]. This communication is a fundamental means to ensure the distribution of tasks and coordination of actions between agents, where main task of an SMA is to develop a standard for communication between agents [16]. This standard is designed for the exchange of information, knowledge and services between agents. We have noticed while creating user profiles in [17] a lack of communication between human actors using the reference ontology. The architecture that we propose in this paper is based on the use of MAS for the analysis of the users’ generated text stream. The system is designed so that each actor (experts, scientists, . . . ) has a cognitive agent allowing to compare the users’ interests with instances of the reference ontology. This ontology will provide specific vocabularies depending on scope for communication between agents and define the concepts and relationships that exist between the words of a formal vocabulary for the agents to use. Among the key areas outlined in the reference ontology, we have computer science, mathematics, physics, chemistry, biology, medicine, geology, economics and logistics (see. Figure1).
We associate to each ontology subject (class) a set of keywords. These keywords are provided by the ontology designers according to domain of users, where each keyword is linked to a set of RSS feeds (i.e. keyword = (RSSlink1, RSSlink2, ..., RSSlinkn)) (see. Figure 2). This structure will help the system to automatically identify user profiles by using Multi Agent System.

This new method of coupling between SMA and ontology proposed in this work will firstly enable to structure a set of domains as a reference ontology, and secondly to facilitate communication between software agents used to automate analysis and processing of users’ messages. This process will then help the system identify user profiles, and then facilitate the recommendation process of RSS feeds.

4. PROPOSED ARCHITECTURE

The architecture we propose is based on cognitive agents that allow interaction with the reference ontology. These agents are in constant communication with a manager agent (MA). Their ability to store informations, analyze messages and manage reference ontology enables to categorize users into profiles, and finally to decide on the RSS feeds that will be recommended for each user profile. The proposed architecture is based on six agents (see. Figure 3):

- Interface Agent (IA);
- Manager Agent (MA);
- Terminologist Agent (TA);
- Scientist Agent (SA);
- Profile Manager Agent (PMA);
- Filtre Agent (FA);
4.1 Interface Agent (IA)

This agent acts as an interface between the user and other agents through Agent Manager. The agent uses the KQML language interface for communication with other agents. It receives the message flow generated by users, then proceeds with the detection of keywords (KW), marking of the KW and then the calculation of the frequency of each KW (for more explanation section 5). Next, it reformulates the result of the analysis and returns the result of treatment to the agent Manager in KQML language.

4.2 Manager Agent (MA)

This agent controls all the operations performed by the system and assigns tasks to agents according to their roles. The agent MA plays an orchestral role between the various agents. Once the agent (MA) receives the request of the agent (IA), it determines the nature of the application to select agents that can handle this request. Requests exchanged between the IA and MA are identified by an identification number.

4.3 Scientist Agent (SA)

Scientist Agent (SA) is an agent associated with a human expert in (Mathematics, sociology . . . ) and communicates with other agents. Its main role is to qualify keywords (KW)s with instances of the reference ontology. Through communication with the agents Manager Agent (MA) and Filter Agent (FA), it allows the profile identification by communication with (TA).

4.4 Terminologist Agent (TA)

Terminologist Agent (TA) is an agent associated with another human expert in terminology who communicates in turn with the agents: Manager Agent (MA), scientist Agent (SA) and Filter Agent (FA), and compares the keywords (KW)s with instances of the reference ontology to help the system perform the categorization of users in their suitable profiles.

4.5 Profile Manager Agent (PMA)

The role of Profile Manager Agent (PMA) is to create, initialize, then treat the user profile after a request from (IA). It can also add a new profile and
initialize resources (attributes and values). It also allows changing the profile setting, and then returning the result of modifications to (IA).

4.6 Filter Agent (FA)

Filter Agent (FA) is a profile filtering agent according to profile interests. When (FA) receives responses from (SA) and (TA), it is able to categorize each user in a profile, and then return the filtering result to (IA) through (MA).

5. ANALYSIS OF MESSAGES

In most Web 2.0 applications (blogs, wiki . . . ), messages are usually short. This makes message parsing more difficult [19] [20]. Therefore, we performed, as already introduced in some works [20] [21], a linguistic analysis of messages generated by users. This analysis consists in extracting the keywords (KW)s of stream text generated by each user. These (KW)s can help characterize the users’ profile: Each KW refers to a generic concept, and the keyword (KW) is considered as an instance of this concept (e.g. dentist belongs to the field of dentistry). In our work, taking into account the need to characterize the profile as soon as possible, the analysis of user text stream is programmed at the Interface Agent (IA) and is based on several treatments (KW detection, KW marking and calculation of frequency of each KW). This process can be seen as a pre-treatment of data and their normal formatting so as to be processed. The main idea is to analyze the text flow of a user and extract meaningful keywords that describe the content. These keywords are stored to be compared with the instances used in the reference ontology. We then associate a weight to each keyword that expresses the importance of each term in the text stream. This weight corresponds to the occurrence frequency of terms. We may notice in any text that many words have slightly different forms, but their meaning remains the same or very similar. This is particularly the case of inflected words. For example, the following words have very similar meanings: transform, transforms, transforming, transformer . . . These words have the same root (lemma). Thus, if we can eliminate the word endings and keep only the root, we obtain identical forms for these words; this is the idea which led to the use of lemmatization. Before the calculation of the weights of terms for each user, we apply three filters on text stream. The first removes spaces and punctuation marks, the second allows converting uppercase to lowercase, and the third allows finding the roots of words (lemmatization). This is the most widely used method to lemmatize words, which was developed by M.F. Porter [22].

6. IDENTIFICATION OF USER PROFILES

After analyzing the text stream and detecting the keywords for each user, we proceed with the identification of profiles. Several methods of user profile identification have been proposed in the literature [23] [24] [25]. According to G. Salton [23], a user profile is based on the Vector Space Model. This vector consists of one or more vectors defined in terms of space. M. Pazzani, J. Muramatsu, D. B. Syskill and Webert define in [24] a set of triples formed by: (keyword, the probability for this keyword to belong to a relevant document, and the probability of belonging to an irrelevant document) are determined for each user. This set constitutes the profile and characterizes the relevance of documents to a user. According to H. Lieberman, N. W. Van Dyke and A. S. Vivacqua [25], the user profile is made from a list of keywords that represents the relevant elements for a user. The process of identifying profiles we present is programmed at the level of SA and TA, and it’s based on the weights calculated for each keyword (e.g. the word ‘J2EE’ is found 13 times in the text stream generated by the user ‘U124’, this means that the weight of ‘J2EE’ term is equal to 13 in the text stream of the user ‘U124’). We distinguish between two types of connections: Online mode and Offline mode. The Online mode calculates the keyword weight during the last connection, while the Offline mode calculates the weight of the keywords since the user’s registration (see Figure 4). After this calculation, the most used keywords by the users are compared with instances of the reference ontology, e.g. the most used keyword by U124 is ‘J2EE’, and this keyword belongs to ‘Computer Science’ class, especially the subject of web development, this means that ‘U124’ belongs to the ‘web development profile’. This process will enable the user’s identification according to the suitable profile, and then facilitate the recommendation process of personalized RSS feeds for each profile.
7. RECOMMENDATION SCENARIO OF PERSONALIZED RSS FEEDS

The recommendation scenario of RSS (Really Simple Syndication) feeds is based on a communication process between agents. The interface agent (IA) intercepts the flow of messages produced by each user, and then proceeds with the analysis of massages (KW detection, KW marking and calculation of frequency of each KW) (section 5). Then, it sends the result of treatment to the Agent Manager (MA) in KQML language. Once the agent (MA) receives messages from the agent (IA), it delegates keywords qualification to the specialist agents (SA, TA) by comparing the keywords (KW)s with instances of the reference ontology (section 6). After that, the agent (PMA) intervenes to create or modify the profile parameters. Then, the agent (FA) categorizes users into their proper profiles before returning the result of recommendation to the profiles identified through the manager agent (MA) (See. Figure 5).

Thanks to this effective communication between different actors, and the data presented at the level of reference ontology, the system will be able to recommend custom RSS feeds for each profile, and then help users to automatically monitor the content of their favorite web sites with no need to consult them.
8. EXPERIMENTATION AND OUTCOMES OF RECOMMENDATION

The proposed system consists of a platform that will allow scientific research to classify users based on profiles. The latter may then receive personalized access to RSS feeds corresponding to their interests. The efficiency of this system will be determined experimentally. We took 30 different scientific categories of users (students, researchers and professors), belonging to the same community (faculty of sciences), and using our E-learning 2.0 prototype. This sampling will allow assessing the possibilities of identification and recommendations. Trying to make the used evaluation as clear as possible, the multi-agent system was performed with varying sets of input data, a dozen of scientific terms. The obtained value for each profile corresponds to the most visited scientific fields by the user in offline mode. This value is associated with a much larger weight when the search is online. Thus, we have established an experimental protocol followed in this section. It takes place in two phases. First, the system is started using as input the terms associated with the present concepts in the reference ontology. Our goal is to examine the mechanisms of structuring and behavior of the system. We ensure as well that vocabulary is structured according to our reference ontology. The system is then launched in fully automatic mode without the user's intervention. This phase allows inspecting the behavior of the system on a purely statistical level. After analysis, our approach seems to elicit categories which are representative of the given sample. The most of RSS-feeds are grouped consistently in the appropriate categories (type of profiles). In particular, let us observe our lattice produced in a more specialized classification of feeds (see Table 1).

Table 1: RSS Recommendation

<table>
<thead>
<tr>
<th>Profiles</th>
<th>Concept</th>
<th>User</th>
<th>Recommended RSS links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Networks profile (20)</td>
<td>CISCO (28)</td>
<td>User06</td>
<td><a href="http://newsroom.cisco.com/rss-feeds">http://newsroom.cisco.com/rss-feeds</a></td>
</tr>
<tr>
<td>Programming profile (62)</td>
<td>RUBY (60)</td>
<td>User11</td>
<td><a href="http://www.camilleouc.com/2016/03/15/20160310-00-00">http://www.camilleouc.com/2016/03/15/20160310-00-00</a></td>
</tr>
<tr>
<td>Web profile (32)</td>
<td>PHP (2)</td>
<td>User15</td>
<td><a href="http://web.devoppes.com/index/rss">http://web.devoppes.com/index/rss</a></td>
</tr>
</tbody>
</table>

Table 1 shows some categories and the corresponding specialization computed by our approach. Just to give an example, the public category "Database" is split in some further subcategories such as "Oracle", "Data warehouse", "SQL", etc, which correspond to some students studying in IT field (offline recommendation), but also to other categories performing this kind of web search at a given time (online recommendation).
9. CONCLUSION

We have presented in this paper a new technique for personalizing RSS feeds in Web 2.0 applications. To implement this technique, we introduced a new method of coupling between the ontological approach and multi-agent system. This model primarily allows efficient communication between different actors in the process of identifying profiles, while benefiting from the messages submitted by users, and using reference ontology in user profiling process. This new approach can help the system to categorize users into profiles respecting their interests centers, and then facilitate the recommendation process of RSS feeds. A future development of this method will target the selection of relevant Web pages whose content is linked to the interests of each profile with the aim to optimize the recommendation process.

REFERENCES:


