Fighting Information Overflow with Personalized Comprehensive Information Access: A Proactive Job Recommender

Danielle H. Lee & Peter Brusilovsky
School of Information Sciences
University of Pittsburgh
135 N. Bellefield Ave. Pittsburgh PA 15260, USA
hyl12@pitt.edu peterb@pitt.edu

***Abstract*—Searching for jobs online is an information intensive activity, because thousands of jobs are posted on the Web daily and it takes a great deal of effort to find the right position. Job search sites require recommender systems to meet diversified information needs: Job seekers who have well-defined careers try to focus on relevant open positions while students who have general and evolving interests want to follow the dominant trends of the job market in order to plan their career path. In this paper, we introduce a comprehensive job recommender system. From the user's perspective, four different kinds of recommendations are implemented. Users of this system can retrieve open jobs with different methods, ranging from exploring to searching.**

Keywords: Job recommender, information retrieval from multiple views, exploratory search

I. INTRODUCTION

The reliance on information is increasing in our daily lives. . When buying an electronic appliance, searching a map, or looking for house rent, information on the Web is widely referred to. However, the amount of available information is increasing sharply and thus the processing of information relevant to each user consequently requires a huge cognitive effort. In order to cope with this information overload problem, personalized information retrieval technology focusing on user's exact needs is required (Brusilovsky, 2001). On the other hand, diversified user populations expect more ways to thoroughly explore various aspects of the information. The problem is how a system can meet these two contradictory requirements at the same time.

A range of *information access technologies* have been invented to help humans find the "needles" of relevant information in the "haystacks" of available information. The traditional toolset of these technologies includes information filtering, information retrieval (search), and hypertext browsing. The continuing growth of information overflow has caused researchers and practitioners to explore *personalized* information access technologies, which take into account the user's individual needs, interests, preferences, knowledge, etc. to increase the user's chances of finding the right information. Adaptive search, adaptive filtering (recommendation), and adaptive hypermedia technologies (Brusilovsky & Tasso, 2004) have demonstrated various ways to personalize the traditional methods of accessing information.

Among these new technologies, *recommender systems* have captured the practitioners' attention, since the industry has moved from mass production to mass customization. Companies running e-business desperately need strategies to persuade visitors to be their customers. Taking into consideration the transaction between a buyer and a seller in the offline, real world, it is natural for a seller to try to grasp the user's interest or intention and recommend services or products suitable to him/her. In the same manner, e-commerce sites not having personal interaction are using recommender systems to understand their users and suggest products (Schafer, et. al., 1999)

However, in the situation where we have a constantly increasing information overflow, no single technology can be considered to be a silver bullet. Each technology has its strong points and weak points. Each can help the users to retrieve some information "needles," but not all of them. The authors argue that the next generation of information access systems should focus on comprehensive personalized information access, i.e., making use of several information access technologies and personalization approaches in a seamless integration. Our earlier experiments with comprehensive personalized information access systems have demonstrated that each technology—when used as a part of an integrated system—contributes to the overall success by helping the user to locate a good portion of relevant items (Brusilovsky, et al., 2005).

In this paper, we report our recent work on comprehensive information access in a different domain – job search. Job search is a relatively new domain for information access, yet a challenging and important one. Since companies mainly rely on the Internet to search for the right human resources, they post their job openings on Web sites and job search systems. A huge volume of job information is now spread over the Web. Finding the right job is harder than finding a relevant news item since job seekers have to consider several aspects of a potential job - categories, locations, education, and salary levels. Therefore, it is even more important to allow them to

Third International Conference on Autonomic and Autonomous Systems (ICAS 2007)
0-7695-2859-5/07 $25.00 © 2007 IEEE

explore information in multiple ways, in order to collect good job openings.

Helping job seekers to find the right job has been the focus of a number of commercial and research-level systems. Simple company-level job advertising systems, which offered basic browsing and search access are now overshadowed by online recruiting services such as careerBuilder.com and monster.com. These services attempt to establish long-term contact with their clients and provide profile-based information filtering to motivate the users to revisit them. Job seekers who want to see job information that is best-fitted to them have to enter personal data, for instance, career purpose, educational background, professional experience, or technical skills as a form of resume. After acquiring plenty of user information, the system then provides what it considers to be possibly relevant recommendation. It takes too much effort for users to give their information for recommendation. Additionally, in the case of users who do not have a precise career purpose or a targeted resume, it is difficult to receive useful recommendations from these sites.

A recommender system using a personalization technology called CASPER was developed to compromise this information overload. CASPER has two stages: produce recommended jobs, on both the server-side and the client-side. In server-side recommendation, it generates a rank-based list of jobs by similarity. The client-side process orders a list of jobs, noting the relevance of each job as calculated by the user's search profile (Smyth, et al., 2002; Bradley, et al., 2003). When this personalized information retrieval is convergent on the user's intentions, it can effectively and efficiently produce targeted information. On the other hand, it does not work well in situations where users have general interests and are uncertain about their job direction. A job information system must provide the space to explore as well as the ability to search, for the benefit of the users.

Unlike the system mentioned above, which relies on just one or two information access technologies, the job recommendation system Proactive, presented in this paper, attempts to bring together a range of technologies. The focus and innovation of our work is twofold. First, we attempted to provide a true integration of these approaches whereby one approach can capitalize on the others and where weaknesses of one approach are balanced by the strengths in others. Second, we attempted to integrate several personalization approaches into one system, using the theories behind recommender systems (Schafer, et al., 1999) and adaptive hypermedia (Brusilovsky, 2001). Proactive is a practical working system developed mainly to help students at the University of Pittsburgh in their job search. It is also used as a vehicle to explore a range of innovative personalized access approaches.

I. THE RECOMMENDATION TAXONOMY

To ensure that a set of information access techniques is comprehensive, it is useful to apply an existing taxonomy. A useful user-focused taxonomy was suggested by Schafer and colleagues (1999). The taxonomy was developed for the context of product recommendation in e-commerce but is equally useful for the very similar context of job recommendation. The two axes of the taxonomy are the degree of automation and the degree of persistence in the recommendation, as shown in figure 1.

Firstly, the degree of automation has nothing to do with which system is used, but is directly related to the user's efforts. *Automatic* means that there is no need for a user to do some explicit efforts to generate recommendations. A recommender using an automatic feature does not obtrusively ask users to do a certain action. On the other hand, it is considered to be *manual* if a customer has to do specific work like answering questions, entering personal information, preferences, or rating information. A second dimension of the taxonomy is its degree of persistence. This is based on the number of user sessions that are needed in order to make a recommendation. If the recommender utilizes the information from only one session with a user, it is *ephemeral* and ephemeral recommenders do not take previous sessions into account. *Persistent* recommendations use information that has been accumulated from previous sessions (Schafer, et al., 1999).

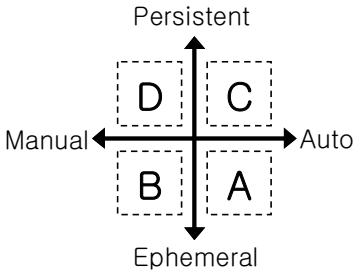

Figure 1. The Recommendation Taxonomy (Schafer, et al., p. 162)

A recommendation, which is automatic and ephemeral (quadrant 'A' in figure 1) does not require any work by the user and is not dependent on them. Hence, the suggestions to any user are the same for all users (Schafer, et al., 1999), although they are able to see the raw data and navigate through it in their own way. Manual and ephemeral recommendations (quadrant 'B') are a list generated by the user's short-term interests. Users give explicit keywords or requests to the recommender, whereupon the system displays the related information but is not capable of remembering these interests from one visit to the other. This representation suggests information that is directly relevant to the user's feedback. Thirdly, recommendation attributed as automatic and persistent (quadrant 'C') may be the most ideal recommendation mechanism from the user's point of view. Users can do their work without any intervention from the system and yet, every time they use it, meaningful data is collected by the system. By interpreting the accumulated data, a recommender can suggest information that matches the user's long-term interests. Lastly, for manual and persistent recommendations (quadrant 'D'), users have to spend some manual effort,, i.e., type in several items of interest or answer some questions, in order to receive a recommendation. Active interaction with the system allows the user to continuously receive a list of data coordinated with the complicated user's mind-set. If users change their

preferences, partially or totally, the system correspondingly produces a new list of information.

II. PROACTIVE

Proactive is an adaptive job recommender system which helps job seekers find relevant opening jobs in multiple ways. This site concentrates on information technology related jobs. At the beginning of a session, Proactive selects and displays jobs posted within 24 hours as shown in Figure 3. By clicking the title of each job, the corresponding job advertisement details are shown. When a user finds an interesting job, he/she assigns it to their favorite list. Based on the properties of the user's favorite jobs, recommendations are generated, so whenever a user designates a job as being a favorite, a new list of 'Recommended Jobs' is created. In addition, users can select their preferred job category and various information regarding jobs, such as location, education level, experience level, company size and industrial domain. After analyzing the user's preferences, the system suggests a group of jobs as 'Preferred Jobs'.

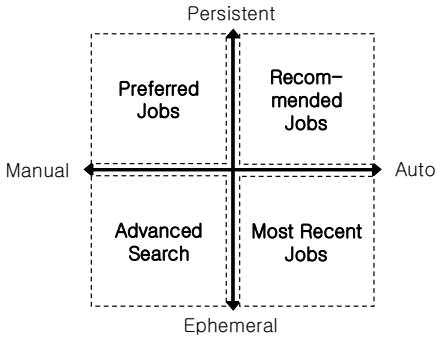

Figure 2. User Interfaces based on the Recommendation Taxonomy

A. *Target user of the system*

This system focuses on two groups of users. The first group is made up of students who want to have an information technology-related career. The people belonging to this group visit job search sites to build their career goals and to plan their future path. In order to meet this need, the authors tried to display job information in various ways, so that students could perceive the dominant trend of the current job market. The characteristics of this group are—no definite career purpose and job description. They have a broad range of preferences and interests.

The second group is for job seekers who already have a job but desire to move on to another job. They have clear career goals and prefer to have their results narrowed down to their predefined interests. However, it is difficult to spend a lot of time looking for good open positions when one is already working. The recommendations based on user preference and profiles are more narrowly developed according to these requirements.

B. *User interface*

The aim of this recommender is to build a comprehensive way to search jobs in order to meet various users' interests or intentions. In order to achieve this goal, the system design is based on the recommendation taxonomy (Figure 2).

The system has four different kinds of interfaces – Most Recent Jobs, Advance Search, Recommended Jobs, and Preferred Jobs. When a user logins to Proactive, they first see 'Most Recent Jobs.' The jobs listed in this page are all of the information that the system collected from one external source, Yahoo! HotJobs within the last 24 hours. This list of raw information contains automatic and ephemeral recommendations. The user's intervention is not necessary and information is uniform for all users. Originally, it is ordered chronologically by the posted time, as illustrated in Figure 2, but the user can modify the list for convenience, for instance resorting the list according to a certain job category or title, company name, location, position type (fulltime, part-time, employee, contractor, etc.), required experience and education level. If he/she types a keyword for job title, company name or location, the search results within these recent jobs are displayed. In 'Most Recent Jobs,' users are able to explore open jobs in whatever manner is convenient and helpful to them, to understand the similarity or differences among job categories, comprehend the current trends of the job market, and thus shape their career path.

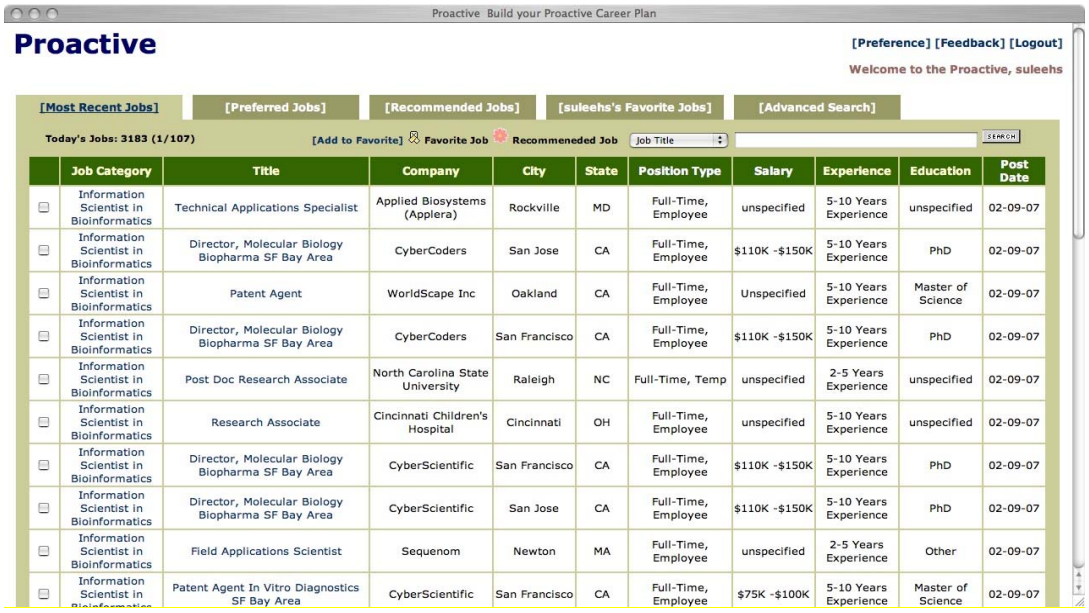

Figure 3. Most Recent Jobs

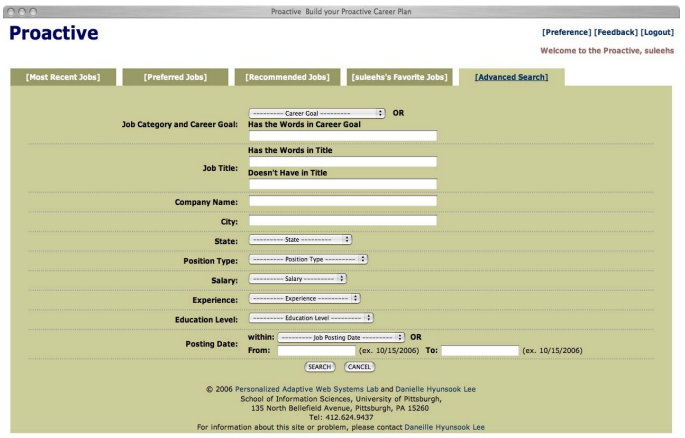

Figure 4. Advanced Search

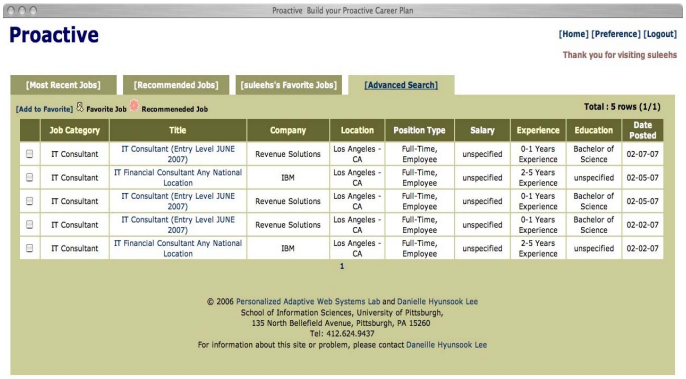

Figure 5. Advanced Search Results Set

In all four pages, users can mark a job as one of their favorites. The Proactive system interprets these favorite jobs as current job interests. By analyzing the characteristics of each facet in the job, case by case, recommended jobs are suggested (Figure 6). Although, users have to add a certain job to the favorite list directly, they don't need to comment about it or give an explicit evaluation. The group of recommended jobs is regenerated periodically and also whenever users change their favorite jobs list. Clicking a button to add a favorite is a way for users to implicitly request similar jobs from the system. Therefore, once users store jobs in their favorite jobs list, they are shown a list of relevant jobs, as suggested by the recommender, periodically. It is a persistent and relatively automatic method to give recommendations to users.

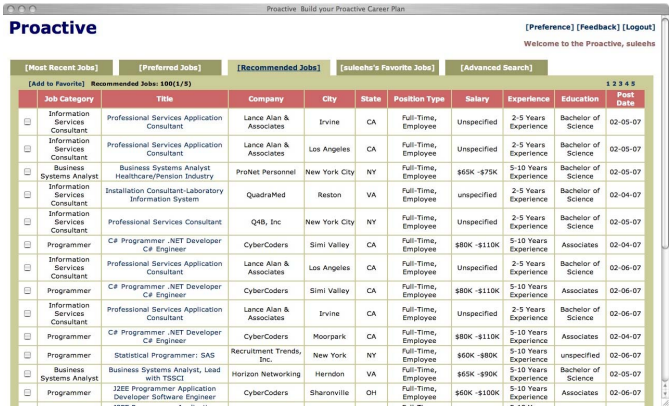

Figure 6. Recommended Jobs

Proactive provides a menu for setting the user's preferences. Users can choose their job categories, preferred location, expected salary level, education level, company size and job type. If they define the preferences, the system generates a new preferred jobs list, which matches these. The system starts to suggest preferred jobs from the time the preferences are described and continues to show preferred jobs in a timely manner afterwards.

The two persistent recommendations - recommended jobs and preferred jobs - are advantageous for giving users recommendations corresponding to their evolving interests. These recommendations can also give users a feeling of control. The user thinks that he/she creates a relationship by investing manual effort and can manage the system as they wish, by defining clear preferences.

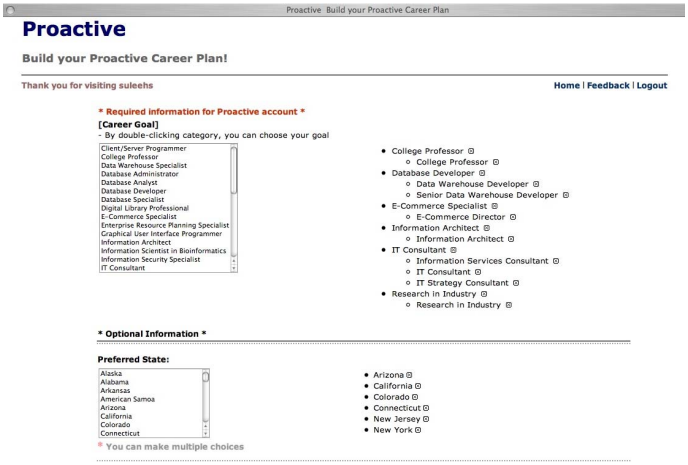

Figure 7. Screen For Setting User Preferences.

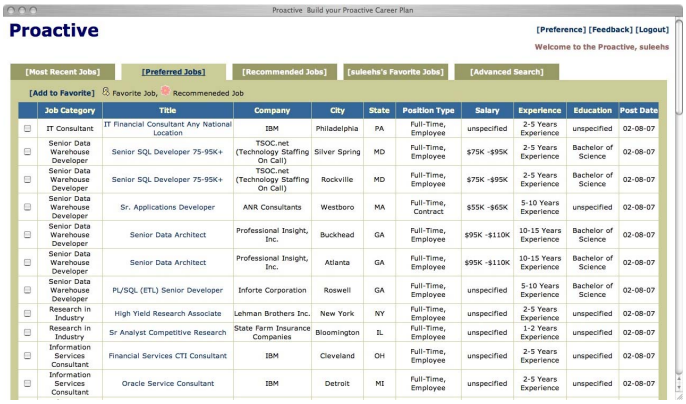

Figure 8. Preferred Jobs

C. *Architecture of the system*

The architecture of Proactive is illustrated in Figure 9.

Proactive consists of five components – Web spider, ontology checker, profile analyzer, preference analyzer, and user interface generator. The Web spider is an automatic parser that acquires job information from an external source periodically. Every two hours, this robot scans an online job search site and collects job information. RSS feeds data are analyzed and an HTML parser extracts corresponding facets in each job case, such as company name, location, expected educational level and experience, etc. The authors plan to add other online recruiting sites in the future.

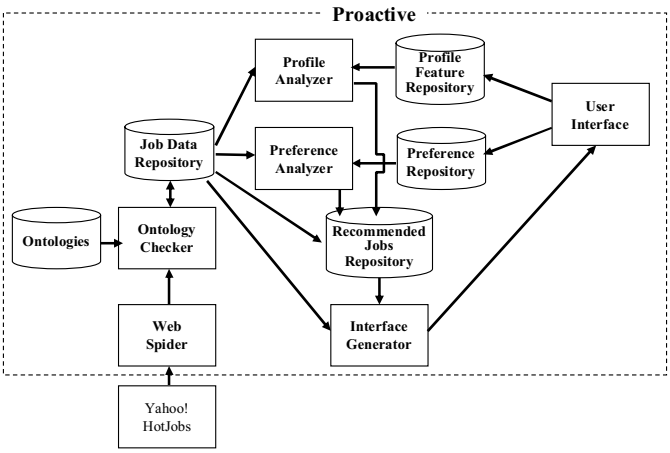

Figure 9. The Architecture of the Job Recommender

Before being stored in the job data repository, the collected information is classified by the ontology checker. In Proactive, predefined ontologies are used to statically represent data. Ontology checkers match data with ontologies and verify the classification. After that, job data is stored in a pre-designated format.

In each of the four interfaces, users are able to mark jobs they are interested in as a favorite job. Users can see not only the list of their favorite jobs, but also recommended jobs, which are based on the properties of their favorite jobs. The component named 'profile analyzer' makes the recommendations. Whenever users change – add or delete an item from their group of favorite jobs, the profile analyzer reads the list and recalculates the weights in them. By comparing the differences of distance in the weights with current open jobs, they generate the list of recommended jobs and store the profile into the profile feature repository. Besides this initial change, the system updates the list every four hours.

The preference analyzer interprets explicitly-defined user preferences and makes a recommendation in the form of a list of preferred jobs. Facets in job information are organized by ontologies and pre-defined groups. After calculating the similarity of jobs to the user's preferences, it displays the preferred jobs list. In the same manner as the profile analyzer, this component recalculates preferred jobs on a regular basis and whenever the user changes their preferences.

So as to fulfill several information requirements, the interface generator makes available the four kinds of user interfaces previously mentioned. Each interface has sorting and paging options, so users can wander around the page using their own strategies. In addition, the user interfaces have some functions that log the user's usage patterns, such as links clicked, what criteria were used to sort job lists, and what search fields and terms were used—as profiling features.

D. *Data representation based on ontology*

In this system, two kinds of ontology, as defined by Yahoo! HotJobs, are used to cluster multi-faceted information into meaningful groups. The first ontology is based on job

category and the second on company information, such as the industry each company belongs to and the company size, according to the number of employees. By using an external ontology, the collected information is organized and maintained in a static and meaningful structure.

An ontology is a relatively comprehensive repository of domain knowledge that explains entities, attributes, relationships and axioms of the domain. It is to help readers understand concepts and structure of the domain (Middleton, 2002). In our system, we use ontologies as a way to categorize jobs and as a knowledge base to define features of each job.

Ontological relationships can help calculate the weight value in each category as training values. The accuracies of recommendations based on ontological relationships are higher than recommendations without ontology (Middleton, 2003). This helps us to provide some structural understanding of the user's interests and reduces the initial effort of acquiring knowledge about users. For a recommender that accumulates a user's usage patterns in order to analyze them in a meaningful way, it is especially hard to generate adequate recommendations to users when there is no initial data or only a small amount of information. Additionally, if users' expectations are not met because of a poor recommendation, they are less likely to visit the site again. This problem is called the cold-start problem and ontology mapping in the recommender alleviates the problem.

III. Conclusion and discussion

In this paper, a comprehensive job recommender is introduced. An online recruiting service is one of the most important e-businesses, because job openings have a wide variety of data and need to be exposed to as big an audience as possible. This system was developed to accomplish two different user requirements: concentrate on a certain kind of information and navigate through a wide range of information. Based on the recommendation taxonomy, multiple ways to access information were provided, from a non-personalized page to a highly user-adaptive page.

We will complete an empirical evaluation of our system in the near future. Discerning which features in our four interfaces can satisfy each user group will be investigated. Furthermore, we are developing an advanced exploratory search interface using visualization and collaborative recommendation. The aim of the search would be to gather information about the user's behavior and to see what kind of features are proprietary to job related applications. Therefore, in the future experiment, user satisfaction and efficiency of information delivery between two systems will be explored. It is expected that these features can be applied to build an open user model of job seekers. A visualization of job information will also be helpful to support the user's understanding of the job market.

IV. Acknowledgements

The work presented in this paper is partially supported by DARPA GALE project and by the National Science Foundation under Grant No. 0447083.

References

[1] Bradley, K. & Smyth, B. (2003) Personzalied information ordering: a case study in online recruitment, *Knowledge-Based Systems*, 16, pp. 269 ~ 275

[2] Brusilovsky, P. (2001) Adaptive hypermedia, *User Modeling and User Adapted Interaction*, 11 (1/2), pp. 87-110.

[3] Brusilovsky, P., Farzan, R., and Ahn, J.-w. (2005) Comprehensive personalized information access in an educational digital library, *Proceedings of The 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, Denver, CO, USA, June 7-11, 2005, ACM Press, pp. 9-18.

[4] Brusilovsky, P. and Tasso, C. (2004) Preface to special issue on user modeling for Web information retrieval, *User Modeling and User Adapted Interaction*, 14 (2-3), pp. 147-157

[5] Middleton, S., Shadbolt, N. R. & De Roure, D. C. (2003) Capturing interest through inference and visualization: ontological user profiling in recommender systems, *Proceedings of the 2nd international conference on Knowledge capture*, Sanibel Island, FL, USA, October 23 - 25, 2003, pp. 62 ~ 69

[6] Middleton, S., Alani, H. & De Roure, D. C. (2002) Exploiting synergy between ontologies and recommender systems, *Proceedings of Semantic Web Workshop 2002 At the Eleventh International World Wide Web Conference*, Hawaii, USA

[7] Schafer, J. B., Konstan, J. & Riedl, J. (1999) Recommender systems in E-commerce, *Proceedings of the 1st ACM conference on Electronic commerce*, Denver, Colorado, USA, pp. 158 ~ 166

[8] Smyth, B., Bradley, K. & Rafter, R. (2002) Personalization techniques for online recruitment services, Communications of the ACM, 45(5), pp. 39 ~ 40