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@sis.pitt.edu

Abstract—Online job search is information intensive activity, because thousands of jobs are posted on the Web daily and it takes too much effort to find out right position. Job search sites require recommender system to meet diversified information needs. Job seekers who have well-defined career direction try to focus on relevant open positions and students who have a general and evolving interest wants to see the dominant trends of job market to plan their career path. In this paper, comprehensive job recommender system is introduced. From user’s perspective, four different kinds of recommendations are implemented. Users of this system can retrieve opening jobs from exploring to searching

Keywords-component; Job recommender, information retrieval from multiple views, exploratory search

I. INTRODUCTION

The reliance on information in our daily lives is getting bigger. 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 immensely increasing and the processing of the relevant information 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 needed (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 contradicted requirements at the same time.

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

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

However, in the situation of constantly increasing information overflow, no single technology can be considered as a silver bullet. Each technology has its strong sides and weak sides. Each can help the users to retrieve some information needles, but not all of them. The authors argue that the next generation of the information access systems should focus on comprehensive personalized information access – i.e., using several information access technologies and personalization approaches in a seamless integration. Our earlier experiments with comprehensive personalized information access systems 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 share 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 an important one. As companies mainly rely on the Internet to find out right human resource, they post the job openings on Web sites and job search systems. 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, or salary levels. Therefore, it is more important to allow them to explore information in multiple ways to collect good openings.

Helping job seekers to find the right job was 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 make the users to revisit them. Job seekers who want to see an up-to-date information about relevant jobs have to enter personal data, for instance educational background, professional experience, or technical skills, to search various jobs, or actually apply to an open position. After acquiring the plenty of information, they then provide possibly relevant information. It takes too much time for users to give their information for recommendation. Additionally, in the case of users who do not have precise career purpose or targeted resume, it is difficult to receive useful recommendations from these sites.

A recommender system using personalization technology called as CASPER was developed to compromise information overload. CASPER has two stages to produce recommended jobs, server-side and client-side. In server-side recommendation, it generates a rank-based list of jobs by the similarity. Client-side process orders a list of jobs depending on the relevance of each job calculated by user’s search profile (Smyth, et al., 2002; Bradley, et al., 2003). This personalized information retrieval convergent on user’s intention can effectively and efficiently produce targeted information. On the contrast, it does not work well in situations where users have general interest and are uncertain about job direction. Job information system has to provide the space to navigate as well as to search for the sake of 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 the innovation of our work is twofold. First, we attempted to provide a true integration of these approaches where one approach can capitalize on the others and where weak sides of one approach are balances by strong sides of others. Second, we attempted to integrate several personalization approaches in one systems using the ideas of recommender systems (Schafer, et al., 1999) and adaptive hypermedia (Brusilovsky, 2001). Proactive is a practical working system developed mainly to help students of University of Pittsburgh in their job search. It is also used as a vehicle to explore a range of innovative personalized access approaches.

1. RECOMMENDATION TAXONOMY

To ensure that a set of information access techniques is comprehensive, it is useful to apply some existing taxonomy. A useful user-focused taxonomy was suggested by Schafer and the colleges (1999). The taxonomy was developed for the context of product recommendation in e-commerce and is equally useful for a 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 system and is directly related with user’s efforts. Automatic means that there is no need for a user to do some explicit efforts to generate recommendation. The recommender using automatic feature does not obtrusively ask users to do a certain work. On the other hand, it is considered as manual that a customer has to do specific jobs like answering questions, entering personal information or preference, or rating information. Another dimension of the taxonomy is the degree of persistence. It is based on the number of sessions taken advantage of to make a recommendation. If the recommender utilizes the information from one session of a user, it is ephemeral and ephemeral recommender does not take previous sessions into account. Persistent recommendations use information accumulated in previous sessions (Schafer, et al., 1999).

![Figure 1. Recommendation Taxonomy (Schafer, et al., p. 162)](image-url)

The recommendation, which is automatic and ephemeral (quadrant ‘A’ in figure 1) is not required to work anything and not dependent to users. Hence the suggestions to each user cannot be different. Even recommendations based on this system are the same to all users (Schafer, et al., 1999), they are able to see the raw data and navigate them in their own way. Manual and ephemeral recommendations (quadrant ‘B’) are to generate list by user’s short-term interest. Users give explicit keywords or request to recommender then system displays the related information but does not remember the interest from one visit to the next. This representation is to suggest directly relevant information to user’s feedback. Recommendation attributed as automatic and persistent (quadrant ‘C’) can be the most ideal recommendation mechanism from user’s point of view. Users can do their works not being intervened by the system and every time they use it, meaningful data are collected by system. By interpreting the accumulated data, recommender can suggest the information fitting into user’s long-term interest. Lastly, for manual and persistent recommendations (quadrant ‘D’), users have to spend some manual efforts, i.e., type in several items of interest or answer some questions to make a recommendation. Active interaction with system makes users to continuously receive data list complying with the complicated user’s mind-set. If users want to change their preference partially or totally, it produces new list of information correspondingly.

II. PROACTIVE

Proactive is an adaptive job recommender system, which helps job seekers to 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 Figure 3. If
clicking the title of each job, the corresponding job advertisement in detail is shown. When a user finds an interested job, he/she assigns it as favorite job. Based on the properties of user’s favorite jobs, recommendations are generated. Whenever user assigns a job as a favorite, the new list of recommended jobs is made. In addition, user can select their preferred job category and various information regarding jobs such as location, education level, experience level, company size or industrial domain. After analyzing the user’s preferences, the system suggests a group of jobs as preferred jobs.

The second group is job seeker who has a job and desire to move to another job. They have clear career direction and prefer results narrowed down to their interest. However, it is hard to spend many times to find good open positions. The recommendations by user’s preference and profiles are developed according to the requirements.

B. User Interface

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

The system has four different kinds of interface – Most Recent Jobs, Advance Search, Recommended Jobs, and Preferred Jobs. Firstly, if a user logsins to Proactive, they can see ‘Most Recent Jobs’ as a front page. The jobs listed in this page are all the data that the system read from external source, Yahoo! HotJobs within 24 hours. This list of raw information is automatic and ephemeral recommendations. User’s intervention is not necessary and information is same to all users. Originally it is ordered by the posted time, illustrated in Figure 2, but user can modify the list for convenience’s sake, for instance resorting the list according to a certain job category or the order of title, company name, location, position type (fulltime, part-time, employee, contractor, et al.), required experience and education level. If he/she types a keyword about job title, company name or location, the search results in recent news are displayed. In ‘Most Recent Jobs’ users are able to explore opening jobs for convenience’s sake and it is helpful to figure out the similarity or differences among job categories, comprehend the current trend of job market, and build their career path.
In all the four pages, users can choose a job as a favorite. Proactive system interprets these favorite jobs as current job interests. By analyzing the characteristics of each facet in job cases, recommended jobs are suggested (Figure 6). Although, users have to add a certain to favorite jobs list directly, they don’t need to comment about it or give obvious evaluation. The group of recommended jobs is regenerated periodically and whenever users change favorite jobs list. Clicking a button to add favorite is for users to implicitly request to the system the similar jobs. Therefore, once they store jobs in favorite jobs list, they are shown a list of relevant jobs suggested by recommender periodically. It is persistent and relatively automatic recommendation to users.

Proactive provide a menu to setting user’s preference. Users can choose their job categories concerned with, preferred location, expected salary level, education level, company size and job types. If they define the preferences, the system generates preferred news matching with them. The system starts to suggest preferred jobs from the time of describing preferences and show the timely changed preferred jobs afterwards.

The two persistent recommendations - recommended jobs and preferred jobs - are advantageous for users to receive recommendations corresponding to the evolving interests of users. These recommendations also can give users a feel of control. A user thinks that he/she creates a relationship by investing manual efforts and manages system as one’s wishes by defining clear preference.

C. Architecture of the system

The architecture of Proactive is illustrated in Figure
Proactive consists of five components – Web spider, ontology checker, profile analyzer, preference analyzer, and user interface generator. Web spider is automatic parser to acquire job information from external source periodically. Every two hours, this robot scans the site and collects job information. RSS feeds data are analyzed and HTML parser extracted each corresponding facets in each job case. The authors have a plan to add other online recruiting site in future.

Before stored into the 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. And then, job data is stored in pre-designated form of information.

In every four interface, users are able to name an interested job as favorite job. Users can see not only the list of favorite jobs, but recommended jobs, which are based on the properties of favorite job cases. The component named ‘profile analyzer’ makes the recommendations. Whenever users change – add or delete – the group of favorite jobs, profile analyzer calculate the weights in them. By comparing the differences of distance in the weights with current open jobs, they generate the list of recommended jobs. Besides, the system updates the list every four hours.

Preference analyzer interprets explicitly defined user’s preference and makes a recommendation as preferred jobs. Facets in job information are organized by ontologies and predefined groups. After calculating the similarity of jobs to user’s preference, it displays the preferred jobs list. In the same manner with profile analyzer, this component recalculates preferred jobs in regular base and every time when user changes their preference.

So as to fulfill several information requirements, interface generator makes four kinds of user interface aforementioned. Every interface has resorting and paging options, so users can wander around the page by their own strategy. In addition, User interface have some functions to collect user’s usage pattern such link click, resorting criteria of job lists and search fields and terms as profiling features.

D. Data representation based on ontology

In this system, two kinds of ontology defined by Yahoo! HotJobs are used to tide multi-faceted information into meaningful groups. First ontology is about job category and second ontology is about company information such as industry each company belongs to and the company size according to the number of employee. By using external ontology, the collected information is organized and maintained in static and meaningful structure.

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

Ontological relationship can help calculate the weight value in each category as training values. The accuracies of the recommendation based on ontological relationship are higher than recommendation without ontology (Middleton, 2003). This helps us to provide some structural understanding of user’s interest and reduces the initial efforts to acquire knowledge about users. Especially, for recommender accumulating user’s usage pattern, analyzing them into meaning way, it is hard to generate adequate recommendations to users without initial data or with a bit of information. Additionally, if users cannot meet their expectation because of poor recommendation, they would not visit the site again. This problem is called as cold-start problem and ontology mapping in the recommender attenuate the problem.

III. CONCLUSION AND DISCUSSION

In this paper, comprehensive job recommender is introduced. Online recruiting service is one of the most important e-businesses, because job opening has wide variety of data and need to be exposed to audience as many as possible. This system developed to accomplish two different user requirements, concentrating on a certain kind of information and navigating wide range of information. Based on recommendation taxonomy, multiple ways to access information were provided, from the least personalized page to highly user-adaptive page.

We will do empirical evaluation of our system in near future. What features in four interfaces can satisfy each user group will be investigated. Furthermore, explorative search will be provided to users. The aim of the search would be to gather information about user’s behavior and see what kind of features is proprietary to job related applications. It is expected that these features can be applied to build open user model of job seekers and feasible to more adaptive job recommender application. Visualization representation of job information also will be provided to support user’s understanding of job market.

REFERENCES


