Proactive: Comprehensive Access to Job Information

Danielle Lee* and Peter Brusilovsky*

Abstract—The Internet has become an increasingly important source for finding the right employees, so more and more companies post their job openings on the Web. The large amount and dynamic nature of career recruiting information causes information overload problems for job seekers. To assist Internet users in searching for the right job, a range of research and commercial systems were developed over the past 10 years. Surprisingly, the majority of existing job search systems support just one, rarely two ways of information access. In contrast, our work focused on exploring a value of comprehensive access to job information in a single system (i.e., a system which supports multiple ways). We designed Proactive, a recommendation system providing comprehensive and personalized information access. To assist the varied needs of users, Proactive has four information retrieval methods – a navigable list of jobs, keyword-based search, implicit preference-based recommendations, and explicit preference-based recommendations. This paper introduces the Proactive and reports the results of a study focusing on the experimental evaluation of these methods. The goal of the study was to assess whether all of the methods are necessary for users to find relevant jobs and to what extent different methods can meet different users' information requirements.

Keywords—Job recommendation, explicit preference, implicit preference, personalized information retrieval

1. INTRODUCTION

The Internet has become an increasingly important source for finding the right employees so more and more companies post their job openings on the Web. Due to the huge number of jobs available on multiple Web sites and the dynamic nature of the information (some job opportunities may be open for just a few days), it is difficult for job seekers to locate the right position within the right time frame. In addition, searching for the right job may require more cognitive effort than merely searching for a Web site or a news article since job seekers need to carefully consider and compare multiple aspects of jobs such as job title, location, expected salary, size and reputation of company, just to name a few. To assist Internet users in searching for the right job, a range of research and commercial systems have been developed over the last 10 years. In this paper, we call them "job search systems."

While the name job search system reflects the main purpose of the systems examined in this paper, it is a bit misleading from the information access point of view. Some subset of these

systems supports search-based access and other systems support alternative information access methods such as hypertext browsing, profile-based information filtering, or proactive recommendation. Each information access method is different in respect of the supported style of user-information interactions. For example, information filtering and recommendation technologies represent a user's job search interests and attempt to proactively retrieve the most relevant job opportunities. All of these methods of information access are known and popular well beyond the area of job search for their ability to lead users to the right information. What looks surprising in this context is that the majority of existing job search systems supports just one, rarely two, major methods of information access. If each method has its value, we can expect that a comprehensive system for job information access (i.e., a system which supports multiple methods) will be more valuable than a system which supports just one or two methods. Why haven't comprehensive systems for job information access been built? Is it too hard to build a comprehensive system for this specific domain? Or, is it simply not necessary because all methods of information access for job searching lead users to the same information -- thus, any single way suffices?

These questions motivated our research into the value of comprehensive access to job information presented in this paper. To answer these questions, we developed Proactive, a comprehensive personalized job search system, which supports all four major methods of information access: browsing, search, filtering, and recommendation. Since its first version become available in 2007 [9], Proactive has provided online access to job information in the area of information technologies for students of the School of Information Sciences, the University of Pittsburgh. In Proactive, multiple methods of information access have its own purpose, support different user needs and deliver unique and relevant information. By integrating several ways of information access and providing personalization, the system as a whole attempts to be more valuable than each of its parts. To complement our observations and answer the research questions noted above, we hosted a controlled user study with 17 prospective job seekers, which is presented in the paper.

2. RELATED WORK

2.1 Principal Information Access Methods

Information science literature recognizes three principal information access methods for finding information to meet users' information needs: ad-hoc information retrieval, hypertext browsing and information filtering [4, 13]. In *ad-hoc information retrieval* (IR), users get access to relevant information by issuing a query to an IR system or search engine and analyzing a ranked list of documents, which are returned as a result. In *hypertext browsing*, a user attempts to find relevant documents by browsing links that connect documents in a collection. In *information filtering* (IF) an information system attempts to recommend documents, which match the user's interests. Information filtering systems, frequently called as *recommender systems*, construct dynamic user profiles by observing user interactions [4].

Each of the major information access methods has its own strong and weak sides. When a user clearly understands his information needs and can express it in unambiguous terms, keyword-based ad-hoc search can deliver an answer in an instant. However, ad-hoc search is known to returns inappropriate results and miss the context of users not expressed in a query. Browsing

allows one to explore information space using hyperlinks. When a user isn't sure exactly what she is looking for, browsing allows figuring out what the available options are, and, frequently, serenditipously find something that the user was not even thinking about when starting the browsing session. In dealing with complex information (such as in job cases), browsing is specifically helpful to acquire a great deal of information and context which is not available in the results of keyword-based search [16]. As the amount of information to browse expands, however, browsing becomes less efficient than ad-hoc search. In addition, both search and browsing are typically driven by a short-term information need and are not able to take into account user's long-term interests [13]. Information filtering and recommendation methods address this shortcoming by providing *personalized* information access.

The nature of personalization in information filtering is determined by the kind of user input, which serves as a source information for recommendation [15]. Most recommender systems take advantage of a single source of information [19]. Some systems recommend better query terms for searches or finer search results through conversational approaches, and these concentrate on customers' explicit interests at the moment [1, 3]. Other recommender system tries to reduce direct interventions into user search process and utilizes implicit user preferences through user feedback [5]. There are also a few systems, which collect users' local preferences through example critiques to estimate the global preference for personalization. This technology utilizes the users' current preferences with improved uncertainty [16]. Another study based on example critique tries to elicit user preferences as they have changed over time [23]. However, as pointed out, these systems correspond to the user's partial preferences and consequently, produce one-sided recommendations. It is hard to find a recommender system with the versatility to cope with multiple inputs. Since there is no system offering perfectly faultless personalization, it is important to acquire user preferences and needs from as many sources as possible.

2.2 Job Recommender Systems

There are two parties involved in job searching processes – employers and job seekers. In order to reduce time and money of employers and to retain transparency of the job market, many companies are adopting up-to-date e-recruiting systems [7, 11]. The e-recruiting systems take advantage of various job-related semantic technologies, as we proposed in this paper. The systems extract various metadata of job candidates (e.g. technical skills, experience levels, degrees, etc.) from online resumes and match the metadata with employers' job requirements. This automated process screens out significant number of applicants (i.e. pre-screening) and leaves a few top candidates for the next stage of screening [8, 17, 20]. Since the e-recruiting was introduced in the mid-1990s, the systems have enjoyed explosive growth, and, as the e-business technologies advance and recruiters gain more e-recruiting experience, the e-recruiting systems have been developed more diversely. Lee (2007) classified the e-recruiting systems into six categories – the general-purpose job board, niche job board, e-recruiting application service provider, hybrid (online and offline) recruiting service provider, e-recruiting consortium and corporate career Web site. All of these systems are mainly serving for employers. A few of them belonging to the general-purpose job board and niche job board are offering services for job seeks, but the services are limited to simple search and navigation of job information and barely onilne resume. On the other hand, for employers, general e-recruiting systems support various recruting related processes such as job requisition management, advertisement, prescreening/assessment of can-

didate, candidate relationship management, and even e-recruiting performance analysis [10]. Nonetheless, some researchers questioned about the performance of the system. They suggested that depending on how to identify job requirements and evaluation criterions, the subsequent results of the pre-screening are different to a large extent. As the result, some of the employers that had experienced disappointing results of the prescreening went back to traditional recruiting process or semi-automated process [7]. In spite of the worries about the performance, it is known that these e-recruitment systems are among the most successful e-business applications [11].

Compared with e-recruiting systems for employers' sake, it is hard to find systems to proactively help job seekers find good jobs and facilitate their application processes. For job seekers, finding a job is a cognitively intensive challenge. Each job case consists of multiple facets (kind of job, salary, location, etc.), and job seekers have specific preferences or restrictions for each facet. For these job seekers, many commercial online recruiting sites are in service. These sites usually provide personal repositories to save interesting jobs and enable the users to search and navigate jobs. The repositories are, however, for users to simply refer to the saved jobs in future, and the systems don't make any attempt to utilize the saved jobs to suggest the users relevant jobs. The search and navigation don't take into consideration of the users' job interests, either. Even though they spent significant amount of time to find a job on the system, their past search or navigation records and further the job interests represented by the records are neglected by the systems. The users have to re-start the information seeking activities from the scratch every time. Such sites also offer electronic resume services for their users. It requires entering a long array of personal information such as career goals, educational background, professional experience and technical skills. In spite of the users' endeavors, their e-resumes are used solely when applying for a job electronically. Users' time and efforts to compose their e-resumes don't give them a guarantee of any information seeking aids such as personalized recommendations. For the user who doesn't have a specific career goal, it is even hard to complete the resume and experience the limited benefit of the follow-on services. Compared with e-recruiting systems for employers' sake, thus, most of the online recruiting sites for job seekers are providing limited information access methods, passively. The systems which provide active aids to find good jobs and facilitate their job searching with multiple information access methods are relatively rare. In particular, recommending a personalized list of jobs to candidates is more complicated than the metadata matching tasks of the e-recruiting systems for employers, since the job seeker-oriented systems should understand users' multi-faceted job interests and build the user models, accordingly. As a result, implementing just a single type of information access may take a lot of efforts due to the need to capture what users are interested in.

One online recruiting site, called CASPER, applied personalization information retrieval techniques to filter out the irrelevant information and provide recommendations based on job seekers' long-term and short-term preferences. When a user enters a search query term, search results matched to the user's current interest are adjusted in relation to the user's accumulated past interests [21]. The results of this personalization technique might be ideal for job seekers whose career purpose is clearly defined.

Tvarozek and Bielikova (2007) provided personalized recommendations integrated with faceted browsing of job cases, in order to support for the users with rather vague career goals. To compute the adaptation, each user's usage pattern for the current session is collected and, according to the semantic similarities, the most relevant facets and jobs are recommended. Their

information access technology is innovative because it combines several technologies including user modeling, facet navigation and semantic web technology. Moreover, the technologies utilize various user inputs – implicit and explicit user preferences and short-term and long term user models. However, the accuracy of the user model would be in question, since the model is built based on user's click-stream logs [22]. Put differently, the data could have high level of uncertainty [12].

3. PROACTIVE

Proactive is a personalized system, which provides comprehensive information access for information technology-related job opportunities in the United States. Proactive consists of four components to collect, organize and display job information – Web spider, HTML parser, ontology checker, and user interface generator. First, the Web spider collects 148 categories of IT-related jobs, for instance, database administrators, database architects, information system managers, etc., from Yahoo!, HotJobs (current, Monster.com). After eliminating the duplicated jobs, the spider adds about 300 new jobs per day on average on Proactive using RSS feeds of the data. To build its internal database, the next component HTML parser applies automatic information extraction from the RSS feed. An ontology checker matches the extracted job information with ontologies and pre-defined taxonomies to verify the metadata classification. Ontologies and taxonomies are used to semantically represent data. The most important ontology provided by HotJobs is to classify job categories. This ontology organizes 148 job categories on 2 levels of hierarchy and clusters similar job categories together. The second ontology classifies the industry of each company into 111 categories. In addition, HotJobs provides categorization of company size according to the number of employees and locations of company headquarters. Job information also includes classifications of required educational levels (15 options to specify degree of job candidates), experience, position type (i.e. full-time employee, part-time employee, contractor, part-time contractor etc.) and salary. The set of metadata fields is also extended with geographic information (longitude and latitude of each city) of each job position, and the geo-data is provided by Google MapTM. This allows calculating distance-based similarity between users' geographical preferences and job locations. Along with these coded metadata fields, Proactive database also manages several free-text fields such as job titles, required technical skills, company names, etc.

On the user side, the last component, the user interface generator produces interfaces that support different information access methods. To ensure that Proactive supports all critical ways of personalized information access, the system design was guided by the recommendation taxonomy proposed by [18]. A discussion of the design goals in the context of the taxonomy is outside the scope of this paper and can be found in [9]. To reflect the presence of multiple information access methods in Proactive, the system interface (Fig. 1) has a tab-based organization: four main tabs (Most Recent Jobs, Matched to Preference, Recommended Jobs, and Advanced Search) correspond to the four ways to access job information; one tab (My Saved Jobs) provides access to a user's personal job repository (acting as a set of bookmarks), where he can save interesting jobs discovered by the tools on the other tabs. In addition, the user interface generator logs all of the user's actions: links clicked, sorted criteria selected, search fields content, etc.
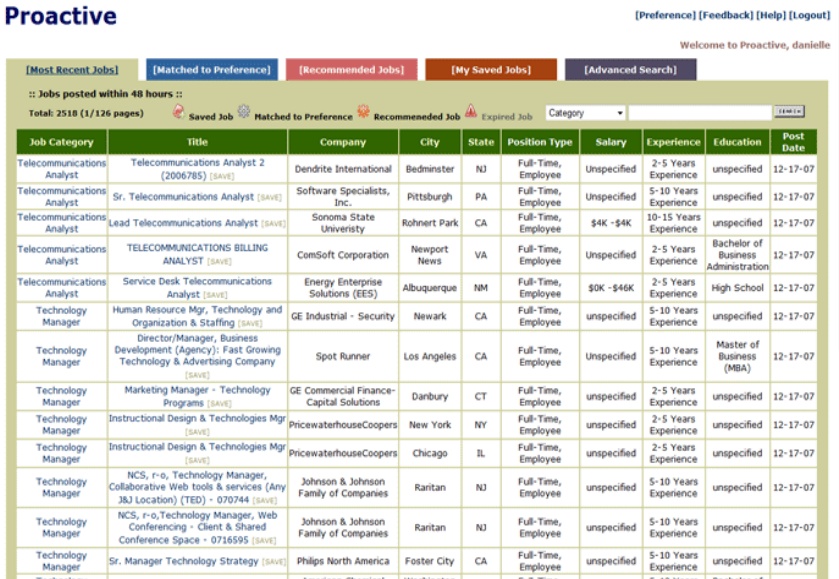
Fig. 1. At the beginning of a session, Proactive opens Most Recent Jobs tab showing jobs posted within the last 48 hours

Proactive attempts to serve for the whole spectrum of job seekers. On one end of the spectrum are users who are just starting to plan for their future careers. Students who use the Proactive system are a typical example of this group. Their career goals are not yet clear, and their job interests are broad. The users in this group visit job recruiting sites to plan their career path and to better understand the skills or qualifications for a job. In order to meet their needs, the Proactive system presents a wider view, helping the users to discover the dominant trends in the current job market. Among the tabs, 'Most Recent Jobs' and 'Advanced Search' work best to deliver a broader view of the job market.

On the other end of the spectrum are experienced job seekers who already have a job or long-term professional experience, but who want to move on to another job. These users have a clearly defined career purpose and specific job interest. When they use a job search site, they try to narrow down the jobs to those well-suited to their preferences. They want to see all jobs matching their preferences, with a minimum expenditure of time and effort. To assist these users, the system attempts to model their explicit and implicit job preferences and directly recommend the most relevant jobs. Such tabs as 'Recommended Jobs' and 'Preferred Jobs' were designed to deliver jobs matched to implicit and explicit preferences. A more detailed description of all tabs is provided below.

3.1 Four Methods of Information Access

In Proactive, the *Most Recent Jobs* tab (Fig. 1)' supports simple *browsing-based access* to job information. It displays a table of jobs posted within the last 48 hours, providing the most important aspects of each job case in several columns. The information presented in this list is not personalized; it is the same for all users. However, the users may explore this information and adjust the view to their needs using several provided tools – column sorting by various metadata

Fig. 2. The Advanced Search tab allows users to specify a number of job parameters

fields (job category, job title, city or state, company name, position type, educational level or required experience), filtering, and page-by-page navigation. Thus, the 'Most Recent Jobs' tab combines a broad view of job information with the freedom to apply a user's own exploration strategy. This tab could be the most helpful for the users to understand the current trends in the job market and to shape their career path.

Advanced Search tab supports traditional *ad-hoc search-based access* to job information. Using the Advanced Search interface, users can specify various job parameters and retrieve a full list of jobs, which match all these parameters (Fig 2). This interface could be helpful to both novice and experienced job seekers. Following typical *advanced search* functionality in modern information access systems, Proactive Advanced Search Tab allows users to use a combination of free-text and metadata search to locate relevant information. Metadata-based search fields allow users to specify desired values of several fields with restricted vocabulary such as job category, industry, company size, etc. Full text search field allow searching for desired keywords in several free-text fields such as job title, city or company name. Using Advanced Search interactively, users can narrow down a list of interesting jobs or get a better picture of the whole job market for a particular job type, city, or certain industry domain.

The *Preferred Jobs* tab in Proactive supports *profile-based information filtering* (Fig. 3). At any moment, it lists jobs that the system recommends to the users on the basis of their explicitly provided profile of interests (called *job preferences* in Proactive). Information filtering can be considered to be an inverted search [2], where a user's interest profile acts as a long-term query that operates on new information added to the system and filters out matching documents. Since user interests, unlike interactive ad-hoc queries, have to be specified in advance and are rarely edited, user interest profiles are more elaborate than interactive queries and represent a broader view of user interests. It is common for profile-based filtering systems to allow the user to specify complex combinations of desired values for multiple fields. Following this trend, the Proactive interface allows users to choose preferred job categories, preferred locations, company size, and industry. The system allows specifying multiple preferences for each facet. A job opportunity will be considered to match the user profile if it matches at least one of the preferences in
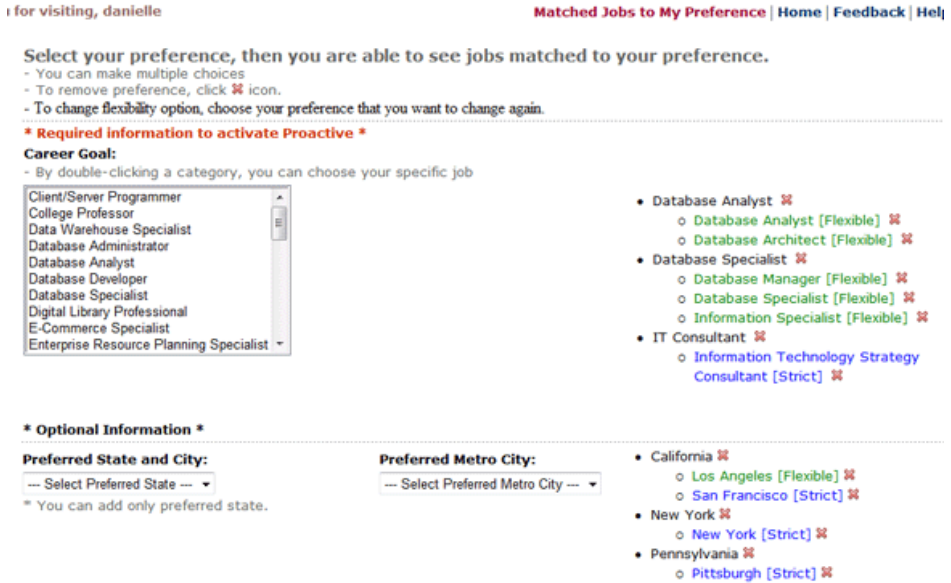
Fig. 3. Proactive offers an elaborated interface to specify long term user job preferences

each aspect where the preferences are provided. For example, a database architect position in a communication equipment company located in San Francisco will match the user profile shown in Fig. 3. To give users more control over the matching process, the strictness of matching (strict or flexible) can be separately adjusted for each specified preference. For example, a user selects the job category "database manager" indicating a flexible preference and selects Pittsburgh as a strict preference (see Fig. 3). In this case, the system will consider jobs that are ontologically similar (i.e., belonging to the same higher-level category) to the "database manager" as relevant, while still preferring exact matches. At the same time, jobs in a location other than Pittsburgh will simply be rejected even if the location is just 10 miles from Pittsburgh.

The three information access tabs introduced above produce three different lists of jobs (recent jobs, jobs matching a query, and jobs matching the user's profile of interests). Once an interesting job is found in any of these lists, the user can save it in his or her *personal repository,* adding optional ratings and comments. The content of this repository is accessible through the *My Saved Jobs* tab. The interface allows the users to rate, comment, or remove saved jobs. In addition, Proactive interprets the act of saving a job as an implicit expression of user interests and it uses the list of saved jobs to generate a list of recommended jobs, in the *Recommended Jobs* tab. The recommended jobs are regenerated every 4 hours. Note that unlike preferred jobs, which are based on explicitly provided preferences, recommendations based on saved jobs use the implicit information channel and do not require any additional actions from the user.

3.2 Ontology-based Information Representation and Recommendations

Many recruiting sites do not offer several methods of information access because posting job openings and maintaining the information is a constant effort and processing information for multiple information accesses is expensive. Many sites are posting job advertisements with little

Table 1. Example of Job Case Facets and the Encoding Values in Proactive

	Job Case #1	Encoded value	Job Case #2	Encoded value
Job Title	ColdFusion Programmer	*not encoded*	Network Consultant	*not encoded*
Job Category	Programmer	4.4	Network Consultant	40.4
Company Industry	Technology	14.5	Technology	14.5
Company Size	Less than 100 employees	1	500 - 1999 employees	3
Job Location	San Diego, CA	32° 42' 55 (Lat.) / 117° 9' 23" W (Lon.)	Atlanta, GA	33° 44' 56" N (Lat.) / 84° 23' 17" W (Lon.)
Educational Level	Bachelor of Science	21	Master of Science	31
Experience Level	5-10 Years	4	2-5 Years	3
Position Type	Full-Time Employee	11	Full-Time Employee	11

or no standardized format. Some sites, such as indeed.com or FlipDog.com utilize the automatic extraction technology for job information. However, most job recruiting sites do not have the advanced technology to extract and organize job information in a fully automatic way.

As explained above, the information organization in Proactive is based on external ontologies and local taxonomies. It allows defining a job case as a *multi-faceted classification*. Each case has several *facets* such as job category, company industry, or company size with values coming from the ontologies and taxonomies mentioned above. These values are encoded (Table 1) and used for ranking search results and generating recommendations. Middleton et al. (2003) demonstrated that the accuracy of recommendations based on semantic (ontology-based) representation is higher than recommendations without ontologies [14]. Proactive uses K-Nearest Neighbor (KNN) algorithm to retrieve and rank recommendations. This approach uses KNN's strength in indexing and comparing the underlying closeness of two points being considered [6]. KNN works well with points located in a multi-dimensional space such as job cases in Proactive.

With the KNN approach, the list of recommended jobs is compiled simply by picking those job cases that are the most similar to job cases in the list of saved jobs, which represent user *implicit job preferences*. In the equations 1 and 2, s_i is a saved job and c_j is a recently added candidate job. i takes on values from 1 to I, where I is the total number of a user's saved jobs and j takes on values from 1 to J, where J is the total number of candidate jobs. Each job case is represented as a vector of 7 facets (job category, company industry and size, job location, educational level, experience level, and position type). In this context, the distance between a saved job s_i and a candidate job c_j can be calculated as a sum of distances between corresponding values for each facet (equation 1). Here $\Delta(s_{i1}, c_{j1})$ denotes the distance between s_{i1} (the value of the first facet in i_{th} saved job) and c_{j1} (the value of the first facet in j_{th} candidate job). As the equation shows, the more similar these two job cases are semantically, the smaller is the distance between them. Finally, the total distance between a certain candidate job and the whole set of saved jobs is calculated as a sum of its distances to each of the saved jobs as shown in equation 2. In this equation, each distance is divided by r_i - the ratings of the saved job. The higher is the rating of a saved job, the smaller is the impact of its difference with the candidate job on the total distance.

$$D_R(s_i, c_j) = \frac{\Delta(s_{i1}, c_{j1}) + \Delta(s_{i2}, c_{j2}) + \cdots + \Delta(s_{i7}, c_{j7})}{r_i} \tag{1}$$

$$\text{Total Distance}_R(s, c_j) = \sum_{i=0}^{n} D_R(s_i, c_j) \tag{2}$$

In a similar way, the list of preferred jobs is compiled by picking job cases that are the most similar to the user's *explicit preferences*. To do so, the corresponding distances in user's preferences (p_{ik}) are compared with the facets of new jobs (c_m) (refer to equation 3 and 4). *i* takes on values from 1 to *I*, where *I* is the number of specified facets and *k* takes on values from 1 to *K*, where *K* is the total number of the preferences which the user specified in each facet. *j* runs from 1 to *J*, where *J* is the total number of candidate jobs. In equation 3, for instance, p_1 is the job category facet and because users can specify multiple preferences for the same property, a user specified two preferences - database administrator and database analyst - and they correspond to p_{11} and p_{21}. These preferences are compared with the job category of a candidate jobs (c_{j1}) with the degree of flexibility of the preferences (whether the preference is 'strict' or 'flexible'). f_{11} is the flexibility degree of database administrator and f_{12} is the flexibility degree of database analyst. In other words, a candidate job is more likely to be chosen as a preferred job when the properties of the candidate job are closer to a user's preferences and the preferences are stricter. The distances of a certain preference facets are summed up by the total number of facets (*k*) and total number of candidate jobs (*j*) (refer to equation 4).

$$D_p(p_k, c_{jk}) = \sum_{i=0}^{I} \Delta(p_{ik}, c_{jk}) \times f_{ik} \tag{3}$$

$$\text{Total Distance}_p(p, c_j) = \sum_{j=0}^{J} \sum_{k=0}^{K} D_p(p_k, c_{jk}) \tag{4}$$

4. THE USER STUDY

The availability of Proactive, a full-fledged job search system that supports several methods of information access allowed us to run a user study comparing several methods in the same context. The main goal of our user study is to explore the value of the *versatility* of Proactive system in a realistic job search context. It means finding evidence that the merger of several information access approaches into one system helped users to find more relevant jobs than each of these technologies alone. Our secondary goal is to investigate the absolute and comparative performance of specific information access technologies. By performance, we mean the ability of a specific technology to find relevant information for users searching for jobs in a pool of real job offerings. In addition, we wanted to explore a few other aspects related to our recommendation mechanisms as well as to solicit direct user feedback.

4.1 Participants and Experimental Procedure

To increase the validity of our study, we specifically recruited University of Pittsburgh students who were genially interested in exploring actual jobs in the information technology area. In total, we recruited 17 participants. The group included 4 undergraduate students, and 13 graduate students. Their majors ranged from media communications to information sciences and telecommunications. Most of our subjects were planning to graduate within one or two semesters and were looking for jobs in the areas covered by the system.

The experiment started with an interview to elicit each participant's job interests. Then, participants were asked to set up his or her preferences explicitly through the Proactive interface (Fig. 3). Once the selection of preferences was completed, they explored Proactive to find jobs of interest using the list of most recent jobs (Fig. 1) and advanced search (Fig. 2). When participants found an interesting job, they saved it along with its 5-point-scale rating (5 points - very good, 3 points - relevant, 1 point – very bad) to define their implicit preference.

Once the participant had a sufficient (for recommendation approach) number of interesting jobs were found and saved, the experimenter used the system to generate a randomly mixed list of preferred and recommended jobs. In this list, the preferred jobs were generated using participants' explicitly-specified preferences and recommended jobs were generated using their implicit preferences encompassed in the saved jobs. The participants were asked to rate the jobs in the combined list using the same 5-point-scale ratings.

Finally, the participants were asked to fill in a questionnaire. The questionnaire asked how eager they were to find a job, and solicited their feedback on the usability of the three explored interfaces (for defining preferences, saving jobs, and conducting advance searches). Note that 'Preferred' and 'Recommended' job views were not used by the participants directly; instead, they rated a random mixture of these jobs in one list. It was done to avoid having the job rating and the type of recommendation impact the user's relevance judgment.

4.2 Analysis and Results

4.2.1 Traditional Technologies vs. Personalized Recommendations

As described, at the beginning of our experiments, we asked participants to set up their preferences explicitly through the interface shown on Fig. 3. On the basis of those preferences, we generated one kind of personalized recommendations, "preferred" jobs. The participants were also asked to find and save relevant jobs using *Most Recent Jobs* (i.e. hypertext browsing) and *Advanced Search* (i.e. ad-hoc information retrieval) views. Based on users' saved job list, as implicit preferences, we generated another kind of personalized recommendations, "recommended" jobs. Note that the browsing and search tools which were used for implicit preferences are both typical for Web information systems. Using a combination of browsing with column sorting and search with detailed query settings, users can discover current job trends: what kind of job is the most popular in specific job category(ies), which company is offering the most jobs, which is the most attractive city for a specific kind of jobs, and so on. While these typical methods are quite effective and powerful, they are clearly not sufficient to allow the users discovering all interesting jobs. In our study, personalized recommendations methods discovered many jobs that the users missed when using traditional methods. In fact, in our study, the majority of relevant jobs was suggested by the Proactive system rather than found by the users themselves (Fig. 4). We compared the total number of good jobs found by traditional technologies and per-
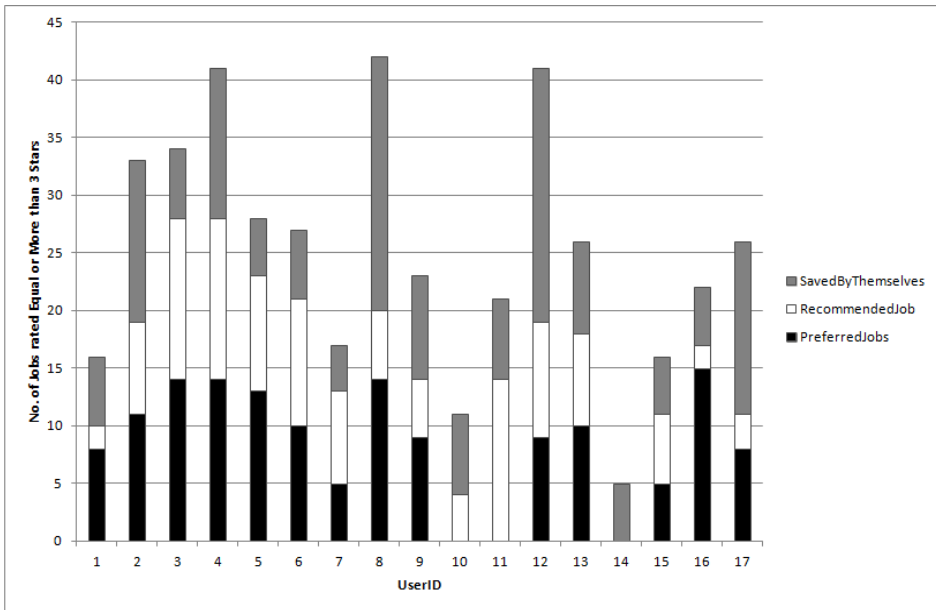
Fig. 4. The number of jobs rated with 3 or more stars

sonalized recommendations (i.e, considering only jobs rated as positive or relevant) using the t-test. The average numbers of the positively-rated jobs for saved jobs were significantly smaller ($M = 9.35$) than the personalized recommendations ($M = 15.89$), $t(16) = 3.29$, $p = .005$. The settings of our study make it hard to directly compare the ratings of jobs retrieved by traditional information access methods and personalized recommendations. When using search and browsing, only positively rated jobs were saved (and irrelevant jobs were ignored) while *all* jobs suggested by recommendation mechanisms had to be rated – both good and bad. Given that it was the participant's decision when to stop browsing and searching for jobs, however, we can at least interpret this data as evidence that each of the recommendation technologies is comparable to the effectiveness of the traditional browsing and search.

4.2.2 Recommended vs. Preferred Jobs: the Quality

As mentioned in the Section 4.1, each participant rated a set of personalized recommendations that consisted of randomly intermixed "recommended" and "preferred" jobs. The preferred jobs were generated using a participant's explicitly-stated preferences. The recommended jobs were generated using saved jobs as implicit preferences. Of 245 recommended jobs, more than 50% (51.1%) were rated positively (from very good to relevant; Fig. 4) while others (49.9%) were rated negatively (Table 2). Out of 238 preferred jobs, 60.9% were rated positively (Fig. 4) and the remaining 39.1% were rated negatively (Table 2). Since our experiment design simulated a realistic job search context without a predefined set of relevant documents, all relevance decisions made by the users were considered as "ground truth", i.e., 51.1% and 60.9% represented the precision of the evaluated methods (Fig. 4). In the domain of recommender and filtering systems, this fraction of positive ratings is considered as very good. Many reported recommender systems have precision less than 25% and are still considered as useful and successful. This

Table 2. Ratings of recommended jobs and preferred jobs

	N	Positive Rating	Relevant Rating	Negative Rating
Recommended Jobs	245	N = 69 (28.2%)	N = 56 (22.9%)	N = 120 (49.9%)
Preferred Jobs	238	N = 70 (29.4%)	N = 75 (31.5%)	N = 93 (39.1%)

demonstrates that both personalized recommender approaches are quite efficient in a realistic job search context.

To test the mean differences of ratings in the two kinds of recommendation, t-test was executed. There was no significant difference of ratings between recommended jobs ($M = 2.82$) and preferred jobs ($M = 2.65$), $t(227) = 1.75$, $p = .082$. While the mean data hints that preferred jobs might be more useful for the users, the current study can't statistically distinguish the quality of recommendations produced using implicit preferences (recommended jobs) and explicit preferences (preferred jobs).

4.2.3 Saved vs. Recommended vs. Preferred Jobs: the Overlap

As demonstrated above, *saved* jobs (found by browsing and search), preferred jobs (generated from explicit preferences), and *recommended* jobs (generated from implicit preferences) can be considered as comparable approaches to finding relevant jobs. However, are these approaches redundant or complementary? Are these approaches simply different ways to discover the same set of good jobs or could each of them help users to discover a unique set of good jobs? To answer this question, we determined how large the overlap is between the sets of jobs produced by the three approaches. We found that the overlap was very small (Fig. 5). The lists of preferred jobs and saved jobs contained only 5 common job cases out of a total of 406 cases (the ratings of the common jobs were identical with the exception of one case). Only 9 job cases out of 483 were found in both the preferred jobs and recommended jobs lists (Fig. 5). Due to the design of the recommendation algorithm, there was no overlap between saved and recommended jobs. Therefore, we observe that each of these three approaches can produce a very unique set of jobs. We can't single out one approach as being significantly better than others; in fact, we can't even drop any of these approaches without eliminating a large fraction of the good jobs. We consider this to be a strong argument in favor of versatility. Our data shows that it is vital for a successful information system to offer several different methods to find relevant information. Traditional
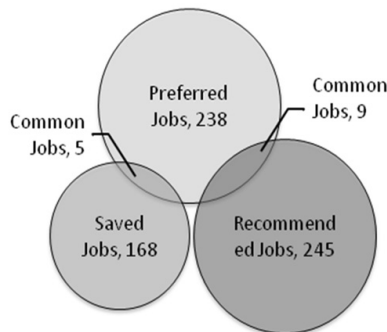

Fig. 5. Different information access approaches generated essentially different pools of job cases with almost no overlap

browsing and search are, by far, not sufficient to find relevant information and have to be bolstered with recommendations. Moreover, just one type of recommendation is not sufficient either – both explicit and implicit preferences should be exploited to find good jobs in the huge number of jobs cases.

4.2.4 Distance-Based Recommendations vs. User Ratings

As mentioned in the Section 3.2, Proactive uses distance-based KNN algorithm. The closest cases are expected to be most relevant and are presented to the user as "recommended" or "preferred" jobs. The quality of such a distance-based approach could be evaluated from two prospects. From one prospect, we can check to what extent the closest cases are really relevant as ranked by the user. This prospect is *analogous to precision* and was assessed in section 4.2.2. From the other prospect (*analogous to recall*), we can check whether the cases highly-rated by the user are really closest to the set of saved jobs (or explicit preferences).

To assess the quality of preferred job recommendations from this prospect, we divided the set of all preferred jobs into three subsets: a set of 'good jobs' (rated as 'very good' and 'good'), a set of 'relevant jobs' (rated as 'relevant') and a set of 'bad jobs' (rated as 'very bad' or 'bad'). To prove the quality of the distance-based approach, we had to show that the set of good jobs is the closest to the explicit preferences and the set of bad jobs is the most distant. We performed ANOVA on the distance-to-user preferences among the three subsets. We found a significant difference in the weight distance among the levels of ratings, $F(2, 235) = 11.00$, $p < .001$. From the results of post hoc pair wise comparisons, Scheffé pairwise test, there was a likelihood of decrease in the distance value as the rating improved (Fig. 6). The distances of good jobs ($M = 17.26$) were significantly lower than the distance in badly rated jobs ($M = 56.06$), $p < .001$; the relevant jobs ($M = 22.73$, $SE = 6.71$) were significantly lower than bad jobs, $p < .001$, as well. There was no significant difference between the good and relevant jobs ($p < .57$). The same process was repeated for the set of recommended jobs with respect to their distances to the saved jobs. However, we found no significant differences among the three levels of rating, $F(2, 242) = .678$, $p = .509$, $\eta^2 = .01$. This result points out that the distance-based prediction of recommended jobs is not as good as prediction of preferred jobs. We discuss this issue in more detail in the conclusions.

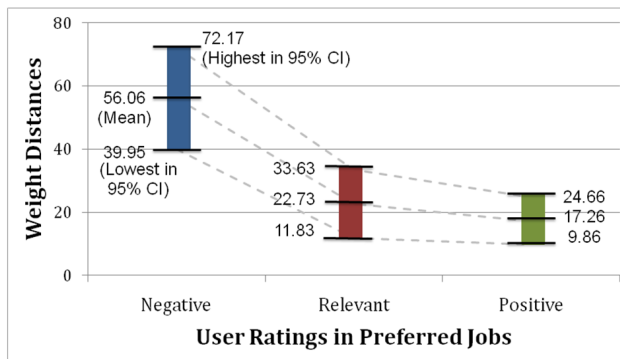We also attempted to evaluate the distance-based approach as a whole (i.e., for recommended

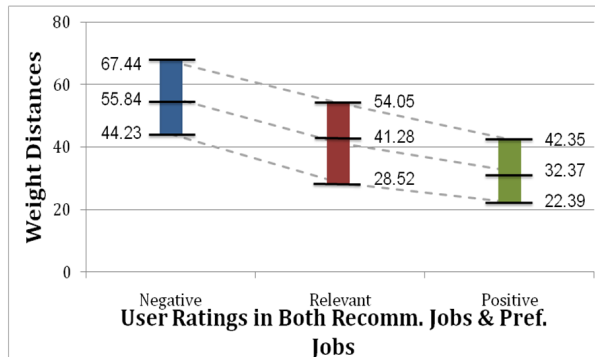Fig. 6. Distribution of weight distance of preferred jobs by ratings

Fig. 7. Distribution of weight distance of preferred jobs and recommended jobs by ratings

jobs and preferred jobs together). A one-way ANOVA revealed a significant group difference by ratings, $F(2, 480) = 4.257$, $p = .015$. To find the pattern of differences on the weight distance, Scheffé post hoc tests were examined. The distances for the set of good jobs ($M = 32.37$) were significantly lower than distances for bad jobs ($M = 55.84$), $p = .005$ (Fig. 7). However, there were no other significant differences for the weight distance among the rating levels (relevant ($M = 41.28$) vs. bad, $p = .09$; good vs. relevant, $p = .34$). 22.39 was the optimal weight distance threshold for a 'good' rating, and 42.35 was the potential maximum threshold range for a non-negative evaluation.

4.2.5 The Volume of User Preferences and Recommendation Quality

The previous sections evaluated the performance of the KNN-based recommendation approach for the whole set of users. However, it is natural to expect that the quality of recommendation may depend upon the volume of preferences specified by the user.

For the preferred jobs, this volume can be measured as the total number of job preferences specified by each user. The mean number of preferred job categories was 14.06 ($SD = 7.99$) and the mean number of other preferences was 6.06 ($SD = 4.337$). The correlation between the total number of specified preferences and the distance between job cases and the set of preferred jobs was significantly negative, $r = -.48$, $p < .001$. The correlation between the number of preferred job categories and weight distance for the preferred jobs was also significantly negative, $r = -.53$, $p < .001$. This data provides evidence that the more preferences specified by the user, the shorter the distance. Given the relationship between distance and quality investigated above, the shorter distance will result in better generated recommendations. In addition, specifying preferred job categories appears to be important for making accurate recommendation.

For the recommended jobs, the volume of preferences can be measured as the number of saved jobs. The correlation between the total number of saved jobs and the distance between recommended jobs and the set of saved jobs was significantly negative, $r = -.17$, $p = .01$. This is a strong evidence that the more jobs are saved, the better the system comprehends the user's implicit preferences. Both findings are very important: the data provides evidence that the quality of recommendations can be improved by investing time in specifying richer preferences or finding more relevant jobs to save.

4.2.6 Subjective Evaluations and Other Results

The subjective evaluation of the system was positive, overall. Participants considered the functionality to save a job as interesting ($M = 4.03$), easy to use ($M = 4.62$) and relaxing ($M = 4.50$). The interface for setting up preferences was perceived as positive (interesting, $M = 3.97$; easy to use, $M = 4.41$; and relaxing, $M = 4.09$), too. The user assessment of the advanced search functionality was less enthusiastic, but still positive ($M = 3.66$). Finally, there was significance in the rating differences between undergraduate students and graduate students. Ratings done by graduate students ($M = 3.23$,) were higher than those done by undergraduates ($M = 2.36$), $t(649) = -7.55$, $p < .001$. This could be connected with the result that graduate students specified larger number of preferences, both implicit and explicit. The number of preferences saved by the graduate students ($M = 23.17$) was significantly higher than that of undergraduates ($M = 16.31$), $t(15) = -7.88$, $p < .001$. The average number of saved jobs was also significantly higher for graduate students ($M = 10.62$) than for undergraduate students ($M = 7.50$), $t(15) = -6.04$, $p < .001$.

5. CONCLUSION AND DISCUSSION

This paper argues for the need to support multiple ways of information access in Web-based information systems. As an example of a comprehensive information access system, we introduced *Proactive*, a job recommender system that supports four different ways of information access for job information (browsing, search, explicit preference-based recommendations, and implicit preference-based recommendations). The evaluation with 17 participants demonstrated that the traditional search and browsing-based access are by far not sufficient to discover relevant information. Each of the explored recommendation approaches was able to gather about the same number of useful jobs as search and browsing combined. Most importantly, there was almost no overlap between three sets of jobs. Thus, the comprehensive nature of the system was important: each information access technology enabled the users to discover a unique set of good job cases, which would not be found without this technology.

More specifically, a relatively simple distance-based approach used in both recommendation algorithms demonstrated reasonable performance. Almost 30% of jobs cases found by each of the recommendation algorithms were judged as good or very good and more than half of the cases were found to be at least relevant.

Consequently, the contribution of the *Proactive* recommender system is twofold. First, *Proactive* demonstrates a way to implement an information system which integrates several information access approaches while demonstrating the importance of such integration. Second, it presents an efficient approach to organize fully-personalized access to information about job cases, which is based on automatic processing, semantic classification, and distance-based recommendation.

However, one possible source of weakness was the attempt to find jobs which are close to the whole set of saved jobs. Given the number of explicitly-specified job types, it is safe to assume that some users were looking for several different jobs. In this context, it could be more appropriate to cluster saved jobs and to recommend jobs which are similar to one of the clusters, not to the whole set of jobs. Another problem may be related to our approach to combine all preference criteria to calculate recommended jobs. In the future studies, we plan to investigate whether a recommendation based on *one* preference criteria is more helpful. The analysis of usage pat-

terns using log data mining is another direction for improving the quality of recommendation. On the user interface side, one possible direction for system improvement is the use of faceted browsing instead of simple list browsing. Faceted browsing might be especially beneficial for less professional users, who may not know how to use regular search efficiently. Since the information in Proactive is organized with several ontologies and taxonomies, faceted browsing will be a most natural extension.

REFERENCES

[1]   Aktas, M.S., M. Pierce, G.C. Fox, and D. Leake, *A Web based Conversational Case-Based Recommender System for Ontology aided Metadata Discovery*, in *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing*. 2004, IEEE Computer Society. pp.69-75.

[2]   Belkin, N.J. and W.B. Croft, *Information filtering and information retrieval: two sides of the same coin?* Commun. ACM, 1992. 35(12): pp.29-38.

[3]   Bridge, D. and F. Ricci, *Supporting product selection with query editing recommendations*, in *Proceedings of the 2007 ACM conference on Recommender systems*. 2007, ACM: Minneapolis, MN, USA. pp.65-72.

[4]   Brusilovsky, P. and C. Tasso, *Preface to Special Issue on User Modeling for Web Information Retrieval.* User Modeling and User-Adapted Interaction, 2004. 14(2): pp.147-157.

[5]   Gadanho, S.C. and N. Lhuillier, *Addressing uncertainty in implicit preferences*, in *Proceedings of the 2007 ACM conference on Recommender systems*. 2007, ACM: Minneapolis, MN, USA. pp.97-104.

[6]   Kleinberg, J.M., *Two algorithms for nearest-neighbor search in high dimensions*, in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*. 1997, ACM: El Paso, Texas, United States. pp.599-608.

[7]   Lang, S., S. Laumer, C. Maier, and A. Eckhardt, *Drivers, challenges and consequences of E-recruiting: a literature review*, in *Proceedings of the 49th SIGMIS annual conference on Computer personnel research*. 2011, ACM: San Antonio, Texas, USA. pp.26-35.

[8]   Laumer, S. and A. Eckhardt, *Help to find the needle in a haystack: integrating recommender systems in an IT supported staff recruitment system*, in *Proceedings of the 47th annual conference on Computer personnel research*. 2009, ACM: Limerick, Ireland. pp.7-12.

[9]   Lee, D.H. and P. Brusilovsky, *Fighting Information Overflow with Personalized Comprehensive Information Access: A Proactive Job Recommender*, in *Proceedings of the Third International Conference on Autonomic and Autonomous Systems*. 2007, IEEE Computer Society. p.21.

[10] Lee, I., *An architecture for a next-generation holistic e-recruiting system.* Commun. ACM, 2007. 50(7): pp.81-85.

[11] Lee, I., *Modeling the benefit of e-recruiting process integration.* Decis. Support Syst., 2011. 51(1): pp.230-239.

[12] Li, J. and O. Zaïane, *Combining Usage, Content, and Structure Data to Improve Web Site Recommendation E-Commerce and Web Technologies*, K. Bauknecht, M. Bichler, and B. Pröll, Editors. 2004, Springer Berlin / Heidelberg. pp.313-315.

[13] Micarelli, A., F. Gasparetti, F. Sciarrone, and S. Gauch, *Personalized search on the world wide web*, in *The adaptive web*, B. Peter, K. Alfred, & N. Wolfgang, Editors. 2007, Springer-Verlag. pp.195-230.

[14] Middleton, S.E., H. Alani, N.R. Shadbolt, and D.C.D. Roure, *Exploiting synergy between ontologies and recommender systems*, in *Semantic Web Workshop 2002, World Wide Web Conference*. 2002, Sementic Web Workshop 2002, WWW2002: Hawaii, USA.

[15] Montaner, M., B. López, and J.L.D.L. Rosa, *A Taxonomy of Recommender Agents on the Internet.* Artif. Intell. Rev., 2003. 19(4): pp.285-330.

[16] Peter, G. UPRE: User Preference Based Search System. 2006.

[17] Ramli, R., S.A. Noah, and M.M. Yusof, *Ontological-based model for human resource decision support system (HRDSS)*, in *Proceedings of the 2010 international conference on On the move to*

meaningful internet systems. 2010, Springer-Verlag: Hersonissos, Crete, Greece. pp.585-594.

[18] Schafer, J.B., J. Konstan, and J. Riedi, *Recommender systems in e-commerce*, in *Proceedings of the 1st ACM conference on Electronic commerce*. 1999, ACM: Denver, CO, United States. pp.158-166.

[19] Schafer, J.B., J.A. Konstan, and J. Riedl, *E-Commerce Recommendation Applications.* Data Min. Knowl. Discov., 2001. 5(1-2): pp.115-153.

[20] Singh, A., C. Rose, K. Visweswariah, V. Chenthamarakshan, and N. Kambhatla, *PROSPECT: a system for screening candidates for recruitment*, in *Proceedings of the 19th ACM international conference on Information and knowledge management*. 2010, ACM: Toronto, Canada. pp.659-668.

[21] Smyth, B., K. Bradley, and R. Rafter, *Personalization techniques for online recruitment services.* Commun. ACM, 2002. 45(5): pp.39-40.

[22] Tvarozek, M. and M. Bielikova, *Adaptive Faceted Browsing in Job Offers.* Proceedings in Informatics and Information Technologies: Tools for Acquisition, Organization and Persenting of Information and Knowledge, 2007: pp.149-160.

[23] *Viappiani*, P., P. Pu, and B. Faltings, *Conversational recommenders with adaptive suggestions*, in *Proceedings of the 2007 ACM conference on Recommender systems*. 2007, ACM: Minneapolis, MN, USA. pp.89-96.

Danielle Lee

Danielle H. Lee is a doctoral candidate at the School of Information Sciences, University of Pittsburgh. She received her Master's degree in Information Science at Syracuse University. She worked for Samsung SDS Co. as a software engineer, Jangan College as an invited professor and Center for Dental Informatics, University of Pittsburgh as a researcher. Her research interests are personalized recommendations using users' social networks, behavioral influence and information propagation in online social networks and knowledge mining of collective intelligence.

Peter Brusilovsky

Peter Brusilovsky is a Professor of Information Science and Intelligent Systems at the University of Pittsburgh, where he also directs Personalized Adaptive Web Systems (PAWS) lab. He has been working in the field of adaptive educational systems, user modeling, and intelligent user interfaces for more than 20 years. He published numerous papers and edited several books on adaptive hypermedia and the adaptive Web.