Supporting Exploratory People Search: A Study of Factor Transparency and User Control

Shuguang Han, Daqing He, Jiepu Jiang, and Zhen Yue

School of Information Sciences, University of Pittsburgh
135 N Bellefield Ave., Pittsburgh, PA, United States
shh69@pitt.edu, dah44@pitt.edu, jiepu.jiang@gmail.com, zhy18@pitt.edu

ABSTRACT

People search is an active research topic in recent years. Related works includes expert finding, collaborator recommendation, link prediction and social matching. However, the diverse objectives and exploratory nature of those tasks make it difficult to develop a flexible method for people search that works for every task. In this project, we developed PeopleExplorer, an interactive people search system to support exploratory search tasks when looking for people. In the system, users could specify their task objectives by selecting and adjusting key criteria. Three criteria were considered: the content relevance, the candidate authoritativeness and the social similarity between the user and the candidates. This project represents a first attempt to add transparency to exploratory people search, and to give users full control over the search process. The system was evaluated through an experiment with 24 participants undertaking four different tasks. The results show that with comparable time and effort, users of our system performed significantly better in their people search tasks than those using the baseline system. Users of our system also exhibited many unique behaviors in query reformulation and candidate selection. We found that users' general perceptions about three criteria varied during different tasks, which confirms our assumptions regarding modeling task difference and user variance in people search systems.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval - Search process; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - *graphical user interfaces (GUI), user-centered design*

Keywords

Exploratory Search; People Search; Interactive People Search;

1. MOTIVATION AND INTRODUCTION

With the help of powerful search engines such as Google and Bing, people have access to numerous web documents. However, there are many occasions when locating the right person is more important than finding the right documents [1]. There are various reasons to motivate users to look for people. For example, researchers may want to find appropriate collaborators for a cross disciplinary project [2]; program chairs of an academic conference may need to search for experts to serve on the program committee [3]; companies may need to locate qualified job candidates [4]; and online question answering websites may need to match the right experts to the incoming questions [5].

People search is not a new research topic. Some earlier works include automatic assigning reviewers for manuscripts [6], finding experts for a given topic [7] and locating potential collaborators [8]. This topic was treated as either an information retrieval or a recommendation problem: related techniques such as content based retrieval [9], collaborative filtering [10], and hybrid algorithms combining both have been used. However, as Terveen and McDonald [11] pointed out, it fundamentally changes the nature of the problem when the returned results are people rather than documents. People are social creatures; "assessing" people is significantly more complex than assessing web documents. As a result, they proposed the concept of "social matching" to emphasize the "social" dimension. "Social matching" systems such as Referral Web [12] and Expertise Recommender [13] are able to return highly relevant candidates who are also socially related to the information seekers.

Motivated by the fact that human decision making is a complex process that requires balancing different factors, recent studies on people search have been focused on better modeling of the people search problem, and more effective combination of multiple factors. Systems such as Smallblue [14], SaNDVis [15] and CollabSeer [16] integrated content matching with social networks to find appropriate candidates. IRIS-IPS [3] went on to take into account the candidates' authoritativeness as one additional dimension, and argued that authoritativeness is useful in people search tasks such as keynote speaker finding. Additional attributes such as credibility, accessibility and trust were also identified as important factors in designing an expert locating system [1]. Hofmann et al. [42] investigated the influence of several contextual factors, including availability, approachability, and physical proximity, on people's searching for similar experts. Those factors were also studied in other tasks such as finding co-workers [43] and helpers [44] in the enterprise environment. However, the combination of those factors has not yet been well-studied. The approaches to fusing multiple factors could be a simple linear combination with fixed parameters indicating the importance of each factor [16], but parameter learning in linear combination method is usually difficult because of the lack of gold standard data. Therefore, other researchers [14, 15, 18] simply visualized multiple factors in an integrated interface and made the factors transparent and controllable to users.

Current approaches have several important limitations. First of all, people search is performed within specific contexts. These contexts can significantly influence the selection of factors and the importance of each factor. For example, close social connection is likely a positive factor in collaborator finding, but it may be

viewed as a negative factor in paper reviewer assignment due to "conflict of interest". High authoritativeness/ reputation can be a positive factor in finding keynote speakers, but may not be that important in finding collaborators.

Secondly, different users may interpret the same task differently based on their own understanding, which could result in diverse judgments on the importance of each factor. However, most of the current studies either ignored the involvement of users [6, 19] or only tried to utilize the global optimized parameters [16, 20, 21] and assumed that users preferences are constant.

Finally, people search itself is a complex process and is influenced by many factors. Most tasks for people search are exploratory in nature [23], in which users only start with vague ideas about for whom they are searching. It is through the exploration of, and comparison among, the candidates that the users gradually form clear criteria for final selection. Therefore, users would prefer process transparency and procedure control. Fully automatic algorithms are often too opaque for users [22].

To resolve the above challenges, we try to support the exploratory process of people search tasks. An interactive people search system is developed to demonstrate and evaluate the idea. Our design of the interface follows the style of a faceted search browser [23] in order to better support users' exploration and information seeking process. We highlight three different factors and allow users full control of the exploration of those factors: topic relevance, candidate authoritativeness and social similarity between users and candidates. These factors are presented in detail in Section 3. Another contribution of this paper is that we exemplified several people search scenarios in which the simple keyword search is insufficient. The people search tasks explored in this study are: finding new collaborators, finding conference mentors, finding external thesis committee members and suggesting reviewers. To evaluate our approach, we conducted a user based experiment. The research questions in this study are:

- Can an interactive people search system equipped with user control and factor transparency perform better in exploratory people search tasks than existing system?

- How do factor transparency and user control influence the people search process?

- How do users rate the importance of each factor in different people search tasks?

- Which factors are important in different people search tasks?

2. RELATED WORKS

People search covers several research topics such as expert finding [19], collaborator discovery [16], reviewer assignments [6], coauthor prediction [20], and job candidate finding [4]. Other topics such as keynote speaker finding and conference program committee member finding were also proposed, but not well-investigated because of the difficulty in evaluation [3].

Expert finding has been studied extensively in the literature on enterprise and organizational knowledge management [1]. Related techniques were not well-studied until the expert search task of TREC, which provided generic datasets and benchmarks for evaluation. A widely adopted method was to construct expert profiles using previous publications or co-occurring texts [19]. Later the method was extended to expert finding in the web environment [24], academic domain [25], and social websites

[26]. Reviewer assignment abilities help conference organizers or journal editors to match manuscripts with appropriate reviewers [6], which were also viewed as an "expert finding" task [27, 28].

More recently, the expert finding problem was studied with regard to social media websites such as Facebook and Twitter. For example, the SearchBuddies system automatically provided responders to questions posted in Facebook's status messages by matching the most relevant social contacts in Facebook [29]. The Cognos system [30] recommends topic experts in Twitter by considering multiple metadata information. The user generated metadata information was discovered to be an important resource for identifying people's expertise. Similar conclusions were also drawn in [31], in which the Chinese version of Twitter was used.

Expert finding in academic settings can also take advantage of coauthor networks. For examples, collaborator discovery has been modeled as a link prediction problem [20] and the authors with high similarity were assumed to have a high probability of collaboration. The similarity between two authors was computed and the most similar pairs were chosen as the predicted links. Network topological features such as the common neighbor measures, the shortest path were usually adopted to measure the vertex similarity. In [20], the authors found that the Katz index has consistently performed best, and the simple common neighbor measures also perform well. If the prediction is for each individual, the coauthor prediction is a personalized recommendation problem. The classic Collaborative Filtering (CF) method was used in [32] for people-to-people recommendation, but CF often suffers the cold-start problem when data is sparse. A hybrid method that combines multiple data sources can help relieve the data sparseness [17].

Previous works [2, 3, 14-16, 33] found that combining multiple factors such as content and social relations will be useful in finding appropriate candidates. The algorithms for combining different factors can be either the simple linear combination [3, 16], regularization based combination [2], or the random walk based method [33]. Those algorithms assumed uniform users' preferences, and they tried to learn global optimized parameters for all users. However, users may have significantly different preferences about the factors in different scenarios [44]. To solve this problem, related works [22, 47] on item recommender system involved users in the search process, and made the users explicitly express their preferences by providing the transparency of algorithm process and the used data sources. Compared to item recommender systems, searching for people is even more complicated [34]. It is unclear whether or not users would benefit from the system, and whether or not users can really use the recommended candidates [35]. A general framework to handle diverse people tasks and different user preferences is still lacking, which is the focus of this project.

3. THE PEOPLE EXPLORER SYSTEM

We developed the PeopleExplorer, an interactive integrated people search system, the main features of which are showcased in Figure 1.

3.1 System Design Rationale

Search box and document surrogate are two basic components in a Search Engine Results Page [36]. Instead of the document surrogate, the *candidate surrogate* provided by our system represents a returned candidate.
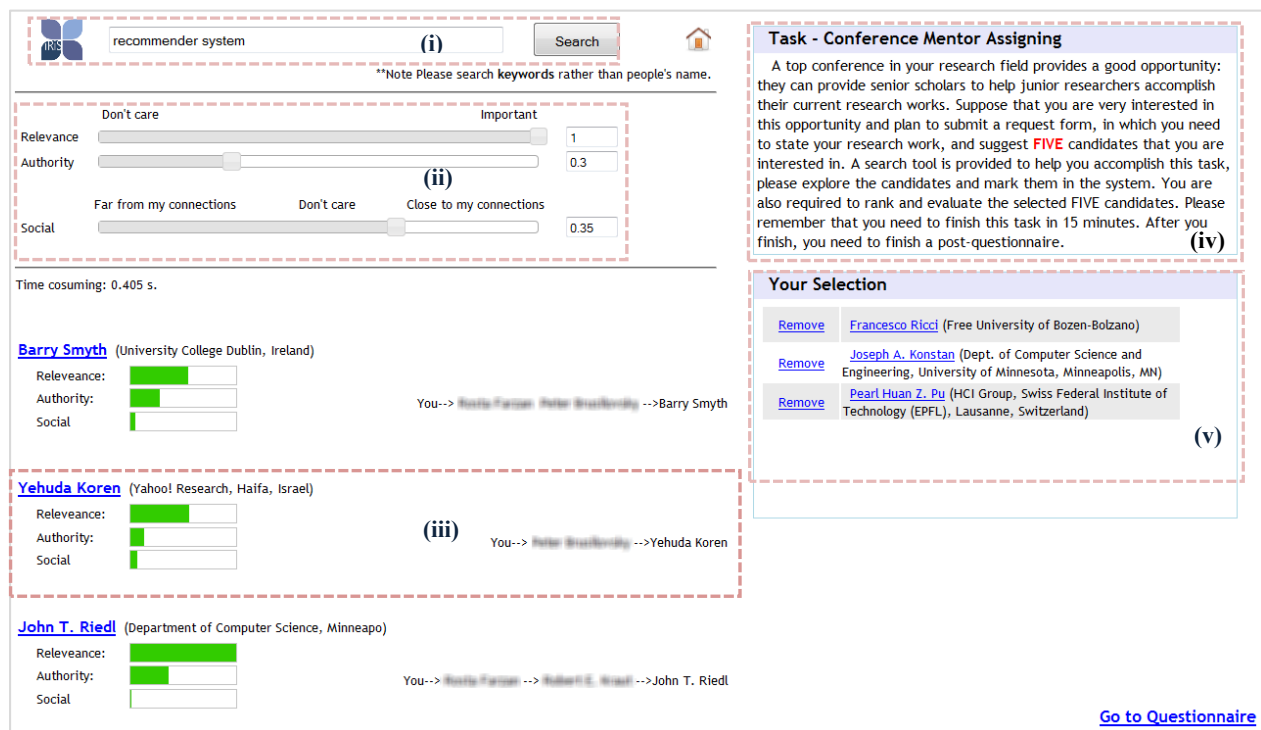
Figure 1: A screenshot of PeopleExplorer: (i) a search box; (ii) three factors: "relevance", "authoritativeness" and "social"; the sliders are used to tune the weights of each factor. In the "*Relevance*" and "*Authority*" sliders, users can tune between "*Don't care*" and "*Important*"; in the "*Social*" slider, users can tune between "*Far from my connections*" and "*Close to my connections*". (iii) Candidate representation, including a visualized score for each factor and the connection(s). Eight candidates are displayed in each page. The author name hyperlinks to the ACM author profile page. (iv) the task description and (v) the selected candidates.

The *search box* is shown as (i) in Figure 1. Rather than soliciting people's names, we asked users to issue keywords describing a subject domain or a topic in the search box. This decision was made to support exploratory searches. Thus, lookup search [23] for known people is not included.

The *candidate surrogate* is provided as seen in Part (iii) in Figure 1. Candidates' names and affiliations are presented at the top. Then, three green bars are used to present the candidate scores on three factors ("relevance", "authoritativeness" and "social similarity"). Our intention is to help users understand why the candidate is returned. Also for each candidate, we provide the possible connection path(s) between the user and the candidate.

To support the exploratory tasks, additional features such as offering workspace and task management [37] are often provided in the general search tasks. PeopleExplorer also offers a task description panel (Part iv in Figure 1) and a workspace panel (Part v in Figure 1) to help users' management of their current progress.

The most important functionality of PeopleExplorer is the user-controlled *sliders*, shown as (ii) in Figure 1. Each of the sliders represents one factor. Users can adjust any combinations of those factors. For the *social* factor, users can specify their needs to find the candidates who are "*far from my connections*" or who are "*close to my connections*". The value for each factor is its weight/ importance, not the threshold value. *Relevance* measures the content relevance between candidates' expertise (represented by candidates' publications) and issued queries. *Authoritativeness* denotes the importance of each candidate in the coauthor networks of our dataset (see Section 4.2). *Social* indicates the social similarity between each candidate and the user (see section 4.3).

3.2 The Dataset

PeopleExplorer uses a dataset containing 151,165 conference papers that were published in the ACM Digital Library between 2000 and 2011. The dataset was automatically crawled from the ACM Digital Library websites. Our dataset contains only the ACM hosted conferences[1]. Each paper's record includes a title, an abstract and the authors. The authors were disambiguated using the ACM author identifiers, which is a unique number assigned to each individual person by the ACM Digital Library. In total 209,592 unique authors were identified. Coauthor relationships were extracted to create a coauthor network. A link between authors is added if they coauthored at least one paper; however, we didn't consider the collaboration frequency.

4. METHODS

Our system considered both the content relevance and the coauthor networks. In terms of using networks, either the network measure using local network information or the network measure using entire network information have been adopted in previous studies [16, 20], but they were usually used separately. Han et al. [17] showed that it is better to combine both, particularly when the target users are junior researchers. Thus, we integrate three types of information: the content relevance (*Relevance*), the local network measure (*Social*), and the global network measure (*Authoritativeness*).

4.1 Content Relevance

Balog [19] formalized the expert retrieval problem as the calculation of the probability a candidate *ca* being an expert for

---

[1] http://dl.acm.org/proceedings.cfm

the given query q, i.e. $p(ca|q)$. To compute $p(ca|q)$ in an academic dataset, we utilized the method proposed by Deng et al. [25] because it can incorporate citation information. In their model, $p(ca|q)$ is computed using Formula (1). Assuming that ca is conditionally independent of q given document d, the formula can be transformed into Formula (2), in which the document d in the collection becomes the bridge to connect q and ca.

$$p(ca|q) \propto \sum_d p(ca|d,q)p(q|d)p(d) \qquad (1)$$

$$= \sum_d p(ca|d)p(q|d)p(d) \qquad (2)$$

To compute $p(q|d)$ we use the query likelihood language model score with dirichlet smoothing. We adapted Indri [2] to build index for documents and compute document score $p(q|d)$ for each query q using #combine in Indri. $p(d|ca)$ denotes the document-candidate association, which is defined as Formula (3). n_d is the number of authors in document d, and A_d denotes the set of authors for document d. $p(d)$ is the prior probability of each document. Although multiple factors could affect the prior probability, here we modeled it as the probability of the document d being cited by researchers, which is proportional to the number of citations, i.e. c_d [25] (see Formula 4).

$$p(ca|d) = \begin{cases} \dfrac{1}{n_d}, ca \in A_d \\ \\ 0, ca \notin A_d \end{cases} \qquad (3)$$

$$p(d) \propto \ln(e + c_d) \qquad (4)$$

4.2 Candidate Authoritativeness

The authoritativeness values were computed using the PageRank algorithm [38]. We directly applied this algorithm to the coauthor networks without any changes. Our assumption is that one author is more likely to be authoritative when he/she has many connections with other authoritative authors. Since PageRank was initially designed for directed networks whereas the coauthor relationship is bidirectional, we deconstructed one unidirectional coauthor link into two directional links [39].

4.3 Social Similarity

We considered the social similarity $S_{u,ca}$ between candidate ca and the user u. Since our dataset might not have sufficient information for the user u, especially for those junior researchers recruited in our later user experiments, PeopleExplorer asks users to provide his/her social connections by entering his/her closely connected colleagues or friends u'. Users can provide more than one colleague or friend, and we represented them as SC_u.

$S_{u,ca}$ is computed using Formula (5), in which $w_{u,u'}$ denotes the importance/ weight of u' to the user u, and C is a normalization constant. $S_j(a, a')$ measures the social closeness between author a and author a'. It is defined in Formula (6). If author a and author a' are coauthors, their social closeness is set to 1, otherwise, it is measured by how many coauthors they have in common. $\Gamma(a)$ denotes a set of coauthors for a. In our experiment, we set $w_{u,u'}$ uniformly to $1/|SC_u|$. Even if user u does not share common neighbors with candidate ca, user u will also receive the score passed from user u' if his/her closely connected connection u' shares common neighbors with candidate

ca. That is why we can see the third connection path in Figure 1 which two authors between the user and the candidate *John Riedl*.

$$S_{u,ca} = \frac{1}{C} \left\{ S_j(u, ca) + \sum_{u' \in SC_u} w_{u,u'} S_j(u', ca) \right\} \qquad (5)$$

$$S_j(a, a') = \begin{cases} \dfrac{|\Gamma(a) \cap \Gamma(a')|}{|\Gamma(a) \cup \Gamma(a')|}, a' \notin \Gamma(a) \\ \\ 1 \qquad , a' \in \Gamma(a) \end{cases} \qquad (6)$$

4.4 Integration

For a given query q, the system needs to rank each candidate ca based on the integration of the scores from three factors: the content relevance score $p(ca|q)$, the candidate authoritative score pr_{ca} and the social similarity score $S_{u,ca}$. The distribution of the scores from each factor is highly skewed: most of the scores are distributed in a tightly focused range and the rest are scattered in other ranges. In order to make the user feel that the change in factors will cause the result to change more smoothly, we computed the logarithmic values of the score from each factor. We interpolated them using three weights: w_e, w_p and w_s for *relevance*, *authoritativeness* and *social similarity*, respectively. w_e and w_p ranged from 0 to 1, but w_s ranged from -1 to 1 because the PeopleExplorer system also recommends candidates who are distant from the user's social connections. Users can specify their preferences/weights (i.e. w_e, w_p and w_s) for each factor in the search interface. The integrated score $P_{ca|u,q}$ is shown in Formula (7).

$$P_{ca|u,q} = w_e \log p(ca|q) + w_p \log pr_{ca} + w_s \log S_{u,ca} \qquad (7)$$

5. EXPERIMENT DESIGN
5.1 The Baseline System

A set of experiments (involving humans) was conducted to evaluate the PeopleExplorer system. The experiment was designed as a comparative study to evaluate the system's effectiveness and efficiency against a baseline system. PeopleExplorer acts as the experimental system (**E**) (see Figure 1), whereas a baseline system (**B**) was obtained by keeping the same interface except removing the three sliders. This difference between the two systems helps us to focus on exploring how the sliders affect user behaviors in the PeopleExplorer system.

The baseline system returns candidates based on the content relevance scores. Here we did not consider using the network based method and the hybrid method as baselines for the following reasons: 1) the network analysis methods do not explicitly model the search query but do provide the same recommendations regardless users' input; 2) the hybrid method needs training data to learn the weights for different factors such as social and content relevance in advance. In addition, users' preferences for different tasks (or even the same task) are different, one-size-fits-all solutions for all users and all tasks do not exisit. If we determine fixed parameters, the performance will be distorted by the way we choose those parameters.

5.2 Exploratory People Search Tasks

As mentioned in Section 1, search tasks are usually performed within certain contexts, in which the factors and the importance of each factor may diverge. In order to demonstrate the ability of PeopleExplorer to support diverse tasks, we simulated four different people search tasks: *Conference Mentor Finding*, *New*

Collaborator Finding, *External Thesis Committee Member Finding* and *Reviewer Suggestion*. The simulated work task situation was treated as a useful way of evaluating Interactive Information Retrieval systems [45].

Task 1: Conference Mentor Finding

A top conference in your research field provides a good opportunity: senior scholars offer helps to junior researchers on their current research work. Suppose that you are very interested in this opportunity and plan to submit a request form, in which you need to include your research statement; and suggest **FIVE** appropriate mentors in their database in whom you are interested. A search tool is provided to help you accomplish this task. Please explore the candidates and mark them in the system. You are also required to rank and evaluate the selected candidates after you finish the task.

Task 2: New Collaborator Finding

Suppose you (and your advisors) are planning to write a project proposal for a funding agency (e.g. NSF, Google Grant). In addition to your current coauthors, you want to find appropriate new collaborators to enhance the diversity of your team. A search tool is provided to help you identify potential collaborators. You need to find **FIVE** candidates and mark them in the system. You are also required to rank and evaluate the selected candidates after you finish the task.

Task 3: External Thesis Committee Member Finding

You need to form a committee for your doctoral dissertation, with at least one external committee member from other departments or other universities. You understand that not everyone will be willing to be a committee member, and you don't want to put all your hopes on one person. Your strategy is to find **FIVE** appropriate candidates and contact them one-by-one until someone accepts. A search tool is provided to help you accomplish this task. Please explore the candidates and mark them in the system. You are also required to rank and evaluate the selected candidates after you finish the task.

Task 4: Reviewer Suggestion

Suppose that you are planning to submit a paper manuscript to a journal. The journal editors ask you to suggest **FIVE** appropriate researchers that might be able to review your paper. Fortunately, the journal provides a search tool to help you explore the reviewers. You are asked to use this tool and mark the candidates in the system. You need to find **FIVE** candidates. You are also required to rank and evaluate the selected candidates after you finish the task.

Figure 2: **Descriptions of four people search tasks.**

These four tasks represent the range of people search activities in academic settings. More importantly, we think these tasks would have varying impacts on the three factors in PeopleExplorer. Our initial expectations about the importance of the three factors are: candidate authoritativeness might be the most important factor in finding conference mentors; social closeness is important in finding appropriate coauthors; both candidate authoritativeness and social similarity are important factors in finding external thesis committee members; and social closeness might be a negative factor in the reviewer suggestion task due to the potential for "conflict of interest".

5.3 Experiment Procedures

The experiment began with an introduction to the study. Then, the participants were asked to fill out a questionnaire to collect background information. Next, the participants were given a 10-minute tutorial about the system, and then performed a training task for 20 minutes to become familiar with the systems. All of the four tasks were completed using our system with no interventions by the experimenters.

After the training session, each participant was asked to work on the four tasks, with two using the experimental system (**E**) and the remaining two using the baseline system (**B**). The order of those four tasks and the order of using the experimental or the baseline system were rotated based on the Graeco-Latin Square design [40]. Figure 3 shows one round of the experiment orders, which consisted of 8 different participants. We recruited 24 participants and did this task order for three rounds. Using Latin-Square, we can ensure that the tasks occurred equally in each order position, which helps reduce the learning and fatigue effects.

Each task was described in Figure 2. The participant needs to propose their specific people search topics based on their domain knowledge. To prevent the tasks from turning into known-people search, we deliberately recruited PhD students who are junior researchers and who normally do not have extensive knowledge in one research domain. We asked the participants to then work on different research topics for each task (including the training task), explaining that they needed to explore five relevant candidates for each task. But at the same time, the participants needed to select those topics that they have some basic knowledge to ensure the quality of the chosen candidates.

	B	B	E	E		E	E	B	B
P1	1	2	3	4	P5	1	2	3	4
P2	2	3	4	1	P6	2	3	4	1
P3	3	4	1	2	P7	3	4	1	2
P4	4	1	2	3	P8	4	1	2	3

Figure 3: **One round of Latin Square of task order and system order. P1, P2, …, P8 denote eight participants.**

At the end of each task, the participant was asked to work on a questionnaire collecting information about their satisfaction with the search results and the system. The questionnaire has three parts. In part 1, participants were asked to rate and rank their preferences for the five marked candidates. In part 2, participants needed to answer five questions asking about the system usability and their satisfaction. In part 3, an open-ended question was asked, in which participants could express their thoughts about the three factors and provide suggestions. Users' interactions with the systems were automatically logged.

5.4 Participants

We recruited participants by promoting the experiments via emails and word-of-mouth. The recruited participants were all PhD students majoring in computer science and information science. The reason for such a limited audience is that the adopted ACM digital library dataset only contains data in this area, and Task 3 is meaningful only for PhD students. In total, 24 PhD students from 8 universities in China (3), United States (4) and Singapore (1) were recruited. Ten were female and 14 were male. 67% of the participants searched for people at least once a week in academic search engines such as Google Scholar and Microsoft Academic Search. 92% of participants searched at least 2-3 times a month. The participants have diverse research interests such as information retrieval, machine translation, computer graphics and computer vision, geographic information system, information seeking behavior and information security.

6. RESULT ANALYSIS AND DISCUSSION

We evaluated the PeopleExplorer system in terms of both system performance and system usability. We adopted the Generalized Linear Models (GLM) in SPSS to test for significant differences in the compared measures. The reported significance p-value is based on the Wald Chi-Square test, and we used the Bonferroni adjustment for multiple tests. The reason for using

GLM is that the model does not require the normality of outcome variables and it also helps to explore the interaction between tasks and used systems.

6.1 System Performance

6.1.1 System Effectiveness

Although the task templates were provided, the participants still needed to propose their own people search topics. Therefore, there is no ground truth for each task, and only participants themselves can judge the relevance of the candidates. After five candidates were marked, the participants were required to re-examine all of those candidates and assign a relevance score (on a five point Likert scale with 1 being non-relevant and 5 being highly relevant) to each of the selected candidates. The assignment of the relevance score to one candidate is independent of the others, so more than one candidate can receive the same relevance score. These final scores were used to evaluate the effectiveness of the systems. This method of using participants' generated content to simulate ground-truth data has also been adopted in several other works [44, 46]. The measures we used are:

- The *average relevance* score over the five selected candidates in each task
- The *number of candidates* and the *number of unique candidates* generated by the system during a task. The more returned candidates and the more *unique* returned candidates, the more effort users needed to put into the search to review and select candidates. We assumed that an effective system should, without losing the relevant candidates, provide fewer candidates so as to minimize users' information search cost.

The results of these two measures are presented in Table 1. The *average relevance* measure tells us that the experimental system returned candidates that were more relevant than the baseline system ($p<0.05$). This means that the experimental system is more effective in helping user to find relevant results.

Measures	Baseline	Experimental
Average relevance	3.96(0.72)	**4.13(0.72)***
Number of candidate	119.0(108.0)	136.5(100.0)
Number of unique candidate	77.7(60.9)	**49.0(34.2)****

Table 1. **The mean (standard deviation) of three measures, * denotes $p<0.05$ and ** denotes $p<0.01$.**

There was no significant difference in the *number of candidates* in the two systems, but the *number of unique candidates* returned by the experimental system is significantly less than that of the baseline system. On average, the candidate pool returned by the experimental system contained only 49 unique candidates. In contrast, the pool returned by the baseline system had 77 unique candidates. Thus, the effort needed to select relevant candidates would be greatly reduced in the experimental system. More importantly, even though the experiment system returned fewer candidates, the participants still could find significantly more relevant candidates. This again demonstrates the effectiveness of the experimental system.

6.1.2 System Efficiency

To assess the system's efficiency, we computed the time spent for each participant to finish each task. Then, we looked into the mean time for all of the tasks as well as for each individual task. However, we found no significant difference between the mean time spent (across all tasks) on the experimental system and that

of the baseline system (see Table 2). In fact, the mean time in the baseline system is even lower than that in the experimental system for all tasks. A possible explanation is that using the sliders, we actually introduced an extra cognitive load because users need to think about formulating/reformulating queries as well as to work out how to set the appropriate values for each slider. This higher cognitive load may slow progress in the task.

We also observed that there is no significant difference for the mean time spent within each task either. However, the mean time spent on each task seems to have different patterns. For example, the mean time spent on task 3 has a large variance, which means the system may be efficient for some users while less efficient for others (the minimum time is 1.02 minutes while the maximum time is 20.72 minutes). However, the mean time spent on task 1 in the experimental system is less than that in the baseline system.

To conclude, the experimental system does not help to reduce the time spent on finding relevant candidates. If we combine it with the system effectiveness, we may conclude that the experimental system helps to find more relevant candidates without introducing significant extra time costs.

Tasks	Baseline	Experimental
1: Conference mentor finding	6.74(2.57)	4.46(3.71)
2: New collaborator finding	5.68(4.42)	5.96(3.81)
3: External thesis committee member finding	5.01(2.94)	7.11(6.36)
4: Reviewer suggestion	6.05(5.90)	6.08(4.64)
Overall	5.87(4.08)	5.91(4.69)

Table 2. **The mean (standard deviation) of time spent on each task (unit: minutes).**

6.2 User Perceptions

User perceptions were analyzed based on the questionnaire offered at the end of each task. As shown in Table 3, Q1 to Q4 are questions about the system usability, and the participants' ratings of the experimental system are significantly better than those of the baseline system for all of the four questions. However, the statistical analysis also shows that there is an interaction between the tasks and the system used ($p<0.05$) on Q4 (Overall user satisfaction). Therefore, we further analyzed users' ratings of the overall satisfaction in each task. The significant difference was found only in task 1 (conference mentor finding, $p<0.01$). Q5 asked about the usefulness of the candidate surrogate: participants felt that it was generally useful, but significantly more useful in the experimental system. Q6 asked about the usages of sliders in the experimental system. The participants gave high ratings to this functionality (mean=4.60, standard deviation=0.61).

Open-ended questions were focused on the participants' feelings about the system and their suggestions for improving the system. 92% of the participants think that the three factors are reasonable and useful because the sliders act as the "*user-friendly filter*", and help them to explore "*diverse*" results under "*control*". In addition, they like the fact that the system can provide highly relevant, but unexpected, candidates. In addition to those three factors, users also suggested adding more factors such as whether the candidates have been active in recent years, or even the candidate's: one participant wrote "*I tend not to select people who have bad temper*".

The participants also provided their ratings of the importance of each factor for different tasks. The ratings are diverse even within

the same task. For example, in the conference mentor finding task, some participants thought that social similarity is not important, whereas others thought that it is the most useful factor. In the external thesis committee member finding task, some participants cared more about the authoritativeness and the expertise of the candidates in the research field, whereas others thought that social similarity should be considered because it is not known if the candidate will "*accept the invitation*" or if the candidate is beyond their social connections.

Questions	Baseline	Experimental
Q1: The system provides me relevant candidates	3.89(0.56)	**4.46(0.64)****
Q2: The system can help me find relevant candidates efficiently	3.58(0.84)	**4.42(0.71)****
Q3: The system is easy to use	3.77(0.90)	**4.31(0.70)****
Q4: Overall, I am satisfied with the system in this task	3.15(0.99)	**4.17(0.76)****
Q5: The display of each candidate helps me understand why I got the candidate	3.77(1.13)	**4.31(0.78)****
Q6: The sliders helps me get better candidates	-	4.60(0.61)

Table 3. **The mean (standard deviation) of users' agreements about user perceptions. ** denotes p<0.01. We used a five-point Likert scale from "strongly disagree" to "strongly agree". In the analysis, we labeled 5 as "strongly agree" and 1 as "strongly disagree".**

6.3 Search Behavior Analysis

In the sections above, we compared only the overall performance of the two systems: we did not look into the details of how the user's query behavior and result selection process will be affected by the user's manipulation of the sliders in the experimental system. This is the focus of this section.

6.3.1 Query Behavior Analysis

Three measures were adopted to analyze differences in the user's query behavior on the two systems: the *number of queries*, the *average query length* and the *query reformulation patterns*. In this project, we analyzed four query reformulation patterns [48]: *new*, *generalization specialization*, and *reconstruction* (see Table 4). For each pattern, we computed the proportion of this pattern over all used patterns. The reported values are the average over all users under both systems. Table 5 shows the results.

	Definition
New	Q_i is the first query or does not share any common terms with Q_{i-1}
Generalization	Q_i share common terms with Q_{i-1}, but Q_i contains fewer terms
Specialization	Q_i share common terms with Q_{i-1}, but Q_i contains more terms
Reformulation	Q_i share common terms with Q_{i-1}, but Q_i has at least one different term with Q_{i-1}

Table 4. Query reformulation patterns and their definitions. Q_i and Q_{i-1} are two consecutive queries in a query session.

The standard deviations in all cases were very high. We can think of several possible reasons including the fact that participants in the same task can choose different topics; therefore, their familiarity with the topic and the topic diversity can significantly influence the behaviors of the participants. The

mean values of *number of queries* and *average query length* show that the participants in the experimental system tended to issue slightly fewer and shorter queries, which might be because they were able to explore more results by manipulating the sliders rather than issuing a new query or specifying their information needs using a longer query – both of which can be very costly. However, no significant differences were detected.

As for the *query reformulation patterns*, we observed significantly less usage of *generalization* and *specialization* patterns, and significantly more *new* patterns in the experimental system but no difference with regard to *reconstruction* patterns. *Specification/ generalization* helped users to narrow down/expand their information needs by adding/removing one or more factors. In the baseline system, this can only be done by changing query terms while users of the experimental system can also utilize the sliders as an alternative way to express their information needs. For example, tuning the "*authoritativeness*" to a high value may indicate that users do not want to restrict their results to an exact match of query terms. People who are authorities on different topics may be also relevant. This can serve as a similar function to the *generalization*.

Measures	Baseline	Experimental
Number of queries	4.91(5.04)	4.28(3.94)
Average query length	2.60(0.89)	2.43(0.75)
Reformulation pattern : new	0.65(0.33)	**0.76(0.29)***
Reformulation pattern: specialization	0.10(0.13)	**0.06(0.11)***
Reformulation pattern: generalization	0.09(0.14)	**0.03(0.09)***
Reformulation pattern: reconstruction	0.17(0.20)	0.15(0.20)

Table 5. **The mean (standard deviation) of six measures. ** denotes p<0.01 and * denotes p<0.05.**

6.3.2 Result Browsing Analysis

Three measures were used to compare system differences in result browsing behavior: the *number of pages viewed* by each user, the *average time spent on each page* and the *(pseudo) average rank positions* over the five selected candidates. Since the average relevance for the experimental system is significantly better than the baseline system, we want to further study which types of candidates resulted in such differences. In the experimental system, when a candidate was selected, its rank position in the returned page and the page number were automatically logged. The final rank is based on the number of pages and the rank position of the candidate. Moving to a new page is the only way to introduce new candidates into the baseline system (which will lower the candidate rank), whereas users in the experimental system can manipulate the sliders (but the rank for each candidate will be recalculated because of the change in returned candidates). By tuning the sliders, users actually go through several pages but these efforts will not be counted towards rank positions. To make a fair comparison, we constructed a pseudo rank for each marked candidate in the experimental system: if the user marked a candidate *ca* under a query *q* in the experimental system, we computed the pseudo rank of *ca* under the condition that users decided not to utilize the sliders but still wanted to mark the same candidate, which is actually the rank of *ca* under query *q* in the baseline system.

Table 6 shows the results of these three measures. There is no significant difference on the number of pages viewed in the two systems, but users in the experimental system spent significantly less time in examining each page. This may be due to that fact

that, although the number of pages presented to the users are approximately the same, the number of unique candidates in the experimental system is significantly fewer than that of the baseline system (see Table 1), which means more candidates in the experimental system were actually repeated. Thus, the participants did not have to check every candidate. There is a significant difference in the (pseudo) average rank position between the two systems. Compared to the baseline system (if we assumed their query patterns are similar), the participants in the experimental system were able to mark those candidates who have lower ranks. Given that the average relevance in the experimental system is significantly better than that of the baseline system; we can infer that the increase of relevance is due to the participants marking those candidates in lower ranks but who had higher relevance. It seems that providing sliders gives users' a freedom to describe their information needs more accurately. In addition, as one participant noted in the open-ended questionnaires, it also helped them to explore the high relevant but "*unexpected results*".

Measures	Baseline	Experimental
Number of pages viewed	14.88(13.5)	17.06(12.4)
Time spent on each page (minute)	0.65(0.70)	**0.39 (0.25)****
(pseudo) average rank position	14.57(12.0)	**61.03(59.0)****

Table 6. **The mean (standard deviation) of three measures. ** denotes p<0.01 and * denotes p<0.05.**

6.3.3 Slider Tuning in the Experimental System

In this section, we focused on the interaction between the participants and the experimental system in terms of using sliders. This part of the discussion concentrates on answers to the following questions: whether the participants used the sliders, and whether they used the sliders consistently. We measured the interaction in terms of *the frequency of slider tuning* (tuning times per minute) for each task. Two consecutive tuning actions are treated as one if the time interval is less than three seconds.

Conditions	Sliders tuning per minute
1: Overall	1.20(0.81)
2: First half of the search session	1.14(0.92)
3: Second half of the search session	1.25(1.12)
4: First task	1.18(0.78)
5: Second task	1.23(0.85)

Table 7. **The mean (and standard deviation) of five measures.**

As reported in Table 7, the participants indeed tuned the slide bars to facilitate their exploration of candidates (average 1.20 times/ minute). Our user logs showed that there was only one participant in the reviewer assignment task (Task 4) who did not utilize the sliders at all. In order to check whether or not the participants used the sliders throughout the experiment rather than simply abandoning sliders after their first use, we compared the slider usages between the first half of the search session and the second half of the search session for each task. In addition, we also compared the slider tuning between the first assigned task and the second assigned task for each participant. Table 7 shows the results. There is no significant difference between the first half of the search session and the second half, nor between the first task and the second task. This indicates that the participants actually used the sliders consistently. In addition, we found that the mean sliders tuning times in the second half time/ task are slightly more than in the first half time/ task. Thus, we can see that users are more likely to use the sliders more after the initial use.

6.4 Factor Importance for Different Tasks

In responding to the open-ended questions in Section 6.2, the participants reported diverse opinions about the importance of the factors during the selection of correct candidates. In addition to the open-ended questions, we also asked the participants to rate the importance of each factor using a question template "*How do you rate the importance of [relevance/ authoritativeness/ social] in this task?"* The participants were called upon to respond this question in a five point Likert scale upon completing each task. The Likert scale selections range from highly important (noted as 5) to the highly unimportant (noted as 1). However, the participants' perceived factor importance may not truly reflect the actual importance of the factors as they might not be accurate. Therefore, we further analyzed how users really make use of the three factors during the search sessions in terms of how they tuned the three sliders. Each time the participant selected a candidate, our system automatically logged the candidate's corresponding parameters, that is, the tuned weight of each of the three factors in three sliders. However, the comparisons among the three factors may be inappropriate because the tuned weights of each factor also depend on the method used to calculate the score for each factor. The values for social similarity can either be positive ("*close to my connections*") or negative ("*far from my connections*"): it is important even when it is close to the negative end (e.g. -1). Therefore, we used the absolute values for the weights of social similarity. In this section, we reported both the perceived factor importance and the tuned factor importance. We plotted the average ratings and tuned weights for each factor in Figure 4 (a) and (b). The user's perceived importance of three factors is all higher than 3.0, and the participants indeed tuned the weights of different factors (only one participant in one task did not). This indicates the necessity of involving all three factors in different people search tasks.

As seen in Figure 4, we found that the participants seldom changed the importance of content relevance, which was stable in both figures across all four tasks, whereas the importance of social similarity and candidate authoritativeness varied a great deal in different tasks. For example, candidate authoritativeness is an important factor while social similarity is the least important factor for Task 1 (conference mentor finding) in terms of both the perceived importance and the tuned importance; however, candidate authoritativeness becomes a less important factor while social similarity becomes a more important factor in Task 3 (external thesis committee member finding). We also found several conflicts between the *perceived* factor importance and the *tuned* factor importance. For example, the participants rated the social similarity factor as a more important factor in Task 2 than in Tasks 3 and 4. However, they perceived this factor as almost equal in importance in Tasks 2, 3 and 4. The use of the system in the first three tasks is within our expectations, but not for the reviewer suggestion tasks (Task 4). Our expectation for Task 4 was that the participants may need those candidates who are "*far away from my social connections*". However, many participants actually selected a large proportion of candidates (41.7%) who were "*close to my social connections*", significantly more than the proportion of candidates (15.0%) who were marked as "*far away from my connections*". This means that some participants actually want the reviewers to be socially close rather than distant. Two reasons mentioned by the participants for this preference were "*it is ok to choose someone you know unless you are coauthors*", and "*personally, I would prefer those non- authoritativeness people, and it is much better if we have personal connections*".
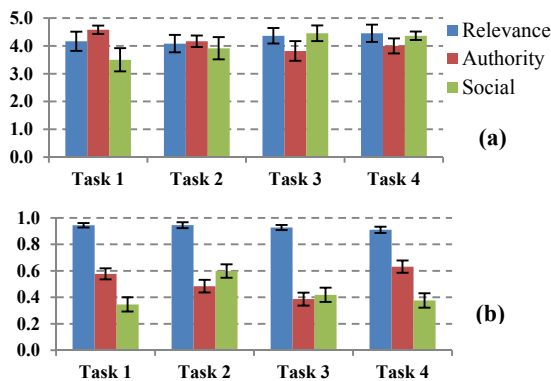
Figure 4: **The average user preferences (with standard errors) on three factors in each of the four tasks: (a) user perceived factor importance; (b) user tuned factor importance**

6.5 Limitations

We do acknowledge several limitations in our system design and experiment setting. First, the target population for the system and the four tasks are junior researchers in computer science and information science fields. The experimental results may not be generalizable to other populations or domains as the considered factors may vary with different users and different subject domains. For example, in the collaborator finding task, senior researchers may be more concerned with finding the people who can help generate influential works (which is not a factor modeled in PeoleExplorer) while the junior researchers may focus on candidates' content relevance and reachability of the candidates. However, we believe that the principle design of the system and the experiment are generic enough to be extended to other user populations and tasks without fundamental changes. Second, the system effectiveness evaluation in Section 6.1.1 was based on the participants' subjective preference judgments on candidates as the ground truth data was lacking. Therefore, traditional evaluation measures such as Precision@N, Recall and Average Precision cannot be adopted to provide more informative evaluations of interactive people search tasks. A test collection therefore is critical to an objective evaluation of the system's performance.

7. CONCLUSION AND FUTURE WORKS

People search has been investigated in many different domains. However, there is still a lack of studies investigating the diverse and exploratory nature of this problem. To resolve those challenges, we designed and developed an interactive people search system called PeopleExplorer. Users can undertake diverse people search tasks and explicitly model their preferences using three sliders. *"Relevance"* measures the content similarity between an issued query and candidates' publication profile; "*authoritativeness*" is computed using the PageRank algorithm in coauthor networks; and "*social similarity*" measures the social closeness between a candidate and the user.

We conducted a user-based experiment to evaluate the usability and performance of our system based on four exploratory people search tasks. Each participant then completed the four tasks using both PeopleExplorer and a baseline system. The results show that: 1) Compared to the baseline system, our system is capable of finding more relevant candidates. At the same time, it requires less or comparable effort such as checking fewer unique candidates and expending approximately the same search time. 2) Users' preferences in different tasks are diverse. In finding conference mentors, users pay more attention to the authoritativeness of the candidates, whereas social similarity is

also important in searching for new coauthors and external thesis committee members. Overall, our results suggest the necessity of modeling task difference and user variance in people search, as well as related topics such as expert finding and link prediction.

The novel contributions of this project are that we modeled people search as an exploratory task and exemplified with four realistic exploratory tasks. We also designed and evaluated an innovative solution to model task diversity and user preferences visually. The experimental results indeed confirmed the necessity of modeling the user variance and task contexts in people search and also demonstrated the effectiveness of our proposed method.

There are several possible future directions. First, in our current system, some participants felt that authoritativeness is vaguely defined and difficult to use. They also suggested considering other types of information such as whether the candidates are still active in their fields. We would also like to incorporate other types of factors in Ackerman, et al., [1]: the accessibility, trust and reputation are important in designing an expert search system. In the future, we would like to further improve our system by considering more factors and allowing users to make choices on which factors to use. Second, in Section 6.2.1, we showed that the number of unique candidates presented in the experimental system is significantly lower than the baseline system. Although within the same time period, users need only to check fewer candidates: however, this comes at the cost of diversity, which actually has been recognized as a very important measure in exploratory search tasks [23] and recommender systems [22]. If the objective is to find as many relevant candidates as possible, it is unclear that our system performs better.

8. REFERENCES

[1] Ackerman, M. S., Wulf, V. and Pipek, V. Sharing Expertise: Beyond Knowledge Management. MIT Press, 2002.

[2] Tang, W., Tang, J., Lei, T., Tan, C., Gao, B. and Li, T. On optimization of expertise matching with various constraints. Neurocomput., 76, 1 (2012), 71-83.

[3] S. Han, D. He., Z. Yue, J. Jiang, W. Jeng. IRIS-IPS: An Interactive People Search System for HCIR Challenge. In the Proceedings of HCIR (2012).

[4] Rodriguez, M., Posse, C. and Zhang, E. Multiple objective optimization in recommender systems. In the Proceedings of Recommender systems 2012 (2008).

[5] Ehrlich, K. and Shami, N. S. Searching for expertise. In Proceedings of the SIGCHI 2008 (2008).

[6] Dumais, S. T. and Nielsen, J. Automating the assignment of submitted manuscripts to reviewers. In the Proceedings of the SIGIR 1992.

[7] Streeter, L. A. and Lochbaum, K. E. An expert exert-locating system based on automatic representation of semantic structure. In Proceedings of the lEEE Conference on AI Applications (1988)

[8] Payton, D., Daily, M. and Martin, K. Dynamic collaborator discovery in information intensive environments. ACM Comput. Surv., 31, 2es (1999), 8.

[9] Dumais, S. T. Latent semantic analysis. Annual Review of Information Science and Technology, 38, 1(2004), 188-230.

[10] Goldberg, D., Nichols, D., Oki, B. M. and Terry, D. Using collaborative filtering to weave an information tapestry. Commun. ACM, 35, 12 (1992), 61-70.

[11] Terveen, L. and McDonald, D. W. Social matching: A framework and research agenda. ACM Trans. Comput.-Hum. Interact., 12, 3 2005), 401-434.

[12] Kautz, H., Selman, B. and Shah, M. Referral Web: combining social networks and collaborative filtering. Commun. ACM, 40, 3 (1997), 63-65.

[13] McDonald, D. W. and Ackerman, M. S. Just talk to me: a field study of expertise location. In CSCW 1998: 315-324

[14] Lin, C. , N. Cao, Liu, S. Papadimitriou, S. Sun, J, Yan, X. SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence. In Proceedings of the IEEE 25th International Conference on In Data Engineering

[15] Perer, A. and Guy, I. SaNDVis: visual social network analytics for the enterprise. In the Proceedings of CSCW 2012: 275-276

[16] Chen, H.-H., Gou, L., Zhang, X. and Giles, C. L. CollabSeer: a search engine for collaboration discovery. In the Proceedings of JCDL 2011: 231-240.

[17] Han, S., He, D., Brusilovsky, P. and Yue, Z. Coauthor Prediction for Junior Researchers. In Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction 2013: 274-283.

[18] Donovan, J., Smyth, B. Gretarsson, B., Bostandjiev,S and Höllerer, T. PeerChooser: visual interactive recommendation. In Proceedings of CHI 2008:1085-1088.

[19] Balog, K., Azzopardi, L. and Rijke, M. d. Formal models for expert finding in enterprise corpora. In the Proceedings of SIGIR 2006: 43-50.

[20] Liben-Nowell, D. and Kleinberg, J. The link prediction problem for social networks. In the Proceedings of CIKM 2003: 556-559.

[21] Yu, K. W. C., Shipeng Yu , Volker Tresp , Zhao Xu. Stochastic Relational Models for Discriminative Link Prediction. In Proceedings of the Advances in Neural Information Processing Systems (2007). MIT Press.

[22] Bostandjiev, S., O'Donovan, J., Höllere, T. TasteWeights: a visual interactive hybrid recommender system. In the Proceedings of Recommender systems 2012.

[23] Marchionini, G. Exploratory search: from finding to understanding. Commun. ACM, 49, 4 2006: 41-46.

[24] Jiang, J. Han, S and Lu, W. Expertise Retrieval Using Search Engine Results. In SIGIR 2008 workshop on Future Challenges in Expertise Retrieval (fCHER), 2008: 11-16.

[25] Deng, H., King, I. and Lyu, M. R. Formal models for expert finding on DBLP bibliography data. In Proceedings of the Eighth IEEE International Conference on Data Mining 2008: 163-172.

[26] Jiang, J, He, D and Ni, C. 2011. Social reference: aggregating online usage of scientific literature in CiteULike for clustering academic resources. In Proceedings of JCDL 2011: 401-402

[27] Mimno, D. and McCallum, A. Expertise modeling for matching papers with reviewers. In SIGKDD 2007.

[28] Karimzadehgan, M., Zhai, C. and Belford, G. Multi-aspect expertise matching for review assignment. In the Proceedings of the CIKM 2008.

[29] Hecht, B., Teevan, J., Morris, M. R. and Liebling, D. SearchBuddies: Bringing Search Engines into the Conversation. Proceedings of ICWSM 2012 (2012).

[30] Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N. and Gummadi, K. Cognos: crowdsourcing search for topic experts in microblogs. In the Proceedings of the SIGIR, 2012.

[31] Jiang, M., Cui, P., Wang, F., Yang, Q., Zhu, W. and Yang, S. Social recommendation across multiple relational domains. In the Proceedings of CIKM 2012.

[32] Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y., Compton, P. and Mahidadia, A. Collaborative filtering for people to people recommendation in social networks. AI 2010: Advances in Artificial Intelligence2011), 476-485.

[33] Tang, J., Wu, S., Sun, J. and Su, H. Cross-domain collaboration recommendation. In Proceedings of KDD 2012: 1285-1293.

[34] Shami, N. S., Ehrlich, K., Gay, G. and Hancock, J. T. Making sense of strangers' expertise from signals in digital artifacts. In the Proceedings of the SIGCHI 2009.

[35] Reichling, T. and Wulf, V. Expert recommender systems in practice: evaluating semi-automatic profile generation. In the Proceedings of the SIGCHI 2009.

[36] Hearst, M. A. Search user interfaces. Cambridge University Press, 2009.

[37] White, R. W. and Roth, R. A. Exploratory Search: Beyond the Query-Response Paradigm 2009.

[38] Brin, S. and Page, L. The anatomy of a large-scale hypertextual Web search engine. In the Proceedings of WWW 1998.

[39] Yan, E. and Ding, Y. Discovering author impact: A PageRank perspective. Information processing & management, 47, 1 2011), 125-134.

[40] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. Found. Trends Inf. Retr. 3, 12 (2009), 1-224.

[41] Smirnova, E., Balog, K. A user-oriented model for expert finding, In Proceedings of the 33rd European conference on Advances in information retrieval, 2011: 580-592

[42] Hofmann, K., Balog, K., Bogers, T., and Rijke, M. 2010. Contextual factors for finding similar experts. J. Am. Soc. Inf. Sci. Technol. 61, 5 (2010), 994-1014.

[43] Woudstra, L., and Hooff, B. 2008. Inside the source selection process: Selection criteria for human information sources. Inf. Process. Manage. 44, 3: 1267-1278.

[44] Yarosh, S., Matthews, T., and Zhou, M. 2012. Asking the right person: supporting expertise selection in the enterprise. In Proceedings of the SIGCHI 2012: 2247-2256.

[45] Borlund, P. 2000. Experimental components for the evaluation of interactive information retrieval systems. Journal of documentation 56.1 (2000): 71-90.

[46] Shah, C., and Gonzalez-Ibanez, R. (2011). Evaluating the Synergic Effect of Collaboration in Information Seeking. In Proceedings of SIGIR 2011: 913-922.

[47] Bostandjiev, S. Donovan, J. and Höllerer, T. 2013. LinkedVis: exploring social and semantic career recommendations. In Proceedings of the 2013 international conference on Intelligent user interfaces (IUI '13): 107-116.

[48] He, D., Göker, A., and Harper, D. 2002. Combining evidence for automatic web session identification. In Information Processing & Management. 38, 5 (2002), 727-742.