Summer 10-6-2011

FINDING AND VALIDATING EXPERTISE

Meredith Taylor

Debbie Richards

FINDING AND VALIDATING EXPERTISE

Taylor, Meredith, Department of Computing, Macquarie University, North Ryde, NSW, 2109, Australia, meredith.taylor@mq.edu.au

Richards, Debbie, Department of Computing, Macquarie University, North Ryde, NSW, 2109, Australia, deborah.richards@mq.edu.au

Abstract

Being able to make timely contact with an expert who is willing and able to answer the problem at hand is important. However finding such an expert using the systems found in many organisations can be difficult because experts are too busy to ensure information is complete and up to date. We have designed, implemented and evaluated a prototype Expert Recommender Systems based on an investigation of the requirements for such a system in two knowledge intensive organizations together with review of the literature and comparison with existing systems. We designed a triangulated approach which combines automated expert profile creation and maintenance, validation by the expert, and feedback on the expert and the system by the searching party. In this paper we present results from a usability study we conducted of the prototype system with a particular focus on the searching algorithm we designed.

Keywords: Expertise Recommender Systems, usability study, searching algorithm

1 Introduction

The knowledge or expertise held by the people in an organisation is a valuable resource. In order to properly access, use, and share this knowledge, many organisations use a type of search engine for experts called an Expert Recommender Systems (ERS). In some cases the inquirer's main interest is in the answer, in other cases the main interest is to find an expert who will handle the problem (Yimam-Seid and Kobsa, 2003). Although ERS allow for fast searching of experts, people often find that there is no way of telling how useful or accurate a recommendation is going to be. A "yellow-pages" type ERS relies on the expert to enter and maintain their own profile in the system and as a result expert profiles are often missing, out-of-date or incomplete. An alternative to self-reporting recommender systems, are fully automated approaches to locate experts such as SAGE (Becerra-Fernandez, 2000) using inputs such as email (e.g., Ehrlich, Lin and Griffiths-Fisher, 2007), bulletin boards (Krulwich and Burkey, 1996), Web pages (e.g. Foner, 2002 and Pikrakis *et al*., 1998), software code (e.g. McDonald and Ackerman, 2000 and Vivacqua, 1999), technical reports (e.g. Crowder, Hughes and Hall, 2002) and the artefacts of social software systems (such as WebLogs and Wikis) and social networks (e.g. Ehrlich, Lin and Griffiths-Fisher, 2007). However, a review of ERSs, including many of these systems, by Sim, Crowder and Will (2006) found problems related to heterogeneous information sources, expertise analysis support, reusability and interoperability.

The key problem of both the yellow pages / self-referral style ERS or the automated ERS is the lack of cross-validation (i.e. between human opinion and hard evidence). In the case of self-referral, what evidence exists that an individual is the expert they claim to be or that the data is up to date? In the case of automated expertise detection, what crosschecks have been made that what has been mined is complete, accurate and/or error-free? Additionally, while many product recommenders do seek and use feedback as part of their reasoning process, ERS typically do not allow searchers to provide any feedback about how useful a recommendation or expert was. An exception is Aimeur et al. (2007) who obtained feedback by ensuring that all searcher and expert interaction is controlled by the system and enforced by having profiles for both searchers and experts. However, their approach is geared towards providing quick solutions to problems within an organization rather than putting people in contact with one another, which we are more interested in. We are not so focussed on accessing *what* people know but on finding out *who* knows what.

Ehrlich *et al*. (2007) considered the social side of finding and contacting experts. They describe SmallBlue (see also Lin *et al*., 2008) an ERS developed for IBM that maps each staff member's social network in order to tell who is connected to whom and where social networks overlap. When someone searches for an expert, along with the list of recommendations, the searcher is given the shortest social path they can follow to contact the person. The idea behind this approach is that experts are much more likely to answer a query if it comes from someone in their social network (or from someone who has been approved by someone in their social network) rather than a complete stranger. As with Aimeur et al's system, this approach is acceptable for an ERS used exclusively by staff members but does not support searching by external parties.

We have developed a triangulated cross-validation approach to locating experts and an accompanying prototype system which includes automation, self reporting and feedback. This approach has three supporting dimensions: 1) Automated searching, where the system automatically generates an expert profile based on home pages, publication databases and other accessible data 2) "self-reporting" and "referral by others", where the experts identified by the first dimension are sent their results and asked to validate them, and are allowed to refer others as experts to the system 3) A feedback mechanism, where experts indicate their preferences/availability, searchers explicitly provide feedback on recommendations they receive and unobtrusive feedback via an insystem contact form which tracks if the searcher contacted the expert. Such unobtrusive methods to gather data to produce searching profiles are commonly used (as in Quickstep by Middleton *et al.*, 2001). Additionally, an expert ranking algorithm which takes into account the searcher's feedback, the search terms and the availability of the expert was developed, evaluated and revised. To test the underlying concepts of this approach, two studies were performed. The first study (Taylor and Richards, 2008) tested one method of automatically generating expertise information, and having the experts validate this information. Our evaluation yielded promising results as 94% of the staff members who responded agreed with the assigned area of expertise. In our second study we evaluated the prototype system and ranking algorithm that was developed to implement the triangulated approach. Presentation of the results from this study is focus of this paper. In the next section, we provide further description of the triangulated approach and then present the evaluation study and results for our prototype system. Conclusions and future work appear in section 5.

2 Approach

A set of requirements were established and specified following interviews with a total of 24 people in two knowledge-intensive organisations, one a university the other a research organization. Interviewees represented key areas of the organisations interested in expertise finding (such as the research office, community outreach, publicity, etc) and involving senior people and project leaders in each of these sections. The questions and findings are reported elsewhere (Richards and Busch, 2009).

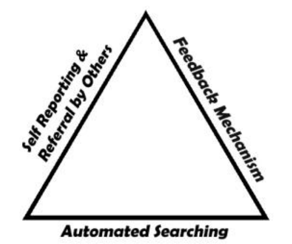


Figure.1. Our Triangulated approach

We also conducted review of the literature and current systems in use in these organizations, in other similar organizations and comparisons of off-the-shelf products available for managing and locating expertise. The resulting Software Requirements Specification (SRS) can be obtained by contacting the authors. We found that individuals used two main strategies to find experts. One could be called a quantitative or data/evidence driven approach. The other was more qualitative and involved using your social networks to find a suitable expert.

To provide both a referral and automated system we have created a triangulated approach. As depicted in Figure 1, the proposed approach uses *automated searching* as a foundation from which the initial data is captured and against which the data is regularly crosschecked. The key inputs to data mining include individual web pages, project/grant repositories, citation indexes (e.g. CiteSeer (http://citeseer.ist.psu.edu/) and publications databases. In the information extraction technique we trialled (Taylor and Richards, 2008), results were sent to each of the 20 identified experts in the Computing Department in an email giving them an opportunity to review and validate the areas of expertise found by the system. The results provided to each expert was a set of RFCD (Research Fields, Courses and Disciplines) as defined by the Australian Research Council. These codes were based on the expert's publications and were used as an externally validated and publically recognized indicator of their research areas. As part of the data collected was the date of the source to assist with identifying currency together with statistics on the level of expertise using simple measures such as the number of publications in that area and term frequency inverse document frequency (TFIDF). To address the issues of external validation, expertise currency and motivation to enter and maintain the data, individuals would be sent the output from data mining at regular intervals, say twice yearly. Given the importance of reputation and track record in the university system we believe academics would be motivated to check what the system says, as it reflects the data in the world about them, and to correct any errors or omissions.

Having the expert confirm and correct the results of automated searching contributes to the second dimension: *Self reporting and referral by others*. The system also allows for others, such as a PhD supervisor, to nominate or refer another person, such as their student. Given that experts are busy people, rather than asking experts to "opt in", by using automated methods as the foundational first step in information acquisition, this system would instead ask them to "opt out" if they do not wish to be registered with the system. It would also be much easier for an expert to review a portfolio which has been automatically generated for them rather than have to format one for themselves. From discussions, interviews and personal experience the key question to be asked of the advice of any recommendation system is “was it useful?” As shown in Figure 1, the third support to our approach is a *feedback mechanism* that allows the searchers of the system to validate the recommendations themselves. Feedback is gathered from both the searcher and the expert. Once an expert is recommended by the system the searcher can choose to contact the expert via a contact form provided by the system. The system would then email this message to the expert along with a link to a form where they can send their initial reply. In this way the expert can provide information about their availability and expertise as well as writing a personalized message to the searcher. If the searcher wishes to contact the expert via another route, such as a phone call, the system would allow them to indicate this by clicking a link or button. If the searcher chooses this option, a message will be sent to the expert informing them that they may be contacted regarding the search terms entered by the searcher. The expert will be directed to a form where they can indicate if they are available and possess the expertise to address the query. The system will then be able to update the expert’s statistics in the system and inform the searcher of the expert’s status if the system has both searcher and expert profiles or if the searcher provides the system with their email address. Thus the system is able to keep track of what recommendations yield success (that is, what recommendations result in an expert being contacted). If the searcher does not indicate in any way that they wish to contact the expert, the system will store the search terms used for a brief period of time and observe if similar search terms also yield unsuccessful results. If so, the system will reevaluate the profiles that are recommended for those search terms. After interacting with a recommended expert, a searcher may provide feedback on the expert whenever they wish. If two weeks has lapsed with no feedback, the system could send the search a reminder to provide feedback.

User models are maintained to capture the needs and preferences of the person looking for someone (e.g., they are a news journalist and need someone to provide an expert opinion for television in layman’s terms in the next hour) and also the preferences, communication skills and availability of the expert (e.g. they have given radio interviews in the past but are away for two weeks). The user models can potentially be populated from data found in other places, such as from webpages maintained by the individual or corporate pages, but the feasibility of this depends on the level of webpage standardization across the organization and the degree to which content management is enforced. Thus in our solution we leave this as potential future work. As part of the user model are statistics relating to the responses from experts regarding their expertise and availability and the requesters satisfaction with the service provided. This feedback mechanism will act as a referral system, suggested by some of the people we spoke with as the type of system they would be interested in, rather than a yellow-pages model. While the system uses feedback statistics for generating and ordering recommendations in response to a query, for privacy and ethical reasons, only the expert will be able to see his/her own statistics regarding their overall usefulness and availability as perceived by the service requesters. We see this personal feedback as potentially useful for professional development and self evaluation. As represented, this additional feed-back/validation pillar provides a triangulated approach bringing together automated machine-based knowledge discovery and manual validation.

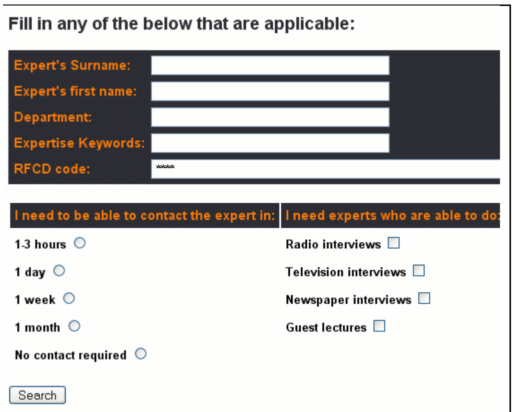
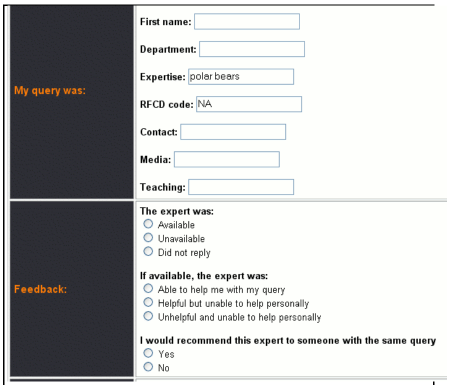


Figure 2: partial initial search screen

3 Prototype Evaluation Study

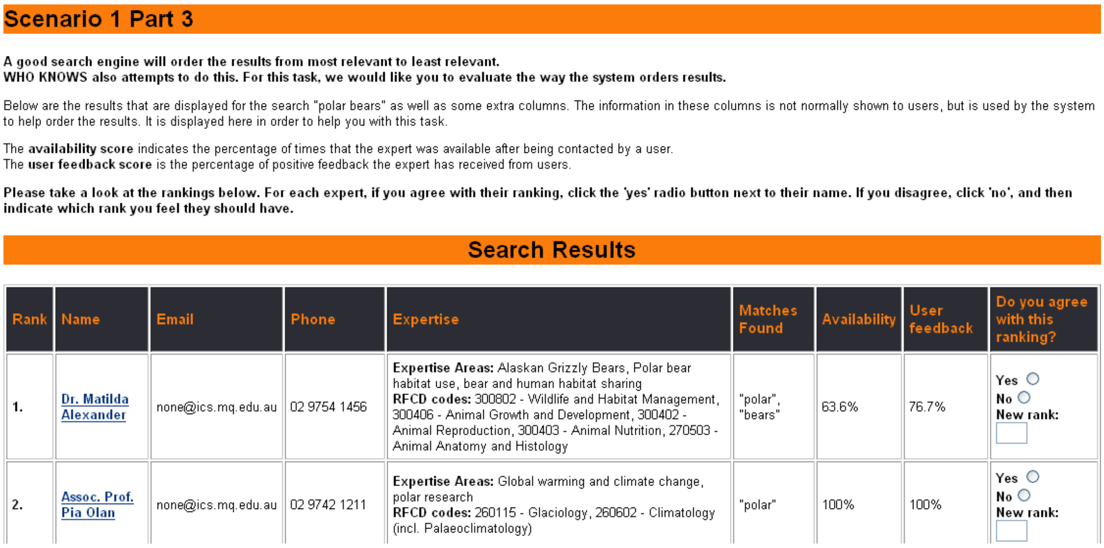
To measure potential user satisfaction and system usability we implemented a prototype ERS, WHOKNOWS? that we populated with a small amount of test data. Four mock expert profiles were put into the system as well as several screens to help searchers find and contact experts. The screens consisted of an initial search screen (Figure 2), a screen to show the search results, a screen showing the full profile for an individual expert, a screen for the searcher to contact the recommended experts, a screen for the searcher to provide feedback on an expert (Figure 3). Not included in the usability study was the screen for the expert to respond to a searcher's request. The expert profiles included the expert's contact details, expertise areas, and type of media contact the expert was available for (such as newspaper and television interviews.) The search function allowed requesters to indicate how soon they needed to contact the expert, which would allow the expert to prioritise their requests.



My query was:	First name: Department: Expertise: polar bears RFCD code: NA Contact: Media: Teaching:
Feedback:	The expert was: Available Unavailable Did not reply If available, the expert was: Able to help me with my query Helpful but unable to help personally Unhelpful and unable to help personally I would recommend this expert to someone with the same query Yes No

Figure 3:Partial Feedback form

WHOKNOWS? and our survey for collecting participant responses to the system were both made available online. Participants were recruited via an emailed advertisement. Invited participants included individuals we had interviewed, but all responses were anonymous. After completing the online consent form, the participants would be shown two scenarios. We felt one scenario was not enough exposure to the system and would not provide sufficient data. Each scenario gave the participant a job occupation and a reason to use the system to search for an expert. After each scenario had been completed, the participants were asked to fill in a questionnaire. The participants were asked to evaluate the algorithm the system uses to rank the experts (Figure 4), as well as their opinion on certain components of the system (such as the contact and feedback pages).



Scenario 1 Part 3

A good search engine will order the results from most relevant to least relevant.
WHO KNOWS also attempts to do this. For this task, we would like you to evaluate the way the system orders results.

Below are the results that are displayed for the search "polar bears" as well as some extra columns. The information in these columns is not normally shown to users, but is used by the system to help order the results. It is displayed here in order to help you with this task.

The **availability score** indicates the percentage of times that the expert was available after being contacted by a user.
The **user feedback score** is the percentage of positive feedback the expert has received from users.

Please take a look at the rankings below. For each expert, if you agree with their ranking, click the 'yes' radio button next to their name. If you disagree, click 'no', and then indicate which rank you feel they should have.

Search Results

Rank	Name	Email	Phone	Expertise	Matches Found	Availability	User feedback	Do you agree with this ranking?
1.	Dr. Matilda Alexander	none@ics.mq.edu.au	02 9754 1456	**Expertise Areas:** Alaskan Grizzly Bears, Polar bear habitat use, bear and human habitat sharing **RFCD codes:** 300802 - Wildlife and Habitat Management, 300406 - Animal Growth and Development, 300402 - Animal Reproduction, 300403 - Animal Nutrition, 270503 - Animal Anatomy and Histology	"polar", "bears"	63.6%	76.7%	Yes No New rank:
2.	Assoc. Prof. Pia Olan	none@ics.mq.edu.au	02 9742 1211	**Expertise Areas:** Global warming and climate change, polar research **RFCD codes:** 260115 - Glaciology, 260602 - Climatology (incl. Palaeoclimatology)	"polar"	100%	100%	Yes No New rank:

Figure 4: Ranking evaluation exercise.

3.1 Ranking Algorithm

The algorithm for ranking the recommended experts was also implemented so it could be assessed by the participants. The algorithm contained the following steps:

1. all experts to whom the criteria entered by the searcher was not applicable were removed.
2. if the searcher entered expertise keywords, the remaining experts were ranked on how many keywords were found in their listed areas of expertise. Experts for whom no keywords were found were removed.
3. the remaining experts were ranked on their combined availability and searcher feedback scores.

Steps 2 and 3 were actually completed together by multiplying the fraction of words in the search query that were also in the expert's listed areas of expertise by their combined availability and searcher feedback scores.

An expert's availability score is recorded by the system (later referred to as the system's availability score) and calculated by the formula

$$A = \frac{available}{contacted} \quad \text{......(1)}$$

where *contacted* is the number of times the expert has been contacted by searchers and *available* is the number of times the expert has been available after being contacted.

The searcher feedback score F is divided into 3 parts. The first part is the searcher's availability score for the expert, *searcherAvailable*. This can be expressed as:

$$userAvailable = \begin{cases} 0 & \text{if the expert did not reply} \\ 0.5 & \text{if the expert replied and was not available} \\ 1 & \text{if the expert was available} \end{cases}$$

The availability score entered by the searcher for the expert is used in addition to the system's availability score for two reasons:

1. In case the expert indicates that they are available or unavailable to the system, but then changes their mind and contacts the searcher separately; and
2. In case the expert does not respond to the system's email, but rather contacts the searcher directly.

The second part of the expert's feedback score is their helpfulness score, *helpfulness* and can be expressed as:

$$helpfulness = \begin{cases} 0 & \text{if the expert was unhelpful} \\ 0.5 & \text{if the expert was unable to help personally but was still helpful} \\ 1 & \text{if the expert was able to help the user} \end{cases}$$

An expert could gain a score of 0.5 if, for instance, they were unable to help, but referred the searcher to another expert who was able to help. The searcher is only required to answer this section if the expert was available.

The final part of the expert's feedback score is their recommendation score, *recommendation,* which indicates whether the searcher would recommend the expert to someone with a similar query. This score can be represented as:

$$recommendation = \begin{cases} 0 & \text{if the searcher would not recommend the expert for a similar query} \\ 1 & \text{if the searcher would recommend the expert for a similar query} \end{cases}$$

The total feedback score, F, is simply the sum of the 3 sub-scores and can be represented as:

F = *searcherAvailable* + *helpfulness* + *recommendation*......(2)

The combined availability and searcher score, s, can then be represented by the formula:

$$s = \frac{A+F}{2} \quad \text{......(3)}$$

The final ranking of the expert is thus represented as the formula:

$$ranking = s\left(\frac{keywords}{query}\right) \quad \text{......(4)}$$

Where *query* is the number of words in the query and *keywords* is the number of query words found in the expert's expertise keywords.

4 Results

As the usability test was online, it was not possible to make sure that the participants completed the whole test. As a result 38 participants started the test, but this number dropped to 28 by the time the participants were asked to evaluate the system's algorithm for ranking experts. The same 28 participants completed the first scenario and filled out the survey, and 22 of those went on to complete the second scenario and fill out the associated survey. There were 11 females (aged 20-69, mean 32) and 17 males (aged 19-69, mean 34). Four females and 9 males had jobs which involved searching for experts. This was important in validating that the prototype was consistent with the searcher's task and mental model.

Expert ID	e1	e2	e3	e4	E5
System's rank	1	2	3	4	5
% agreement	100	54.55	36.36	45.45	40.91
Average rank	1	2.727	4.227	3.136	3.909
StDev	0	0.883	0.973	0.889	1.109
Adjusted rank	1	2	5	3	4
Mode	1	2	5	4	5
Confidence. Interval 95%	(1,1)	(2.36 3.10)	(3.82, 4.63)	(2.76 3.51)	(3.45 4.37)

Table 1: Participants ranking of experts

4.1 Evaluating the system's ranking algorithm

Although 28 participants completed this section, several entered invalid data, such as giving two experts the same rank, or entering rankings that were not in the range of 1-5. After these responses were removed, there were 22 participants remaining. All 22 participants agreed with the system's ranking for the first expert, meaning that all participants thought that the highest ranked expert deserved to be ranked highest, 13 participants agreed with the system's ranking for the second expert, 8 agreed with the system's ranking for the third expert, 10 agreed with the system's ranking for the fourth expert and 9 agreed with the system ranking for the fifth expert. While all participants agreed with the system's rank for expert *el*, as shown in Table 1, this level of agreement is significantly less for each subsequent rank with a second highest level of agreement of 54.55% for *e2*. A t-test performed on the participants' rankings for each expert other than *e1* (with the system's rank taken as the population mean, µ) show p to be significantly less than 0.01 in each case (*e2* t=3.864, *e3* t=5.919, *e4* t=-4.557, *e5* t=-4.615 df-21). This leads us to reject the null hypothesis that there is no difference between the average participant's rankings and the system's rankings.

To explain this discrepancy between the system's ranking and the participants' rankings, we must look at the ranking exercise itself which involved finding an expert in "polar bears". As seen in Figure 5, the expert ranked highest had both the words "polar" and "bears" in the description of her expertise areas, and thus was also ranked highest by all the participants. The remaining experts had only either "polar" or "bears" in their expertise descriptions. As no special weighting is given to individual search words, the remaining experts were ranked on their combined availability and user feedback scores. It was for this reason that the expert ranked third had expertise in "Bipolar disorder", an availability score of 100% and a user feedback score of 66.7%. This expert was ranked last, on average, by the participants, because his area of expertise had nothing to do with bears or the concept of 'polar' that was used in the search query (unlike the expert ranked second). One participant commented that results with whole word matches should come before results with partial matches (i.e. matches with 'bears' and 'polar' should come before matches with 'bipolar'). While it would not be difficult to add this functionality to the algorithm, in this particular example however, an expert with expertise in 'polar bonds' (a term in Chemistry) and the same feedback and availability scores as the expert currently ranked third by the system (who was matched on 'bipolar'), they would still be ranked higher than the two experts who matched 'bears' (who were ranked fourth and fifth). This is because even though the context of the word 'polar' is wrong, it was still a whole word match, and the expert's combined feedback and availability scores were higher than the combined feedback and availability scores of each of the two experts who were matched on the word 'bears'. One major issue participants had with the rankings was that the context of the search terms was not taken into account. When a human is assessing the relevance of an expert to a search query, they will not only look at the presence of the query words in the expert's expertise description, but also look at how the context of the search words in the query compare to the context of the search words in the expertise description.

An expert with expertise in glaciers and polar regions would be ranked higher than someone with expertise in bipolar disorder or polar bonds if the search query was 'polar bears'. This type of context matching is not a simple task for an ERS to perform as it would require a domain specific ontology for every domain of expertise and cost far more in processing time than a simple string matching algorithm. Another issue some participants had with the rankings was that both words in the search query were given equal significance. Some participants felt an expert whose expertise matched the word 'bears' should be ranked higher than an expert who matched the word 'polar' (when it was in the right context) and some participants felt the opposite should be the case. In most cases, this seems to be a matter of personal preference. For instance, if the query used in this example, ‘polar bears’, did not return any complete matches, would a searcher have more success contacting an expert on Kodiak bears, or an expert on polar regions? On one hand, an expert in Kodiak bears may know experts on other kinds of bears and be able to give the searcher a referral. On the other hand, an expert in Polar Regions may know experts in polar wildlife and may also be able to give the searcher a referral. The easiest solution to this problem would be to add an advanced search option, where the searcher could indicate the words in the query that they feel to be most important. Experts who were matched with these words would be ranked higher than those that matched other words in the query, regardless of their availability and feedback scores.

In general, most participants appeared to consider the expertise areas the most important indicators of rank for this example. One participant commented that they would not rank an expert with high feedback and availability scores in an unrelated field higher than an expert with poor availability and feedback scores in the field they were looking for. While the most qualified expert, in this example, was ranked highest, this may not always be the case. If an expert’s availability and feedback scores are 0, for instance, they will be ranked last by the system, regardless of their expertise area. The way to overcome this is to simply display all complete query word matches first, ranked by their availability and feedback scores, and then display the partial matches. These partial matches would be ranked by the usual combination of availability score, feedback score, and fraction of matched keywords, and additionally any extra weighting of keywords that the user may have specified.

A final issue, pointed out by one participant, was that experts who were unavoidably unavailable (due to sickness, or long service leave, for instance) would find that their availability score would drop and they would start getting ranked progressively lower, and would not be contacted even when they became available. The proposed expert response form allows an unavailable expert to indicate their first available date after being contacted by a searcher. This information could be displayed in their profile so a searcher knows not to contact the expert before the date shown. After the date has passed, this information would be removed from the expert’s profile and searchers can assume they can contact the expert as usual. Experts should also be able to enter dates that they will be unavailable in their profiles before being contacted by a searcher. The ranking algorithm could then be further altered to take any unavailable dates into account when ranking experts, by either not displaying experts who are unavailable at the time, or by giving them the lowest possible rank (either in the complete matches or partial matches section) regardless of feedback and availability scores. The same participant also pointed out that an expert who had been contacted once and responded once (an availability score of 100%) would be ranked higher than an expert with the same area of expertise who had responded 99 times out of 100 (an availability score of 100%). This is an important point and should be taken into account by the algorithm. Thus, an improved ranking algorithm would be the following:

1. Same as step 1 in section 2.
2. If the searcher entered expertise keywords, the experts for whom no keywords were found are removed.
3. a) Experts whose expertise description yielded matches with all query words are listed first and ranked using a modified version of equation 4 in section 2 :

$$ranking = s\left(\frac{keywords}{query}\right) f(numContacted)\, g(currentDate, unavailableDates) \quad \text{.........................(5)}$$

Where *keywords/query* = 1; *f(numContacted)* is a function that takes in the number of times an expert has been contacted and outputs a real number; and *g(currentDate, unavailableDates)* is a function that takes the current date (or the last available date the searcher has indicated that they can receive a reply from an expert) and any dates an expert has indicated that they will be unavailable, and returns 0 if *currentDate* matches with any dates in the set *unavailableDates* and 1 otherwise. Thus *ranking* = 0 for any expert who is unavailable

OR

b) The experts with partial matches would be ranked using the following modified version of equation 5:

$$ranking = s\left(\frac{keywords}{query}\right) f(numContacted)\, g(currentDate, unavailableDates) * \sum_{i=0}^{i=n} weight(queryWord[i]) \quad \text{............} (6)$$

Where *weight(x)* returns the user-specified weight of search word *x* and *queryWord*[*i*] returns the *ith* word found in both the query and the expert's expertise description, I = 1,2,…*n*.

4.2 Evaluating the Search Page

For the search page (Figure 2), participants were asked to indicate their level of agreement on a 5-value Likert scale for the statements: 1) I found it easy to understand how to search for the expert on the search page and 2) The search options on the search page were specific enough for me to search for the expert I needed. For both scenarios, the majority of the participants agreed with both statements. In fact 71.43% of participants either agreed or strongly agreed with the first statement and 78.57% of participants either agreed or strongly agreed with the second statement for scenario 1. For scenario 2, 95.45% of participants either agreed or strongly agreed with the first statement and 90.91% of participants either agreed or strongly agreed with the second statement. This percentage increase can be explained partly by the fact that the participants would have a better grasp of how the system works after completing the first scenario and starting the second and partly by the fact that six of the participants who completed the first scenario did not complete the second scenario. Out of the 22 participants who completed both scenarios, 81.82% agreed with the first statement after completing scenario 1 and 86.36% agreed with the second statement after completing scenario 1, showing increased satisfaction perhaps due to increased familiarity.

A few participants found the layout of the search page confusing initially, and some said that they would have preferred less options and an "advanced search" option instead. A few also disliked the drop down list of RFCD codes, although this option would be unlikely to be used by someone who didn't have a specific code in mind.

4.3 Evaluating the Results Page

The Results Page provided a list of experts with their name, email address, phone number and description of their areas of expertise. Participants were asked to rate: 1) the experts' details were sufficient for me to tell if I needed to contact them, 2) it was clear to me how I could use the system to contact the experts on this page and 3) after reading this page I understood how to provide feedback on the expert. As with the results for the search page, levels of agreement with each statement rose for the second scenario. Percentage agreement for each statement for scenario 1 was 89.29%, 85.71% and 60.71% and for scenario 2 was 95.45%, 95.45% and 86.36%, respectively. This percentage increase can be partly attributed to the smaller number of participants who completed the second scenario. Three of the six participants who did not complete the second scenario strongly disagreed with one of the three statements; the other 3 participants who did not continue chose 'neutral', 'disagree', or 'agree' in response to the statements. The large increase in agree and strongly agree responses for the third statement can most likely be attributed to the participants gaining experience in using the system after they completed the second scenario. Many participants thought the instructions on how to submit feedback were not very clear initially. To rectify this, there should be a separate button for each expert that, when clicked, would take the searcher immediately to the feedback page for that expert. In reality, however, a searcher is probably not likely to provide feedback on an expert immediately, but rather after some time has elapsed and they have been sent a reminder email by the system that includes a link to the feedback page for the expert they contacted.

4.4 Evaluating the Expert Profile Page

The Expert Profile Page provided specific and further details for an individual expert. For the first scenario, 92.86% of participants either agreed or strongly agreed that the details of the expert were sufficient for them to tell if they needed to contact them, 71.43% of participants either agreed or strongly agreed that they found the categories on the page easy to understand, and 89.29% of participants either agreed or strongly agreed that it was clear to me how I could use the system to contact the expert on this page. For the second scenario these percentages of agreement were 95.45%, 90.91% and 95.45%, respectively. The differences in the percentages can mainly be explained by the smaller number of participants who completed the second scenario. Out of the 22 participants who completed both scenarios, 100.00% agreed or strongly agreed the details were sufficient after completing scenario 1, 81.82% agreed or strongly agreed the categories were easy to understand and 100.00% agreed that it was clear how to contact the expert.

A couple of participants commented that they wanted more details on the experts' areas of expertise on their profile page, such as a list of research projects they had taken part in or papers they had written. This data, if available, would be shown or linked to on each expert's profile page in a complete version of the system where the experts were real people. These responses show that some people wish to have access to all available information about an expert before deciding whether to contact the expert, and it is therefore important for organisations to constantly keep track of this data.

Another participant commented that they were not able to discern how helpful an expert was going to be by viewing their profile. This is a difficult problem to fix, as the feedback information the system uses to rank an expert is not displayed for ethical and practical reasons. Many experts would not be happy with their details in a system that displays to the public what other people think of them. If an expert saw that they had an average feedback score of 20%, for instance, they may become upset and ask for their profile to be removed from the system. While an expert should be allowed to know what their feedback score is, showing this information to all users of the system would not be appropriate. Showing an expert's availability information to the public, however, would probably be acceptable. It would be beneficial to both the searcher and the expert to have the expert's availability information displayed, as the searcher will know that they might not have much luck if they try to contact the expert, and the expert will not have to be constantly rejecting requests for help from searchers.

4.5 Evaluating the Contact Page

The Contact Page provided an email form with headers and body to allow contact to be made with the expert. For the first scenario 85.71% of the participants agreed or strongly agreed that 1) it was clear who the email was being sent to and 2) that they would use the feature in the future if it was available. 89.29% agreed or strongly agreed that 3) the form was easy to fill out. For the second scenario agreement, 90.91% of the participants agreed or strongly agreed with all three statements. The most promising result was the high percentage of participants who indicated for both scenarios that they would use the feature in the future if available. None of the participants disagreed with this statement, although one participant commented that they could imagine copying the email address and sending their own email to the expert. A number of participants did not like the first line in the email body: "Dear <TITLE SURNAME>, **DO NOT CHANGE THIS LINE". This was placed there to allow a message to be sent to several experts at once, although filling this line out for the searcher when they only want to contact one expert could lessen the confusion considerably.

4.6 Evaluating the Feedback Page

We asked participants for their level of agreement with the statement "I was able to adequately express my feelings about the expert on this page" (see Figure 3). 60.71% of participants agreed or strongly agreed with the statement after completing scenario 1, and 63.64% agreed or strongly agreed with the statement after completing scenario 2. This value is quite low compared to the other sections. Some participants thought the feedback options available on this page were too rigid, especially for scenario 1, when the recommended expert actually recommended another expert, but wasn't of any help otherwise. The section of the feedback form that requires a Yes/No answer (*I would recommend this expert to someone with the same query*) would be especially hard to fill out in a situation such as this. Another participant commented that the additional comments section was the only place where an expert's performance could be evaluated (with the other sections evaluating the expert's immediate response and availability). The feedback page was structured in this way to avoid making people fill out too many sections, as they would be unlikely to provide feedback regularly if this was the case.

Free text comments are hard to quantify, however, and no score has been derived from the additional comments to use in ranking the expert. Adding another section to indicate how satisfactory an expert's performance was could be a step towards solving this problem. It could allow the user to give a Yes/No response to the statement "I was satisfied with the expert's performance" or have them rate their satisfaction with the expert's performance on a scale of 1-5 with 5 being very satisfied and 1 being very dissatisfied. There are some issues with this method, however. One person's satisfaction with another person's performance can be very subjective. One person may think an expert performed excellently, while another may think they performed poorly, even if they gave the same performance in both cases. If free text was used to evaluate an expert's performance, the searcher could choose the comments to be sent to the expert so they can see exactly what the searcher was dissatisfied with. A Yes/No response, or a rating out of 5 would not give the expert a good idea of exactly what the searcher thought, and would therefore not be able to improve. A second option would be to show each expert their feedback scores and comments on a private part of their profile.

5 Conclusions and future work

As a key part of our approach is the combination of self-reporting/referral and automated searching through available data. Some data can only be obtained via self-reporting (e.g. indicating if you are available to do media interviews or guest lectures). However, information about which units one teaches, expertise areas, grants, etc. can be gained from personal websites and internal databases. An outstanding issue would be how to reconcile differences between these sources and between the outputs of automated searching and self-reporting.

In the current approach, experts' profiles are generated once and updated periodically instead of dynamically at query time. A future implementation of the system could provide dynamic generation of experts' profiles to provide the most up-to-date information on experts. Ideally the system would select the most appropriate algorithm and information source to meet the searcher's requirements and intelligently infer expertise need from searcher activity. A more sophisticated keyword matching technique incorporating a more sophisticated ontology should be used to allow for misspellings and partial similarity matching. The search results could contain a link to each expert's social network (possibly represented as a diagram) to show what people the expert is associated with. This would be especially useful for people to find alternative ways of contacting an expert. We are interested to investigate further other potential uses within our approach of social network data, as in SmallBlue (Lin *et al*, 2008) and social network analysis (Scott, 1991) as some of our interviewees wanted to find someone based on the experience of someone they personally knew and trusted.

A current social networking site that has ERS-like features is LinkedIn (http://www.linkedin.com/). Users can create personal profiles for free and enter details about their profession and expertise areas. They can create a social network by adding other members as contacts, thus gaining access to all their contacts' social networks as well. LinkedIn users self-report their expertise and ask members of their social network to provide positive references or recommendations for them. People who have profiles in the system can ask questions and answer questions posed by other members. Members can earn 'expertise points' by answering questions and having the person who asked the question rate their answer as the best. In terms of locating experts, this question-answer feature caters mostly to people who have *information need* rather than *expertise need* (Yimam-Seid and Kobsa, 2003.) LinkedIn and our approach both use self/other-reporting and rating answers in LinkedIn is like our use of feedback for ranking recommendations, we however also use data mining to validate and find expertise. LinkedIn allows members to contact experts either through their (or a contact's) social network or by paying money to upgrade their account and then sending an in-system email to them (InMail). Our system uses an in-system contact form, but also makes experts' contact details available to all searchers so they can choose to contact the expert another way if they wish. Our prototype system allowed experts to indicate what kind of media contact they would allow (such as newspaper and radio interviews), and requesters were able to select how soon they needed to contact the expert, which allowed experts to prioritise their requests. Experts should also be able to indicate what dates they are unavailable and prevent requesters from sending emails during those times. An in-system email account could help cut down on unwanted traffic in an expert's personal or work email account. The system could order the emails by how soon the requester needed a response, and highly sought experts could be sent summary emails informing them of the number of requests they have.

ERS's have the potential to save people valuable time and effort searching for an expert. The fact that many people still prefer to find experts themselves rather than using the available technology points to a development gap in these systems that needs to be filled. The positive responses of participants to the prototype evaluation is a good indication that our triangulated approach to finding experts has the potential to produce ERS's that people can and want to use.

References

Aïmeur, E., Onana, F., Saleman, A. (2007). HELP: A Recommender System to Locate Expertise In Organizational Memories. In ACS/IEEE Int.l Conf. on AICCSA '2007, Amman, Jordan, 866-874.

Becerra-Fernandez, I. (2006). Searching for experts on the Web: A review of contemporary expertise locator systems. ACM Trans. Internet Technol. 6, 4 (Nov. 2006), 333-355.

Crowder, R., Hughes, G., Hall, W. (2002). An agent based approach to finding expertise. In Proc. of 4th Int.l Conf. on Practical Aspects of Knowledge Management, Heidelberg, Germany, 179-188.

Ehrlich, K., Lin, C.-Y., Griffiths-Fisher, V. (2007). Searching for experts in the enterprise: combining text and social network analysis. In: GROUP '07: Proc. 2007 Int.l ACM conf. on Supporting group work. ACM, New York, NY, USA, pp. 117-126.

Foner, L. (2002). Yenta: A Multi-Agent Referral-Based Matchmaking System. In Proceedings of the 1st Int. Conf. on Autonomous Agents Marina del Rey California, 2002, pp. 301-307.

Krulwich, B. and C. Burkey, (1996). The ContactFinder Agent: Answering Bulletin Board Questions with Referrals Proc. 13th Nat. Conf. on AI . Vol 1, Portland, Oregon, pp. 10-15.

Lin, C.-Y., Ehrlich, K., Griffiths-Fisher, V., Desforges, C. (2008). SmallBlue: People Mining for Expertise Search. IEEE MultiMedia, vol. 15, no. 1, Jan-Mar 2008, pp. 78-84.

McDonald, D.W. (2001). Evaluating Expertise Recommendations. In Proc.Int. Conf.on Supporting Group Work (GROUP'01), September 30-Oct. 3 2001, Boulder, Colorado USA. pp. 214-223.

McDonald, D. W., Ackerman, M. S. (2000). Expertise recommender: a flexible recommendation system and architecture. In Proc. of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00), December 2-5 2000, Philadelphia, Pennsylvania, USA, 231-240.

Middleton, S. E., De Roure, D. C. D., Shadbolt, N. R. (2001). Capturing knowledge of user preferences: ontologies in recommender systems. In Proc KCAP'01. ACM Press, NY, USA, 100-7.

Pikrakis, A., T. Bitsikas, S. Sfakianakis, M. Hatzopoulos, D. DeRoure, S. Reich, G. Hill, and M. Stairmand, (1998). MEMOIR – Software Agents for Finding Similar Users by Trails. In Proc. 3rd Intl.Conf. on Practical Application of Intelligent Agents and Multi-agents, London, 453-466.

Richards, D. & Busch, P. (2009). Combining Information Systems and Computer Science Approaches to Identifying Sources of Expertise, Proc. ACIS'09, Melbourne, 702-711.

Scott, J., (1991). Social Network Analysis: A handbook, Sage Publications, London, U.K.

Sim, Y., Crowder, R., Wills, G. (2006). Expert Finding by Capturing Organisational Knowledge from Legacy Documents. In Proc. Int.l Conf. on Comp. & Comm. Eng. (ICCCE '06) KL, Malaysia.

Taylor, M and Richards, D. (2009) Discovering Expertise from Publication Data, In Richards, D, Kang, (eds.) Proc. Pacific Knowledge Acquisition Workshop (PKAW08), LNAI 5465, 218-230.

Vivacqua, A., (1998). Agents for Expertise Location. In Proc. AAAI Spring Symp. on Intelligent Agents in Cyberspace, Stanford, CA, 9-13.

Yimam-Seid, D., Kobsa, A. (2003). Expert Finding Systems for Organizations: Problems and Domain Analysis and the DEMOIR approach. Jrnl of Org.l Comp. & Electronic Commerce, 13:1-24.