Searching for Experts in the Enterprise: Combining Text and Social Network Analysis

Kate Ehrlich
IBM Research
Cambridge, MA 02142
katee@us.ibm.com

Ching-Yung Lin
IBM Research
Hawthorne, NY 10532
chingyung@us.ibm.com

Vicky Griffiths-Fisher
IBM Global Business Services
London, UK
vicky.griffiths-fisher@uk.ibm.com

ABSTRACT
Employees depend on other people in the enterprise for rapid access to important information. But current systems for finding experts do not adequately address the social implications of finding and engaging strangers in conversation. This paper provides a user study of SmallBlue, a social-context-aware expertise search system that can be used to identify experts, see dynamic profile information and get information about the social distance to the expert, before deciding whether and how to initiate contact. The system uses an innovative approach to privacy to infer content and dynamic social networks from email and chat logs. We describe usage of SmallBlue and discuss implications for the next generation of enterprise-wide systems for finding people.

Categories and Subject Descriptors
H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work, Evaluation/methodology

General Terms
Management, Human Factors

Keywords
Social Network Analysis, Social Networks, Expertise Location, Collaboration, Communication, CSCW

1. INTRODUCTION
It is well established that we rely on a personal network of friends and colleagues to get trusted information [7, 23] to help filter, interpret and make sense of information [12] and as brokers who introduce or refer us to new people [5]. Though personal networks are invaluable for getting quick answers, they are not always sufficiently large or diverse to reach everyone directly who has the right information. The limited reach of a personal network is especially pronounced for new hires who have not had the time to build a network or for people who have recently assumed new or different roles that require a different set of connections than the ones they have.

GROUP'07, November 4–7, 2007, Sanibel Island, Florida, USA.

Researchers have argued for some time that personal networks should be augmented with technology, called expertise locator systems, to help find the right person (e.g. [1, 3, 19]). Existing systems have focused on search algorithms or methods of acquiring data that can be searched (see [4] for a review). But searching for people is not like searching for documents. People operate in an organizational and social context which circumscribes what they know and how they know it [8]. Moreover, while companies vary in their tolerance for "cold calls", even internally, someone receiving a request for information is more likely to respond positively to the request if it comes from a friend rather than a stranger. Thus, expertise locator systems should reflect the social contexts in which people are embedded [11] to help evaluate potential experts but also to facilitate the path to conversation. In earlier research, Ehrlich [11] argued that people balance several social and organizational factors before deciding who is the most suitable and responsive person to approach. Because these factors and their significance vary according to the searcher and the context of search, expertise locator systems should provide sufficient information for the user to make the determination of suitability and responsiveness.

Lin [17] proposed an enterprise social networking system, called SmallBlue, that unlocks the business intelligence of 'who knows what?' and 'who knows whom?' residing in an organization, without requiring explicit involvement of individuals. The aim of SmallBlue is to find experts, communities, and networks in large companies through data mining, information retrieval and social network analysis techniques. SmallBlue helps users manage their personal networks, and reach out to their extended network (the friends of their friends) to find and access expertise and information. In SmallBlue, expertise search is utilized through Google-like keyword searches with a returned ranked list of the top N experts of the keywords. The network structure is also used to determine the shortest path from the user to the selected expert. In addition to the requisite search interface, the system addresses issues of interpretation and responsiveness through two additional features we have called Social Distance and Expertise in Context. Social distance provides explicit, personalized information about the shortest and alternate paths that link the user to the seeker. Expertise in context provides additional information about selected experts drawn from their participation in open forums, communities, blogs and social bookmarking systems.

The purpose of this paper is to describe early user studies of the SmallBlue system. Before describing SmallBlue we review relevant research and present a brief summary of a study that informed some of the design decisions. In Section 4, we briefly introduce the SmallBlue system, which has been deployed within IBM and adopted by over 1700 people within 6 months of availability. It can be used to search the expertise and social network of more than 150,000 employees. In Section 5, we describe the results of a preliminary user study of SmallBlue and conclude with discussion of implications and future directions.

2. PREVIOUS RESEARCH

Expertise location has often been approached as a social matching or recommender system e.g. [19, 24, 27] which derives information about experts from data provided by the user, typically in the form of profiles, or from mined data. One approach typified by systems such as *HelpNet* [18] asks users to provide the information by filling in profiles which they then have to maintain. Another approach found in systems such as NASA's *Expert Finder* [6] process employees' published documents such as resumes and corporate newsletters to create expertise profiles. There are also hybrid approaches such as the systems described in [14] in which basic contact information and reporting structure are created and maintained from personnel records but supplemented with data mined from user generated blogs and social bookmarking systems .

A different approach creates systems that are more like extended personal networks. Some of these systems e.g. *ContactMap* [30] maps a person's own network in a way that makes important demographic factors such as geographic or organizational affiliation or interest group more salient. This method of display helps people manage their network of contacts to be more aware of people who are important but contacted infrequently. The value of these systems lies in providing visualization and other features to help people manage their existing contacts rather than in identifying the experts who are outside the user's personal network.

Other systems go beyond the personal network – the people known personally to the user – to combine expertise location methods and social networks. For example, *Referral Web* [16] combines social networks and collaborative filtering to create personalized recommendations and personalized referral paths from the user to the designated expert. This system helps identify people who will respond to requests but it doesn't necessarily provide users with enough information to make their own determination of the expert's likely responsiveness to a request for information.

Other systems have used graph-based ranking algorithms to extract both social network and expertise rankings [9]. These systems facilitate expertise finding. However, users still need to manually update their profiles which make it hard to build scalable systems that can also be easily maintained over time.

CommunityNet [25] integrates automatic email content and social network analysis to find out how experts collaborate and to recommend experts to users within their own personal network. Initial experiments using the Enron email dataset of nearly 0.5 million emails showed promising results in finding experts. However, *CommunityNet* did not look beyond the personal ego net for expert finding. Song *et al.* proposed using dynamic graphs called *ExpertiseNet*s to model the relational and evolutionary expertise for mining, retrieval and visualization [26].

Although a few systems incorporate social network data, many have shied away from presenting whole social networks that go beyond the user's direct connections, principally because of difficulties gathering the data and privacy concerns. This is unfortunate because social networks contain a rich set of data about "who knows who" not just "who knows what" which is critical for general knowledge sharing. Previous research has revealed the importance of the underlying social network structure for understanding patterns of collaboration and information sharing [7, 10, 15].

3. PRELIMINARY STUDY

People in professional services are a key target group for expertise locator systems. To gain a better understanding of how people in professional services might leverage their social network to get information we conducted a series of interviews with 7 project managers who worked in a global consulting practice. The data from these interviews were synthesized to identify the major elements of the work environment and how searches are conducted. We validated the initial findings with an additional focus group of 15 project managers. We focus here on some of the comments we received that pointed to when and how this group went to their personal networks for information.

3.1 Work Environment

Consulting practitioners have little free time and are under pressure to execute, often in full visibility of the client. Their focus is firmly on delivery activities. When faced with information or a knowledge gap they take what they feel is the most expedient route to filling it and getting the job done. They want relevant, useful, easy to access content, quickly, and will reject sources that are cumbersome or time consuming to use unless they have to use them. Consequently, they rely heavily on personal networks and their hard-drive as primary sources of information, assets and resources.

3.2 Importance of personal networks

It was readily apparent from the interviews that when it came to searching for information, people were a primary source of knowledge / asset sharing

"... It really comes down to knowing who knows what and being able to get hold of them.".

Our findings revealed that typically the first port-of-call for sourcing content is through personal networks rather than going online.

"...You need to look through 100s docs to get the useful material."

"...You spend effort wading through documents when you know a 2 minute conversation with an expert would point you in the right direction."

3.3 Limitations of personal networks

Access to expertise is particularly important for getting specific, differentiating information on delivery activities, for accurate timescales and project estimates. Our study revealed several challenges in getting access to the right expertise. As projects became longer, practitioners' personal networks were growing slower than when projects had a shorter cycle.

"*... your network does diminish if you don't actively try and move around on projects...*"

But practitioners still depended on a personal connection or introduction to a peer for a timely response.

"*… You tend to need a way into speaking with somebody, maybe 'so and so recommended you and said you might be able to help out with this usually gets a fairly quick response*".

When it came to searching for people, we heard stories about needing to go through 3, 4 or even 5 connections before reaching the right person. One person told us of going to someone they knew who then gave them a name of another person who further introduced them to a third person and so forth. As each new person was contacted there was a delay in reaching them further impacting the timeliness of the resulting information. While our informants could tolerate some indirection, when it got to the 4th new contact or beyond, the time delay was unacceptable.

Although this was an informal study it reinforced the importance of using other people as a source of information, of engaging the other person in conversation not just getting an answer, and of the extreme time pressures of their work.

4. OVERVIEW OF SMALLBLUE

SmallBlue, proposed by Lin [17], is an expertise search and social network analysis suite that automatically captures and visualizes social networks. It enables users to find people with specific knowledge or skills in an enterprise. The current paper provides a high level overview; for more details including the architecture and design see [17].

SmallBlue is made up of the SmallBlue social sensor system and a suite of 4 web based user tools:

- **SmallBlue Ego** tool displays the user's personal network (not described in this paper)
- **SmallBlue Find** tool provides an interface to a relevance ranked list of people who match the search terms. Millions of search terms are currently indexed in SmallBlue.
- **SmallBlue Reach** tool provides dynamic profile information similar to Fringe [14] to help see the expert's knowledge in context to assess their suitability
- **SmallBlue Net** tool which displays the social network of the top experts matching the search term or the social network of any group/community of people within company

The SmallBlue social sensor analyzes outgoing email and chats on users machines and creates information search indices and personal social network indexes which are sent to the SmallBlue Server for large-scale network and search indices aggregation.

4.1 Data acquisition

Many existing expertise locator systems acquire data by having individuals fill out profile information, by extracting information from online databases, or deriving artificial intelligence algorithms from existing sources. Those sources could be "public" such as co-authored documents, patents or user-generated content in blogs, wikis and social tagging systems. Data can also be acquired from private sources such as email, chat and calendar entries, which contribute semantic information as well as social network data. In deciding which data source to use, we quickly ruled out user authored profiles because it would have taken too long to build a critical mass of information and would not have included the social network information we needed.

The issue of whether to use public or private data sources was more complex. Private data such as email logs contain rich information to derive what one knows and who one knows These data also address issues of (a) coverage – everyone uses email so data can be collected from everyone not just the people who have authored documents or other data; (b) maintainability – new email is constantly being generated; (c) ease of use – people are already using email so other than asking users for permission to use their data there is no additional work required. The disadvantage of using private data is that capturing and using the data may violate privacy. There is some evidence suggesting that the majority of users are willing to share some private data under the right circumstances [2]. However, it can be challenging to get people to provide those data without sufficient clear and enforceable safeguards.

In SmallBlue the decision was made to address the privacy issues and collect email data. By addressing the privacy issues, it would be easier to scale SmallBlue over the long term because of the better coverage of email data. To address privacy concerns, SmallBlue developed a strong set of policies that restricted what data could be collected, how data could be used and what information was available to users (see [17] for a detailed description). SmallBlue relied on aggregated and inferred information which prevented any user from ever seeing a direct relationship between any person in SmallBlue, their email, and, the information being displayed. In other words, the system never kept or displayed any information about who communicated with whom about what. Privacy and rights of users are critical to the success of the SmallBlue system. Without adequate attention to these privacy issues users will quickly stop using the system which will dry up the pipeline of data for searches and the system will lose value.

Data were only collected from people who had opted into SmallBlue after reading the privacy policies. There was no requirement or coercion to join; users could opt out at any time and have all their data removed. New users could try SmallBlue at any time before opting in. When a user opted in, they only had to specify the location of their email archives and chat history – users could include one or the other or both -- and the tool would extract and index the data. The real email or chat data never left the users' machine.

4.2 Search

SmallBlue Find is a search engine which returns a relevance ranked list of people by interpreting a search string and mapping it onto related keywords. The search engine aggregates the results for all the keywords and ranks them based on relevance weighting and aggregated social network structure. SmallBlue then generates a list of people who best match the search term. Only the top 100 people are displayed, with 10 people per page. Figure 1 shows a sample output of a search. The display shows the person's picture along with their job title, role, and online status. Users can also filter a search within a business division, a country, a community, a group, or/and a specific social distance. These filters are turned off by default and the search engine returns the name of anyone within the company who matches the search criteria.

Consider the case of someone who has been asked by their manager to gather information about second life, the 3D virtual environment that has been creating buzz on the internet and business press. Entering "second life" into the search window would bring up the display shown in Figure 1 which shows 6 of the 10 people from the first page, along with their picture and other information. To the right of each picture there is text in red which shows the degrees of separation between the person displayed and the user. The two examples in bold say "My collaborator or contact" indicating a personal contact and "Ask: Vicky" indicating that I can reach the person indirectly through Vicky. This information is personalized for each user. Anyone entering the same search term at approximately the same time period would see the same set of names in the same order. The specific degrees of separation as well as the names of the intermediate people would be different.

Figure 1. Relevance ranked search results and personalized degrees of separation for the search term "second life"

4.3 Social Distance

To increase the likelihood of the user contacting someone new, and having that person respond, SmallBlue displays the minimal number of intermediate people to reach the person as a form of "six degrees of separation" [20, 29]. Paths of three or less were shown in the initial relevance ranked display with the line, "Ask <person>" or "Ask <person1> => <person2>.

Most expertise location systems do not provide information about how to reach someone through existing contacts. Yet as we saw in our preliminary study and from existing literature, personal networks are the most common method for reaching experts. When it comes to doing a "cold call" to someone new, even in the some company, it helps to have some context. In some companies it is a recognized and accepted practice to call other people to get information that is needed for a project or assignment. And as long as it is a reciprocated practice – that is anyone can call and be called by anyone else – it works well.

Information about social distance also appears in another tool within SmallBlue, called Reach described below. From the Find page, users can click on a name or picture to navigate to the Reach page, which Expertise in Context

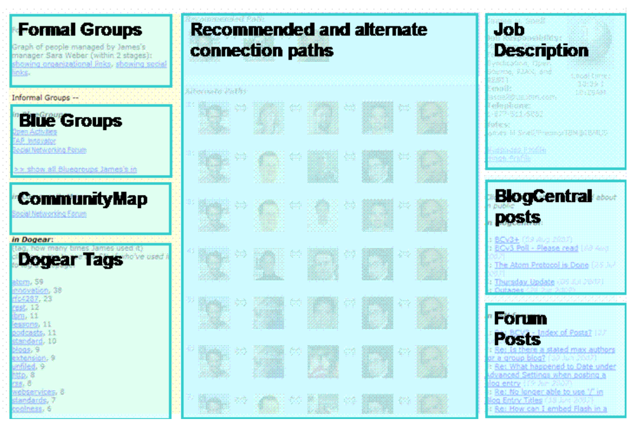
We have argued that an expertise locator system should provide users with contextual information to help them make their own determination of the appropriateness of a candidate's expertise, and likely responsiveness. In our case, this information is provided through SmallBlue Reach and SmallBlue Net.

4.4 Expertise in Context

4.4.1 SmallBlue Reach

SmallBlue Reach (Figure 2) displays current public information about a selected person. This information includes a list of the shortest paths to reach the person and up to 16 alternate paths. The page also displays information about the expert's interests and activities, drawn from a variety of data sources that are public within the enterprise. This information can help a user assess the suitability of a candidate by seeing what they might have authored in a blog, what social tags they have used or which pages they have bookmarked all of which create an impression of the person's current interests and knowledge.

Figure 2. Sample Reach page displaying shortest path and feeds from user authored blogs and social bookmarks



On the left side of the page shown in Figure 2, the Blue Groups and CommunityMap indicate the communities the expert has joined. Community membership is voluntary. So, assuming the names are meaningful, the list provides cues to the person's interests and perhaps their expertise. If someone were a member of a community called "Second Life" we might infer that they had enough interest in virtual worlds to make the effort to join the community. The Dogear tags displays any tag the expert has used more than 3 times in an enterprise social tagging system called Dogear [21]. This information can provide useful cues to the expert's potential interests along with their actual bookmarks which are displayed on the page but not shown in Figure 2. The list of communities and tags are links that when clicked will take the user to a page showing the social network of all the members of the selected community or the selected tag. We describe the social network view below under SmallBlue Net.

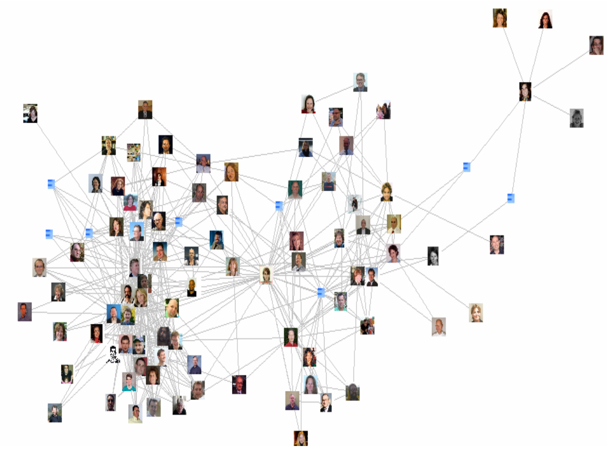
On the right hand side of the page, the job description shows the person's formal role which is a good indication of how they might be using their knowledge. For instance, a manager might be seen as having less practical knowledge of a technical area than a subordinate who is actively applying their knowledge. In IBM, anyone can post a blog at Blog Central or a comment at one of the many Forums that are available. SmallBlue picks up this information and displays the title of the 5 most recent postings along with the posting date. The content of these posts can be quite informative of the expert's current interests. Beneath the list of Blog and Forum entries are the actual pages that someone has bookmarked and beneath that is the person's self-described expertise. The information about expertise is provided voluntarily by any employee as part of their Blue Pages entry. Blue Pages is a corporate directory that contains standard employee information such as job title, location, contact information, reporting structure which is mostly sourced from employee records. There are also some additional fields which the employee is at liberty to fill out. The self-described expertise is one of these optional fields. If present, it can provide information that is highly informative of a person's expertise and credentials. Because this information is provided voluntarily and generally not updated, it is not a reliable single source for selecting an expert.

4.4.2 SmallBlue Net

SmallBlue Net displays the social network of people associated with a topic or the people in a community/group. Although we conventionally think of an expert as someone who has the most knowledge of a particular topic, sometimes we want to find the person who knows the expert rather than the expert themselves or we want to find the person who others think is the expert which is associated with the person at the center of the social network for a topic..

For these reasons, it is sometimes useful to see the whole social network. Figure 3 displays the social network of the same 100 people that were found by the Find tool in response to the search term, "second life". Mousing over one of the thumbnail pictures in SmallBlue Net brings up the same information about role and business unit as displayed in SmallBlue Find. In addition to the traditional network clustering view, the interface provides a map view which displays the geographic location of all the people in the network (not shown in Figure 3).

Figure 3: Social network for "second life" from SmallBlue Net



There are several ways of viewing the data. For instance, instead of seeing the nodes displayed as pictures, users can see the nodes color coded by business unit. This is an especially useful way to see how information is getting shared and propagated in the organization and provides insight into the structure and cohesiveness of emergent communities. Users also have the option of seeing who the hub of the network is and who is providing a bridge between separate clusters.

Figure 4: Social Network for "second life" showing key hub

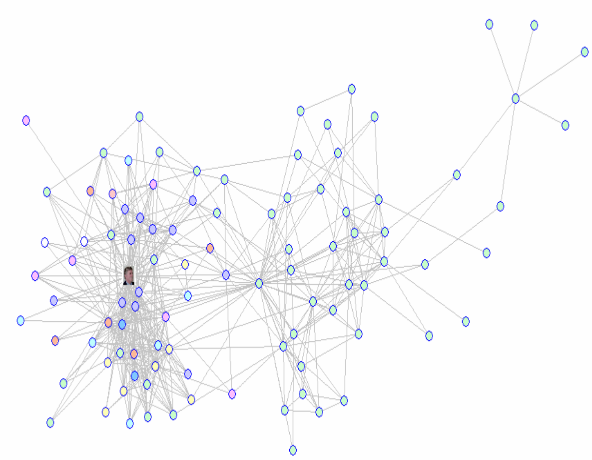


Figure 4 displays the business code view with the hub highlighted by a thumbnail picture. The hub is often but not necessarily the same as the person ranked 1 in SmallBlue Find. He or she is, however, the person most connected to the top 100 people associated with the search term, when the hub is also the most highly ranked person it means that the community agrees with the inferred designation of expertise. When the hub is not the most highly ranked expert it might be indicate of someone who is connected with lots of experts and therefore in an important role as intermediary or broker to those people.

5. USER FEEDBACK

We explored the usage of an early version of SmallBlue in three ways: (1) interviews conducted about a month after the system had been released, (2) log data collected after the system had been in use for about 3 months and (3) data from an online survey administered at 3-4 months. We addressed several key questions in the evaluations:

- How rapidly are people adopting SmallBlue? Getting rapid adoption is crucial for building up enough data to create a reliable, robust system.
- Were users satisfied with how SmallBlue dealt with privacy? If users are satisfied with privacy this would indicate that it is feasible and practical to build an expertise locator system by extracting content and social network from communication data and to do so at a level that can scale.
- Were people really using SmallBlue or just downloading it out of curiosity? This is difficult to assess because the real need for SmallBlue is unpredictable and highly variable across users
- Were users satisfied with SmallBlue in general and with the accuracy and usefulness of the responses

5.1 Adoption

The data for the number and rate of adoption of SmallBlue were very encouraging. Users were enthusiastic about the tool and recognized the value of its direction and purpose. We got comments such as, *"This has enormous potential", and "I like this because I am now starting to see people I know"*. The tools were regarded as especially useful by people who were in new roles and who had an on-going need to build up a new social network of contacts. For instance, one person commented, "*I would use it to find people in other development teams that I need to interact with e.g. to find someone on another team with specific expertise*". People were also using it to find people in areas of interest outside their main job. For instance, one person who was interested in healthcare applications told us that he could never have found the people or communities without the tool. Another person who was looking at new job opportunities in the company used the tool to find all the people in common with a potential future manager who could then is used for references; she got the job. Another person, who had been using SmallBlue for several weeks said, *"It has already helped me find some SMEs I couldn't have found otherwise."*

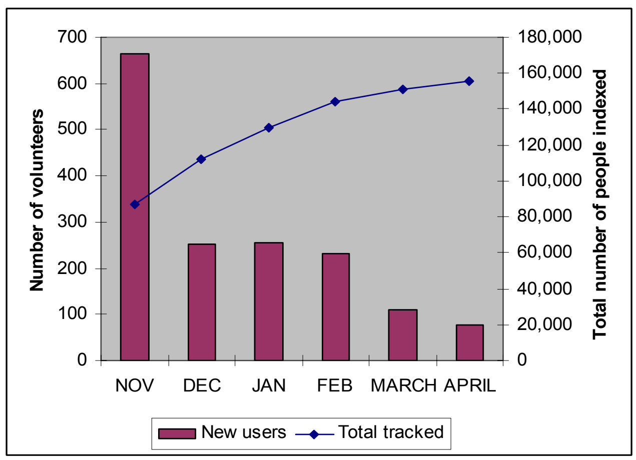
It was important to the overall usage of SmallBlue that we get enough people to sign up early for us to build a large enough corpus of data for our inference engine and create meaningful searches for the users. Building the initial data set is a challenge for any expertise locator systems

In October 2006, SmallBlue was made available to a small group of people internally, through email invitations and demos to targeted groups of consulting practitioners and friendly users to make sure that the system was behaving reliably and that we had identified and corrected major technical problems before making the system more widely available. Initially, we had about 150 people sign up. SmallBlue was then made available to a general audience in November through an internally managed early adopter program.

Figure 5 shows the number of people who opted into SmallBlue over the first 6 months. The total number of people who are indexed in SmallBlue comes from two sources. A) Each new volunteer brings in their direct connections. B) In addition, the data from each person's email is updated regularly which can add additional new people. As more people volunteer, the rate of new, unique people they bring in declines. In the first month or so of usage, there was a ratio of around 112 names added to SmallBlue from both sources, for every new volunteer. By the end of April that number had dropped to around 94.

With nearly 700 people signing up in November alone we had enough data to feed the inference engine and build the aggregated store of information that would disguise individual data. Almost all the growth in adoption came through word of mouth and curiosity. By May 2007, about 6 months after the system had been released internally, there were over 1600 people who had opted in – that is allowed us to mine their email – and more than 150,000 people who were indexed. In addition, we offered a 'free trial' where people could use SmallBlue without signing up and several hundred people did so. We take this rapid adoption as indicative of the need for an expertise locator tool and as something of an endorsement of SmallBlue we had built. Currently, there are about 4~5 million of unique email and chat indices in SmallBlue.

Figure 5. Number of volunteers and total number of people indexed



5.2 Interviews

We conducted semi-structured phone interviews with 11 people who at that point had been using the tools for only a few weeks to get an early assessment of whether we were headed in the right direction. The interviews included questions on general background such as tenure in the company and in their present job as well as their motivation for trying SmallBlue. Most people indicated that they were trying it out of curiosity. But one of the people we interviewed told us that she had just started in a new job. We also asked how often they needed to find new people. 30% said daily, 40% said weekly and another 30% said monthly or less. In other words, finding people is not a daily activity for these people.

A principal goal of the interviews was to assess the perceived accuracy of responses. To get at this, we asked each of our interviewees to enter a query on a topic they were familiar with and provide comments and feedback on what they found along the way. When the results came back, we asked them to tell us how many people they knew on the first page of 10 names and their perception of the quality of the results. We also asked them how likely they were to go to one of the people they didn't know and what they would need that weren't already in the system, to help them contact the person. Users were prompted for feedback on features or comments, in addition to any spontaneous comments. Most people responded that they knew 6 or 7 out of the 10 displayed on the first page which was a good indication of accuracy. When asked to comment on the accuracy most people indicated satisfaction with the results.

After the exercises, users were asked to rate "*their level of satisfaction*" on a 5 point scale (5 high) and to answer how they anticipated using it in the future. The mean ratings are shown in Table 1. At that time of the interviews, the Reach feature had not yet been implemented.

Table 1: Level of satisfaction with the tools where 1 = Very Dissatisfied and 5 = Very Satisfied (N = 11)

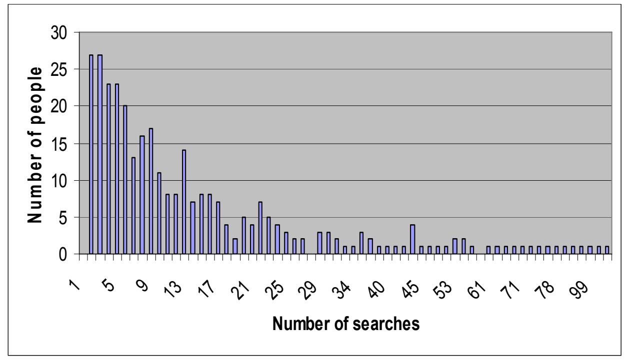
	Mean	S.D
Search	4.10	0.74
Social Networks	3.70	1.06
Overall Accuracy	3.35	0.71
Installation and Documentation	3.27	1.01
Overall	3.93	0.17

Although informal, these ratings provide some additional support for our system. The level of satisfaction with the accuracy of the system was lower than we had hoped. However, during the interviews, several people commented that they could not evaluate the accuracy of the system for people they did not know. They were not disturbed by the number of unknown names that appeared just unable to evaluate them. We also asked each person to indicate their level of comfort with how we had addressed privacy issues. No-one expressed concern. Indeed several people indicated that the importance and value of sharing information and knowledge overrode issues of privacy in this setting.

5.3 Log Data

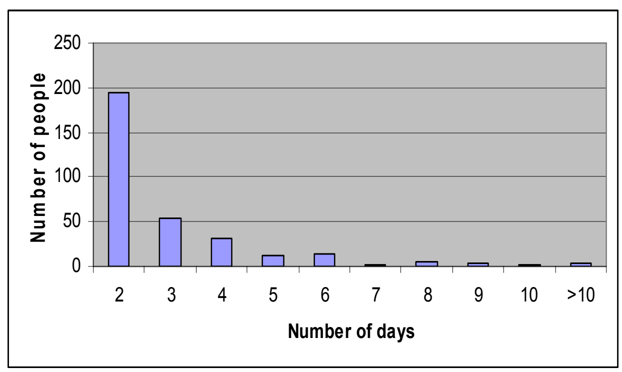
About 3 months after the system had been deployed there were 1255 unique users who had done at least one search. Of these, 323 (26%) people used the system for 2 or more days. These people also conducted a number of searches as shown in Figure 6; 145 (45%) of this group had 10 or more searches. These data point to an interesting usage pattern. Where the majority might just be trying out the system, as is typical with a new technology, the people who used it more than 2 days were conducting multiple searches. We didn't collect sufficient data on the reason for search, but we believe that some people are looking for people to contact but others are just looking to develop awareness, familiarity with others in the organization.

Figure 6 . Frequency of search for people who used SmallBlue 2 or more days



It was also interesting to see whether SmallBlue was used once or over an extended period of time (Figure 7). The majority of people (74%) did use SmallBlue for only one day. But there were still a large number of people who continued to use it over the 3 months in which we collected usage data. We explored the usage patterns further and found that although most of the multi-day people used SmallBlue for consecutive days, there were several instances when there was an elapsed time of 40 days or more from first to most recent use, during which time they would use the system occasionally. These observations suggest that at least for some users, searching for people was an occasional and also unpredictable activity.

Figure 7 . Number of people who used SmallBlue for 2 or more days



The Reach tool was added several months after the system was initially deployed. Nevertheless, over the period of our study, there were 455 people who used it an average of 5.6 times each. One person said, *"The paths are really helpful … you can see who your personal nodes are"*. This person found the network view of all of the paths an especially helpful feature for identifying the key connectors in her extended network. Another user also told us that the way she would use the Reach feature was to say to the person she wanted to contact, *"You don't know me but I found your name in the system where I see that you are a colleague of <person1> who works with <person2> who knows <person3> who I know.*" These users also emphasized the importance of seeing themselves in the social network maps.

One user described the need to see herself on the map as similar to the “You are here” marker on a map in a shopping mall which provides a way for the user to see a path from their current place to a destination. By marking the individual in the social network visualization we are, perhaps, providing a similar service and helping the user see the social path to another person. In general, the social network view was less well understood or used than the other features although some users found it to be an invaluable way to get insight into how knowledge was distributed in the organization. In general, most users find it hard to read and interpret social network diagrams. Some researchers have cautioned against the diagrams arguing that the value of these diagrams has not been well established for most users [27].

5.4 Surveys

We collected data from 42 early adopters who completed an online survey that ran for about 5 weeks in January and February 2007. The main purpose of the survey was to provide us with general feedback on the system and point the way to further improvements. The results were generally encouraging. 74% had used the system more than once and 36%, were using it every couple of weeks. 69% agreed or strongly agreed that they would continue using the system beyond the trial period. One of the key questions was whether people would really use it or only install the software and play around with it. A good indicator of real usage was whether users would actually contact someone since that would mean that they found a person who could be useful and they found a way to reach them. Of the people who responded to the survey 34% indicated that they had contacted at least one person.

On the important question of privacy, only 3% of the respondents reported any level of dissatisfaction with the way the system handled privacy. This was an important finding and provided strong evidence that systems based on email mining can be adopted.

One of the key questions was whether people would really use it or only install the software and play around with it. We saw in the log data that there were a substantial number of people who used the system over several days and even over an extended elapsed time. The results of the survey data provided additional support for real usage. Of the people who responded to the survey 34% indicated that, in the past 3 months, they had contacted at least one new person as a result of using the system.

6. DISCUSSION

The goal in designing SmallBlue was to provide a robust, scalable system that would provide people with visibility and insight into expertise beyond the limits of their own personal networks. We were faced with several challenges – balancing privacy and utility, creating something flexible to lots of different ways of working yet robust enough to accommodate growth, making something that is easy and simple while also having a rich set of features, and, using technology to create a system that mimics and supports the subtle social nuances of human communication. We followed the principle that finding the right expert is necessary but not sufficient for the purpose of engaging the expert in the kind of dialog and discussion necessary to gain the required knowledge.

SmallBlue addressed these challenges by designing a system around the principles of good search, social context and approachability that together define a new approach to expertise location that places emphasis on the social context of use [13]. These principles resulted in a set of search and visualization tools that relied on a combination of public profile information and dynamic private communication data to infer social network relationships and expertise. Results from an early deployment of the system indicated that a large number of people had been using the main features to find people, but also to gain insight and awareness of others in the organization as well as follow through to contact new people.

Any expertise locator system is only as good as the quality and especially, the quantity, of data that goes into it. We were fortunate to have had a large number of people volunteer their communication data soon after SmallBlue was released. This was an important achievement because it validated the willingness of people to contribute their data and it meant that users joining the system would have people they could search.

There have been several approaches to the design of expertise locator systems. Those systems which incorporate social network data (e.g. [16, 19] do so primarily to filter choices so that the user is only presented with experts who are within a few degrees of separation. SmallBlue took a different direction. Instead of filtering the search results, SmallBlue presents an unfiltered list of experts along with information about the degree of separation. In this way, the user can decide how to use the social connection information. SmallBlue extends previous research by providing innovative approaches to privacy and by providing visibility into the distribution of knowledge within established and emerging communities.

6.1 Content and privacy

SmallBlue used stored communication records and basic directory information to extract expertise and relationship data. These data sources are representative of a broad population and readily available but hardly accessible for privacy reasons. Where many expertise systems fall short, however, is in not adequately representing peoples’ interests and activities. We include access to output that users have created using the new social software tools such as blogs, wikis and social tagging. Until we understand how this kind of information contributes to a determination of expertise and social distance, we have provided it “as is” to users.

One of the main challenges of SmallBlue was to reliably extract and reveal content and social network data from communication logs in a private and secure manner. The privacy solution adopted by SmallBlue only extracts data from people who have explicitly opted into the system and takes the unusual approach of only using content that has been authored by the volunteer. It also disguised attribution to any data source through aggregation and inferencing the data. We don’t know whether it was this approach or for another reason, but since the system has been launched we have received no complaints about privacy either directly from our users or in any of the feedback even though there were opportunities to voice concern. There has been a steady rate of people signing up and volunteering their data, coming almost entirely from word of mouth. We don’t believe that users would recommend the system to others if they had a serious concern about privacy.

6.2 Who is the expert?

Although we have talked throughout this paper of "finding experts" in fact, SmallBlue is much more about finding people than identifying the real experts in an organization. Those people who are broadly regarded as experts, might not in fact, be that hard to find. It is the individual who has some small but important knowledge who is harder to locate. SmallBlue helps to identify those people by showing people associated with a search term in a social context as well as in a relevance ranked search list. Moreover, we distinguish between those people "who know what" and the people "who know who" [12]. The latter are very important in an organization because they can act as brokers or gatekeepers who filter out information and people who are not relevant to the group, and, who can protect people who may get overloaded with requests [5].

Sometimes we want to find the person whose expertise is defined by what they know, other times by the person who is central in a network. In many cases this is the same person. But as we used the system and observed the kind of search terms others used, we found several instances where the person who appeared as the most central in the social network, was not the highest ranked in and vice versa. Moreover, we could also see which people were acting as brokers and linking between groups or sub-groups. These people were sometimes in the top 6 ranking but not necessarily. The value of these differences between ranking and network centrality is that there is not just one kind of expert. Depending on the circumstances and the need, users may sometimes want to seek out the person who has the most knowledge, in other cases seek out the person who is best connected in the community of knowledge and in yet other cases reach out to the person who is brokering and connecting people.

6.3 Fostering collaboration and community

As noted by Ackerman et al [4] the goal of sharing expertise has value beyond the immediate need to find a particular person. Awareness of "who knows what" in an organization is key to fostering the collaboration and knowledge sharing that drives growth [10]. It can be especially hard to build awareness of people from other parts of an organization, geography or specialization even though these people provide important knowledge for globally distributed teams [22]. One of our users said it best when she told us that contemporary work environments often feel like being at a party where someone has turned out all the lights. You have a sense that there are other people in the room but you can't see them and so you don't know who to turn to. We hope to see more systems that will illuminate the other people in the room.

Following earlier research [28], we observed a variety of social network patterns that are indicative of emergent community structures. One common pattern was a visualization showing small clusters of people who were largely defined by business unit and mostly disconnected from each other. Another common pattern was an intact community with a very dense core group of people in the middle but from multiple business units. We have not yet undertaken a systematic study of these and other patterns. But they provide fascinating insights into the underlying organizational structure reminiscent of the patterns seen by social network theorists and practitioners [10].

6.4 Limitations

Prior methods of finding people are inadequate especially for the consulting professionals who were a part of our target audience. When they needed to find someone who could help supply information in connection with client work, they would initially turn to someone they knew in their own network. However, they frequently were unable to find the right person either in their network or through a recommendation from their network. In part this reflects the growing diversity of information that is applied to client work but it is also symptomatic of the time it takes to build a broad enough personal network to tap into the right sources. Although any expertise locator cannot address all the queries someone might have, it should fill the gap between the limitation of personal networks and the inadequacy of simple corporate directories. Even though SmallBlue was able to index a large number of people in a relatively short amount of time, one of the most common drawbacks we heard from our early adopters was that there were insufficient people in the system. SmallBlue was deemed insufficient if a search returned too few people from a sought after business unit or if there was no good path to the expert. This is primarily a problem of coverage. One of the challenges of creating an enterprise wide system is that it has to be valuable across diverse needs. This requires that we get participation from a large number of people from each unit and geography. We observed that after one person from a new division or geography signed up, several others from the same unit would be in the system shortly afterward, suggesting that some of the growth in participation occurs through word of mouth. It is an open question whether an expertise locator system, such as SmallBlue, which only requires no effort by users beyond an initial authorization, can continue to grow organically or whether it will require more intentional promotion to reach the next level of critical mass.

6.5 Future Work

There is clearly much more research needed to understand the reasons how and why people use technology to find experts rather than rely solely on their personal network. One of the things that surprised us from the user feedback was the relative infrequency with which people actually contacted someone new as a result of using SmallBlue yet still found the system to be of value. We don't yet know whether an expertise locator system augments personal networks by providing additional information which someone can use to search within their network, or whether it supplements personal networks by providing an alternate route to expertise. We also have much to learn about what people are looking for when they are looking for experts or how the kind of additional information that SmallBlue provides can be of use in that search.

An expertise locator system should be good at identifying the right experts in the enterprise. It should also be good at helping people reach the expert. By addressing both of these requirements directly, we believe SmallBlue represents a new approach to expertise location which we hope will spur further research into the fascinating realm of how we find and engage new people in conversation.

7. REFERENCES

[1] Ackerman, M.S.: Augmenting Organizational Memory: A Field Study of Answer Garden. *ACM Trans. on Information Systems* 16, 3, (1998) 203-224.

[2] Ackerman, M.S., Cranor, L.F., Reagle, J.: Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proc. 1st ACM Conference on Electronic Commerce,* ACM Press New York, NY, USA (1999), 1-8.

[3] Ackerman, M.S., McDonald, D.W.: Answer Garden 2: merging organizational memory with collaborative help. In *Proc. CSCW 1996,* ACM Press New York, NY, USA (1996), 97-105.

[4] Ackerman, M.S., Wulf, V., Pipek, V. (eds.): Sharing Expertise: Beyond Knowledge Management. MIT Press 2002.

[5] Allen, T.J.: *Managing the flow of technology*. MIT Press 1977.

[6] Becerra-Fernandez, I.: Searching for Experts with Expertise-Locator Knowledge Management Systems. *ACL 01 Workshop on Human Language Technology and Knowledge Management* 1, (2001) 9-16.

[7] Borgatti, S.P., Cross, R.: A Relational View of Information Seeking and Learning in Social Networks. *Management Science* 49, 4, (2003) 432-445.

[8] Brown, J.S., Duguid, P.: *The Social Life of Information*. Harvard Business School Press 2000.

[9] Campbell, C.S., Maglio, P.P., Cozzi, A., Dom, B.: Expertise identification using email communications. In *Proc. IKM 2003,* ACM Press (2003),

[10] Cross, R., Parker, A.: *The Hidden Power of Social Networks*. Harvard Business School Press Boston, Mass 2004.

[11] Ehrlich, K.: *Locating expertise: Design issues for an expertise locator system*. In: Ackerman, M.S., Volkmar, P., Wulf, V. (eds.): Beyond Knowledge Management: Sharing Expertise. MIT Press, Cambridge, MA 2003

[12] Ehrlich, K., Cash, D.: The Invisible World of Intermediaries: A Cautionary Tale. *Computer Supported Cooperative Work* 8, 1, (1999) 147-167.

[13] Erickson, T., Kellogg, W.A.: Social translucence: an approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 1, (2000) 59-83.

[14] Farrell, S., Lau, T.: Fringe Contacts: People-Tagging for the Enterprise. In *Proc. Workshop on Collaborative Web Tagging,* (2006),

[15] Hansen, M.: The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly* 44, (1999) 82-111.

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

[17] Lin, C.-Y., Griffiths-Fisher, V., Ehrlich, K., Desforges, C.: SmallBlue: People Mining for Expertise Search and Social Network Analysis. *IEEE Multimedia Magazine* Oct.-Dec, (2007)

[18] Maron, M., Curry, S., Thompson, P.: An inductive search system: Theory, design, and implementation. *IEEE Transactions on Systems, Man and Cybernetics* 16, 1, (1986) 21-28.

[19] McDonald, D.W., Ackerman, M.S.: Expertise recommender: a flexible recommendation system and architecture. In *Proc. CSCW 2000,* ACM Press (2000),

[20] Milgram, S.: The small world problem. Psychology Today, Vol. 2 1967 60-67

[21] Millen, D.R., Feinberg, J., Kerr, B.: Dogear: Social bookmarking in the enterprise. In *Proc. SIGCHI 2006,* ACM Press (2006),

[22] Mockus, A., Herbsleb, J.D.: Expertise browser: a quantitative approach to identifying expertise. *Proceedings of the 24th International Conference on Software Engineering* (2002)

[23] Nardi, B.A., Whittaker, S., Schwarz, H.: It's not what you know, it's who you know: Work in the information age. First Monday 2002

[24] Reichling, T., Schubert, K., Wulf, V.: Matching Human Actors based on their Texts: Design and Evaluation of an Instance of the *ExpertFinding* Framework. In *Proc. Proceedings of the Group05 Conference,* (2005), 61-70.

[25] Song, X., Lin, C.-Y., Tseng, B.L., Sun, M.-T.: Modeling and predicting personal information dissemination behavior. *Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (2005)

[26] Song, X., Tseng, B.L., Lin, C.Y., Sun, M.T.: ExpertiseNet: Relational and Evolutionary Expert Modeling. In *Proc. Intl. Conf. on User Modeling,* Springer (2005), 24-30.

[27] Terveen, L., McDonald, D.W.: Social matching: A framework and research agenda. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 3, (2005) 401-434.

[28] Tyler, J.R., Wilkinson, D.M., Huberman, B.A.: *Email as spectroscopy: automated discovery of community structure within organizations, Communities and technologies*. In: Huysman, M., Wenger, E., Wulf, V. (eds.): Communities and Technologies. Kluwer, BV, Deventer, The Netherlands 2003

[29] Watts, D.J., Strogatz, S.H.: Collective dynamics of'small-world networks. *Nature* 393, 6684, (1998) 409-410.

[30] Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., Hainsworth, J.: ContactMap: Organizing communication in a social desktop. *ACM Trans. Comput.-Hum. Interact.* 11, 4, (2004) 445-471.