Find an Expert: Designing Expert Selection Interfaces for Formal Help-Giving

Sharoda A. Paul
GE Global Research
2623 Camino Ramon, San Ramon, CA.
sharoda@gmail.com

ABSTRACT

A critical aspect of formal help-giving tasks in the enterprise is finding the right expert. We built and evaluated a tool, *Find an Expert*, to examine what the most important expert selection criteria are for help-seekers and how to represent them in expert selection interfaces for formal help-giving tasks. We observed users' expert selection decisions and found that the diversity of topic expertise and the amount of related experience were significantly important in helping users decide which expert to contact. Through self-reported data from users, we found that in addition to expertise and experience, expert accessibility indicators, like online availability and language proficiency, were considered important criteria for selecting experts. Finally, publicly-displayed crowd-sourced ratings of experts, while deemed useful indicators of expert quality by help-seekers, raised concerns for experts. We conclude with suggestions regarding the design of expert selection interfaces for formal help-giving tasks.

Author Keywords

Expert selection; expert search; formal help-giving; design; collaborative troubleshooting.

ACM Classification Keywords

H.5.3. Information Interfaces and Presentation : Group and Organization Interfaces - *Computer-supported cooperative work*

INTRODUCTION

A key aspect of troubleshooting tasks within the enterprise is finding the right expert for a given problem. While expert finding in the enterprise has been broadly examined in the HCI literature, most studies have focused on informal help-giving tasks of knowledge workers, *e.g.*, seeking colleagues for gathering information on a particular topic [24] [30]. In contrast, our goal was to understand expert selection in formal help-giving tasks, such as collaborative troubleshooting of equipment issues in an industrial setting.

CHI'16, May 07-12, 2016, San Jose, CA, USA.

DOI: http://dx.doi.org/10.1145/2858036.2858131

With the proliferation of mobile devices in the enterprise, technicians performing equipment maintenance increasingly engage in real-time collaborative troubleshooting with remote experts [19]. We were interested in understanding how to design expert search tools for collaborative troubleshooting tasks. More broadly, we wanted to examine how to best represent experts in expert selection interfaces for formal help-giving.

We took a systems research approach [4] and created a new tool, called *Find an Expert*, to examine and understand expert selection behavior. *Find an Expert* allows technicians to enter a query describing the problem and returns a list of relevant experts through text mining of case resolution logs. Our goal for creating *Find an Expert* was to examine the nuances of expert selection decisions in formal help-giving tasks. In this paper we present insights gained from evaluating the *Find an Expert* user interface to answer the following research questions with respect to formal help-giving – 1) How do users of expert finding systems weigh expert quality and accessibility factors in making expert selection decisions? and 2) What are the most important expert selection criteria and how to represent them in expert selection interfaces? We answer these questions in the context of collaborative troubleshooting of equipment issues in an industrial organization.

The rest of the paper is organized as follows – we first provide background on expertise seeking in the enterprise and expert finding tools. Then, we describe the expert identification method in *Find an Expert* and then focus on the design rationale and features of the expert selection interface. Next, we present results from evaluating *Find an Expert* with experts and field technicians from a large industrial company. Finally, we discuss our findings and provide implications for designing expert selection interfaces for formal help-giving tasks in the enterprise.

BACKGROUND

Finding experts in the enterprise has long been an area of interest within HCI. McDonald & Ackerman [16] conducted one of the first empirical studies of expertise seeking in the enterprise. Through an extensive field study in a software firm, they found that expert location had two major phases – *expert identification* and *expert selection*. Expert identification is the problem of knowing which individuals in the organization have the required skills to solve a problem. Expert selection is the problem of appropriately choosing among people with the required expertise. We use this framework to review the literature on expert-finding in the enterprise.

Expert indentification

Expert identification in the enterprise is a difficult problem because of the nature of expertise itself, and how people's expertise grows and changes over time [16]. Qualitative studies of expertise seeking have found that both historical data and personal networks [18] are leveraged by expert seekers. In recent years, several technical solutions have been developed to build expert finding systems. Since employees often use historical artifacts to identify experts [16], algorithmic expert identification from employee-generated content has been a popular technique for designing expert-finding tools. Expert identification has been attempted using email communications and chat messages [10] [28], curated user profiles [21], employee-authored work documents, and social network contributions (e.g. SmallBlue [14], Referral Web [13]).

Two primary approaches have been proposed [1] in the information retrieval literature for expert identification from employee-authored documents – 1) *candidate-based approach*, which builds a profile for a candidate expert based on the relevance of the information generated by them to the user's query and 2) *document-based approach*, which retrieves all documents relevant to a user's query and then finds the link between documents and the authors of the documents. Recently, these information retrieval approaches have been augmented with text mining techniques, such as topic modeling of employee-authored documents. Topic modeling approaches address the drawbacks of the profile-based and document-based approaches proposed by Balog et. al [1]. The profile-based approach underperforms the document-based approach due to two reasons. First, noisy documents may get added to the candidate profile [1] which, though associated with the candidate, do not represent the candidate well. Second, when supporting documents do match the candidate well, they often do not match the topic (i.e., the whole query concept) that the user is interested in. While the document-based approach addresses the first drawback, it only relies on query terms matching documents and does not consider any semantic concepts underlying queries and documents. [17]. Momtazi and Naumann [17] show that topic modeling based on Latent Dirichlet Allocation (LDA) outperformed several profile and document-based approaches for expert finding. We built on the LDA-based topic modeling approach to determine topic expertise of experts, as described in detail in the next section under Expert Identification.

Expert selection

Understanding expert selection criteria is important for designing effective expert finding tools. Studies [29] have found that making expert selection criteria visible in expert recommendation interfaces can significantly improve usability and user satisfaction. An important aspect of expert selection is balancing source quality vs. accessibility. A review [12] of 72 studies of expertise seeking in various domains found that the results were mixed when it came to whether quality or accessibility has a larger impact on source selection. Hence, further research is needed to gain a nuanced understanding of which characteristics of an expert are key to expert selection decisions, and how those characteristics should be represented in expert selection interfaces. Furthermore, studies of expert selection have typically focused on informal help-giving. For instance, Shami et. al [24] examined how enterprise knowledge workers weigh online profiles generate by the IBM SmallBlue system to infer the suitability of an expert for informal help-giving. They found that participation in social software, social connection information, and self-described expertise in the corporate directory were the most important signals used by help-seekers. Yarosh et al. [29] conducted a follow-up study in which they enhanced the IBM SmallBlue system to provide additional information about experts. They found that over a wide range of informal help-seeking tasks the top five most useful pieces of information about an expert were - company division, expertise summary, job responsibilities, job title, and geographic location. Similarly, a diary study [30] of knowledge workers seeking informal help in the same organization revealed that group, expertise, role, location, and experience, were important helper-selection criterial; however these characteristics of experts were often not adequately represented by help-giving tools. The existing research shows that a variety of factors about experts are used in expert selection decisions and that different expert selection criteria are prioritized depending on the task and user context.

We were interested in examining whether the findings from past studies would generalize to the domain of formal help-giving. Formal help-giving is different from informally assisting colleagues in the enterprise. When experts' job is to provide troubleshooting help, help-giving interactions are often documented, and experts' reputation is tied to performance evaluations. Thus, the motivations, processes, and outcomes of formal help-giving differ from those of informal help-giving and few studies have examined these difference in terms of designing expert finding tools. The existing studies of informal help-giving in the enterprise provide useful jumping-off points, and we drew on them to iteratively design and evaluate the expert selection interface for *Find an Expert*.

FIND AN EXPERT

Find an Expert was designed for technicians who perform maintenance and repair of industrial equipment at a large, global industrial organization. Today, when faced with technical issues in the field, technicians seek help from product experts by filing a case in a case management tool. Experts' formal role is to provide technical help, and they are measured on the quality and timeliness of help provided for resolving cases. Designated experts are manually assigned to relevant cases and provide case resolution through the case management system. Experts draw on their technical expertise and their engineering knowledge, along with enterprise documentation (e.g., instruction manuals, drawings etc.) to provide help. More details on the environment and culture of help-giving can be found in Paul and Bolinger [19] .

With the proliferation of mobile devices, the organization wanted to augment the case resolution system with tools for real-time collaborative troubleshooting which would allow technicians to directly contact experts for problem solving [19]. The first step of real-time collaborative troubleshooting was enabling help-seekers to identify the right experts to contact for a given problem encountered in the field.

Find an Expert allows technicians to find experts relevant to a given problem by leveraging existing case resolution logs for expert identification. Technicians type in keywords describing the problem they need expertise on. In response, the tool presents a ranked list of relevant experts, along with information about the quality and accessibility of experts.

Next, we briefly describe the expert identification technique and then focus on the expert selection interface.

Expert indentification

Given a query describing a problem encountered in the field, *Find an Expert* returns the experts who have topic expertise in that query. For this, we determine how likely is candidate ca an expert on the input query q, denoted by $P(ca|q)$. Per Bayes Theorem, we can calculate this as –

$$P(ca|q) = \frac{P(q|ca)P(ca)}{P(q)} \quad (1)$$

where $P(ca|q)$ is the probability of candidate ca generating the query q; $P(ca)$ is the prior probability of candidate ca, and $P(q)$ is the probability of query q. Considering that $P(q)$ is constant and $P(ca)$ is a uniform distribution over all candidates, $P(ca|q)$ depends primarily on $P(q|ca)$.

To calculate $P(q|ca)$, we use topics as hidden variables to model the relationship between user queries and experts. Latent Dirichlet Allocation (LDA) [5] is a popular topic modeling technique that models documents as a mixture of topics and represents each topic as a probability distribution over words. We use LDA to extract topics from the corpus and then use these topics to model the relationship between candidates and words in the query, similar to Momtazi and Naumann [17]. However, Momtazi and Naumann [17] assume that a given query maps to a single topic. In contrast, we assume that problems occurring in the field, as expressed by queries, are mixtures of topics. Hence, we calculate the probability of candidate ca generating the query q as –

$$P(q|ca) = \prod_{w \in q} \left(\sum_{t \in T} P(w|t)P(t|ca) \right) \quad (2)$$

where $P(w|t)$ is the probability of word w belonging to topic t; $P(t|ca)$ is the probability of expert ca generating topic t.

Expert ranking

The expert identification process consists of the following steps (Figure 1) – 1) Perform LDA on the case log corpus to extract latent topics, 2) Create an expert-word matrix that contains the probabilities $P(w|ca)$ for all experts ca and all words w in the vocabulary, using topics inferred in step 1. 3) When a query comes in, calculate the probability $P(q|ca)$ for all experts per equation 2, and 4) Rank experts based on $P(q|ca)$ and present the top ranked experts to the user.

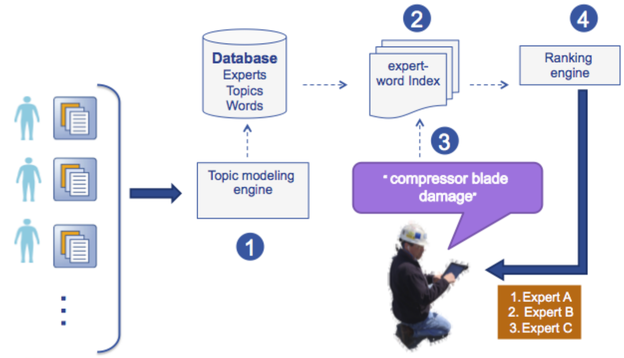

Figure 1. Find an Expert architecture

Steps 1 and 2 are performed offline and the resulting expert-word probabilities are stored in a database. When a technician types a query, the probability of each expert generating the query is calculated in real-time and all experts are ranked by this probability. The tool currently displays the top 5 experts for the given query, with an option to view more experts if needed.

We used a corpus of 22,431 randomly-selected cases pertaining to power plant equipment maintenance, submitted by 2,386 technicians and assigned to 301 experts. Each case contained the case submitter name (technician), assignee name (expert), description of problem (entered by technician), and comments (resolution provided by expert), along with other meta-data about the case. Each case was tagged by the technicians with "group", "section", and "component" keywords pertaining to the parts of field equipment that the problems were related to. The group, section, and component tags were selected from drop-down lists.

Expert selection interface

The *Find an Expert* interface (Figure 2) allows technicians to enter a query and presents a ranked list of experts. Technicians can review the presented experts and select experts to contact using mobile collaboration tools. We drew on past research on expert selection interfaces as well as iterative design and evaluation with product experts and technicians to design the user interface. Below, we outline the different aspects of the interface and the design rationale behind including them. We also highlight the design alternatives tried and the design lessons learnt during iterative prototyping.

Prior work points to two characteristics of experts that are important for expert seekers - quality and accessibility. Hertzum [12] found that expert selection decisions are not governed by a principle of least effort (i.e. selecting the most easily accessible source); instead, expertise seekers aim to strike a balance between quality and accessibility. To examine how enterprise help-seekers weighed quality vs. accessibility of experts, we included indicators of both in the *Find an Expert* interface.

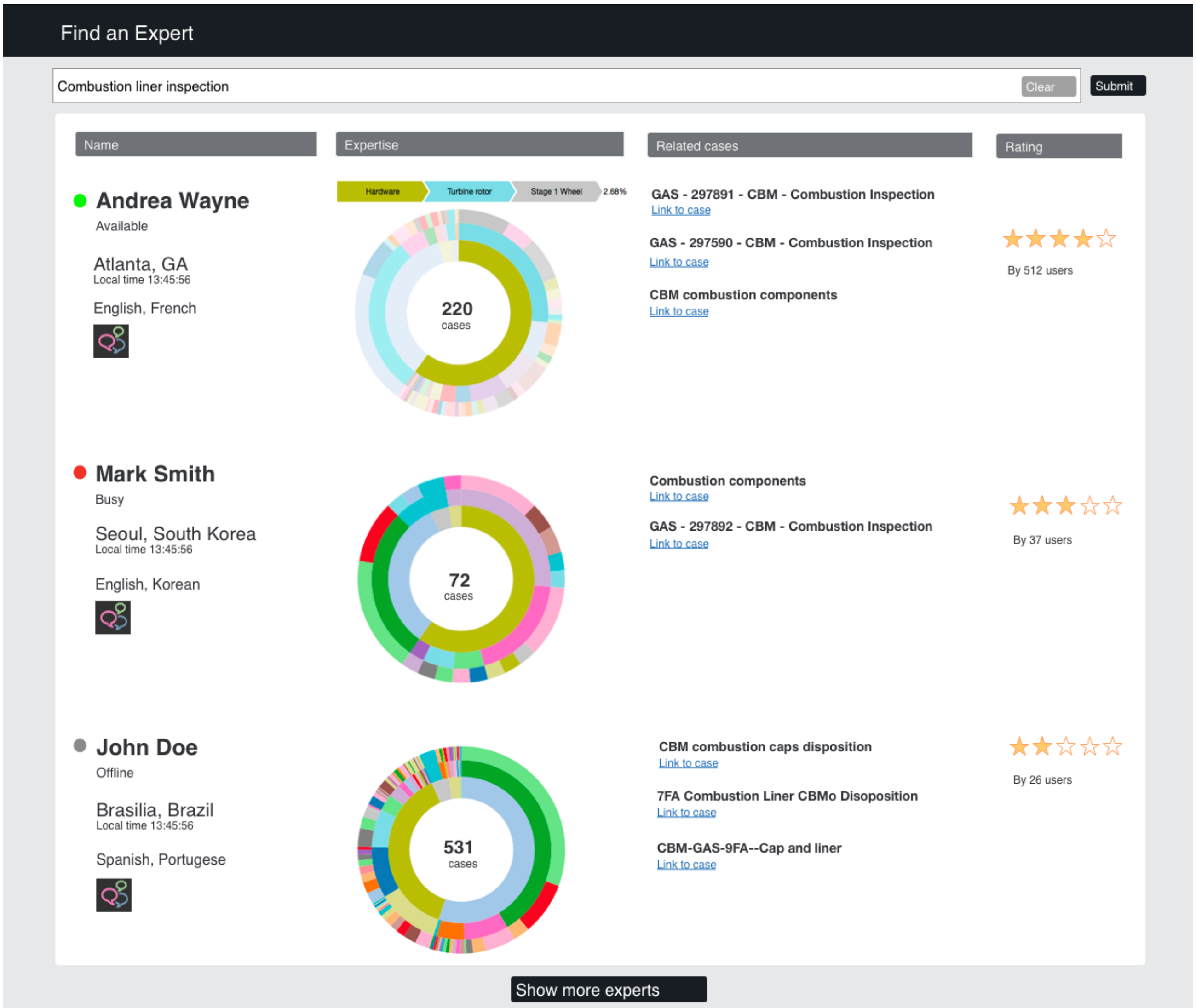

Figure 2. Find an Expert interface. The number of experts displayed for each query is configurable; here we show 3 experts per query.

Expert quality

When people are used as expertise sources, source quality has two important components - reliability of the information and relevance to the information need [12]. The reliability of the information provided by an expert is often judged by referring to historical artifacts created by those experts. For instance, software engineers use work artifacts such as records of change history of software logs to identify possible experts [16].

To indicate reliability of the expert, we display the expert's query-independent expertise based on all the cases they've resolved in the past. A key question here is how to visually represent expertise in the expert selection interface. Most expert-finding systems based on textual analysis of expert-generated data show textual representations of expertise, such as tag-clouds [24] or wordclouds [7]. Wordclouds have been used to represent text-based models of users due to their ability to support impression formation [22]. Thus, in our early prototypes, we depicted topic expertise by displaying a word-cloud for each expert which was generated using their top 50 words by LDA score [7]. However, preliminary feedback from users revealed several drawbacks of word clouds in depicting expertise. The wordclouds did not have good discriminative power for helping technicians compare similar experts. Also, the wordclouds did not represent the expert in terms of the mental model of the technicians. Thus, our preliminary findings reflected some of the drawbacks of wordclouds in representing text-based models identified in prior work [9].

We found that technicians used the domain taxonomy to describe their own expertise (e.g. "steam field engineer" or "control systems expert") and to formulate queries for expert search (e.g. "gas turbine bucket damage"). This hierarchical taxonomy contains group (e.g. hardware), section (e.g. rotors), and component (e.g. wheel) of the technology that they are experts in. Cases in the case resolution system were also tagged with keywords from this taxonomy. Hence, we decided to use these keywords as *units of analysis* [9] that mapped to the expertise mental models of technicians and experts. We replaced the word clouds by interactive sunbursts [6]. The sunburst supports visualization and exploration of hierarchical data [26] and has been used for exploring topic-based models of textual data [25]. In *Find an Expert*, the sunburst of an expert visually summarizes his cases by group, section and component. Hovering over a section of the sunburst displays a bread-crumb trail above the sunburst showing the percentage of the experts' cases that pertained to that group, section, or component (Figure 2). Preliminary evaluations suggested that both technicians and experts preferred the interactive sunbursts to the word clouds.

In addition to overall expertise, we added a rating feature to indicate and capture the reliability of an expert. This feature allows expert-seekers to rate an expert on a five-star scale. Past research found that it would be important to include users' self-perception or mutual ratings of expertise in expertise profiles [21]. There have been some studies of crowdsourced reputation systems for consumer websites like StackOverflow [15] and Quora [20], which allow users to rate the quality of content produced by other users through social voting and gamification mechanisms. However, there have been few studies of enterprise or professional crowdsourced reputation systems. The closest is LinkedIn's Skills Endorsement feature, which allows users of the professional social network to tag themselves with topics representing their areas of expertise and for their connections to provide "social proof" via an "endorse" action [3]. However, how users perceive and use this feature has not been examined in detail. In contrast to Skills Endorsement, *Find an Expert* allows user's to rate the overall quality of an expert as opposed to rating them on specific skills. We were interested in examining how users would react to crowdsourced ratings as a reputation mechanism and whether these ratings would be perceived as useful indicators of expert quality.

Finally, to indicate how relevant the expert is to the information need, we display query-specific experience of experts by listing up-to three most relevant cases resolved by each expert generated by the topic modeling algorithm. Users can follow the link to the case in the case resolution system where they can read the details of the case. Hence, we combine query-dependent and query-independent expertise indicators to provide as much information as possible about experts' quality.

Expert accessibility

Expert accessibility has been found to be an important expert selection criteria for help-seekers during informal help-seeking tasks. Shami et al. [24] found that expert seekers use the level of participation of employees on an enterprise social network as a way of gauging willingness to help. They also found that help-seekers judge accessibility based on how much profile information employees have provided in internal social network profiles. Other studies [12] have found that having a social tie with an expert or having prior experience working with them increases perceived accessibility. Drawing on past research [8] [30], we display several indicators of an expert's accessibility.

We provide a link to the expert's enterprise social network profile where users can view their org chart, role, group affiliations, and social network connections and contributions. This link is represented by an icon of the company's internal social network and is displayed along with other availability information about the expert. Our prior work [19] and discussions with technicians suggested that availability is a crucial aspect of formal troubleshooting, especially in a global organization where experts are geographically dispersed and in different timezones. Thus, the *Find an Expert* interface displays experts' local time, location, and languages spoken to indicate their potential availability and accessibility.

In early versions of the prototype we showed an expert's "busyness" level as an accessibility indicator. We used their current caseload, in terms of number of open cases in the case resolution system, as a proxy for busyness level. However, experts pointed out that number of cases was not always an accurate indication of busyness level. Experts with a large number of cases may not be actively working on them, while a few complex cases might keep an expert busy. Thus, experts suggested allowing user-entered busyness level in the form of an online status rather than automatically inferring busyness level. In preliminary evaluations, technicians mentioned that instead of caseload, they would use the online presence information to decide whether to contact an expert. One technician remarked,

> "If you appear online and available, I will call you if you're the best expert even if you have a large number of cases."

Hence, instead of displaying a system-generated busyness-level, *Find an Expert* allows experts to set their online status to Available, Busy or Offline. This status could also be derived from enterprise presence systems.

EVALUATION

We evaluated *Find an Expert* to examine how the design of the interface supported expert selection tasks. We wanted to learn how technicians would weigh expert quality vs. accessibility factors in making expert selection decisions and what was the best way to represent different aspects of experts. We conducted evaluations with two user groups - experts and technicians. Interviews with experts were aimed at getting feedback on whether *Find an Expert* represents experts adequately. Interviews with technicians were focused on answering the research questions, and hence in this paper we primarily present findings from technicians, adding findings from experts as appropriate.

Participants

We conducted 60-minute interviews with 5 product experts (all male) and 14 technicians (1 female). The gender ratio was representative of the technician population at our organization. All participants were selected randomly from among employees interested in evaluating advanced technology products. The age range of participants was 27–65 years. Experts had 12–37 years of experience resolving cases while technicians' had 5–40 years of field experience. All participants were located in the United States at the time of the study and were English-speaking. All experts used the case resolution system to resolve cases as their primary job responsibility. All technicians used the case resolution system to file cases, though their frequency of use varied. 35% of technicians used the case resolution system a few times a month, while an equal number used it few times a week (28%) and a few times a year (28%).

Methods

In the interviews with experts, we demonstrated the tool using five queries generated by each expert and asked them how well the interface represented their ability to resolve technicians' problems. Interviews with technicians were more detailed and consisted of three parts. In part one, we introduced and demonstrated *Find an Expert*. In part two, we asked each participant to try out the tool using five queries describing recent cases they had filed. As participants reviewed the results of each query, we asked them to select upto two experts who they felt they could contact for the given problem based on the information presented. As participants explored the tool, we observed their usage, and for each expert selection decision, we asked them why they had selected the particular expert for the given query. Finally, we asked each participant to take a survey that asked them to rate how useful they found different pieces of information presented about experts and why.

The evaluation was conducted with a prototype version of *Find an Expert* as the tool had not been rolled out across the organization yet. Hence, we randomly generated the expert ratings. Also, we restricted the interface to show only 5 experts for each query; the 'Show more experts' feature was disabled to ensure uniform conditions for the user study. Observational data were logged using the tool and researchers noted answers to questions asked during part two of the study. Qualitative data were analyzed using grounded a theory approach [27] while quantitative data were analyzed using the statistical software R. Next, we present the findings of the evaluation.

FINDINGS

Overall, both experts and technicians were enthusiastic about *Find an Expert* and agreed that it would be useful, especially for novice technicians who were not familiar with the expert pool. Experts verified that *Find an Expert* shows the most important aspects about them for technicians to make expert selection decisions.

Next, we examine the findings from technicians' evaluation with respect to observed tool usage and self-reported data.

Tool usage

Technicians' queries ranged from 1 word to 8 words. Examples of queries were "steam turbine journal taper", "hydrogen seal casing", and "1st stage shroud block tuning pins". Though we conducted the evaluation with 14 participants, tool usage data was logged only for 9 participants due to a technical glitch with data logging. We analyzed a total of 46 queries from 9 technicians for the regression modeling (8 technicians executed 5 queries each and one technician executed 6 queries). This resulted in 230 (i.e., 46x5) expert selection decisions.

We first used observed data to understand which factors affected whether an expert was selected or not. For each query executed by technicians, they were asked to name their top 2 choices of experts to contact out of the 5 experts presented by the tool. To examine which factors affected whether an expert was selected or not, we conducted multiple logistic regression with the binary dependent variable being whether an expert was selected (1) or not (0). The independent variables were the characteristics of an expert displayed by the *Find an Expert* interface, namely - online status, location and local time, languages spoken, total number of cases, number of related cases, number of stars in the rating, and number of users who had rated the expert.

Expert's location was coded as 1 (same country) or 0 (different country) based on the technician's current location. We did not separately include timezone since location and timezone are related and technicians were concerned about whether experts were on shift or not, as opposed to the exact timezone they were in. All technicians were native English-speakers; hence, we coded an expert's language as 1 (speaks English) or 0 (doesn't speak English). We wanted to examine whether technicians preferred experts with diverse expertise, depicted by a more 'busy' sunburst, or whether they preferred experts with a narrowly-focused expertise. We used the number of sections in the sunburst of an expert (mean = 63.3, median = 50.5) as a proxy for expertise diversity, and included it as an independent variable.

There was sufficient variability between the five experts displayed for each query. Across all queries, 61% of the time locations of the experts' displayed were both local and remote. For each query, 35% of the experts displayed were 'available', 32% were 'busy', and 33% were 'offline', on average. For a given query, average range of expertise diversity was 115; thus, there was a wide range of expertise diversity among experts.

Each technician was asked to select up to two experts for each query. Since we received more than one response from each participant, the observations were not independent. We used a multilevel logistic regression model with random intercept for subjects [11] using R's *glme* function. However, the results showed that the variance and standard deviation of the random effect was 0, which suggested that there was no participant-level variation, and all the observed variation in the data was among-observation variance. Hence, we revised our model to a generalized linear model using R's *glm* function. The results are displayed in Table 1.

Diversity of expertise and related cases were the only factors about experts that significantly affected whether they were selected for a query. For each point increase in the diversity of overall expertise (as measured by the sections in the sunburst), likelihood of being selected as an expert increases by 0.008 ($p < 0.05$). Thus, experts with more diverse expertise were more likely to be selected, though the strength of the effect was low. In general, we observed that all technicians interacted with the sunburst to explore the expertise of experts presented to them. We observed participants using the sunburst to get an overall impression of experts' exper-

Parameters	*Estimate*	*Standard Error*
Intercept	-2.133^{**}	0.674
Related cases	0.515^{***}	0.152
Expertise diversity	0.008^{*}	0.004
status Busy	0.331	0.363
status Offline	-0.467	0.381
Location	-0.21	0.406
Languages	0.311	0.418
Total cases	0	0.001
Rating stars	0.023	0.11
Rating users	0.005	0.003

Table 1. Regression results for observed data. Note $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. $n = 9$ **participants**

tise and to compare the experts presented. For instance, some technicians remembered the colors that represented the areas of expertise pertaining to their query and then compared the size of sections with that color across different sunbursts.

Number of related cases was also significant in predicting whether an expert was selected. For every unit increase in related cases, probability of being selected as an expert increased by 0.515 ($p < 0.001$). We observed that technicians often quickly scanned the list of presented experts by interacting with the sunbursts and then looked more closely at the related cases solved by the experts deemed relevant.

Online status, languages, location, total cases resolved, the number of stars in the rating, and number of users providing the rating were not found to be significant in predicting whether an expert would be selected among the top two experts.

Self-reported data

The post-task survey was taken by all 14 technicians. We asked participants to rate each piece of information in the *Find an Expert* interface in terms of how useful it was for making expert selection decisions. Ratings were on a scale of 1 (not useful) to 5 (very useful). Figure 3 shows the average rating for each item.

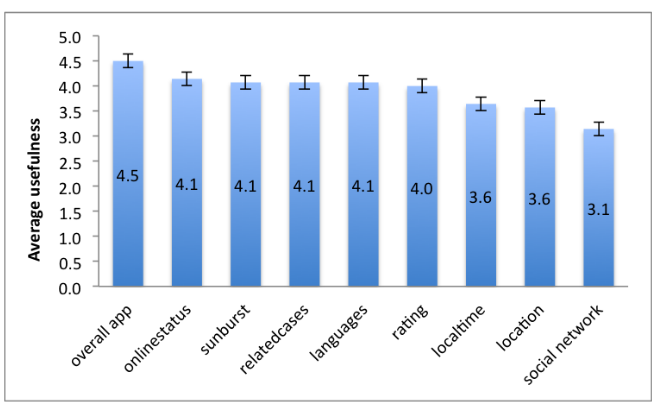

Figure 3. Survey data from evaluation of Find an Expert. Responses were on a 5-point scale with 1 = Not useful, 3 = Neutral, and 5 = very useful. Bars show mean rating with standard error. $n = 14$ **participants.**

Overall, the app was rated high on usefulness (average score = 4.5). The top 5 most useful indicators of expertise were the sunburst visualization of expertise, related cases, crowdsourced rating, online presence, and languages spoken by the expert. All of these received an average score of 4.0 or above.

Validating the findings from observed data, the sunburst and related cases were deemed useful indicators of expert quality. The sunburst visualization was rated useful (score ≥ 4.0) by 79 % participants. Participants reported finding the sunburst an effective way to visually explore the overall expertise of an expert. As one technician explained,

> "[The sunburst] was an easy way to sort out and navigate the expertise level without text. I really liked the graphical interface."

Another participant mentioned that he enjoyed the interactivity of the sunburst in exploring the history of cases resolved,

> "The sunburst was very useful as it showed me [the expert's] range of expertise and I could explore the total cases and percentage of cases in each category."

The remaining technicians (21%) rated the sunburst as "neutral". They felt that the sunburst provided broad areas of expertise while they were more interested in query-specific expertise. One technician suggested that we highlight the sections of the sunburst relevant to their query.

Related cases received a score ≥ 4.0 by 79% users. One technician said,

> "The related cases was very useful as it gave me an in-depth look at what [the experts] have worked on. I would use it as the first criteria to determine who to call."

Another participant mentioned that related case would help him explore further the topics and related problems in the case resolution system before deciding whether he needed to call an expert. He said,

> "I would use it to research similar cases and contact those who have resolved related cases."

Importantly, participants found the combination of query-independent and query-dependent expertise to be very powerful in making expert selection decisions. One participant said,

> "The visualization with the actual cases pairs well together. The visualization can be vague to a degree whereas the case titles are very specific."

Crowdsourced ratings as an indicator of expert quality was rated useful (average score = 4.0). 79% technicians rated the crowdsourced rating useful (score $\geq$ 4.0) for evaluating experts. Though technicians in early prototyping suggested more contextualized ratings and the ability to add comments, the 5-star scale was well-received in the final evaluation.

It was important to gauge the reactions of both experts and technicians with respect to the ratings. We found that experts and technicians responded differently to the ratings feature. All experts were comfortable with receiving ratings but raised concerns about the potential effects of received ratings on reputation and performance review. Hence, some experts suggested making the ratings privately available instead of being publicly displayed. Experts were also concerned about potential unfairness of reviews and had several suggestions for mitigating this, such as rating the quality of the solution provided rather than the expert, allowing ratings by only those technicians who had implemented solutions, allowing experts to respond to ratings, and allowing "moderators" to moderate unfair ratings and comments.

Most technicians were comfortable with providing ratings; however, some technicians mentioned that they would rather provide feedback privately to managers. Technicians felt that ratings would give an idea about the quality of the answer that could be expected from an expert. One technician, who has also lately been providing assistance with resolving cases, said,

> "I like [the rating feature]. I understand that it can be skewed but I think it's useful because people will be fairly honest. I can say this based on the ratings I get on my cases. Field engineers don't give a 1 or 2 unless I've truly dropped the ball."

Technicians who rated the ratings lower than 4.0 were concerned about known drawbacks of crowdsourced ratings, such as preferential attachment [2], which might result in well-rated experts getting overwhelmed by requests. Combining ratings with a queuing and load-balancing algorithm could help balance requests across experts.

With respect to accessibility features, expert's languages and their online presence received the highest ratings (average score ≥ 4.0). Having a common language was important for communication, however, technicians felt that using language proficiency of experts as a selection criterion may be more important for technicians who did not speak English. Online status was important to most technicians as this indicated the availability of the expert to solve their problem right away.

Location and local time were rated lower (average score = 3.6). Validating our prior findings during early evaluations, technicians preferred online status to location and local time. As one technician said,

> "If they are online, it doesn't matter where they are from."

It was more important to know whether they were on shift and how much longer they would be working, so technicians could anticipate how soon they would receive a resolution, and follow-up advice, if necessary.

Interestingly, technicians rated the link to the expert's social network profile the lowest among all factors (average score = 3.1). Only 14% of technicians rated the social network link as "very useful". The rest of the technicians did not feel the need to examine the role, organizational structure, or social network contributions of experts. In their explanation of their ratings, 57% participants said that social network information was not useful or relevant. This is in direct contrast to findings from studies of informal help-giving that found that social network contributions and richness of profile on social networks as the most important factors affecting expert selection decisions [24]. Technicians deemed the social network profile and contributions of experts to be largely irrelevant to their ability to resolve technical issues.

Finally, we asked technicians if there was additional information they would've liked to see about experts in helping them make expert selection decisions. The only request was to provide a direct link within the tool to contact an expert.

DISCUSSION & DESIGN IMPLICATIONS

Quality vs. accessibility of experts

Both observed and self-reported data from the evaluation show that the most important expert selection criterion for formal expert-finding was the quality of the expert. The most important quality factors were overall expertise (as represented by the interactive sunbursts), evidence of related experience (as represented by related cases), and crowdsourced judgements of expert quality (as represented by ratings). The combination of overall expertise and evidence of past experience related to the query was considered a powerful and effective way to convey experts' expertise in the *Find an Expert* interface.

Online presence and expert's language skills were rated the most useful accessibility criteria for formal help-giving. An interesting finding was that though these accessibility factors were rated as useful, they were not found to significantly impact expert selection decisions in observed data. We think the reason for language skills not showing up significantly during observations was that all technicians were English-speaking and 80% of experts displayed by *Find an Expert* spoke English. Further study is needed to validate the language-related results for a different mix of English and foreign-language users. The results indicate that while accessibility of experts is important, expertise is the primary criterion for selecting an expert.

While research on informal help-giving found accessibility indicators such as social connectedness and social network participation important for help-seekers, our results were contradictory. Technicians enjoyed working with known experts; but they did not feel the need for social connectedness information in the *Find an Expert* interface. Though past research [30] found group and role information useful, only 21% technicians wanted to view the org chart. Finally, technicians did not feel the need to view experts' social network profile and contributions. Overall, these findings indicate that expert quality, online availability, and communication skills are more important to represent in expert selection interfaces for formal help-giving, than social connectedness and approachability of experts.

Some of these findings might seem intuitive given the formal roles and incentives of enterprise troubleshooting. However, our study empirically validates the expert selection criteria for formal help-giving and contributes to a nuanced understanding of how different expert selection criteria are prioritized. Such an understanding can lead us to design better expert search and recommendation interfaces for the enterprise.

Representing quality and accessibility in user interfaces

Our evaluation uncovered some valuable findings regarding the representation of human expertise as derived from textual data models. We found that visualizations that map to domain taxonomy were preferred to word-based representations of expertise. Further, our study validates that the sunburst is an intuitive visualization that allows quick navigation, impression formation, and comparison of experts. However, some improvements can be made to our implementation of the sunburst, as suggested by participants. One suggestion was to replace the percentages of cases, in the bread-crumb trail displayed when a user hovered over a section, with absolute numbers, as the latter would be easier to interpret. When an expert was very diverse, the sunburst had too many colors which made it hard for participants to search and recall areas of expertise visually. Hence, the color-scheme needs to be carefully designed for scalability and recall. Also, busier sunbursts led to smaller sections which were harder to explore by hovering over. To overcome this, designers could create thresholds for elimination of sections. Finally, the hover interaction may need to be re-designed for touch screens. We implemented one version [23] of the interactive sunburst; future work should address how this implementation can be improved to make the sunburst more usable for representing human expertise.

In terms of representing related evidence, links to details of cases are useful as they allow expert-seekers to dig deeper into related material and to validate the expertise of the suggested expert. Finally, simple levels of user-defined availability are preferred to automatically-inferred busyness levels.

The findings about availability and approachability indicate that formal and informal help-giving systems have different design considerations. For informal help-giving [24] users need to gauge the willingness and approachability of potential help-givers, along with expertise. In formal help-giving systems, expertise, experience, and availability are more important considerations that sociableness or approachability of experts. Similarly, in a formal help-giving system, users might contact the most knowledgeable experts irrespective of their busyness level or location, whereas employees seeking informal help might consider how busy an expert was or whether it was appropriate to contact them.

Finally, rating people on their expertise is a delicate issue for formal help-giving. When experts are in a formal help-giving role, crowdsourced ratings have the potential to impact performance review. However, our study showed that help-seekers are enthusiastic about crowdsourced reputation systems for experts. Thus, a thoughtfully designed rating feature is needed that incorporates contextual information and moderation to make the feedback meaningful for experts and trustworthy for help-seekers. Further study is needed to examine the tensions between private and public ratings and the effects of crowd-sourced ratings on formal reputation-building and performance review mechanisms.

While our study focused on a particular source of textual evidence for inferring expertise, we believe our findings are generalizable to enterprise expert-finding systems based on other kinds of textual data such as expert-authored documents, emails, and technical reports etc.

One limitation of our study is that we used an artificial scenario as opposed to observing real world usage of *Find an Expert* during collaborative troubleshooting tasks. However, this scenario-based evaluation methodology, used by similar studies [24], allowed us to tease apart the nuances of expert selection decisions. In future work, it will be interesting to validate our findings using real-world usage data from *Find an Expert*. Additionally, future work could systematically evaluate whether variations in the placement (e.g., left vs. center) and presentation (e.g. text vs. visual representation) of a given piece of information affect users' prioritization of that information in expert selection decisions.

CONCLUSION

Few studies in HCI have examined how expert-seekers evaluate experts for formal help-giving tasks. Through the design and evaluation of *Find an Expert* we provide insights about which aspects of an expert's quality and accessibility are important for help-seekers in collaborative troubleshooting tasks. Our work informs the design of expert selection interfaces by highlighting 1) how help-seekers weight quality and accessibility aspects of experts in expert selections decisions, and 2) the design considerations for expert selection interfaces for formal help-giving in the enterprise.

ACKNOWLEDGEMENTS

We are grateful to Chris Cava for his help in accessing case data. Thanks to Arnie Lund for his support, and to study participants for their time. Finally, thanks to Abhishek Srivastav for providing encouragement, review, and critical feedback.

REFERENCES

1. Krisztian Balog, Leif Azzopardi, and Maarten De Rijke. 2006. Formal models for expert finding in enterprise corpora. In *Proceedings of ACM SIGIR*. 43–50. http://dl.acm.org/citation.cfm?id=1148181
2. Albert-Lszl Barabsi and Rka Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286, 5439 (1999), 509–512. DOI: http://dx.doi.org/10.1126/science.286.5439.509
3. Mathieu Bastian, Matthew Hayes, William Vaughan, Sam Shah, Peter Skomoroch, Hyungjin Kim, Sal Uryasev, and Christopher Lloyd. 2014. LinkedIn Skills: Large-scale Topic Extraction and Inference. In *Proceedings of the ACM Conference on Recommender Systems (RecSys '14)*. ACM, New York, NY, USA, 1–8. DOI: http://dx.doi.org/10.1145/2645710.2645729
4. Michael S. Bernstein, Mark S. Ackerman, Ed H. Chi, and Robert C. Miller. 2011. The Trouble with Social Computing Systems Research. In *Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 389–398. DOI: http://dx.doi.org/10.1145/1979742.1979618
5. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *the Journal of machine*

Learning research 3 (2003), 993–1022. http://dl.acm.org/citation.cfm?id=944937

6. Mike Bostock. 2012. Sunburst Partition. Web. (13 November 2012). Retrieved September 15, 2015 from http://bl.ocks.org/mbostock/4063423.
7. Kevin Robert Canini, Bongwon Suh, and Peter L. Pirolli. 2011. Finding credible information sources in social networks based on content and social structure. In *Proceedings of the IEEE Third International Conference on Social Computing (SocialCom)*. 1–8. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6113088
8. Munmun De Choudhury, Hari Sundaram, Ajita John, and Dore Duncan Seligmann. 2007. Context aware routing of enterprise user communications. In *IEEE PerCom Workshops*. 39–44. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4144790
9. Jason Chuang, Daniel Ramage, Christopher Manning, and Jeffrey Heer. 2012. Interpretation and Trust: Designing Model-driven Visualizations for Text Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 443–452. DOI: http://dx.doi.org/10.1145/2207676.2207738
10. Leonard N. Foner. 1997. Yenta: A Multi-agent, Referral-based Matchmaking System. In *Proceedings of the First International Conference on Autonomous Agents (AGENTS '97)*. ACM, New York, NY, USA, 301–307. DOI: http://dx.doi.org/10.1145/267658.267732
11. John Fox and Sanford Weisberg. 2015. Mixed-Effects Models in R, An Appendix to An R Companion to Applied Regression, Second Edition. Web. (23 January 2015). Retrieved September 15, 2015 from http://socserv.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Mixed-Models.pdf.
12. Morten Hertzum. 2014. Expertise Seeking: A Review. *Information Processing and Management* 50, 5 (Sept. 2014), 775–795. DOI: http://dx.doi.org/10.1016/j.ipm.2014.04.003
13. Henry Kautz, Bart Selman, and Mehul Shah. 1997. Referral Web: Combining Social Networks and Collaborative Filtering. *Commun. ACM* 40, 3 (March 1997), 63–65. DOI: http://dx.doi.org/10.1145/245108.245123
14. Ching-Yung Lin, Nan Cao, Shi Xia Liu, Spiros Papadimitriou, Jimeng Sun, and Xifeng Yan. 2009. SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence. In *Proceedings of IEEE ICDE*. 1483–1486. DOI: http://dx.doi.org/10.1109/ICDE.2009.140
15. Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design Lessons from the Fastest Q&a Site in the West. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2857–2866. DOI: http://dx.doi.org/10.1145/1978942.1979366
16. David W. McDonald and Mark S. Ackerman. 1998. Just talk to me: a field study of expertise location. In *Proceedings of the ACM conference on Computer supported cooperative work*. 315–324. http://dl.acm.org/citation.cfm?id=289506
17. Saeedeh Momtazi and Felix Naumann. 2013. Topic modeling for expert finding using LDA. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 3, 5 (2013), 346–353. DOI: http://dx.doi.org/10.1002/widm.1102
18. Julian E. Orr. 1996. *Talking about Machines: An Ethnography of a Modern Job*. ILR Press. https://books.google.com/books?id=8ZBFI_i1nqUC
19. Sharoda A. Paul and Joseph W. Bolinger. 2015. Ask an Expert: Mobile Workspaces for Collaborative Troubleshooting. In *Hawaii International Conference on System Sciences (HICSS)*. 442–451. DOI: http://dx.doi.org/10.1109/HICSS.2015.60
20. Sharoda A. Paul, Lichan Hong, and Ed H. Chi. 2012. Who is Authoritative? Understanding Reputation Mechanisms in Quora. In *Proceedings of Collective Intelligence Conference*. http://arxiv.org/abs/1204.3724
21. Tim Reichling and Volker Wulf. 2009. Expert Recommender Systems in Practice: Evaluating Semi-automatic Profile Generation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 59–68. DOI: http://dx.doi.org/10.1145/1518701.1518712
22. A. W. Rivadeneira, Daniel M. Gruen, Michael J. Muller, and David R. Millen. 2007. Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 995–998. DOI: http://dx.doi.org/10.1145/1240624.1240775
23. Kerry Rodden. 2013. Sequences Sunburst. Web. (21 October 2013). Retrieved September 15, 2015 from http://bl.ocks.org/kerryrodden/7090426.
24. N. Sadat Shami, Kate Ehrlich, Geri Gay, and Jeffrey T. Hancock. 2009. Making sense of strangers' expertise from signals in digital artifacts. In *Proc. ACM CHI*. 69–78. http://dl.acm.org/citation.cfm?id=1518713
25. Alison Smith, Timothy Hawes, and Meredith Myers. 2014. Hierarchie: Interactive Visualization for Hierarchical Topic Models. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. 71–78. http://nlp.stanford.edu/events/illvi2014/papers/smith-illvi2014b.pdf

26. John Stasko, Richard Catrambone, Mark Guzdial, and Kevin McDonald. 2000. An Evaluation of Space-filling Information Visualizations for Depicting Hierarchical Structures. *International Journal of Human-Computer Studies* 53, 5 (Nov. 2000), 663–694. DOI: http://dx.doi.org/10.1006/ijhc.2000.0420

27. A. Strauss and J.M. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications. https://books.google.com/books?id=wTwYUnHYsmMC

28. L.A. Streeter and K.E. Lochbaum. 1988. An expert/expert-locating system based on automatic representation of semantic structure. In *Proceedings of the Fourth Conference on Artificial Intelligence Applications*. 345–350. DOI: http://dx.doi.org/10.1109/CAIA.1988.196129

29. Svetlana Yarosh, Tara Matthews, and Michelle Zhou. 2012. Asking the right person: supporting expertise selection in the enterprise. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. 2247–2256. http://dl.acm.org/citation.cfm?id=2208382

30. Svetlana Yarosh, Tara Matthews, Michelle Zhou, and Kate Ehrlich. 2013. I need someone to help!: a taxonomy of helper-finding activities in the enterprise. In *Proceedings of the ACM conference on Computer supported cooperative work*. 1375–1386. http://dl.acm.org/citation.cfm?id=2441931