Loughborough
University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

THE DESIGN AND EVALUATION OF EKE, A SEMI-AUTOMATED EMAIL KNOWLEDGE EXTRACTION TOOL

Abstract

This paper presents an approach to locating experts within organisations through the use of the existing indispensable communication medium and source of information, email. The approach was realised through the email expert locator architecture developed by the authors, which employs email content in the modelling of individuals' expertise profiles. The approach has been applied to a real world application, EKE, and evaluated using focus group sessions and system trials. In this work, the authors report the overall findings obtained from the focus groups sessions. The aim of the focus group sessions was to obtain information about the participants' perceptions, opinions, underlying attitudes, and recommendations in regards to the notion of exploiting email content for expertise profiling and to tackle the socio ethical challenges associated such an adoption. The paper provides a review of the various approaches to expertise location that have been developed over the past few decades, to link expertise seekers with internal experts and have provides the findings on the end-users' perspectives on the usability and functionality of EKE and the socio-ethical challenges raised by its adoption from an industrial perspective.

Keywords

1. INTRODUCTION

In this highly competitive global economy, knowledge, technology, skills, and communication have become the most critical determinants of success that will drive businesses forward. Hence, it is essential for companies to be equipped with the right technology that will help its employees communicate effectively with each other sharing ideas, experiences, learning skills, and knowledge. This requirement led to the

development of expertise finding tools which help users in locating experts with particular skills in order to help them solve problems. Such tools usually exploit information not generally exploited by knowledge workers.

Until very recently, finding expertise required a combination of individual, social, and collaborative practices. Mostly, it involves contacting individuals one trusts and asking them for referrals to other people (Campbell et al, 2003). However, huge costs are involved in following pointers to experts including efforts repeated by different people looking for the same answers, miscommunication that leads to the wrong expert and time pressures that lead to taking the advice of the not-so-expert who happen to be found quickly (Campbell et al, 2003).

Moreover, as organisations grow and become more geographically dispersed, the knowledge they generate also grows. Consequently, the difficulty of finding where and with whom the knowledge resides increases. In the last twenty years or so, a class of knowledge management software termed "expertise locating/ finding systems" has appeared in order to improve the quality of expertise finding and to speed up expertise finding process. An expertise locating system can be viewed as a repository that holds pointers to experts. Employees at various locations, around the world, can query this repository in the search for experts who might help resolve their problem. A clear definition of an expert is hard to find. For the purposes of this research, the authors will consider an expert to be an individual displaying an in-depth understanding of a given domain derived from training and experiences who can reliably perform specific observable behaviours relevant to a task within the domain (Shanteau & Stewart, 1992; Ericsson, 1996).

The aim of this work is to introduce and evaluate EKE (Email Knowledge Extraction), an expert finding tool developed by the authors as part of an ongoing research project at Anon. The paper starts with a summary of existing approaches to expert finding focusing on those that exploit email content, and highlighting the associated limitations. This is followed by an outline of the EKE tool and the underlying concept. The paper then presents the evaluation findings of the EKE tool using focus groups to establish people's

comprehension and acceptance of the EKE concept. The paper concludes with future work that remains to be tackled in the rich field of expert locators.

2. EXPERTISE LOCATOR APPROACHES

The task of expert finding is a complex endeavor given that it strives to identify persons with specific experience or expertise. Unfortunately, experts are inherently difficult to find. They are dispersed, expensive and culturally isolated. What is even more difficult is determining whether the knowledge they possess qualifies them to be "experts" in the subject under consideration. Even if their knowledge is sufficient, the type and the degree of expertise differs from one expert to another, adding a further layer of complexity. This is further accentuated, according to Maybury (2007) by expert seekers who generally have poorly defined search requirements. They are unaware of experts' past experiences and thus are not fully capable of distinguishing a good expert from a bad one. Moreover, their complex problems often require the combined wisdom of multiple experts.

Over the past two decades, various approaches to expert finding have developed. This section outlines requirements for and the challenges to expert finding. It presents the current state of the art in expert finding and highlights commercially available solutions. The idea of expert finding has been introduced by researchers with the aim of assisting users to discover subject matter experts in order to learn from their knowledge. To help fulfil this goal, expert finders require a range of information relating to individuals and their expertise (Yimam-Seid and Kobsa, 2003; Sim and Crowder, 2004). These include:

➢ The ability of the person to answer the user's queries or meet their search criteria,
➢ The level of the individual's knowledge or experience,
➢ A list of other individuals who could serve the same purpose,
➢ Performance of the individual in comparison with others in the field,
➢ Contact information of individuals who hold the required knowledge.

The need for techniques that gather and make such information accessible, have stimulated research into expert finding. Early efforts were focused on mechanisms such

as helpdesks, knowledge directories, and expert databases used to store and share knowledge from employees with others. Generally, these mechanisms depend on the manual entry of expertise data. However, manually entering expertise information is a time-consuming and arduous process. Individuals should voluntary and continuously maintain their expertise profiles reflecting the changes in their skills. Failure to do so leads to profiles becoming outdated. On the whole, when such mechanisms take on a large number of employees with inaccurate profiles, they loose credibility and thus speedily become ignored by the knowledge seekers. Generally, expertise databases are difficult to maintain, quickly outdated, and reflect self-reporting biases (Maybury, 2006).

The creation of the World Wide Web has added many dimensions to the field of expertise finding. Experts now have an international platform to express their interests and competencies through the development of personal web pages. Web searches may lead information seekers to the web pages of potential information providers. However, the success of these searches is dependent on two important elements. The first element relates to whether the keywords queried are present in the experts' web pages. The second element is dependent on whether personal web pages are regularly updated by the experts to reflect changes in their skills. Yimam-Seid and Kobsa (2003) question the adequacy of utilising search engines to trace experts, in light of having a large number of hits returned. Such a procedure would entail traversing the results returned, selecting the most appropriate expert, and determining their accessibility; hence it is greatly time consuming. The problem is further exasperated by lacking the facility to ensure whether the content is valid or not.

To overcome these hurdles, researchers have proposed the use of automated tools to identify, profile, and recommend experts facilitating the competence discovery process. From the review of the available literature, the authors have identified the following functions that expert finders need to support:

➢ *Discover:* Process general information available in organisational sources such as emails, instant messages, and publications created by individuals with the explicit purpose of identifying individuals with expertise.

➢ *Categorise*: Process the information sources of individuals to categorise their areas of expertise.

➢ *Authenticate*: Obtain a verification of areas and level of expertise. This can be accomplished through human assessment, or automated user feedback mechanisms.

➢ *Recommend*: Return a list of experts on a given area ranked by the level of expertise.

Numerous attempts have been made by researchers in both academia and industry to semi-automate/ automate the process of finding the right expert. These include: HelpNet (Maron, 1986), Expert/ Expert-locator (Streeter & Lochbaum, 1988) , ContactFinder (Krulwich and Burkey, 1996), Agent Amplified Communication (Kautz et al, 1996), Yenta (Foner, 1997), Phoaks (Terveen, 1997), Expertise Browser (Cohen et al, 1998), MEMOIR (Pikrakis et al, 1998), Expertise Recommender (McDonald and Ackerman, 1998), MIT's Expert Finder (Mattox et al, 1999), InfoScout (Prasad and Nagendra, 1999), SAGE (Becerra-Fernandez, 1999) and the KCSR Expert Finder (Crowder, 2002), Tacit Knowledge Systems' ActiveNet (Tacit, 2011), AskMe (AskMe, 2011), Autonomy's IDOL (Autonomy, 2011), Triviumsoft's SEE-K (Triviumsoft, 2011), MITRE's ExpertFinder (MITRE Cooperation, 2011), MITRE's XpertNet (MITRE Cooperation,2011), MindServer Expertise (Recommind, Inc. , 2011), and many others. These systems can be distinguished from one another by the information source(s) they use to infer expertise including: self disclosure, emails, knowledge repositories, program codes, bulletin boards, web pages, and documents (e.g technical reports, resumes, briefings).

There have been numerous and continuing attempts to design and implement expertise location systems that employ email messages as a vital data source to infer expertise from. Email has several characteristics which makes it suitable for this application. Email is a well established communication and collaboration medium, used on a daily basis by knowledge workers worldwide. According to statistics gathered by Royal Pingdom for the year 2010, an average of 294 billion email messages was sent per day. Email supports key knowledge processes such as knowledge creation and sharing. Moreover, it creates an electronic record of these knowledge processes making it possible to track and link daily workflows to the people involved.

Expert finding systems that exploit email have been an active research area for several years. In 1993, Schwartz and Wood exploited email for the purpose of inferring common interest relationships between people. Schwartz and Wood (1993) explore graphs produced from email communication patterns to discover shared-interest relationships amongst people and obtain a list of people who share a given interest. The system returns a list of people in no particular order. The user is then left with the challenge of searching this list by selecting people who are known to have similar interests to the experts being sought. To conduct a search, the user is required to have a prior social network with people who have expertise in their desired queries. This would prevent a novice user from taking advantage of this system. Moreover, it is questionable how useful a system would be to someone with an existing social network capable of providing them with aid.

In 1997, Kanfer et al developed the *Know-who* system which manages the information received through emails. The system keeps a log of senders email addresses from whom the individual has received emails. To identify an expert on a given topic, the individual emails the system with their desired question. The system then processes this question and searches the emails of individuals shown in the log of senders as part of the individual's social network. This results in a response to the individual in the form of a list of expert(s) contact details who may be able to provide assistance. One obvious drawback of the Know-who system, is that it is not capable of identifying experts outside the user's social network, thus impeding the process of expertise assistance (Kanfer, et al.

1997). Moreover, in 1997, Foner presented Yenta. The system determines the user's interests by analysing their outgoing electronic mail and files. It tries to form clusters of users with similar interests. Users can send messages to individuals in the cluster or to the entire cluster (Foner, 1997).

In 2001, Sihn and Heeren presented *XpertFinder*, a system that manages the information users send via email. Only the user created content and the address fields of emails are analysed and assigned to predefined subjects of a subject tree. Within each subject area, *XpertFinder* anonymously highlights the people who regularly communicate on the relevant area. To identify expert(s) on a given topic, the individual emails the system with their desired query. The system then processes the query, completes the selected recipients email addresses, and forwards the email (Sihn and Heeren, 2001).

In 2003, Campbell et al. at IBM presented e^3 (email expertise extraction), that exploits email content and social networks (Campbell, 2003). Campbell et al. (2003) compared in two different organisations a content based approach that analyses only email content with a graph based method that examines social networks from email communications and concluded that in the e^3 system the latter outperformed the former. However, the size of the networks studied is very small and does not reflect the characteristics of social networks in practice (Zhang, 2007).

The systems currently available for expert identification by email in the commercial market include: IBM's SmallBlue (Lin et al, 2009), Tacit's ActiveNet (Software, 2011), AskMe Enterprise (AskMe, 2011), and Corporate Smarts' Intelligent Directory (Corporate Smarts, 2005).

IBM's SmallBlue (Lin et al, 2009) project (which is also part of the Atlas software suite) applies artificial intelligence algorithms for purpose of uncovering participants' social network "who they know" and also the expertise of those participants "what they know". The application analyses employees' email outbox, however only to employees within the same company. The application also analyses employees' outgoing instant messaging

chat transcripts and their profile information. The inferred social network for each participant is private and displayed as a visualization in Ego for each user. SmallBlue comes with a search engine which enable users to search for expertise based on a topic and displays the results in the context of the wider enterprise social network.

Tacit's ActiveNet software (formally known as Tacit KnowledgeMail System), autom.atically processes users' emails (unpublished expertise) and artifacts (published expertise) for the creation of user profiles. Tacit does not perform social network analysis. Employing statistical (i.e. frequencies of nouns and noun phrases) and linguistic methods (i.e. context of surrounding words from unstructured text and time stamp of their appearance), the user's level of expertise on a particular topic is automatically ranked. Tacit allows individuals to create a public and a private profile. The public profile is available to be searched and directly provide the expert's contact detail. Private profiles can similarly be searched, however the expert's identity is not revealed without consent from the expert. This poses the disadvantage of an increased workload for the expert who potentially has to handle many requests for their contact information. Users query the system by performing a keyword search on both the public and private profiles. Depending on the profiles matched, the returned result of the search is either an ordered list of experts or a list of relevant topics based on the frequency, intensity, and history of the query topic within the person's expertise profile. From the topics returned for the private profile match, the system provides the searcher with the facility to email the related experts, thus protecting the experts' privacy. The disadvantages of the ActiveNet system include the lack of sufficient benchmarking against other systems and the substantial cost involved with implementing and maintaining the system.

AskMe is a tool which automatically processes documents including resumes, email, Q&As (Questions and Answers), and external publications uploaded to the system. The system stores user requests and answers to queries raised through the system, which are then made available to other users in the organisation. Experts are identified and ranked based on the answers they provide. AskMe allows two types of profiling, auto profiling and dynamic profiling (Maybury, 2006). Auto profiling refers to the identification of expertise through the extraction of keywords from documents uploaded through the

system or through a third party repository (e.g. email or document repository). On the other hand, dynamic profiling refers to the identification of expertise from the user's solutions and publications. The documents are sent to FAST (Fast Search and Transfer, www.fastsearch.com) to identify keywords. Expertise validation is dependent on the history of information sources and expert qualifications/certifications. Users have full control over their expertise profile (i.e. can modify keywords in their profiles). To deal with privacy issues, AskMe allows content tagging thereby limiting the access to sensitive information. Rules can be implemented to control who has access to what type of information and how it is disseminated. Users can then find experts by typing in a keyword or phrase in a web based interface. The returned results include categories such as Q&As, publications, and people. The people's category includes a summary of the expert's record consisting of the number of answers and publications on the topic, along with their average response time. Similar to ActiveNet, the disadvantages of the AskMe system include the lack of sufficient benchmarking against other systems and the substantial costs involved with implementing and maintaining the system.

Corporate Smarts' Intelligent Directory is a system which employs communication network algorithms to identify experts. It identifies important ideas and communication networks from corporate email by analysing parameters including the flow, subject matter, frequency and context of the email communication. The system stores the network and the normalised content of the email communications. Users query the system by entering the related terms in the search facility provided. The return result is an ordered list of employees' names and contacts.

Expert locator systems have been implemented in a variety of organisational domains however, the successful implementation of such systems undoubtedly requires much more than the efforts on the part of the system developer. It requires as stated by Maybury (2007) user involvement, user/culture centered design, clear purpose, realistic goals, measured usage and benefit, simplicity, ease of use, incremental deployment, appropriate privacy, incentives for use, and effective marketing, communication, and training. The authors' review of the literature has shown there to be no research available

in the public domain showing the evaluation of commercial systems. Indeed, for expert systems to have a major impact, performance evaluation should be conducted, addressing a fundamental creed of performance evaluation stating that if something is worth doing, it is worth doing well.

3. EKE: A SYSTEM FOR LOCATING EXPERTISE FROM EMAILS

This section outlines the approach developed by the authors to identify individual competencies from email messages. The Email Knowledge Extraction (EKE) system, reported in detail in Anon et al, (2006a), aims to allow users to locate experts who may possess specialised knowledge that users seek.
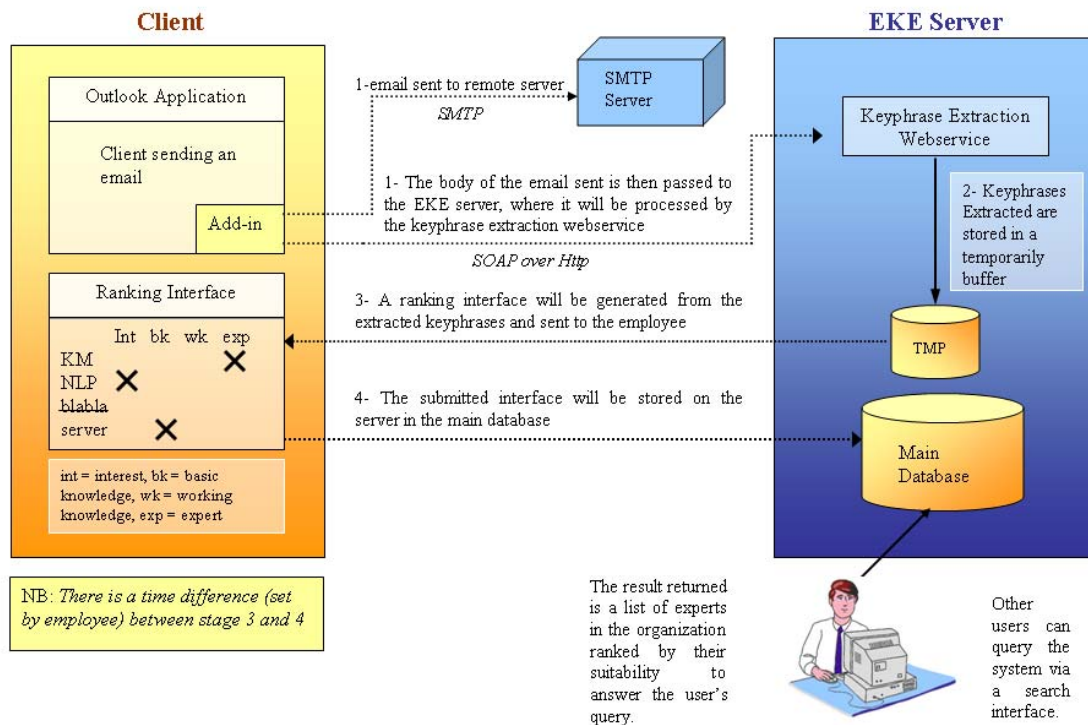

Figure 1: EKE Generic Architecture

Figure 1 illustrates the workflow of the EKE architecture. The system comprises of client side and server side software. An Outlook plugin implements the client side software. The plugin captures email content sent by users and forwards it to the server side software. The server side software is composed of a web service based keyphrase extraction engine that identifies topical terms representing users' competencies from the

email body. The terms are directly returned to the user and graphically displayed for verification (i.e. is the term returned indicative of their skills or not) and ranking (i.e. their subjective level of expertise in the returned proficiency area). Once the user verifies and ranks the returned terms they are stored in the respective user's profile in a repository on the system server. The repository is available to be queried by all users within the organisation who have a need to locate others with the necessary experience to answer their questions.

From a technical viewpoint, identifying keyphrases suggestive of sender skills and experiences represent the main challenge of utilising email message content for expert identification. A detailed overview of EKE's keyphrase extraction engine is provided in Anon et al (2006). The engine was built using the Natural Language ToolKit (NLTK), a collection of libraries and programs for symbolic and statistical natural language processing for the Python programming language. In brief, the extraction algorithm comprises of two stages. The first stage employs training in which a part of speech (POS) tagging model was created. In the second stage keyphrases are extracted from email messages with the help of the speech-tagging model created in stage one.
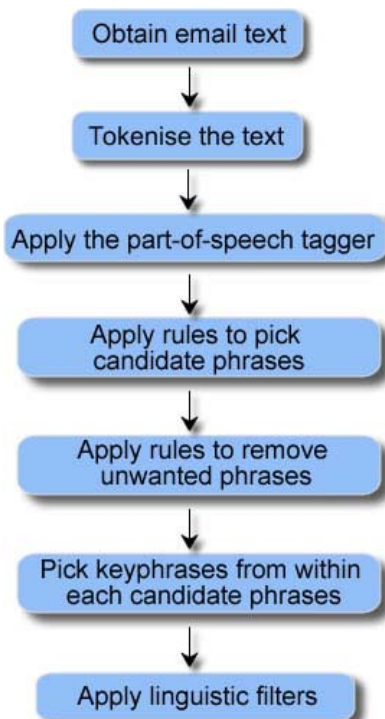
Figure 2 Stages of the extraction process

Figure 2 shows the keyphrase extraction steps of stage two. The email message sent by the user is captured by the system and the text in the email body is fed into the keyphrase extraction engine. The text is then divided into tokens using regular expression rules and tagged by their parts of speech with the help of the POS model created. Keyphrases are identified by applying rules which were manually set by the author to group all occurrences of specific sequences of tags together. The rule is formed from a sequence of grammatical tags which are most likely to contain words that make up a keyphrase. Once sequences of tags are collated, more rules are applied to remove a subset of non relevant phrases. Keyphrases are then chosen from the identified candidate phrases. The process concludes with the use of linguistic filtering to extract the most important keyphrases. This results in a set of lines, each containing a sequence of tokens representing key competencies.

4. RESEARCH DESIGN

The overall research adopted a systems development methodology. Although many methodologies exist, they all recognise three main stages: concept development, system building, and system evaluation (Brustein, 2002). The concept development stage of the project involved the initial development of the system requirements, followed by a comprehensive literature review in the field of expert locators (reported in detail in Anon et al, 2006a). A domain analysis of this data identified shortcomings in the research subject and further aspects that need to be addressed (also reported in detail in Anon et al, 2006a). This in turn led to an identification of further developments to the project requirements. In the system building stage, the expertise locator tool – EKE – was developed, and an evaluation of the keyphrase extraction engine was conducted (Anon et al, 2006b). The development process was iterative and evolutionary. Individual programme units were developed, tested, and integrated to obtain the final system. Evaluation of the complete keyphrase extraction engine was essential to assess its performance. The engine was evaluated for the effectiveness of extracted keyphrases using the recognised f-measure performance measure (Anon et al, 2006b).

During the overall system evaluation stage, the stage reported in this paper, the system availability was expanded to target potential users. This section outlines the research design of the study, and the reasoning behind an evaluation approach to testing EKE. The abstract and necessarily unmanageable nature of knowledge (Abou-Zeid, 2007, p. 615) means positivist, critical or post modernist approaches to testing are not suitable because they require the existence of an objective, independent reality. An interpretivist approach was then reviewed for this research, and in particular the evaluation method. Evaluation is oriented towards assessing and improving any given object, programme, system, theory and most other entities (Stufflebeam and Shinkfield, 2007, p. 4). There are two broad categories of evaluation: formal and informal (Clarke, 1999, p. 1), distinguished by the means with which they are conducted. Informal evaluation occurs on an almost daily basis, judging the value or worth of a purchase for example. Formal evaluation, however, is a disciplined form of inquiry that applies to the collection and analysis of information (Lincoln and Guba, 1986, p. 550).

Within formal evaluation there are two different formative evaluation and summative evaluation (Stufflebeam and Shinkfield, 2007, p. 25). Formative evaluation focuses on the process of improvement and about identifying strengths and weaknesses (Clarke, 1991, p. 7). Summative evaluation, on the other hand, is concerned with post process decision-making and has a focus on providing information to make a decision for action (Clarke, 1991, p. 8). The evaluation technique chosen for this study was formative, because the aim is to improve (Stufflebeam and Shinkfield, 2007, p. 23) EKE. It might appear that summative evaluation is more appropriate, with its focus on the completion of a programme, process or product, but summative evaluation does not allow the clarification of goals or debate surrounding the nature of any implementation (Clarke, 1991, p. 8). Given the evidence, the evaluation approach was deemed the best by the authors, as it allows participants to determine the place of EKE among existing tools.

The evaluation consisted of focus group sessions and a demonstration of EKE. Focus groups were conducted to evaluate the system from an industrial perspective. Moreover,

the system was evaluated by potential end-users selected from within the school of Informatics at Anon University in the United Kingdom. The system was evaluated for functionality, robustness and ease of use. Comments, concerns and errors were communicated on detection. The feedback was either resolved or earmarked for further action. Moreover, key socio-ethical challenges involved in the implementation and use of the developed system as identified by the authors (Anon et al, 2007) were further investigated to establish the participants' views on these issues. The data collected was analysed using observer impression. In particular, the recursive abstraction method was used. The method was chosen to provide a compact summary that would have been difficult to accurately discern without the preceding steps of distillation. A detailed description of the EKE evaluation is provided in section 4.1.

4.1 EKE Concept Evaluation

Due to the organisational environment in which EKE will operate, it was essential to establish the end-users' perspectives on the usability and functionality of the EKE tool and on socio-ethical challenges raised by its adoption from an industrial perspective. To accomplish this, two focus group sessions were conducted. The focus group technique encouraged creative thinking and stimulated discussion through group interaction. To encourage participation, an invitation and a product brief were emailed to potential organisations. Those keen to take part in the study received an information pack and a draft agenda of the focus group. A total of 12 people participated in the group evaluation sessions. The first group (Group A) comprised five participants from SAP, one of the largest European software enterprises whilst the second group (Group B) comprised seven participants from the UBS (United Bank of Switzerland) one of the leading financial firms. Group A was conducted over the telephone (teleconference) while Group B was conducted using the technique of focus group interviewing.

In both focus groups, the sessions commenced with a presentation that provided the participants with background information about the project. The participants were then shown a video of the EKE tool in operation. This was followed by a group discussion in which various unstructured, open-ended questions were addressed. After obtaining the

participants' consent, the responses were recorded using a tape recorder in order to obtain a complete record of the discussion.

Participants were asked to conduct a SWOT (Strengths, weaknesses, opportunities, and threats) analysis of the tool by noting down the internal and external factors that would facilitate/hamper its adoption. Following this, the groups were asked to conduct a brainstorming session to identify any improvements which they felt could be made to the EKE tool. A summary of the exercise was fed back and confirmed by the participants to be an accurate account and amended where required. The session concluded with the participants filling in a questionnaire regarding socio-ethical issues. Results from this evaluation are detailed in Section 5.

5. EKE CONCEPT EVALUATION FINDINGS

This section provides the overall findings obtained from the discussions of the two focus groups. As stated earlier, the main reason for conducting these focus groups was to establish industrial end-users' perspective on usability and functionality of EKE and on socio-ethical challenges raised by its adoption.

5.1 FOCUS GROUPS FINDINGS

The views detailed in this section were gathered from open-ended questions, comments made during a demonstration at a focus group, and remarks made during group discussion.

Knowing that EKE analyses the email body, opinion was divided as to the inclusion of the full email chain (forwarded and replied to emails) during the analysis phase. It was argued by some that the keyphrases contained in the email chain are often relevant to the users' areas of interest/knowledge and as such, the email chain should be included in the analysis. Others felt that the email chain may not necessary depict the sender's area of knowledge. Additionally, the domain independent nature of EKE, which allows the identification of non-industry specific terms, generated discussion. Some argued that this approach was beneficial as it increased the scope of the knowledge areas that can be identified. However, others argued that pre-feeding the system with a list of industry

specific terms would reduce the surplus of non-relevant terms that may be identified. Similarly, it was argued that the system could be improved through the use of domain ontologies, which help add context to the keyphrases identified. When designing EKE ontologies were considered, but the added complexity, efforts, and expenses required to create and maintain such a list where deemed excessive for large organisations. Moreover, creating and maintaining domain specific ontologies is difficult, because it requires a deep understanding of the industry under consideration. In addition, ontologies cannot be easily adapted for reuse in a different domain and have shown to cost large amounts of money to develop. An example is the Gene Ontology, a well known ontology in the biomedical field, is known to have cost at least an estimated $16 million at the end of (Good et al., 2006).

The issue which attracted much discussion during the focus groups was the adoption challenges faced by EKE. There were a number of trends which emerged. One of these trends is related to the level of user involvement. Some participants felt that the level of user involvement in the selection and categorisation of specific keyphrases is appropriate, while others did not. This can be related to the desire to minimise the number of interruptions generated by EKE. As such, those who viewed the level of user involvement as inappropriate made suggestions of how these could be reduced. Although there was consensus in the groups that EKE's technique of only prompting users once for each identified keyphrase was significant, there still remains the potential for further reduction. Two suggestions were made. One suggestion was to extract keyphrases relevant to users' areas of knowledge, also through the adoption of an ontology based approach. The second suggestion was that users should be able to select when and how often these keyphrases are to be categorised. For privacy reasons, the groups put forward that only work related emails should be analysed and that the system should be configurable to allow for this flexibility. Due to the nature of work undertaken by Group B participants, where information is private within subgroups of the organisation, it was felt that EKE's configurability could be extended to provide different levels of access for different users.

The groups recognised that user training and education are vital to the successful adoption of EKE. Users should understand and be comfortable with using EKE from the outset, otherwise there is a risk of alienating employees before they have learnt about EKE and its benefits for them and for their organisation.

The individual opportunity provided by EKE and which the groups agreed on, included increased employee productivity through a reduction in time wasted looking for the right employees to help solve problems. An added opportunity provided by EKE is better organisational communication networks instigated through EKE. Such networks can lead to increased positive collaboration after the initial contact.

The common fear amongst the groups was the negative impact of EKE on their work. Though, both had different reasons for their concerns. With regards to their respective organisations, it was felt that there was a risk of malicious use if outsiders were able to gain access to the EKE database.

Although the primary role of EKE is to help employees locate other members of staff with the right knowledge in the organisation, one other application for EKE was uncovered. This application was activity recording, which can help employees keep a personal record of their activities. Self selection and categorisation by employees can be easily extended to allow activity recording. Moreover, two suggestions for further work were made. These included the addition of instant messaging as a source of information, and the use of calendar appointments to uncover knowledge areas of the meeting attendants.

Overall, the groups felt that EKE is well designed and easy to use and would greatly aid employees to find the right employees, especially in large organisations. This was attributed to the close integration of the tool with existing technologies and work

practices. One group compared EKE with a similar web based tool which their organisation currently uses, and favoured EKE's approach which encourages participation. The use of the well established communication medium, email, was seen by the groups as a flexible source of knowledge areas. This allows EKE to be easily adopted by organisations into their existing email systems.

7. CONCLUSION

The paper has reviewed the various approaches to expertise location that have been developed over the past few decades, to link expertise seekers with internal experts. As discussed in the paper, the first generation of such systems emerged from the use of helpdesks as formal sources of knowledge, and comprised knowledge directories and expert databases. Microsoft's SPUD project, Hewlett-Packard's CONNEX KM system, and the SAGE expert finder are key examples of this genre. In the second generation of expertise locators, companies offered personal web pages where employees could advertise their expertise internally or externally, but frequently became out of date. The third generation of approaches relied primarily on secondary sources for expert identification. Web applications studied browsing patterns/activities in order to identify experts. With this type of application, if the user knew a particular expert, the user could ask the system to reveal the browsing path of that expert, relevant to the user's query. Among other disadvantages, if an employee did not know an expert, the user must ask the system to identify one or more experts. The fourth generation have tried to utilise social networking methods along with natural language processing and artificial intelligence techniques in order to analyse stored knowledge, seeking to identify expertise and experts. The major drawback of the fourth generation approaches is that potential expert listings are unordered when presented to a user seeking experts, requiring significant user effort to identify the best expert. Such systems identify experts by textual analysis but rarely support expert selection by users. In addition, such systems fail to present the varying degrees (or levels) of expertise that people possess and tend to assume a single level of expertise. It is thus entirely the user's responsibility to systematically process the returned results in order to identify the most suitable experts for answering specific queries.

With what could be seen as the fifth generation of expert locators in the development of EKE, the results have provided the end-users' perspectives on the usability and functionality of such a tool and the socio-ethical challenges raised by its adoption from an industrial perspective. In a domain that has published little in the way of results on the evaluation of such techniques and tools, this paper has provided both an expert locator concept and an approach of evaluating an expert location tool. The main themes that were raised during the evaluations were, user involvement, privacy, productivity, and integrated tools. These themes would benefit from further investigation with a view to using them as areas for measurement in future evaluation studies of expert locating tools.

The study also highlighted the need for such tools to be multifunctional and go beyond just providing knowledge sources. An example extension to the system was activity recording, which could help employees keep a personal record of their activities. Interestingly, this approach would lead to a system that could build a profile of employees working habits. Over the decades 'big brother' type systems have provoked many negative responses from those involved, but overtime have employees reduced their uneasiness towards such systems, which could lead to more integrated systems and potentially better information and knowledge management tools. The research findings have also provided areas for future research, in particular, the potential willingness of employees to use more integrated systems to aid in more efficient and effective working processes, but with the trade off against a reduced level of personal privacy.

8. REFERENCES

Abou-Zeid, E. 2007, "Towards a Design Theory of Autopoietic Knowledge Management Support Systems", *Managing Worldwide Operations and Communications with Information Technology*, ed. M. Khosrow-Pour, IGI Publishing, Hershey, Philadelphia, pp. 614-617.

AskMe [Homepage of AskMe], [Online]. Available: http://www.askmecorp.com/ [Last accessed in 2011, 05/01].

Autonomy [Homepage of Autonomy], [Online]. Available: http://www.autonomy.com
    [Last accessed in 2011, 02/01].

Becerra-Fernandez, I. 1999. "Searchable Answer Generating Environment (SAGE): A
    Knowledge Management System for Searching for Experts in Florida", Proceedings
    of the 12th Annual International Florida Artificial Intelligence Research Symposium
    (FLAIRS), Knowledge Management Track, Orlando, Florida.

Campbell, C.S., Maglio, P.P., Cozzi, A. & Dom, B. 2003, "Expertise identification using
    email communications", twelfth international conference on Information and
    knowledge management, New Orleans, LA, pp. 528.

Clarke, A. 1999, *Evaluation Research: An Inrtoduction to Principles, Methods and
    Practice,* Sage Publications, London.

Cohen, A.L., Maglio, P.P. & Barrett, R. 1998. "The Expertise Browser: How to Leverage
    Distributed Organizational Knowledge", presented at Workshop on Collaborative
    Information Seeking at CSCW'98, Seatle, WA.

Corporate Smarts [Homepage of Corporate Smarts], [Online]. Available:
    http://www.corporatesmarts.com/ [Last accessed in 2005, 16/09].

Crowder, R., Hughes, G.V. & Hall, W. 2002, "An Agent Based Approach to Finding
    Expertise", The 4th International Conference on Practical Aspects of Knowledge
    Management, eds. D. Karagiannis & U. Reimer, Springer-Verlag, Vienna, Austria,
    pp. 179.

Ericsson, K. A., 1996. "The road to excellence". Mahwah, NJ.: Lawrence Erlbaum
Assoc.

Foner, L.N. 1997. "Yenta: A Multi-Agent, Referral-Based Matchmaking System", The
    First International Conference on Autonomous Agents, Marina del Rey, CA, pp. 301.

Good BM, Tranfield EM, Tan PC, Shehata M, Singhera GK, Gosselink J, Okon E,
    Wilkinson M: Fast, cheap and out of control: A zero curation model for ontology
    development. *Pacific Symposium on Biocomputing* 2006, :128-139.

Kanfer, A., Sweet, J. & Schlosser, A. 1997. "Humanizing the Net: Social Navigation with a 'Know-Who' Email Agent", Denver, Colorado.

Kautz, H., Selman, B. & Milewski, A. 1996. "Agent Amplified Communication", Thirteenth National Conference on Artificial Intelligence. Portland, OR, pp. 3.

Krulwich, B. & Burkey, C. 1996. "Learning user information interests through extraction of semantically significant phrases", Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access, Stanford, CA.

Lin, C., Cao, N., Liu, S., Papadimitriou, S., Sun, J., and Yan, X. 2009, "SmallBlue:Social Network Analysis for Expertise Search and Collective Intelligence", IEEE 25th International Conference on Data Engineering, pp. 1483-1486.

Maron, M.E., Curry, S. & Thompson, P. 1986. "An Inductive Search System: Theory, Design, and Implementation", *IEEE Transaction on Systems, Man and Cybernetics,* vol. 16, no. 1, pp. 21-28.

Mattox, D., Maybury, M. and Morey, D. 1999. "Enterprise Expert and Knowledge Discovery", in proceedings of th 8th international Conference on Human-Computer Interaction (HCI International'99), Munich, Germany, pp 303-307.

Maybury, M, 2006. Expert Finding Systems. Technical Report MTR 06B000040, MITRE Corporation

Maybury, M. 2007. Discovering Distributed Expertise, AAAI Fall Symposium Regarding the "Intelligence" in Distributed Intelligent Systems (RIDIS) .

McDonald, D.W. & Ackerman, M.S. 1998. "Just Talk to Me: A Field Study of Expertise Location", pp. 315.

MITRE Corporation [Homepage of MITRE Corporation], [Online]. Available: http://www.mitre.org/ Last accessed in 2011, 05/01].

Pikrakis, A., Bitsikas, T., Sfakianakis, S., Hatzopoulos, M., DeRoure, D., Hall, W., Reich, S., Hill, G. & Stairmand, M. 1998. "MEMOIR: Software Agents for Finding Similar Users by Trails", Proceedings of the 3rd International Conference on the Practical Applications of Agents and Multi-Agent Systems ({PAAM}-98), eds. H.S. Nwana & D.T. Ndumu, London, UK, pp. 453.

Prasad, M. V. Nagendra & Anagnost, T. 1999. InfoScout: An Active Recommender Agent, Madison, Wisconsin.

Recommind, Inc [Homepage of Recommind], [Online]. Available: http://www.mitre.org/ Last accessed in 2011, 08/01].

Shanteau, J. & Stewart, T.R. 1992. "Why study expert decision making? Some historical perspectives and comments", Organizational Behavior and Human Decision Processes, vol. 53, no. 2, pp. 95-106.

Schwartz, M.F. & Wood, David C. M. 1993. "Discovering Shared Interests Using Graph Analysis", Communications of the ACM, vol. 36, no. 8, pp. 78-89.

Sihn and Heeren. 2001. "Xpertfinder - expert finding within specified subject areas through analysis of e-mail communication", In EUROMEDIA 2001: Sixth Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications, pp. 279.

Sim, Y. W. and Crowder, R. 2004. *"Evaluation of an Approach to Expertise Finding".* In: 5th International Conference on Practical Aspects of Knowledge Management, Vienna, Austria, Vienna, Austria.

Streeter, L.A. & Lochbaum, K.E. 1988. "An Expert/Expert-Locating System Based on Automatic Representation of Semantic Structure", Fourth IEEE Conference on Artificial Intelligence Applications Computer Society of the IEEE, San Diego, CA, pp. 345.

Stufflebeam, D.L. & Shinkfield, A.J. 2007, *Evaluation Theory, Models, and Applications,* Jossey Bass, United States.

Tacit Software 2007, [Homepage of Tacit], [Online]. Available: http://www.tacit.com/home.asp [Last accessed on 2011, 05/02].

Anon 2006a

Anon 2006b

Anon 2007

Terveen, L., Hill, W., Amento, B., McDonald, D. & Creter, J. 1997, "PHOAKS: a system for sharing recommendations", Communications of the ACM, vol. 40, no. 3, pp. 59-62.

Trivium, 2007, [Homepage of Trivium], [Online]. Available: http://www.triviumsoft.com [Last accessed on 2011, 05/02].

Yimam-Seid, D. & Kobsa, A. 2003, "Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach", Organisational Computing and Electronic Commerce, vol. 13, no. 1, pp. 1-24.

Zhang, J., Ackerman, M.S. & Adamic, L. 2007, "Expertise networks in online communities: structure and algorithms", Proceedings of the 16th international conference on World Wide Web, ACM, New York, NY, USA, pp. 221.
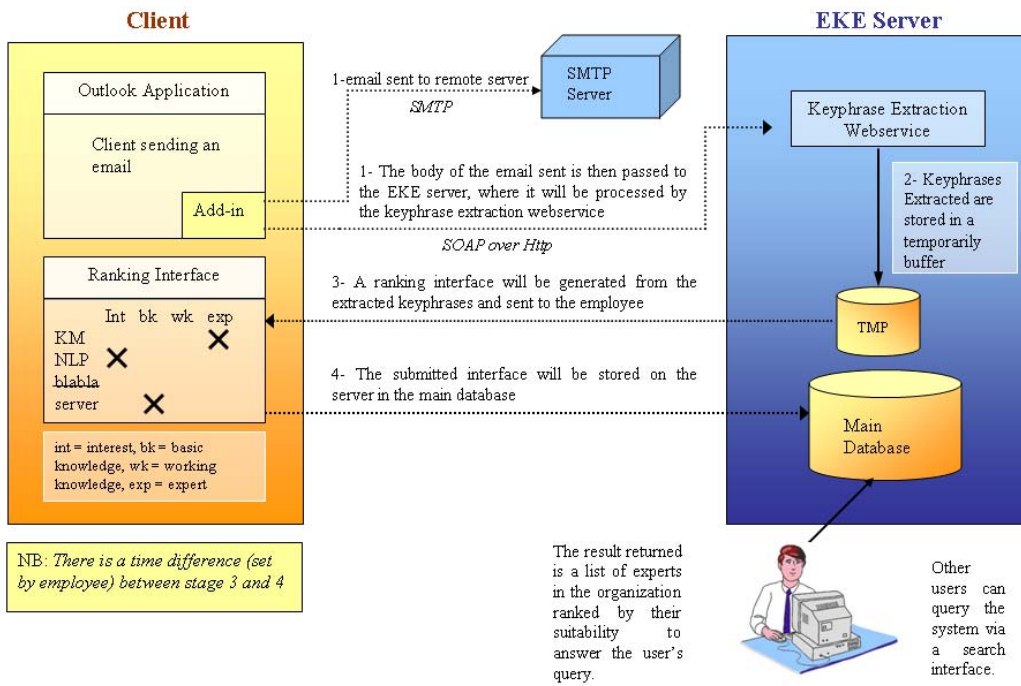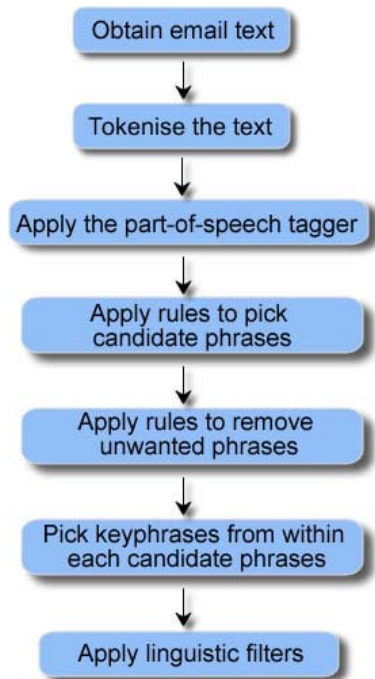
Figure 1

```
Obtain email text
        ↓
Tokenise the text
        ↓
Apply the part-of-speech tagger
        ↓
Apply rules to pick
candidate phrases
        ↓
Apply rules to remove
unwanted phrases
        ↓
Pick keyphrases from within
each candidate phrases
        ↓
Apply linguistic filters
```

Figure 2