IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB12801

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2* edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

Building a Tool for Expertise Discovery

Sara Tedmori, Thomas Jackson, & Dino Bouchlaghem
Loughborough University, Loughborough, Leicestershire, UK, LE11 3TU, T 44 (0) 1509 635649,
{s.m.j.tedmori, t.w.jackson, n.m.bouchlaghem}@lboro.ac.uk

Holger Adelmann, AstraZeneca R&D Charnwood, Loughborough, Leicestershire, UK, LE11 5RH, holger.adelmann@astrazeneca.com

Rama Nagaraju, Eureka Moment Software, 9284 Haden Ln., West Chester, OH 45069, rnagaraju@eurekamomentsoftware.com

ABSTRACT
There is increasing interest in systems that aid employees to find those with the expertise they require. This paper discusses the evolution of expert finding tools, with particular reference to solutions that exploit email sources and identifies related gaps. The authors then propose Email Knowledge Extraction (EKE), a system for expertise discovery which addresses the issues highlighted by gap analysis.

1. INTRODUCTION
In working environments, people are put in situations where they need to make a decision or look for information to resolve an ambiguity or a complexity. Early studies on information seeking behaviour show that people searching for information prefer asking other people for advice rather than searching through a manual for information (Bannon, 1986). A study by Kaurt and Streeter (1995) back up this perception by showing that people were the most valued and used sources of help in software development projects.

Campbell et al. (2003) state that people ask others they know to find someone with a particular skill or experience, following pointers until an appropriate person is found. They also argue that there is a huge cost involved in following pointers to experts. These costs include 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).

Research has shown that employees learn more effectively by interacting with others and the real value of information systems lies in their ability to connect people to people, so they can collaborate with each other (Bishop, 2000; Cross and Baird, 2000; Gibson, 1996; Wellins et al., 1993). Searching for the right piece of information becomes a matter of searching for the right person to refer to. This has lead to the interest in systems, which connect people to others by making people with the necessary expertise available to those who need it, when they need it.

In this paper the authors identify the email communication system as an information source that could be utilised to locate experts within an organisation. The authors discuss the evolution of the expert finding approaches (section 2), with particular focus on expert finding agents that exploit email content as evidence of expertise (section 3). The gaps associated with the current approaches of agents which utilise email are then highlighted (section 4), and finally the authors propose an architecture for an email knowledge extraction system to aid knowledge location within the workplace (section 5).

2. TRADITIONAL EXPERT FINDING APPROACHES
The traditional way of providing automated expert assistance relies on the development of expert databases that require users to manually register and enter their expertise data. Expert databases suffer from many drawbacks. Firstly, maintaining a manually built database requires intensive and expensive labour. Secondly, unless the users regularly update their details to reflect changes in their expertise profiles, the systems will soon become out of date and inaccurate. Thirdly, expertise descriptions are usually incomplete and general, in contrast with the expert related queries that are usually fine-grained and specific (Yimam-Seid and Kobsa, 2003).

The other problem with traditional expert systems is the ability to search and successfully locate the required information stored within the system. Large global enterprises sometimes have disparate expert databases that are sometimes restricted to one region and do not enable the employee to take full advantage of the global expert resource (Adelmann, personal communication). Yimam-Seid and Kobsa (2003) note that using search engines to locate an expert is ineffective. This is due to the fact that the search process is based on a simple keyword matching task, which may not always lead to relevant experts. The task can be very time consuming when a large number of hits are returned. Moreover, Yimam-Seid and Kobsa argue that it is entirely the user's task to extract and compile all the required data to identify the best expert (Yimam-Seid and Kobsa, 2003).

Most importantly, traditional expertise assisting technology adds an extra work load to people's work as they have to maintain their profiles on top of everything else they do. Hence, people are less likely to use it. Expertise software must therefore be integrated into existing business processes. The drawbacks of the traditional approaches coupled with advances in information technology has resulted in a shift towards systems that automate or semi-automate the process of discovering expertise.

3. EXPERT FINDING SYSTEMS EXPLOITING EMAIL
The International Data Corporation (IDC) has predicted that 35 billion emails will be sent globally every day by the end of 2005. IDC's Email Usage Forecast and Analysis report further estimates that the number of emails sent annually in Western Europe will be 1.6 trillion in 2005 (Mahowald and Levitt, 2002). With so many email messages being sent each day, it seems logical that a percentage of them will contain key phrases that will help identify experts within organisations.

From an academic prospective, attempts to develop systems that exploit email to augment the process of finding the right expert can be traced back to the work of Schwartz and Wood in 1993. Their system deduces shared-interest relationships between people. To avoid privacy problems, they decided to analyse the structure of the graph formed from *"From:/To:"* email logs, using a set of heuristic graph algorithms. The output of the system is a list of related people with no essential ranking order. A user searches the system by requesting a list of people whose interests are similar to several individuals known to have the interest in question. This implies that the person should have beforehand a social network with the appropriate contacts relevant to their query and that a novice can not properly take advantage of the system.

Since 1993 there have been several research projects to identify experts from email communication. For example, *The Know-who* system is an email agent that helps to manage the information the users receive through emails (Kanfer et al., 1997). A *Know-who* agent monitors all email messages received by the user and maintains a list of all those from whom the user received email message(s). Based on the content of email communication with the people in the user's social network, it responds to the user's natural language query with a name(s), email address, and confidence level of the person(s) most likely to answer the question (or with a reference to another person who might know the answer). One potential limitation of *Know-who* is that it only identifies people within the user's social network. This makes it unfeasible to identify individuals outside the user's social network with common interests, thus impeding the process of expertise assistance.

Sihn & Heeren (2001) presented *XpertFinder*, a system which analyses email communication of users for the preparation of expertise profiles. The part of the message entirely created by the sender and the address fields of emails are analysed and allocated to predefined subjects with the aid of a subject area tree. Within each subject area, *XpertFinder* allows anonymous highlighting of the people who are frequently communicating. Users submit their requests by emailing the *XpertFinder* system, which in turn completes the selected recipients email addresses and forwards the email. Experts are identified both by high communication intensity (e.g. whether or not they decide to reply to users' queries if they were forwarded to them) as well as communication contacts in specific subject areas (Sihn and Heeren, 2001). Systems similar to *XpertFinder* are hard to share and reuse because they are based on a predefined subject tree. They are labour intensive to build and require ongoing maintenance.

Commercial systems for expert identification using emails include: *Tacit's ActiveNet* (Tacit, 2005), AskMe Enterprise (Ask Me, 2005) and Corporate Smarts' Intelligent Directory (Corporate Smarts, 2005). All of which extract keywords and phrases from users' emails and electronic documents. The information is placed into an expertise profile and distilled into a searchable database in order to enable users to query the system and find relevant people.

Unfortunately, with regards to the commercial systems, no sufficient data is available on how these systems perform. Most of the system information is only available provided in the form of white papers serving as marketing tool to promote an organisations product and point of view. To avoid the dilemma of lack of sufficient data and to help analyse the existing systems, the authors have conducted domain analysis in order to identify opportunities for improvement.

4. GAP ANALYSIS OF EXISTING SYSTEMS

To analyse the existing systems and the newly emerging technologies, domain analysis is needed. Domain analysis can be defined as the process of identifying, capturing, and organizing domain knowledge about the problem domain with the purpose of making it reusable when creating new systems (Prieto-Diaz and Arrango, 1991). A domain model of expert finding systems has been proposed by Yimam-Seid and Kobsa (2003). This model was used by the authors in order to acquire and consolidate information about applications in the expert finding systems domain, with the intention of identifying the gaps of existing technologies that particularly exploit email as the basis for expertise recognition. The authors have identified five gaps, namely (1) an expertise profile gap, (2) an expertise matching gap, (3) an expertise representation gap, (4) a user control gap, and (5) a cultural and management gap. In the following sections, the authors will describe each of these shortcomings and suggest some ways to tackle them.

4.1 Expertise Profile (model) Gap

The core of expert finding systems heavily relies on the *expertise profile (model)* and on how accurate these systems are in their expertise matching process. *Expertise profile (model)* refers to information specific to an individual such as the individual's skills, interests, expertise, personal details, et cetera. Common to most systems is the automatic extraction of key phrases from within the body of emails and the creation of the users profiles, such as Know-who email agent (Kanfer *et al.*, 1997), Ask me (Ask Me, 2005), ActiveNet (Tacit, 2005), and Corporate Smarts' Intelligent Directory (Corporate Smarts, 2005). It is important to look at key phrases and not only keywords because sometimes a combination of keywords provides a more meaningful explanation. In ActiveNet, a user profile consists of a list of noun phrases from the sent items. In Corporate Smarts' Intelligent Directory, a term becomes searchable when it is used in email communication among a group of people. This term will then be added to the user's profile.

Admittedly, extracting key phrases that describe the individual's expertise from an email body poses an immense challenge. Emails are freestyle text, not always syntactically well formed, domain independent, of variable length, and on multiple topics (Tzoukermann et al. 2001). Moreover, the authors were unable to find an empirical evaluation on how effective these systems are in their key phrase extraction process from the email text. The potential key phrases extracted should give some sort of indication of skills and experience traded in the exchange of emails. Such key phrases ought to disclose skills such as technical expertise, management skills, industry knowledge, education and training, work experience, professional background, knowledge in subject areas and so forth. This requires an evaluation criterion that specialises in measuring the accuracy of these systems in terms of how many key phrases are correctly identified, in order to build a more accurate expertise profile.

4.2 Expertise Matching Gap

When a user queries the system, the system needs to match the user's needs with the expertise profiles by using retrieval techniques. It needs to measure similarity between an expert's expertise and a user's request. A search facility is usually provided for users to enter several keywords. However according to Liu (2003), it can suffer from the following drawbacks:

- Some relevant experts are missed
- Some irrelevant experts are retrieved
- Too many experts are retrieved
- Too few experts are retrieved.

These problems need to be addressed by correctly matching the user's needs with the expertise profiles to ensure that relevant experts are not overlooked and irrelevant experts are minimized.

4.3 Expertise Representation Gap

Following expertise matching, the system needs to represent the output to the user. The major drawback of most systems (Schwartz and Wood, 1993; Tacit's ActiveNet) is that the output is presented to the user with no relevant order. The reason behind this is the mechanism employed to rank the identified experts. McDonald and Ackerman (1998) distinguished between two stages in finding expertise within organizations, expertise identification and expertise selection. Some systems only go as far as expert identification through merely textual analysis. Rarely do they support expertise selection and this is an area for further development.

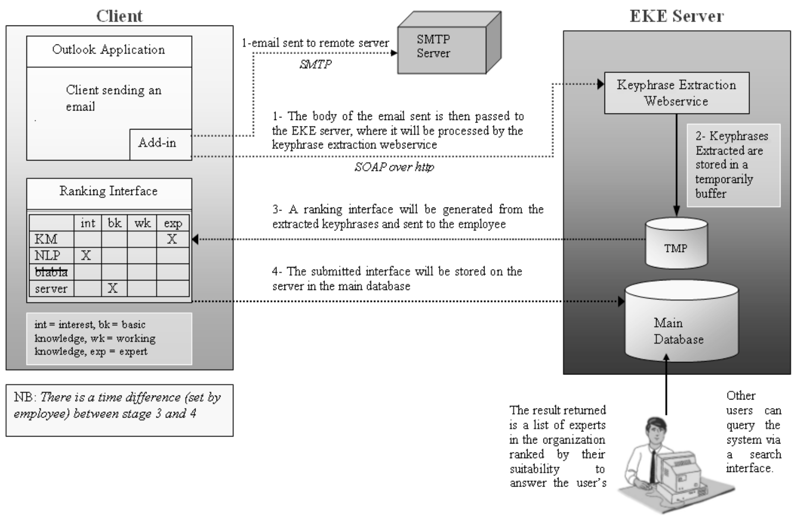
4.4 User Control Gap

Some systems provide the users with the facility to edit their profiles to reflect changes to their expertise. Others like Corporate Smart's allow their users to use system filters to allow its users to select the email message that they do not wish to include in the system sift. However, if a user fails to select a certain message, some of the personal interests which might be regarded as private by the users, could be published in the public domain. This situation requires system features that preserve and protect the privacy of the individual users through enabling them to control the system in how it uses their emails

4.5 Cultural and Management Gap

Although information technology can aid in storing, disseminating, and accessing lots of data and information, it neither creates or guarantees

Figure 1. EKE Generic Architecture



the ongoing creation of knowledge nor promotes knowledge sharing. Technology alone is not sufficient to achieve success (Cross and Baird, 2000). Many well-planned knowledge management (KM) initiatives have been unsuccessful as they fail to acknowledge the cultural and management change dimensions of KM. Changing organizational culture is not an easy task. The challenge is to get people sharing knowledge instead of hoarding it. Thus when embarking upon a KM programme, organisations need to tackle issues such as trust, privacy, motivation, and the barriers to sharing knowledge.

5. AN OVERVIEW OF THE PROPOSED SYSTEM

The primary aim of this research is to provide a fully automated and highly scalable system that uses the knowledge sent via email to ensure that:

- Expertise and knowledge is able to be located quickly and easily.
- Expertise and knowledge is available to the people who need it.

As the name suggests, Email Knowledge Extraction (EKE) is a tool that mines the information contained in employees' emails. EKE automatically finds interest areas by picking out key phrases from an employee's e-mail messages. For ethical and privacy reasons, and to overcome the user control gap, each individual has the option of authorising whether he/she wants his/her knowledge in each area made public.

This paper is a continuation of work reported in a previous submission by the authors (Jackson and Tedmori, 2003) to IRMA International Conference in which a pre-written program called KEA was used to extract the keywords from the email messages. It was noted, however, that after further analysis, the keywords extracted from KEA were occasionally incoherent and did not communicate knowledge fields within the organisation. In light of this, an alternative design is proposed which is concerned with modularity and reusability. Figure 1 shows the newly proposed generic structure of EKE.

One of the key elements of EKE is the ability to capture email messages before they are sent to the server, so individual keyword extraction profiles can be deployed rather than generic ones that apply to the whole organisation. Thus, there is a need to design "email interceptor software" that intercepts the messages before they are sent to the remote email server and retrieves the email content. A software plug-in will be used for this task.

In order to minimize processing overhead on the client machines, as soon as the email content is retrieved, the added plug-in will issue an http request to a web service passing to it the email content. On the server, the web service runs extracting key phrases from the email content and storing them in a temporary buffer. In order to build a good quality expertise profile and to overcome the expertise profile (model) gap, the web service has to be intelligent so that it extracts meaningful key phrases that identify knowledge holders within the organisation. The key is separating knowledge from noise. The extraction web service uses natural language processing. It picks key phrases purely based on the grammatical part of speech tags that surround these phrases, using a predefined set of rules. A rule is a sequence of grammatical tags that is likely to contain words that compose a key phrase. The approach used here does not use a controlled vocabulary, but instead chooses key phrases from the email text itself.

At a certain point in time, a server side application collates all of the extracted keywords and displays them to the user for their approval. The user has to specify the extracted keywords as private or public and rank them using a scale of three to denote their expertise in that field (e.g. basic knowledge, working knowledge, or expert). The keywords accepted by the user are then stored on a main database on the server. The keywords in the database can then be retrieved based on user's queries. Finally, the need to design an interface for searching the main database and an interface to output the results of the queries to the users comes into play. The result returned is a list of experts in the organisation ranked by their suitability to answer the user's query.

7. SUMMARY

The gap analysis model used in this paper has enabled information about applications in the expert finding systems domain to be consolidated and has identified gaps in existing technologies. The analysis has shown that the core of expert finding systems rely heavily upon the expertise profile model and, depending on how accurate the model is, determines the systems ability to match expertise. The key element behind the expertise profile model is its ability to extract relevant key phrases that match the sender's expertise.

The analysis has added to the body of knowledge within the expert finding domain and has enabled a proposed architecture to be presented for review.

REFERENCES

AskMe, http://www.askmecorp.com, accessed on 16-09-2005

Bannon, L. J. (1986), *Helping Users help each other*. In User Centered System Design, New Perspectives on Human-Computer Interaction', Lawrence Erlbaum Associates, Hillsdale, NJ.

Bishop, K. (2000), *Heads or Tales: Can Tacit Knowledge Really be Managed*. Proceeding of ALIA 2000 Biennial Conference, Canberra, pages 23-26.

Campbell, S.C., Maglio, P.P., Cozzi, A., Dom, B. (2003), *Expertise identification using email communications*. Proceedings of the twelfth international conference on information and knowledge management, New Orleans, LA, pages 528-531.

Cross, R. and Baird, L., (2000), *Technology Is Not Enough: Improving Performance by Building Organizational Memory*. MIT Sloan Management Review, Vol.41 No.3, pages 69-78.

Corporate smarts, http://www.corporatesmarts.com, accessed on 16-09-2005

Gibson, R. (1996) *Rethinking the Future*. Nicholas Brealey Publishing, London.

Jackson, T.W. and Tedmori, S. (2004), *Capturing and Managing Electronic Knowledge: The Development of the Email Knowledge Extraction*. Innovations Through Information Technology, Khosrow-Pour, M. (ed.), Idea Group, IRMA, New Orleans, USA, pages 463-466.

Kanfer, A., Sweet, J., and Schlosser, A. E. (1997), *Humanizing the net: Social navigation with a 'know-who' email agent*. Proceedings of the 3rd Conference on Human Factors and the Web, Denver, Colorado.

Kaurt, R. E. and Streeter, L. A. (1995), *Coordination in Software Development*. Communications of the ACM volume 38, Issue No. 3, pages 69-81.

Liu, P. (2003), *An Empirical Investigation of Expertise Matching within Academia*. PHD thesis, University of Leeds.

Mahowald R. P. and Levitt M. (2002), *Worldwide Messaging Applications Forecast and Analysis, 2002-2006: Getting the Message Out.* IDC White Paper.

McDonald, D.W. and Ackerman, M.S.(1998), Just talk to me: A field study of expertise location*. Proceedings of the 1998 ACM Conf. On Computer-Supported Cooperative Work, Seattle, WA. New York: ACM press, pages 315-324.*

Prieto-Diaz, R. (1987), *Domain Analysis for Reusability*. Proceedings of COMPSAC'87: The Eleventh Annual International Computer Software & Applications Conference, Tokyo, Japan, pp:23 – 29.

Sihn, W. and Heeren, F. (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, pages 279-283.

Schwartz , M. F. and Wood, D. C. M. (1993), *Discovering shared interests using graph analysis*. Communications of the ACM, volume 36, issue No.8, pages 78-89.

Tacit, http://www.tacit.com, accessed on 16-09-2005

Tzoukermann, E., Muresan, S., Klavans, J.L, (2001), *GIST-IT: Summarizing Email using Linguistic Knowledge and Machine Learning*. In Proceeding of the HLT and KM Workshop, EACL/ACL.

Wellins, R. S., Byham, W. C., and Wilson, J. M. (1993), *Empowered Teams: Creating Self-directed Work Groups that Improve Quality, Productivity, and Participation.* The Jossey Bass management, San Francisco.

Yimam-Seid, D. and Kobsa, A. (2003) *Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach*. In *Journal of Organizational Computing and Electronic Commerce* Vol.13, No.1, 1-24.

Related Content