Exploiting Email: Extracting Knowledge to Support Knowledge Sharing

Sara Tedmori

AstraZeneca R&D Charnwood
Bakewell Road
Loughborough
Leicestershire
LE11 5RH
United Kingdom

Centre for Innovative and Collaborative Engineering (CICE)
Loughborough University
Loughborough
Leicestershire
LE11 3TU
United Kingdom

Loughborough University

Thesis Access Form

Copy No …… **Location**……

Author Sara Tedmori

Title Exploiting Email: Extracting Knowledge to Support Knowledge Sharing

Status of access [OPEN / RESTRICTED / CONFIDENTIAL]

Moratorium period: ……years, ending …… / 200……

Conditions of access proved by (CAPITALS): ……

Director of Research (Signature) ……

Department of Civil & Building Engineering

Author's Declaration: I agree the following conditions:

OPEN access work shall be made available (in the University and externally) and reproduced as necessary at the discretion of the University Librarian or Head of Department. It may also be copied by the British Library in microfilm or other form for supply to requesting libraries or individuals, subject to an indication of intended use for non-publishing purposes in the following form, placed on the copy and on any covering document or label.

The statement itself shall apply to **ALL** copies:

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Restricted/confidential work: All access and any photocopying shall be strictly subject to written permission from the University Head of Department and any external sponsor, if any.

Author's signature …… **Date** ……

Users declaration: for signature during any Moratorium period (Not Open work):
I undertake to uphold the above conditions:

Date	Name (CAPITALS)	Signature	Address

Certificate of Originality

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgments or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a higher degree.

Author's signature .....

Date .....

EXPLOITING EMAIL: EXTRACTING KNOWLEDGE TO SUPPORT KNOWLEDGE SHARING

By
Sara Tedmori

A dissertation thesis submitted in partial fulfilment of the requirements for the award of the degree Doctor of Engineering (EngD), at Loughborough University

[May 2008]

AstraZeneca R&D Charnwood
Bakewell Road,
Loughborough,
LE11 5RH,
United Kingdom

Centre for Innovative and Collaborative Engineering (CICE)
Loughborough University
Loughborough
Leicestershire
LE11 3TU
United Kingdom

ACKNOWLEDGEMENTS

I would like to express my heartfelt thanks to a number of people who have supported me and made this research possible.

I would like to thank my academic supervisors: Dr. Thomas Jackson and Prof. Dino Bouchlaghem. Tom's continual guidance, support, and encouragement throughout the duration of this project have been invaluable. Dino's assistance and positive attitude have helped me stay on course. I consider it a privilege to have had the opportunity to work with them and share their valuable knowledge and expertise. Sincere thanks go to my ex-director of research, Prof Chimay Anumba, for his guidance and encouragement.

Many thanks go to AstraZeneca and Holger Adelmann for sponsoring this research and providing me with this opportunity.

I am also grateful to the CICE Centre Staff in particular Jo, Colette, Sara, and Nadine for the administrative assistance. I would also like to thank Jeff Pitchers for his help through out my research and Mr. Samra for his technical assistance.

I owe my loving thanks to my husband, Nijad, for helping me through the difficult times and reassuring me that I had done the right thing.

Special gratitude goes to my father, Jamil, my mother, Rawaa, and my brother, Dany. Their unconditional love and support kept me going and spurred me on to greater heights. I would not and could not have done this without them.

I would like to heartily thank: Ashraf Al-Najdawi for being my younger brother. Khusvinder Gill for his endless support during tough times until the present day, and for providing me right amount of coffee and lunch breaks. Steve Smith for his help and friendly cooperation.

Amro and Nuha for their long lasting friendship, assistance, and the fun times we have had over the years. Yiehya and Raghida for their support and for keeping my husband busy.

I also extend my thanks to my friends for reminding me there is more to life then work and providing me with the much needed distractions. Special thanks go to everyone at Holywell Park and especially to Jatinder Gill, Lezan Hawizy, Alexanda Alecu, Senay Mihcen, Hisham Abu-Saymeh, Tarek Alhmiedat, Karen Yunqui LI, and Yanning Yang.

I am also indebted for my long distance family members including my two lovely aunts, Rouba and Lina; my beautiful cousin Rasha and her three baby girls that provided my mom with the needed distractions whilst I was away; my uncle Hasan for his support and for the fantastic holiday.

Dedicated to my father Jamil,
my mother Rawaa,
my brother Dany,
and my loving husband Nijad
For their love, inspiration, and belief

ABSTRACT

Effective management of knowledge assets is key to surviving in today's competitive business environment. This is particularly true for large organisations, where employees have difficulties identifying where or with whom the knowledge lies. Expertise is one of the most important knowledge assets and largely resides in the heads of employees. Many attempts have been made to help locate employees with the right expertise; however, the existing systems (often referred to as expertise finding systems) carry several flaws. In organisations, there are several potential sources where expertise evidence might be found. These sources have been used by the existing approaches to profile employees' expertise. Unfortunately, there has been limited research showing whether these sources contain useful evidence of expertise. Moreover, the majority of existing approaches have not been designed to integrate with the organisations' work practices; nor have they investigated the socio-ethical challenges associated with the adoption of such systems. Therefore, there is a need for expert finding systems that utilise useful sources of expertise and integrate into existing work practices. Through industry involvement, this research has explored and validated email content as a source for expertise profiling. This thesis provides an overview of the traditional and current approaches to expertise finding. The development and implementation of the EKE (Email Knowledge Extraction) system which tries to overcome the aforementioned challenges is presented. EKE has been evaluated by end-users from both industry and academia. The evaluation results suggest that EKE is a useful system that encourages participation, and that in many cases may assist in the management of knowledge within organisations.

KEY WORDS

Email, Expertise Identification, Expertise Location, Knowledge Management, Keyphrase Extraction, Socio-ethical.

PREFACE

This thesis is a result of the research conducted between 2004 and 2008 in partial fulfilment of the requirements of an Engineering Doctorate (EngD) at the Centre for Innovative and Collaborative Engineering (CICE), Loughborough University. The research programme was supervised by CICE, funded by the Engineering Physical Sciences Research Council, and sponsored by AstraZeneca, a major international pharmaceutical firm.

The EngD is a four year, industrially relevant doctoral training programme. The EngD offers a radical alternative to the traditional PhD, geared to training the research managers of the future. It provides candidates with a good balance of academic training, industry experience, and high level research. The EngD is examined on the basis of a thesis containing at least three (but not more than five) research publications and/or technical reports. This discourse is supported by one journal paper, three conference papers, and one un-submitted paper. For ease of reference, the papers have been numbered from 1 to 5 and are located in Appendices A to E. These papers are an integral part of, and should be read when referenced in conjunction with, the thesis.

USED ACRONYMS/ ABBREVIATIONS

BK	Basic Knowledge
CICE	Centre for Innovative and Collaborative Engineering
EK	Expert Knowledge
EKE	Email Knowledge Extraction
EngD	Engineering Doctorate
FG	Focus Group
KEWS	Keyphrase Extractor Web Service
LC	Leicestershire Constabulary
LR	Literature Review
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
P	Prototyping
PHP	PHP: Hypertext Preprocessor
POS	Part of Speech
Q	Questionnaire
R&D	Research and Development
SMTP	Simple Mail Transfer Protocol
TFxIDF	Term Frequency Inverse Document Frequency
UBS	United Bank of Switzerland
VSTO	Visual Studio Tools for Office
WK	Working Knowledge

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 INTRODUCTION

This chapter provides an introduction to the subject of expert locators, conducted in partial fulfilment for the award of Engineering Doctorate at Loughborough University. It presents the background to the research undertaken, sets out the context of study and offers justification for the research. The main aim and objectives of the project are set out, before describing the remaining structure of the thesis. A summary is provided for each of the papers that have been published over the four years of the EngD. These papers should be read in conjunction with the discourse.

1.2 BACKGROUND TO THE RESEARCH

"What is knowledge?" The definitions and debates about the nature of knowledge can be traced back to the days of Plato and Aristotle. Despite the fact that we all have a fairly adequate everyday notion of what knowledge is, *knowledge* has proved an elusive concept (Jashapara, 2004) that is hard to define.

Knowledge is often undistinguishable from information or data; hence, the three terms are often used interchangeably (Davenport & Prusak, 1998). There have been numerous works trying to clearly distinguish between these concepts, due to the fact that organisational success and failure is often dependent on knowing what you have, what you need, and what you can and cannot do with each of them (Awad & Ghaziri, 2004; Nonaka & Takeuchi, 1995; Davenport & Prusak, 1998; Alavi & Leidner, 1999). The efforts to differentiate between data, information, and knowledge have resulted in a plethora of definitions.

Alavi and Leidner (1999) argue that one cannot differentiate between information and knowledge based on the content, structure, accuracy, or utility of the supposed information or knowledge. Rather, information becomes knowledge once it is processed in the minds of individuals, and becomes information again once it is articulated and communicated to others. The recipient can then cognitively process this information and convert it back to knowledge. Unlike data (raw facts and figures), knowledge is continuously recreated and reconstituted through dynamic interactive and social networking activities (Newell et al., 2002).

Humans must do all the work to derive knowledge from information as they derive information from data. However, while data can be found in records and transactions, and information in messages, knowledge is obtained from individuals or group of knowledge holders, or sometimes in organisational routines (Davenport & Prusak, 1998).

The most common notion of knowledge in the literature (Desouza, 2003; Duffy, 1999; Newell et al., 2002; Nonaka & Takeuchi, 1995; Tiwana, 2000; Zack, 1999a) is based on the writings of Gilbert Ryle (1949) and Michael Polanyi (1966) who distinguished between two types of knowledge: tacit and explicit.

- Tacit or implicit knowledge (know-how): is in our heads and is difficult to express, thus making it difficult to communicate and share with others. Tacit knowledge is personal and deeply rooted in each individual's actions,

experiences, ideas and emotions including the individual's hunches, skills, talents, intuition, and so on. The human mind is the storage medium of tacit knowledge. Tacit knowledge can be obtained from organisational routines and culture.

- Explicit knowledge (know-what): is represented by a physical medium (e.g. books, documents, paper). It can be documented, articulated, and easily communicated to others formally and systematically in the form of data, scientific formulae, product specifications, manuals, and universal principles. Explicit knowledge is easier to identify than tacit knowledge.

The concept of knowledge is not new. What is new is the treatment of knowledge as the key corporate asset that must be managed (Beijerse, 1999; Davenport & Prusak, 1998; Zack, 1999b), in order to gain the ultimate competitive advantage in today's organisations. According to Davenport and Prusak (1998), there are three factors that contribute to the inefficiency of how knowledge is managed in organisations.

- Information about where to find knowledge is often incomplete.
- The same knowledge can be found in many different places and at different levels of detail.
- People prefer asking a person in the office next door rather than locating someone elsewhere in the organisation more suited to answer their questions.

The premise of knowledge as an asset that needs to be carefully managed has led to an explosion of knowledge management (KM) initiatives within organisations. KM is concerned with allowing employees to collaborate and innovate by providing them with the necessary knowledge to effectively fulfil their role (Havens & Knapp, 1999). As companies try to implement KM strategies, they learn that due to the unique characteristics of knowledge, managing knowledge is different and more difficult than managing tangible assets. Stewart et al. (2000), identify three of these unique characteristics. Firstly, knowledge is intangible and can go unnoticed or forgotten, unless regularly updated or replaced to assure that the knowledge is accurate, valid, and reliable. Secondly, the useful life of knowledge is not so obvious, in comparison to the useful life of a physical asset which is usually predictable and manageable. Finally, the scope of knowledge is limited, focused, and closely tied to the specific type of activities to which it is applied.

Success and innovation in organisations hinges on many factors, including the ability to get answers to queries quickly and effectively. However, as modern organisations continue to generate vast amounts of valuable information, the number and complexity of information sources used to store this information grows, making it difficult to locate and utilise in a meaningful way. This, in turn, increases the challenge of locating critical information efficiently. It has been estimated that the volume of data stored in the worlds databases doubles every 20 months (Witten & Frank, 1999). Moreover, it has been estimated that knowledge workers spend fifteen to thirty five percent of their work days looking for information, and at least 50 percent of online searches are not successful (Feldman, 2004). As a consequence, in order to achieve a greater competitive advantage, businesses need tools to aid their employees to locate relevant information in a speedy manner.

Motivated by the advances in information technology, a quick fix solution is to store all information in a database that can be reused by current and future employees. Methods of querying a database, including search engines, offer appropriate solutions to many common queries. However they often fail to give adequate responses to questions that require insight or thinking. Knowledge workers are often left frustrated trying to find the required information. Ultimately, the overall productivity of the organisation is hampered by the lack of effective and speedy access to the required information. Stewart (1997) agrees that the attempt to store all corporate knowledge on one huge server is destined to fail and continues to argue that the true value of information systems is connecting people to people, so they can share what expertise and knowledge they have.

Expertise, one of the most important knowledge assets, is defined by Hayes-Roth et al. (1983) as "*knowledge about a particular domain, understanding of domain problems, and skill about solving some of these problems.*" Expertise is normally stored in people's heads and is difficult to codify. The work of Stenmark (2001) notes that the codification process often fails due to three reasons. Firstly, people are not completely aware of their tacit knowledge. Secondly, on a personal level people do not need to make tacit knowledge explicit in order to use it, since they are capable of using it without thinking. Thirdly, if tacit knowledge provides an important competitive advantage, then there is little reason to share it with others. Moreover, should people decide to make their tacit knowledge explicit, not only would they find it difficult and time consuming, they will discover that they will not directly benefit from it. If tacit knowledge is codified, it is most likely to result in knowledge getting out of date. In addition, people may experience problems in understanding the codified knowledge due to the fact that documented knowledge lacks contextual richness (Lyon, 2000).

Studies on information seeking behaviour have continuously shown that people searching for information favour asking people in their personal communication network before using formal sources (Bannon, 1986; Kraut & Streeter, 1995). When employees easily and quickly communicate insights and knowledge with their colleagues, they drive business success. This is why it is crucial for businesses to provide their employees with the tools that enable them to do so. To assist in the attainment of this goal, the idea of expertise locators which can help get the right information to the right people at the right time gained favour.

1.3 PROBLEM OVERVIEW

Many organisations invest in their knowledge assets by recruiting knowledgeable people and then enhancing this investment by training them. Companies hire for experience more often than for intelligence or education because they understand the value of knowledge that has been developed and proven over time (Davenport & Prusak, 1998). The significance of knowledge as an asset becomes clear when a knowledgeable employee leaves the organisation taking with them, their skill, intuition, insight, wisdom, experience, and informal social networks. According to the Delphi Group's study on more than 700 U.S. companies, almost (42%) of corporate knowledge is tacit, locked inside employees' heads (Hickens, 1999). Expertise, such as skills and know-how, is a key component of tacit knowledge. Expertise is shared when people communicate with each other, otherwise it will go unnoticed.

Employees are often working on very similar projects without even realising it. Almost everyday, they are faced with problems that they need to solve in order to accomplish their work. Some of them try to find answers to their problems by searching for information on the web, however this process can be very frustrating and time consuming, and despite the huge number of search engines, most people don't know where to look (Vroom et al., 2004). Moreover, most valuable information is not documented and thus cannot be catalogued and indexed in organisational repositories, and can only be obtained by asking the right people (Yu & Singh, 2003). Thus, the knowledge held by employees needs to be captured and made available to others to learn from in order to build on past experiences and practices, instead of constantly reinventing the wheel. If people are able to find the relevant person or group of people to ask whenever they have a problem, significant time and effort can be saved. The problem of finding the right information and managing information in corporate databases can be reduced to the problem of finding the right people to refer to.

A relatively new approach to expertise location is based on keyphrase identification from electronic mail (email) messages. Email is part of our daily lives, it has achieved wide spread acceptance for personal communication and has become a fundamental prerequisite for doing any kind of business (Bertolotti & Calzarossa, 2001). However, as more organisations adopt email as their primary method of communication, the majority neglect the fact that email content contains data of business decisions, actions and transactions. These emails are subject to legal restrictions, and as such the data they contain are not exploited. Even though the sender and receiver can easily store and index the messages for future reference, in most cases, the knowledge within an email is only shared between the sender and receiver, and is therefore not fully utilised within an organisation. Both public and private organisations are discovering that they have an obligation to preserve the knowledge that resides in emails in order to make use of it and an equal obligation towards their employees' personal privacy.

1.4 THE INDUSTRIAL SPONSOR

AstraZeneca is a major international healthcare business, and one of the top five pharmaceutical companies in the world, engaged in the research, development, manufacture and marketing of prescription pharmaceuticals and the supply of healthcare services. The company employs over 12,000 research and development (R&D) personnel and spends a total of £7 million a day on researching and developing new medicines designed to fight disease in important areas of healthcare. Within the UK its R&D functions are situated across eight sites, with the research described in this thesis associated primarily with the Charnwood R&D site in Leicestershire.

The company prides itself on its creativity and its innovative approach to drug development, the mantra taken from the AstraZeneca website reads:

"At AstraZeneca, innovation is about more than just research. We aim to stimulate continued creativity throughout our organisation by maintaining a culture in which our people feel valued, energised and rewarded for their ideas and contribution to our success - ideas which can make a difference in all aspects of our business"(AstraZeneca, 2007).

There has been an increasing recognition of the importance of collaboration throughout the business. As a result, the business has undergone a restructuring process to further enable collaborative working patterns. The result of this restructuring is a flatter organisational structure substituting the earlier hierarchical structure which was believed to hinder effective drug development.

AstraZeneca R&D Clinical at Charnwood is committed to producing high quality medicines. They have recognised that KM may be able to offer tangible benefits within the Clinical environment of the company. KM is designed to capture, disseminate, and exploit knowledge within an organisation. However, the technological advances over the last 20 years have made a significant impact on knowledge management. This is why organisations such as AstraZeneca must continuously adopt new technologies and improve their practices, in order to increase the likelihood of innovation.

1.5 SCOPE OF THE RESEARCH

This research project is primarily concerned with exploring whether email content can be exploited to locate expertise. Although various expert locator systems have been proposed throughout the literature, there is very little evidence showing how these systems will perform in the real world. Email has become the primary means of communication in many organisations, as it enables the transfer of knowledge quickly and effectively. This knowledge, if properly used, can help expertise seekers in finding experts. It also provides an appropriate platform for overcoming some of the limitations of conventional approaches to expert finding (as discussed in section 2.3).

1.6 AIM AND OBJECTIVES

The aim of this research is to explore the potential of using email messages to determine and locate expertise within organisations.

The specific objectives are to:

- review literature on approaches to expert finding;
- determine the potential value of email in supporting the process of expert finding;
- develop an expert locator system, which is named EKE (Email Knowledge Extraction);
- evaluate the use of EKE; and
- identify enablers and barriers to the use of EKE as an expertise locating system.

1.7 RESEARCH DESIGN

The research project involved several stages, namely the concept development stage, the system building stage, the system evaluation stage, and a concluding stage that addressed key socio-ethical challenges involved in the implementation and use of the system. Details of the adopted research methodology can be found in Chapter 3. The findings from each stage of the research were presented at international conferences and in academic journals that were chosen in accordance to the topic that each stage

addressed. This was done with the intention of seeking feedback from practitioners and researches in the field, that may aid in formulating and validating the concepts studied.

1.8 STRUCTURE OF THE THESIS

This thesis is organised into five chapters, structured as follows:

Chapter 1 describes the background to the research, defines its aim and objectives, discusses the justification and scope of the research and introduces the steps undertaken for this research. The chapter concludes by outlining the papers completed during the course of this research, and included in this thesis.

Chapter 2 reviews related work on the subject area of expert finding. This review includes topics related to the evolution of expert finding approaches, expert finding applications exploiting email, and keyphrase extraction and evaluation techniques. Chapter 2 also highlights the contribution of the research to the area of finding experts from email messages.

Chapter 3 describes and justifies the research methodology chosen for this research. It also presents and justifies the research methods used to achieve specific objectives.

Chapter 4 presents the research undertaken to meet the aim and objectives, and associated task breakdown. This includes details of the development of EKE, its deployment at the Department of Information Science at Lougborough University, and its evaluation.

Chapter 5 presents the main research findings, with a particular focus on EKE's evaluation.

Chapter 6 presents the impacts and implications of the research on the project sponsor and the wider industry. It suggests future work and provides a conclusion derived from the research.

Appendix A to E contain five peer-reviewed published papers, which are the primary outcome of this research. These papers are an integral part of the thesis and thus should be read alongside the discourse. Where appropriate, references to the papers are provided throughout the main body of the thesis. Full bibliographical references are provided at the beginning of each publication in the appendix.

Appendix F includes other support information such as the questionnaires administered to police offers at Leicestershire constabulary, academics at Loughborough University, and participants during the focus group sessions. A report detailing the complete analysis of the focus group discussions and questionnaire analysis concludes this section.

Appendix G provides a reference list of the publications that resulted from this research project.

1.9 SUMMARY OF PAPERS

Table 1.1 summarises five of the publications resulting from this research and included in this thesis. The table contains information regarding the title, status of the paper, and the place of publication of each paper. In addition, a brief description showing how each paper contributed to the fulfillment of the research aim and objectives is provided. Each

paper can be identified by the unique document number assigned to it and by its location within the appendices.

Table 1.1: Summary of Papers

ID	Title	Journal/ Conference	Status	Description
Paper 1 Appendix A	Building a Tool for Expertise Discovery	Proceedings of the 17th Information Resources Management Association International Conference, Washington D.C., USA	Published	This paper discusses the evolution of expert finding approaches, with a particular reference to solutions that exploit email sources. It identifies related gaps to provide a more informative and wider understanding of the current state of play. The paper concludes with an introduction to EKE.
Paper 2 Appendix B	Information Seeking and Sharing Behaviour of a UK Police Force	Proceeding of the 8th European Conference on Knowledge Management, Barcelona, Spain	Published	This paper documents the analysis of a questionnaire survey that looks at how individuals at Leicestershire Constabulary seek information and how they share information once it's located. This is important in order to determine if an expert locator system would work within their organisation.
Paper 3 Appendix C	Locating knowledge sources through keyphrase extraction	Knowledge and Process Management Journal	Published	This paper presents an automated process for keyphrase extraction from email messages for the purpose of identifying and capturing expertise. The effectiveness of the extraction system is tested.
Paper 4 Appendix D	Expertise Location Using Keyphrases in Electronic Mail: Socio-ethical Challenges	Proceedings of the Australian Conference for Knowledge Management & Intelligent Decision Support	Published	This paper addresses the key socio-ethical challenges involved in the implementation and use of email expertise locator system, EKE.
Paper 5 Appendix E	Evaluation of an Email Knowledge Extraction System		Unpublished	This paper presents a brief overview of the EKE system evaluation study conducted in the Information Science Department at Loughborough University. The socio-ethical challenges associated with EKE's adoption are also explored.

1.10 SUMMARY

This chapter has provided a general introduction to the subject domain and justified the need for the research. The structure of the thesis was presented and an outline provided of each of the published papers that are to be read alongside the discourse. Chapter 2 details the background to the research.

2 REVIEW OF RELATED WORK

2.1 INTRODUCTION

This chapter provides a background against which the proposed study was conducted. It details the results of the literature review carried out, identifies the gaps in the literature.

The chapter commences by briefly acknowledging the need to connect people to other people so that they can collaborate. The subject of expert finding is explored, with emphasis on the role of email in knowledge work and its potential value for expertise location. Keyphrase extraction and evaluation techniques are discussed due to their recognised connection to the subject domain.

2.2 THE NEED TO CONNECT PEOPLE TO PEOPLE

Over the last several decades, many reports (Hiltz, 1985; Lang et al., 1982; Mintzberg, 1973; Pelz & Andrews, 1966; Allen, 1977) have indicated that people searching for information prefer to consult other people, rather than to use on-line or off-line manuals. Allen (1977) found that engineers and scientists were roughly five times more likely to consult individuals rather than impersonal sources such as a database or file cabinet for information. In spite of the advancements in computing and communications technology, this tendency still holds; people remain the most valued and used source for knowledge (Cross & Sproull, 2004; Kraut & Streeter, 1995).

Unfortunately, finding individuals with the required expertise can be extremely expensive (Maltzahn, 1995; Campbell et al., 2003), as it is time consuming and can interrupt the work of multiple persons. Moreover, a common problem with many businesses today, large and small, is the difficulty associated with identifying where the knowledge lies. A lot of data and information generated and knowledge gained from projects reside in the minds of employees. Therefore the key problem is, how do you discover who possesses the knowledge sought?

In the search for the solution, information systems have been identified as key players with regards to their ability to connect people to people to enable them to share their expertise and collaborate with each other (Bishop, 2000; Cross & Baird, 2000; Gibson, 1997; Lang, 2001). Thus, the solution is not to attempt to archive all employees' knowledge, but to link questions to answers or to knowledgeable people, who can help find the answers sought (Stewart, 1997). This has led to the interest in systems, which help connect people to others that can help them solve their problems, answer their questions, and work collaboratively.

Cross et al. (2001) reviews (Allen, 1977), (Burt, 1992), (Erickson, 1988), (Schön, 1993), (Walsh, 1995), (Weick, 1979), (Weick, 1995), (Blau, 1986), (March & Simon, 1958), and (Lave & Wenger, 1991) and summarises the benefits of seeking information from other people. These benefits include:

- provision of solutions to problems;
- provision of answers to questions;
- provision of pointers to others that might know the answer;

- provision of pointers to other useful sources;
- engagement in interaction that helps shape the dimension of the problem space;
- psychological benefits (e.g. confidence, assurance);
- social benefits (e.g social approval for decisions, actions);
- improvement in the effectiveness with which a person advances their knowledge in new and often diverse social contexts;
- improvement in efficiency (e.g. reduction in time wasted pursuing other avenues); and
- legitimation of decisions.

Cross (2000) identifies five categories that these benefits fall under: (1) solutions (know what and know how); (2) meta-knowledge (pointers to databases or other people); (3) problem reformulation; (4) validation of plans or solutions; and (5) legitimation from contact with a respected person.

Now that it has been demonstrated that the idea of connecting people to people has been recognised as the way forward for the information community to overcome some of its inherent problems (e.g. finding the right information), the next step is to examine the expert finding approaches in more detail.

2.3 TRADITIONAL EXPERT FINDING APPROACHES

The traditional means of providing automated expert assistance is through the development of expert databases (also known as “knowledge directories”, "knowledge maps"). In expert databases, users have to manually register and enter details about their areas of expertise. The literature reveals that these traditional help systems suffer from numerous shortcomings (Maltzahn, 1995; Yimam-Seid & Kobsa, 2003), including:

- the significant waste of organisational resources (especially financial and time resources) due to the creation and maintenance (i.e. updating) of expert databases;
- the dependency of these systems on the willingness of the users to regularly allocate extra time to provide detailed description of their expertise and update their profiles regularly to avoid becoming outdated and inaccurate;
- the rapid rate of change in the modern organisations’ environments resulting in expert databases becoming quickly outdated;
- the absence of temporal time dimension which takes into account whether the expert is currently practicing the required skill or whether they were involved in it years ago;
- the lack of information on the experts availability;
- the simple keyphrase search facility provided with expert databases, do not sufficiently meet users’ search goals;
- the user’s difficulty to identify specific areas of expertise; and

- the incomplete and general expertise descriptions on one hand, and the specific, qualitative expert related queries on the other hand.

The emergence of the World Wide Web has led to the development of personal web pages where individuals provide information about themselves and about their interest/ expertise areas. Conducting a web search may lead us to such web pages. However, this is dependent on whether the keyphrases queried are present in the experts' web pages. In addition, experts need to regularly update their pages to reflect the changes in their expertise profile. Yimam-Seid and Kobsa (2003) argue that using a search engine to trace an expert may be ineffectual, because of the large number of hits returned. This can be problematic in that a significant period of time is required to traverse the results returned, select the most appropriate expert, and determine their accessibility. Moreover, it is not possible to ascertain the validity of the content. The shortcomings of the traditional expert finding systems alongside the advancements in information technology have resulted in the introduction of systems that automate or semi-automate the process of finding the right expert.

2.4 AUTOMATED EXPERT FINDING APPROACHES

Expertise location systems connect people to people and link people to information about people. The literature review has revealed that there have been numerous attempts to semi-automate/automate the process of finding the right experts, including: HelpNet (Maron et al., 1986), Expert/Expert-locator (Streeter & Lochbaum, 1988), ContactFinder (Krulwich & Burkey, 1996), Agent Amplified Communication (Kautz et al., 1996), Yenta (Foner, 1997), Phoaks (Terveen et al., 1997), Expertise Browser (Cohen et al., 1998), InfoScout (Prasad, M. V. Nagendra & Anagnost, 1999), MEMOIR (Pikrakis et al., 1998), Expertise Recommender (McDonald & Ackerman, 1998), MIT's Expert Finder (Vivacqua, 1999), SAGE (Becerra-Fernandez, 2000) and the KCSR Expert Finder (Crowder et al., 2002). The purpose of developing these systems was to catalog knowledge competencies, incorporating information source(s) not typically captured by human resources systems, in a way that allows the information to be queried at a later date from across the organisation (Becerra-Fernandez, 2006).

One main feature that distinguishes expert finding systems from each other is the information source(s) that they use as the basis for expertise recognition. These information sources include: emails (Kanfer et al., 1997; Kautz et al., 1997), bulletin boards (Krulwich & Burkey, 1996), program codes (McDonald & Ackerman, 1998; Vivacqua, 1999), Web pages (Cohen & Prusak, 2001; Foner, 1997; Kautz et al., 1997), and technical reports (Crowder et al., 2002; Mattox et al., 1999; Streeter & Lochbaum, 1988). As the main focus of this research is to investigate the potential of email as a source for expertise profiling, a comprehensive review of how these expert finder tools work is outside the scope of this thesis. For such a review, please refer to (Yimam-Seid & Kobsa, 2003).

2.5 EMAIL, KNOWLEDGE WORK, AND EXPERTISE LOCATION

This section explores the role of email in knowledge work, focusing on its potential value for expertise location. Email is an important knowledge channel (Lichtenstein, 2004), and collaboration tool (Garcia, 2006), actively used by organisational knowledge

workers worldwide. In December 2006, email was reported as the most common Internet activity, used daily by ninety-one percent of U.S. internet users (Pew Internet & American Life Project, 2007). The organisational dominance of email is also demonstrated by a recent study reporting that email messages involve an estimated average of 15.8 megabytes of archive storage per corporate end-user per day (Derrington, 2006).

Despite email's popularity and ubiquity, there is little research on the value that email provides to organisational KM strategies. Email enables greater knowledge work than possible in earlier technological eras (Jackson & Burgess, 2003; Whittaker et al., 2005). It enables knowledge creation (Ducheneaut & Belloti, 2003), knowledge sharing and knowledge flow (Bontis et al., 2003). According to Lichtenstein & Swatman (2003), employees are motivated to use email for knowledge work for reasons including:

- email messages attract workers' attention;
- email is well integrated with everyday work;
- email discourse provides a context for sense-making about ideas, projects and other types of business knowledge;
- email enables the referencing of work objects (such as digital documents), and provides a history via quoted messages;
- email's personalised messages are appealing, meaningful and easily understood;
- email encourages commitment and accountability by automatically documenting email exchanges;
- email is collected in inboxes and organisational archives, email represents valuable individual, collective and organisational memories that may be tapped later; and
- email discourse facilitates the resolution of multiple conflicting perspectives which can stimulate an idea for a new or improved process, product or service.

Email provides several important, often unexploited, opportunities for expertise-finding. Knowledge in email can be accessed and reused directly (Swaak et al., 2004) or can serve indirectly as a pointer to an expert (Balog & de Rijke, 2007; Campbell et al., 2003). A recognised definition of an expert is someone who possesses specialised skills and knowledge derived from training and experience (Shanteau & Stewart, 1992).

Traditionally, email clients were designed for the reuse of personal knowledge archives. For example, folders are popular structures for organising email messages so that they assist owners with knowledge reuse. This facility was highlighted by a recent study of Enron's publicly available email archive, where significant folder usage was employed (Klimt & Yang, 2004). Employees often search personal email archives seeking knowledge, in preference to searching electronic knowledge repositories (EKR) (Swaak et al., 2004), raising questions about the effectiveness of EKRs for reuse, an issue first raised by (Markus, 2001). Swaak et al.'s (2004) study also found that employees prefer to find an expert to help them with their knowledge-based concern, rather than searching directly for knowledge. The next section describes automated expert finder tools that exploit email as evidence of expertise.

2.6 EXPERTISE LOCATORS EXPLOITING EMAIL

This section reviews several existing approaches that exploit email for expertise location, and highlights their deficiencies. In practice, very few organisations exploit their email message content and very few software applications enable such exploitation.

Schwartz and Wood (1993) were the first to exploit email to deduce shared-interest relationships between people. To overcome privacy concerns, the structure of the graph formed from "From/To" email logs is analysed, using a set of heuristic graph algorithms. The output of the system was a list of people with no essential order. A user searches for people by requesting a list of those whose interests are similar to several others known to have the interest in question. This implies that the user should have a prior social network with the appropriate contacts relevant to their queries and that a novice can not properly take advantage of the system.

The *Know-who* (Kanfer et al., 1997) system as is an email agent that helps manage the information users receive through emails. *Know-who* maintains a list of all those whom the user received emails from. Based on the content of email communications 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 can only identify individuals within the user's social network. This makes it unfeasible to identify individuals outside the user's network with common interests, thus impeding the process of expertise assistance.

Sihn and Heeren (2001) presented a system called *XpertFinder*, which analyses email communication of logged-in 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 will complete 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.

Current commercial systems for expert identification by email include: Tacit's *ActiveNet* (Tacit Software, 2005), *AskMe Enterprise* (AskMe, 2005) and Corporate Smarts' *Intelligent Directory* (Corporate Smarts, 2005). Tacit Software (formally known as Tacit Knowledge Systems) is a firm with a product that purports to transform enterprise email into a shared knowledge resource. The company's knowledge-email product ("ActiveNet", known previously as "KnowledgeMail") automatically processes users' emails (unpublished expertise) and artifacts (published expertise) for the creation of user profiles. However, ActiveNet only allows employees to store their public profile of expertise. When a user enters a request for expertise, the system displays the available employees' contact details in order of closest match with the user's query. The strength of the match is determined by the frequency, intensity, and history of the topic within the person's expertise profile. ActiveNet does not enlist knowledge within the electronic mail messages to make decisions about expertise, but adopts a more

subjective approach by soliciting expertise from users themselves, thus allowing employees to nominate themselves as experts if so desired.

Campbell et al. (2003) proposed a system for email expertise extraction (abbreviated as e^3) that exploits email content and communication patterns. e^3 locates all email messages on a topic and builds an expertise graph by analysing the email messages exchanged between every sender and recipient pair for the topic correspondence. The expertise graph can be analysed by employing a modified version of the Hyperlink-Induced Topic Search (HITS) algorithm to identify experts. However, the size of the networks studied is very small and does not reflect the characteristics of social networks in practice (Zhang et al., 2007).

Next how email messages can point to experts by reviewing several representative approaches based on keyphrase identification is illustrated.

2.7 KEYPHRASE EXTRACTION TECHNIQUES

As the approach taken in this thesis is dependent on extracting keyphrases from email content, this section introduces its importance and outlines the major techniques used in the literature. The main technical challenge when utilising email content for expert identification is the extraction of keyphrases that provide a good indication of the sender's skills and experience. Such keyphrases ought to disclose skills such as technical expertise, management skills, industry knowledge, education and training, work experience, professional background, knowledge in subject areas, etcetera. To date, current systems in the marketplace have failed to achieve this objective, mainly due to the technical difficulty in identifying keyphrases that represent an email sender's knowledge areas, as emails are freestyle text, not always syntactically well formed, domain independent, of variable length, and based on multiple topics (Tzoukermann et al., 2001). Moreover, knowledge is not necessarily represented in one message, but in an email thread discussion.

Several methods have been proposed for the automatic extraction of keyphrases (Barker & Cornacchia, 2000; Frank et al., 1999; Krulwich & Burkey, 1996; Turney, 2000). The two main techniques are domain dependent and domain independent. Domain dependent techniques employ machine learning and require a collection of documents with keyphrases already attached, for training purposes. Furthermore, the techniques (both domain dependent and domain independent) are related to linguistics and/or use pure statistical methods. A number of applications have been developed using such techniques.

There are many weaknesses with the current approaches to automatic keyphrase identification, several of which are discussed here. The extraction of noun phrases from a passage of text is common to most approaches (Hulth, 2003; Tzoukermann et al., 2001). However, a disadvantage of the noun extraction approach is that, despite the application of filters, many extracted keyphrases are common words likely to occur in numerous emails in many contexts. Therefore it is important to distinguish between nouns that comprise of common words and more specific nouns that are more likely to represent the keyphrases of interest. Moreover, Hulth (2003) pinpoints two common drawbacks with existing algorithms. The first drawback is that the number of words in a keyphrase is limited to three. The second drawback is that the user must state the number of keyphrases to extract from each document (Hulth, 2003). A thorough

discussion of existing approaches, together with their merits and pitfalls, is provided Paper 3, section 2, Appendix C.

2.8 KEYPHRASE EVALUATION TECHNIQUES

There are two main approaches to evaluating automatically generated keyphrases (Jones & Paynter, 1999). The first approach adopts the standard information retrieval metrics of precision and recall to reflect how well the phrases generated by the system match keyphrases that are considered to be 'relevant' (e.g. author phrases, the list of phrases usually found at the beginning of many articles such as academic journals). This set of relevant keyphrases is also known as the 'Gold Standard' keyphrases.

The second approach gathers subjective keyphrase assessments from humans. Previous studies involving human phrase assessment (Barker & Cornacchia, 2000; Chen, 1999; Turney, 2000) follow essentially the same methodology. Subjects are provided with a document and a phrase list and asked to assess in some way the relevance of the individual phrases (or of sets of phrases) to the given document. A study by Jones and Paynter (1999), shows that authors do provide good quality keyphrases and thus can be used as the 'Gold Standard' against which other keyphrases can be compared.

2.9 CONTRIBUTION TO KNOWLEDGE

Despite ongoing attempts at utilising emails as an expertise profiling source, the research into email expert locators as it stands today suffers from many shortcomings. No one has taken a step back to access the situation and investigate the appropriateness of email content for expertise location by confirming the claim that valid expertise indicators exist in email content. Previous research in the academic arena primarily consists of theoretical studies (i.e. not followed by real world testing), with the available literature being very unclear on the exact functionality of the existing tools. As such, little consideration has been given to how well these systems perform in relation to identifying experts and to the factors that govern the mainstream adoption of these applications. Moreover, the research available on the performance of keyphrase extraction algorithms shows that there is significant scope for improvement in terms of the percentage of keyphrases successfully identified. Commercial solutions on the other hand remain outside the public and research domains and as such can not be used to validate the theoretical studies. There is also limited material addressing the potential sensitivities of employees with respect to their identification by the systems as topical experts. The author has contributed to overcoming this problem by publishing the findings of this research in academic outlets. This research will significantly add to the body of knowledge available by addressing the shortcomings outlined above. In this research:

- It has been shown that people are still having difficulties identifying sources of new information.
- The majority of people queried recognise that the EKE concept would be of benefit to them.
- Key areas to consider when designing systems similar to EKE have been identified.

- The author has experimentally shown that email contains valid expertise indicators making it possible to profile users' knowledge based on the emails they send.
- The author has developed, implemented, and tested an expert locator tool which exploits email content to classify the user's area of expertise in a real world environment. This provided valuable feedback into the validity of previous theoretical assumptions.
- The author has developed a new keyphrase extraction algorithm which is evaluated and shown to outperform existing algorithms.
- From the literature, key socio-ethical challenges that might be faced in the adoption of EKE or similar systems have been identified. These challenges have been further explored in real world studies, leading to an enhanced understanding of end users' perceptions.

2.10 SUMMARY

This chapter has provided an overview of the relevant research that has been conducted. The research reviewed the need to connect people to people, the evolution of expert finding technologies, the techniques for extracting keyphrases from textual content, and the evaluation techniques to assess how the extraction techniques perform. This serves to provide a knowledge foundation from which to learn and to build upon, ensuring the research conducted for this thesis adds to rather than duplicates existing or other ongoing work.

3 RESEARCH METHODOLOGY

3.1 INTRODUCTION

Research methodology refers to the procedural framework within which research is to be conducted (Remenyi & Williams, 1995). This chapter outlines the research methodology and explains the rationale behind the chosen path. The choice of the research methodology is important in that it should support the aim and objectives outlined within Chapter 1. The chapter then details the research methods that were used in this work.

3.2 RESEARCH STRATEGY

Research can be classified as quantitative or qualitative. Quantitative research (sometimes referred to as scientific or positivist) originated from natural sciences and is based on the principle that phenomena can be explained by objective and factual measures that strive to reduce researcher's bias (Morgan & Drury, 2003). Quantitative research uses empirical approaches since it focuses on objective rather than on subjective measures. These include methods such as structured interviews, self completion questionnaires, and structured observations. On the other hand, qualitative research (sometimes referred to as interpretivist or humanistic) originated from social sciences and is based on the principle that the complex nature of the phenomena makes it necessary for the researcher to adopt methods that will bring them closer to information sources. Furthermore, qualitative research enables the researcher to: interact with participants, question data, and draw upon the participants' past experiences (Morgan & Drury, 2003). Qualitative research employs the use of qualitative approaches to understand and explain social phenomena. These include methods such as unstructured interviews, conversations with participants, field notes, documents, and participant observation data (Myers, 1997).

The objectives of this research project require the use of both qualitative and quantitative strategies. The research project involves capturing the user requirements, then building and developing a system to meet those requirements, thereby having the features of an inductive approach through which theories can be generated. This was accomplished by the use of qualitative strategy. On the other hand, the project involves the processes of evaluation and testing of the system developed, establishing therefore the need for a deductive approach through which theory can be tested. This was accomplished by the use of quantitative strategy. Both sets of features are inherent in the system development methodology, a commonly used methodology in social science. System development methodology was used to achieve the main objective of this research project, which is the design and development of an expertise locator system exploiting email content for the purpose of identifying expertise in the organisation.

A system development methodology refers to the framework that is used to structure, plan, and control the process of developing an information system (CMS, 2005). More than twenty different systems development methodologies exist. Among these are Rationalised Unified Process, Extensive Programming, Reflective Systems Development, Prototyping, etcetera (Avison & Fitzgerald, 2003). Selecting the appropriate methodology is essential for the success of any research. The review of the

various software development methods provided a strong case for the selection of prototyping. For more details about prototyping see section 3.5.3.

3.3 METHODOLOGICAL CONSIDERATIONS

One vital factor that had an impact on the design of this research was the industrial context of the project. The research adopted in the thesis can therefore be considered practice driven. This type of research can be distinguished from conventional approaches by certain features. The features (Zmud, 1998) which were observed in the context of this research project are outlined below:

- The research topic: In practice driven research, the research topic is largely defined by the industrial sponsor and not by the academic research team. In this project, AstraZeneca largely contributed to defining the research topic.
- The end point of the research: At the commencement of a practice driven research project, the end point of the research may not be exactly specified. This point is normally revisited by the research team and the sponsor throughout the project's period. Such a feature was observed throughout the EngD project as the project's goals were revisited.
- The nature of the phenomena: The boundaries of a practice driven research project are not confined by a well defined research model. Rather, it is framed by the current understanding of both the sponsors and the research team of the phenomenon under study. Again, this particular characteristic was witnessed during the project.
- The research design: Unlike conventional approaches where the research team is solely responsible for the research design, practice driven research is designed jointly by the research team and sponsor. AstraZeneca suggested certain amendments to the research design based on their understanding of the phenomenon and the questions of interest to them.

Several benefits can be obtained if practice driven research is adopted. To start with, the topic researched is selected by the sponsoring company making it extremely applicable to practice. Moreover, as both industry and academia normally contribute to the research, the results obtained would be of interest to both parties, therefore enhancing the value of the research.

Practice driven research, however, poses major difficulties. These difficulties arise not only in data collection but also in communicating results to sponsors. The difficulties which arose during this project are listed below:

- Gaining access to research sites: This involves identifying likely sites and convincing potential participants to take part in the research. In spite of the project's direct association with a sponsoring body, which can sometimes serve to validate the research efforts, unfortunately, this was not the case in this particular research project. This is due to AstraZeneca's email system, which is outsourced to IBM. Thus, alternative sites had to be sought and therefore further difficulties were encountered.
- Gaining access to participants at a research site: This mainly relies on locating an internal facilitator who understands the phenomenon and at the same time can

aid in identifying the participants. This problem was encountered when attempting to organise focus groups.

- Maximising information from informants: This becomes a significant issue in situations involving discussions with participants as in interviews, workshops, etcetera. This is due to the significant amount of time that could be wasted if the questions of interest were not carefully designed to maximise the information obtained from the respondents. Hence, the author had to pay careful attention when preparing the focus group discussion guide.

3.4 ADOPTED RESEARCH METHODOLOGY

The end point of the EngD project, as set out by the industrial sponsor, involved the design and development of an expertise locator system that exploits email content for the purpose of identifying expertise in the organisation. The system development methodology was therefore seen as the most appropriate methodology to adopt in order to achieve the research objectives as discussed in section 3.2. Although many system development methodologies exist, all of the models recognise three main stages of system development. These are: concept development; system building; and system evaluation (Burstein, 2002). The stages of system development do not necessarily follow a linear path. Rather they are of interactive and dynamic nature and may be continually revisited and/or one or more of these stages may sometimes be omitted (Hasan, 2004). These stages were adopted in this research. Each stage involved certain actions, consisting of investigation, analysis, design, programming, testing, implementation, maintenance and evaluation (Davis & Olson, 1985; Humphrey, 1989; Kendall & Kendall, 1988; Olle et al., 1991). A brief outline of these stages and the processes within each stage is given in Table 3.1.

Concept development stage: Systems are developed as solutions to users' needs. The concept development stage involves capturing and prioritising these needs. Following this, the researcher must locate, understand, analyse, synthesise, interpret, and apply existing knowledge to identify the limitations of current systems, develop the system requirements, and develop meaningful research objectives. An extensive literature review is usually conducted in this stage.

In this research project, this stage involved the initial capture and development of the users' requirements. Through various meetings and emails between key project members, the user requirements were elicited. The members consisted of managers from AstraZeneca who were acting as the end customer in the project consortium and of the EKE academic research team at Loughborough University. The user requirements were checked to ensure that they were not only actionable, measurable, testable, but also tackled business needs, and provided enough information to be translated into system requirements. From the user requirements, the first draft of the system requirements was produced.

This was followed by a literature review to get an in-depth understanding of the subject of the research. A domain analysis identified shortcomings in the research subject and further aspects that needed to be addressed. This in turn led to the identification of further developments to the projects requirements. This stage resulted in the development of a conceptual design of the system.

Table 3.1: Research Stages and Processes

Stage	Actions/ Processes
Concept Development	Investigation • Initial development of user requirements • Literature review Analysis • Domain analysis System Design and Requirements • Development of system requirements
System Building	Programming • System development • Further system development Unit Testing • Subsystem (Keyphrase extractor) evaluation 1 • Subsystem (Keyphrase extractor) evaluation 2 System Testing • Integration of individual program units to obtain the final system
System Evaluation	Evaluation • EKE system evaluation • EKE concept evaluation

System building stage: Based on the conceptual design, a system is developed as a proof of concept to demonstrate that a system is viable and can be built. This may involve the iterative development processes of analysis, design, implementation, testing, and evaluation. Results from the testing and evaluation can feed back into the system building stage.

Once the concept development stage was completed, the stage of system building commenced. This stage started with the system development, which encompassed the development of the keyphrase extraction unit. The development process was iterative and evolutionary. Individual system units were developed, tested and integrated to obtain the final system. Unit performance testing of the keyphrase extraction unit was conducted as described in section 4.3.2. First, the evaluation measured the performance of the keyphrase extraction unit without making any use of WordNet, a semantic lexicon for the English language. Then the evaluation measured the performance of the keyphrase extraction unit employing WordNet at the filtering stage of the keyphrase extraction process. In light of the evaluation results obtained, alternative approaches to further improve the process in order to obtain higher performance metrics needed to be

explored. This eventually resulted in additional system development. Once unit testing was performed, the system's units were combined and tested as a group.

System evaluation: The system is tested as a whole during the system evaluation stage. The system availability is expanded to target potential users. During this stage, the complete system was piloted by technologically competent end-users, who were frequent and experienced email users, selected from within the school of Informatics at Loughborough 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 obtained was either resolved or marked for further action. Focus groups were used to evaluate the system developed from an industrial perspective.

A point to note is that in addition to the stages of system development methodology that the research project passed through, the project underwent a concluding stage that addressed the key socio-ethical challenges involved in the implementation and use of the system.

3.5 METHODS/TOOLS USED

The success and validity of any research depends on the appropriate selection of research methods (Fellows & Liu, 2003). This subsection presents the overall research methods used for this study and offers a justification for using them. The choice of research methods (e.g. literature review (LR), questionnaire (Q), prototyping (P), focus group (FG)) in relation to the research objectives and associated tasks is provided in Table 3.2. In addition, the table points out the main research outputs. Further information regarding the research undertaken and outcomes are elaborated in Chapter 4.

Table 3.2: Research Road Map

OBJECTIVES	WORK TASKS	ACTIONS/ PROCESSES	METHOD	OUTPUT
1- Review literature on approaches to expert finding	1- Review latest expert finding	Initial development user requirements Literature review Domain analysis Development of system requirements	LR	Paper 1
2- Determine the potential value of email in supporting the process of expert finding	2- Determine the potential of email as a source of information for expertise recognition	Literature review	Q	Paper 2
OBJECTIVES	**WORK TASKS**	**ACTIONS/ PROCESSES**	**METHOD**	**OUTPUT**

Continuation of objective 2	3- Determine the potential for email expert locator to work in organisations	Further development system requirements	Q FG	Paper 2
3- Develop the expert locator system, EKE	4- Review expert finding systems that exploit email content and their limitations	Literature review Further development of system requirements	LR	Paper 1
	5- Implement a tool that extracts keyphrases from users' email	Literature review Further development of requirements System development	P	Paper 3
	6- Identify a performance measure to test the performance of the keyphrase extraction subsystem	Literature review	LR	Paper 3
	7- Measure the performance of the key concept extractor tool	Subsystem (Keyphrase extractor) evaluation 1 Subsystem (Keyphrase extractor) evaluation 2 Further development of requirements	P	Paper 3
4- Evaluate the use of EKE	8- Evaluate the EKE application	System Evaluation	P	Paper 5
5- Identify enablers and barriers to the use of EKE as an expertise locating system	9- Explore user satisfaction	System Evaluation	P FG Q	Paper 5
	10- Critical reflection on socio-ethical issues involved in the use of such a system	Literature review	LR Q	Paper 4 & 5

3.5.1 LITERATURE REVIEW

Conducting a literature review is an essential component of any research project. Literature review is not only conducted for the purpose of identifying what has already been researched and trying to build upon it, but also to formulate sharper and more specific research questions. Other benefits from reviewing the literature could include: identifying appropriate methodologies that could be potentially used in the research project, identifying knowledge gaps in the subject area, and highlighting areas which need further research.

For this project, the literature review was carried out at the beginning and throughout the various stages of the research. The initial review of literature examined the needs and benefits of connecting people to other people so that they can collaborate. It has also reviewed the evolution of the expert finding approaches. This helped identify the gaps in the current approaches to expertise finding systems and justify the proposed approach. In addition, a literature review of keyphrase extraction techniques and keyphrase evaluation technique was carried out due to their connection to the subject domain. This enabled identifying key weaknesses in the existing keyphrase extraction approaches and assisted, therefore, in the implementation of the developed system's keyphrase extraction technique.

3.5.2 QUESTIONNAIRE

Questionnaires may be either self administered or read out by the interviewers. In a self administered questionnaire, the respondent is in charge entirely for understanding the questions, completing, and depositing the questionnaire. The cost and level of skills associated with this method are relatively small. Unfortunately, there is normally little interaction with the interviewer and therefore little assistance can be provided (Fink & Kosecoff, 1998). However, such drawbacks can be overcome by paying particular attention to issues such as the wording and layout of the questions.

Questionnaires which are not self administered such as those conducted via the telephone are known to suffer from a major drawback resulting from the implications of personalising the relationship over maintaining the subject's anonymity (Ibert et al., 2001).

Questionnaires used in this research project involved a web questionnaire directed to Leicestershire Constabulary, a questionnaire that was handed in at the end of focus groups discussion, and finally the questionnaires administered during the system evaluation at Loughborough University. All questionnaires used in this research project were self administered questionnaires, with limited assistance available from the researcher, if necessary. In terms of resources, self administered questionnaires are cheap to administer and can be completed within a short time. Through using this type of questionnaires, confidentiality of responses and anonymity of respondents are maintained. Moreover, bias can be reduced by administering the questionnaire in a standard manner.

3.5.3 PROTOTYPING

Prototyping is a systems development method in which a prototype (an early representation of a final system or product) is built, tested, and then reworked as necessary until an acceptable prototype is finally achieved, from which the complete system or product can be developed (Worth & Greenough, 2005). Prototyping is an iterative trial-and-error process which involves several processes as depicted in Figure 3.1. The processes involved are similar to those covered in Table 3.2 and therefore will not be presented again in this section. Prototyping is most appropriate in situations where not all of the project requirements are known in detail ahead of time or in situations where the functional requirements may change frequently and significantly. It allows faster development of application software, which is undertaken through several iterative stages.

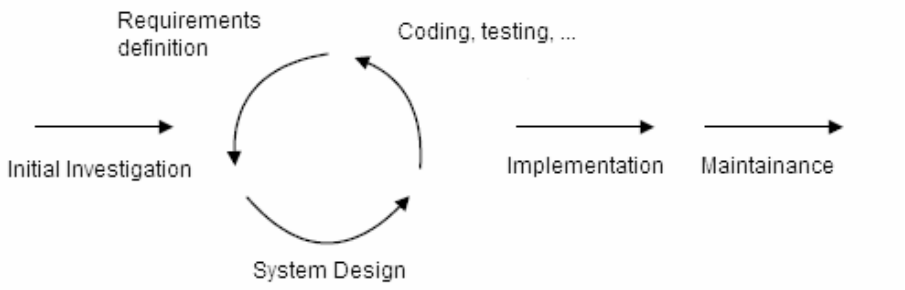


Figure 3.1: Prototyping Cycle (Source: CMS, 2005)

3.5.4 FOCUS GROUP

Krueger and Casey (2000) define a focus group as "a carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, nonthreatening environment". A focus group is a form of group interview involving several participants and a moderator, whose role is to facilitate the discussion. Some researchers make a distinction between the focus group and group interview techniques. According to Byrman and Bell (2003), this distinction can be attributed to three plausible reasons. Firstly, while group interviews normally cover a wide range of issues related to the subject/theme in question, the aim of focus groups is to undertake an in-depth exploration of the particular subject/theme. Secondly, while one main purpose behind conducting group interviews (as opposed to one to one interviews) is to minimise the resources associated with undertaking them, such a purpose does not normally feature as a key objective in focus groups. Thirdly, in focus groups, the researcher is interested in how the interviewees discuss the topic as a group rather than as individuals, which is the case in group interviews. Unfortunately, the distinction is not well defined; hence the terms are used interchangeably.

Focus groups are of qualitative nature. They are concerned with the exchange of ideas, feelings and experiences on a specific topic. A focus group is used to gain richer and deeper insights into a particular research issue through listening to and learning from a group of knowledgeable individuals (Greenbaum, 1998). What is important is the interaction within the group and the joint construction of meaning (Bryman & Bell, 2003). Focus groups can be used to stimulate participants to raise issues which they perceive important. The use of focus group is particularly relevant where testing responses to new products is required in order to determine the products' acceptance by consumers. A well planned and conducted focus group encourages self-disclosure through the facilitation of synergy and other group dynamics. There are a number of practical issues to consider when conducting focus groups (Bryman & Bell, 2003). These involve:

Tape recording and transcription: As with interviewing in qualitative research, focus group discussions need to be recorded and transcribed.

Number of groups: It is desirable to repeat the focus groups several times with different people to minimise group bias. It has been recommended that a minimum of three focus groups should be conducted (Krueger, 1994), however constraints on resources such as time can be limiting factors.

Size of groups: A focus group of size six to ten is normally recommended (Morgan, 1998; Krueger, 1994), though examples can be found of both smaller and larger group sizes.

Level of moderator involvement: The role of the moderator is to guide the focus group and not to take over the discussion.

Selecting participants: Whether to use natural grouping or to use stratifying criteria in order to select people who do not know each other is arguable. Selection of participants should be dependent upon the objectives of the study.

Asking questions: No standard approach on how focus group questions should be structured is available. According to Krueger (1994), some researchers favour to use one or two general questions to encourage debate, with the moderator participating when necessary, while others prefer to use more structured questions.

Telephone focus groups: In focus groups, participants usually meet face to face in a comfortable setting. However, attracting potential participants to take part in a focus group can be particularly difficult, as this implies bringing people together to meet at a specified time in a certain location. Conducting focus group discussions using telephone can be more convenient to participants and can therefore increase the likelihood of participation. Telephone focus groups are usually less expensive and have a higher potential for participation than face to face focus groups. Some have argued that telephone focus groups are less threatening and intimidating and as a result may have a higher clarity of language. Their primary disadvantage is the lack of nonverbal communication richness. Telephone focus groups are usually shorter, have fewer participants (usually 4-6 participants), and fewer questions than face to face focus groups (Krueger & Casey, 2000).

In this research, two focus groups were used to establish end users' comprehension and acceptance of the EKE concept. It was felt that conducting two focus groups would overcome the impact of single group bias, and provide the necessary feedback. The first focus group (Group A) was conducted over the telephone, while the second focus group (Group B) was conducted face to face. The researcher was the facilitator. Group A and B comprised 5 and 7 participants respectively.

3.6 SUMMARY

This chapter has introduced the research strategy of the adopted methodology, which is the system development methodology. It provided a justification for the adoption of the defined research methodology. The methodological considerations for this EngD project were discussed in this chapter. A discussion of the specific research methods that were adopted in this research was also presented. The choice of research methods are mapped out against research objectives and associated tasks along with research output in Table 3.2.

4 THE RESEARCH UNDERTAKEN

4.1 INTRODUCTION

This chapter describes the research undertaken during the four years of the EngD programme to meet the aim and objectives described in section 1.5.2. Since the research was conducted using the System Development Methodology (described in Chapter 3), the discussion in this chapter is structured in accordance with the project's stages. These stages include concept development, system building, and system evaluation stages. The work tasks (see Table 3.2) accomplished in each stage are listed alongside the corresponding section titles. It should be noted that the activities conducted in each stage were not undertaken strictly in a sequential order. In fact, some of these activities overlapped with others or required multiple iterations. To avoid repetition, references are made to papers, sections of papers, or other information in the Appendix. It is strongly recommended that the reader consults these as requested and then returns to the main body of the thesis.

4.2 STAGE 1: CONCEPT DEVELOPMENT (TASKS 1 – 3)

Stage 1 was undertaken to confirm the need for expert finding solutions within organisations and ascertain whether the proposed solution would be accepted and suitable for knowledge workers. This stage included an investigation and analysis of the literature, followed by proposed system design and requirements phase.

4.2.1 INVESTIGATION AND ANALYSIS

At the commencement of the project, the requirements were partly defined by the project's industrial partner, AstraZeneca. The requirements were further refined and amended following the literature review, which was conducted to identify:

- the need to connect knowledge workers to each other so that they can collaborate;
- the state of the existing expert finding tools; and
- the role of email in knowledge work and its potential value for expertise location.

Investigating these areas was important to better understand expert locators. The first few months of this project were dedicated to reviewing the literature. This provided the author with adequate knowledge of the technologies and approaches employed in past and present research. A discussion and analysis of the key areas reviewed are provided below.

4.2.1.1 The Need to Connect Knowledge Workers to Each Other

People acquire and communicate knowledge using different sources (e.g. books, online resources and other people). It has been shown that people are the most indispensable source of knowledge. This can be attributed to the fact that when people attempt to find information, they prefer to ask other people rather than to look through records of information (Allen, 1977; Hiltz, 1985; Lang et al., 1982; Mintzberg, 1973; Pelz & Andrews, 1966).

Literature has identified the benefits of seeking information from other people. As discussed in Chapter 2, these benefits can be categorised into five groups (Cross et al., 2001), namely:

- Finding solutions: One of the benefits is the potential to find a solution to the problem under consideration.
- Meta-knowledge: This provides the knowledge seeker with pointers to alternative valuable sources of information.
- Problem reformulation: The interaction associated with seeking information from other people allows viewing the problem from a different perspective.
- Validation of plans or solutions: Seeking information from other people allows ideas to be discussed and hence validated.
- Legitimation: Seeking information from an expert in the field and citing them as having reviewed the solution increases the credibility of the proposed solution.

The difficulties associated with using people as sources of information are numerous. For instance, finding a knowledgeable person can be both time consuming and disruptive. To overcome these problems, expert finding aids have emerged.

4.2.1.2 The State of Expert Finding Aids

The idea of finding people with particular skills and knowledge has been identified to date back to the 1980s (Maron et al., 1986). This concept has gained precedence in organisations in recent times. A list of systems that have been developed is provided in section 2.4. Existing systems can be differentiated based on the information source(s) they utilise in order to profile the users' areas of knowledge. These sources include email, bulletin boards, programme codes, web pages, and technical reports. This research is concerned with expert finding systems employing email content as the source of expertise indicators. Paper 1, sections 1 – 3, Appendix A reviews the development of expert finding approaches with particular reference to solutions exploiting email sources.

A domain analysis was carried out to investigate existing systems and newly emerging technologies. The research described in Paper 1 uses the domain model proposed by Yimam-Seid and Kobsa (2003) to identify the gaps of current technologies that exploit email as evidence of expertise. The findings of the domain analysis as documented in Paper 1, section 4 revealed a number of problems, namely

- an expertise profile gap;
- an expertise matching gap;
- an expertise representation gap;
- a user control gap; and
- a cultural and management gap.

Having identified the limitations of expert finding systems that exploit email, the next step was to further validate the use of email content as a viable source for expertise finding. The potential of using email to support expert finding is discussed below.

4.2.1.3 Role of Email in Knowledge Work

Email is an important knowledge medium; it is an essential collaboration tool that is widely used for internal and external communications (Garcia, 2006). In spite of the popularity of using email and the attempts to develop tools to locate experts using email, there is surprisingly little research into its potential as a source for expertise profiling.

To fill this gap, the author conducted an online questionnaire to investigate whether it is possible to profile employees' knowledge from emails they send. The questionnaire was administered to academics from the Research School of Informatics and from the Civil and Building Engineering Department both based at Loughborough University, and to employees at an IT firm in the US. The results were published Tedmori et al. (2006). A total of 13 responses from academia and 9 responses from industry were received. Each respondent had to analyse 20 emails from their email outbox, where they had generated, organised, shared, or leveraged knowledge. Respondents had to determine if the selected emails contained keyphrases that described their areas of interest or knowledge. If the email contained such keyphrases, respondents had to specify whether these were general or specific to their interest or knowledge. An example and a definition were provided to the respondents to help them distinguish between keyphrases that are general to their knowledge and keyphrases that are specific to their knowledge. Below are these definitions:

- *General Keyphrase:* A keyphrase is general when it is applicable to or affecting the whole or the majority of employees in the organisation (e.g: Aeroplane, Helicopter).
- *Specific Keyphrase:* A keyphrase is specific when it is concerned specifically with the subject specified (e.g: Airbus320, Bell 206 JetRanger).

They were then asked to indicate their skill level in these areas as "basic knowledge", "working knowledge", or "expert knowledge". Below are the definitions that were provided:

- *Basic Knowledge (BK)* means having a passing familiarity of basic issues, practices, developments, etc., and a general understanding and appreciation of their broad implications.
- *Working Knowledge (WK)* means having a good knowledge of related concepts, theories, principles, standards, frameworks, procedures, etcetera.
- *Expert Knowledge (EK)* means having a thorough understanding as to why and how things operate.

As shown in Table 4.1, 59% of the respondents' emails in the academic sector and 73% of the respondents' emails in the industrial sector did contain keyphrases that could be used to profile their knowledge. In the academic sector, 35% of emails contained general keyphrases and 65% contained specific keyphrases. In both sectors, considering the emails that are general to the respondents' interest or knowledge, there is no significant variation in the level of skill. Whereas, considering emails that are specific to the respondents' interest or knowledge, a considerable difference was noticed. In the academic sector for instance, the significant proportion of skill level is "EK".

Table 4.1: Keyphrases

ACD					
59% have keyphrases					
35% General			65% Specific		
BK	WK	EK	BK	WK	EK
28%	34%	38%	8%	14%	78%

IND					
73% have keyphrases					
41% General			59% Specific		
BK	WK	EK	BK	WK	EK
25%	42%	33%	9%	57%	34%

Using the categories of knowledge in the questionnaire, employees indicated whether the keyphrases found in the email text were either general or specific to their interest or knowledge fields. This indicates that it is possible not only to identify employees with a general knowledge about various areas, but also to identify employees with knowledge in specialised areas. For instance, most people engaged in building construction have a general knowledge of the construction processes, but only a few are specialised in the environmental impact of construction processes.

While this research did not suggest that expertise profiling is best supported by email, it had been found that in email communications where knowledge development and creation occur, it is possible to profile users' knowledge based on the emails they send.

A second questionnaire survey was undertaken to understand the information needs of potential users which can be supported by the use of an expert locator and to ascertain whether the current habits of potential users would provide a basis for successful implementation of the purposed system. The results obtained from analysing this questionnaire are detailed in Paper 2, Appendix B. The specific objectives of the survey were to: (1) understand the behaviour of potential users when attempting to seek and share information; and (2) validate the appropriateness of the proposed concept.

The questionnaire consisted of 31 questions, divided into 6 sections and was distributed to 150 employees working at Leicestershire Constabulary (LC). The questionnaire was completed by 44 participants yielding a 29% response rate. The results indicated that:

- The majority of respondents (59.1%) believed that a minimum of 21% of their time could be saved if they knew where to look for the information required.
- The amount of time wasted can be attributed to information overload, problems with search facilities, incorrect titling/labelling, constant change within the force, etcetera.
- Specific websites and online databases were found to be the most frequently utilised sources, when searching for information.
- The majority of respondents (75%) indicated that they "sometimes/frequently" have difficulties identifying sources of new information.
- A large proportion (90.9%) indicated that they "sometimes/frequently" consult their colleagues when conducting a search.
- A significant relationship was found between the frequency of knowing who to contact and the number of years spent working at LC.
- Participants found that the task of searching for experts was difficult.
- The majority of participants recognised that the purposed concept would be of benefit to them.

- The majority of participants expressed their willingness to use the proposed system if it were to be utilised by LC.

4.2.2 SYSTEM DESIGN AND REQUIREMENTS

Having established that it is possible to identify employees' areas of knowledge from the emails they send, Figure 4.1 illustrates the generic concept of the proposed system model developed from the review and analysis processes.

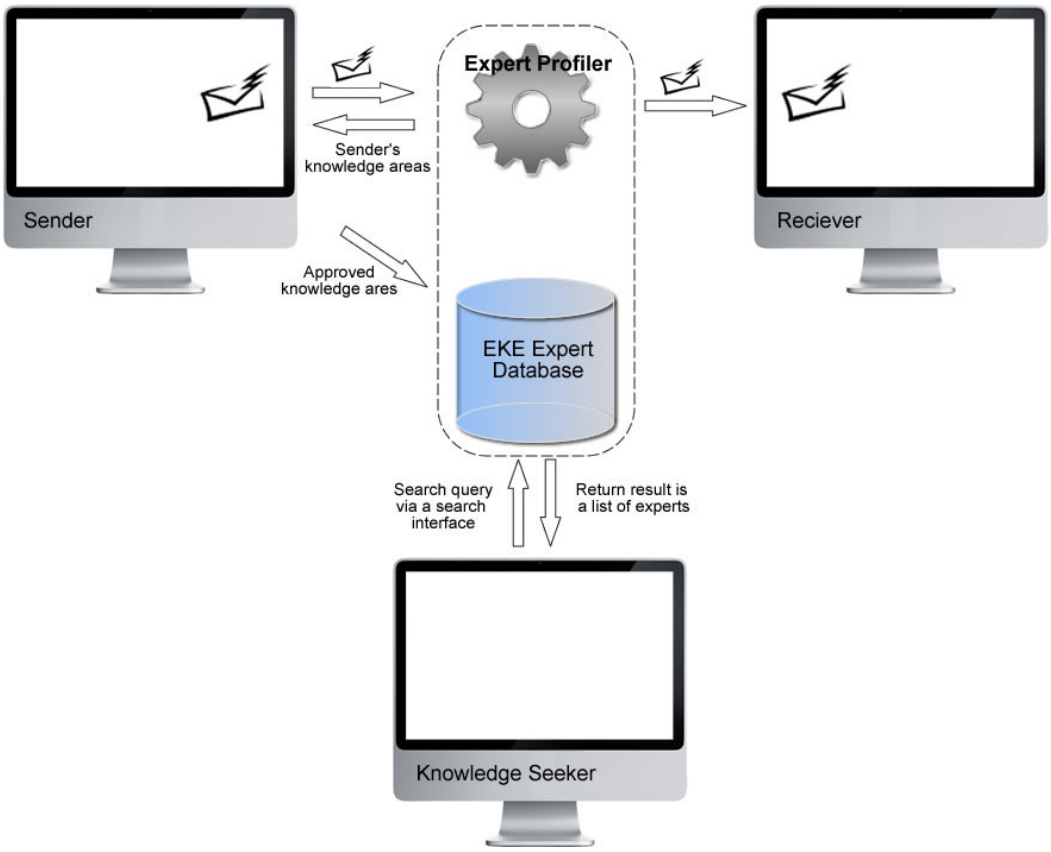


Figure 4.1: The Generic Concept of the Proposed System

The requirements obtained from AstraZeneca and the findings from the investigation and analysis phase, specified that the system will work by capturing the email content before the email is sent to the server. This information will be used to create individual profiles of the users' knowledge, rather than general ones applicable to the whole organisation. To protect users' privacy, only information approved by the user will be used to create these profiles. The profiles will be stored in a centralised location accessible by the whole organisation. Other users requiring assistance with a particular problem can search this repository. This will result in a list of experts ranked by their suitability to help resolve the problem.

4.3 STAGE 2: SYSTEM BUILDING (TASKS 4 – 7)

To explore the potential of the proposed system and to gain insight into its operation in organisations, a prototype was developed. As stated in Chapter 3, Stage 2 of the research involved the development of the proposed system. The system building stage included the actions of programming, testing and evaluation, and integration. Such a stage consisted of four primary phases.

In the first phase (Phase 1), off-the-shelf software, which provided the functionality identified in the concept development stage, was obtained. The E-mailRelay and KEA applications were used. E-mailRelay was setup as a store and forward message transfer agent to enable the message to be processed and the keyphrases to be extracted before being sent to the remote server. KEA was used to extract keyphrases from email messages. Before the keyphrases were extracted, a user specific keyphrase extraction model was built using the KEAModel builder. The model was built using training documents provided by AstraZeneca from the pharmaceutical domain with keyphrases already assigned. Evaluation of this architecture using emails obtained from AstraZeneca showed that the existing keyphrase extraction software KEA was not appropriate for the task at hand. The keyphrases returned from KEA were not representative of the knowledge fields the emails conveyed. This can be attributed to the way KEA works. The approach taken by KEA involves the selection of candidate phrases and the calculation of two feature values for each candidate phrase. One of these features is the Term Frequency Inverse Document Frequency (TFxIDF), which is a measure of a phrase's frequency of use in the document compared to its rarity in general use. The TFxIDF measure is best suited for larger electronically stored documents (e.g journals), and not with considerably shorter email messages.

Following the unpromising results obtained from using KEA to extract keyphrases from emails, an alternative approach was investigated during Phase 2. Numerous techniques were explored as detailed in Paper 3, section 2, Appendix C. The results showed that off-the-shelf systems were not suitable for the project's purposes; hence a bespoke system had to be developed (see Figure 4.2). A detailed description of the proposed system is outlined in Paper 1, section 5.

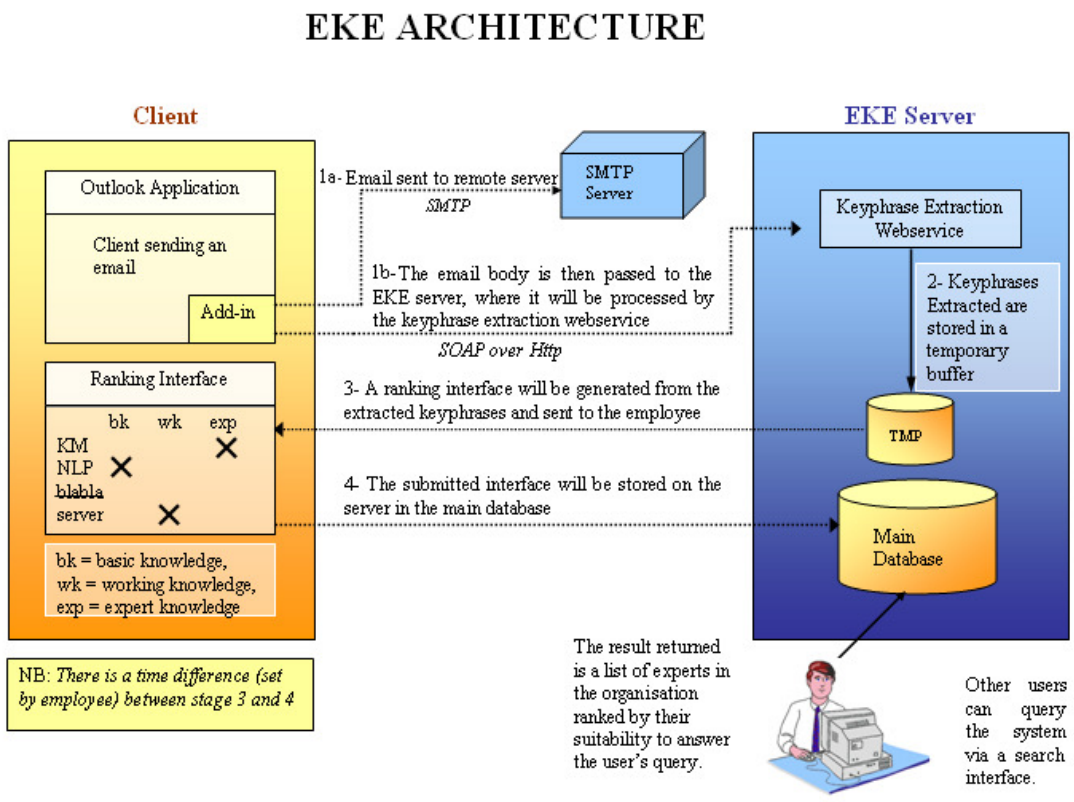


Figure 4.2: EKE Generic Architecture

The proposed system consists of two main components, a client side component and a server side component. The client side software is implemented as an outlook plugin and is responsible for capturing the email content. The email content is then passed to the server side component. The server side component consists of a bespoke web service based keyphrase extraction engine, which will be in charge of extracting keyphrases from the email body. These keyphrases are stored in a temporary database where, at a later point, they are returned to the user for validation and approval. Once validated and approved, the keyphrases are stored in the respective user profile in the main database on the system's server. The user profiles stored on the server can then be queried by other users searching for help with a particular problem. The result of the search query is a list of experts on the topic.

The performance measure of the bespoke keyphrase extraction algorithm was calculated as described in Paper 3, section 4. Although the results obtained were higher than reported performance measurement results for other systems, it was felt that the keyphrase extraction algorithm could be further improved due to the presence of undesirable common words in the extracted keyphrases.

Phase 3 addressed the issue of common words through the use of more filtering techniques. WordNet, a well recognised linguistic tool, was used to reduce the occurrence of common words. This was achieved by comparing extracted keyphrases with the word database provided by WordNet. If the extracted keyphrase was found in WordNet, it was ignored; otherwise it was stored in the temporary database as in stage 2. However, the improvement in the extracted keyphrases was not significant enough to justify the performance decrease in processing time. Details of the keyphrase extraction unit evaluation are described in section 4.3.2.

In light of this, alternative approaches were explored. It was theorised that a higher and timelier user involvement in the selection and validation of keyphrases would result in improved user profiles. In Phase 4, the user validation and approval of keyphrases was modified from a relatively infrequent interval to a real time response. The keyphrases extracted were immediately returned to the user for verification (see Figure 4.2). The verified keyphrases were directly stored into the main database removing the need for a temporary database. The following section describes the implementation of Phase 4.

4.3.1 PROGRAMMING

Having established the requirements for the prototype system, this section details the technical implementation. The system consists of two main components, client side and server side software. The client side software consists of the EKE add-in developed using Visual Studio .Net 2003 and the C# programming language. C# is developed by Microsoft and is based on Microsoft.Net framework. It is an object oriented programming language designed for building a wide range of enterprise applications.

The server side software consists of a web server hosting the keyphrase extraction web service, the EKE Database Server, and the Expert Search Engine. These client and server side components are discussed in detail below. Figure 4.3 shows the software used to develop EKE.

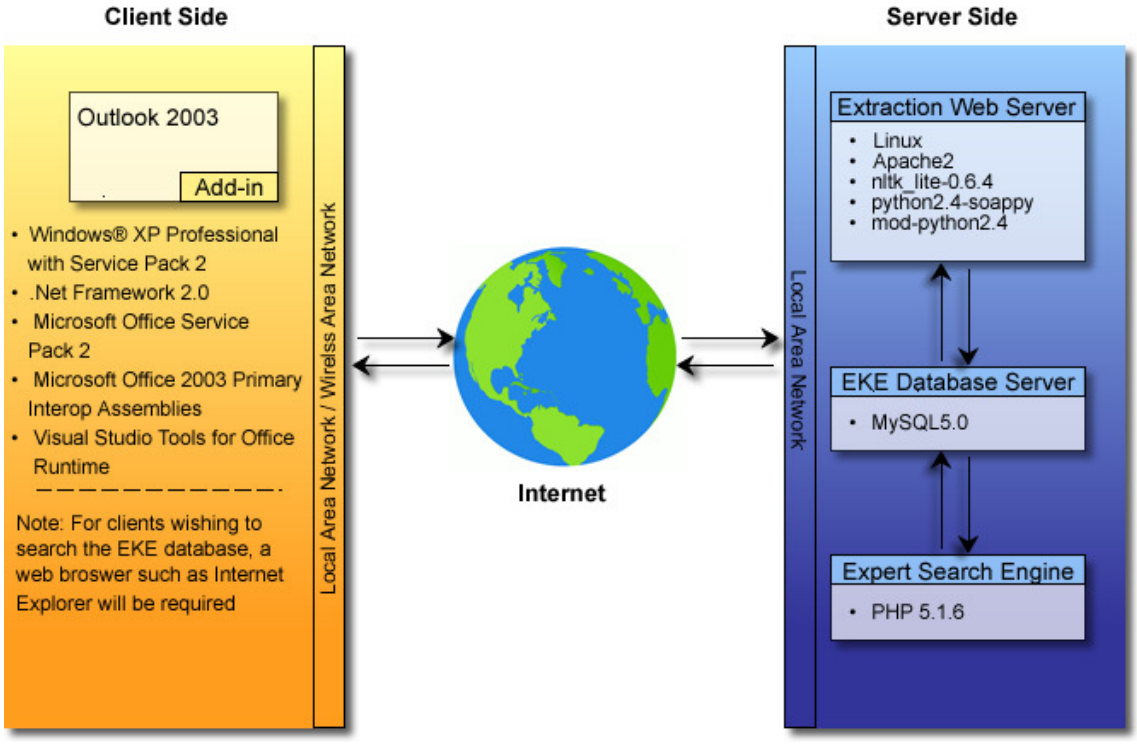


Figure 4.3: Detailed System Architecture

4.3.1.1 EKE Add-in

An Outlook add-in was developed to capture the body from emails sent by users before they are sent to the SMTP (Simple Mail Transfer Protocol) server. The captured body is sent to the Keyphrase Extractor Web Service (KEWS) for processing as discussed in section 4.3.1.2. The EKE add-in is also responsible for receiving the processed information from the KEWS and displaying it to the user for validation. Once validated by the user, the information is sent by the EKE add-in to the SQL Database Server for storage.

Software Development

The EKE add-in was developed in C# using the Visual Studio 2005 integrated development environment with the Visual Studio Tools for Office (VSTO) add-on. VSTO provides developers with the tools necessary to build interoperable applications for Microsoft Office. For this research purposes, VSTO was used to develop an add-in for Microsoft Outlook 2003 to capture email objects (i.e. email body and email sender address).

To run VSTO applications on target computers, the following prerequisites are required:

- .Net Framework 2.0;
- Microsoft Office 2003 with Outlook;
- Microsoft Office 2003 Service Pack 1 or later;
- Microsoft Office 2003 Primary Interop Assemblies; and
- Visual Studio Tools for Office Runtime (same version as .Net Framework).

Figure 4.4 describes the sequence of activities involved in the execution of the EKE add-in. Due to the requirements for prerequisites, the setup program for the EKE add-in

has been designed to automatically check whether the prerequisites exist. The missing prerequisites are automatically installed.

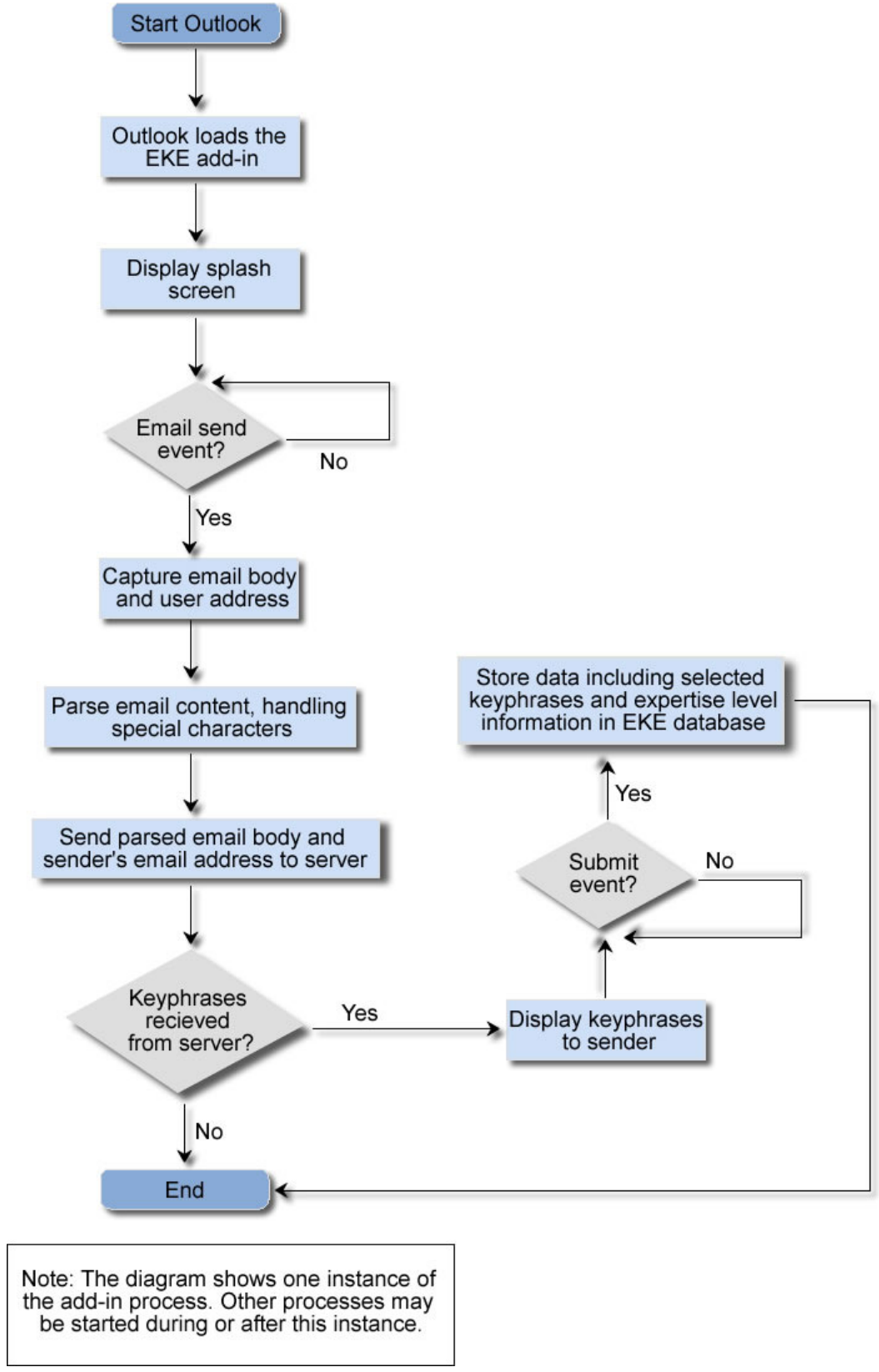


Figure 4.4: EKE Add-in Activity Diagram

Interface Design

The interface was designed to conform with established usability guidelines (see Table 4.2).

Table 4.2: Usability Guidelines

Interface Design	

Basic Interface Guidelines	• Use consistent and contrasting colours across all interfaces. Avoid bright colours to prevent fatigue.
	• Use conventional interface controls which users are familiar with.
	• Provide links to information regarding:
	  o Confidentiality and privacy
	  o Sources of help (e.g. contact information)
	• Use simple language. Avoid technical jargon.
	• Provide context and orientation information.

Navigation	• Provide clear navigation mechanisms. • Provide some contextual information about where the user is and allow them to easily navigate back. • Label links clearly. • Reduce the number of links between the point of entry and the desired destination.
Layout	• Use consistent layout throughout. • Avoid horizontal scrolling and limit vertical scrolling (e.g. paging). • Avoid overuse of multimedia objects (images, videos, flash files) to improve performance. • Use appropriate type-faces and font sizes throughout the interfaces.
Data Entry	• Use standard data entry controls the users are familiar with. • Provide information on types of input expected (e.g. date format). • Validate all data entry and provide the user with confirmation of entry or with friendly error messages.
For related literature see Carroll, (1997), Dix et al., (2003) and Sharp, (2007)	

A walk through of the interface designed following the above guidelines is provided below. As soon as the user loads Outlook, they are presented with an introductory splash screen, indicating the EKE add-in has loaded (see Figure 4.5). The users' interactions with outlook will continue as normal after this point. They will be able to type and send an email as shown in Figure 4.6.

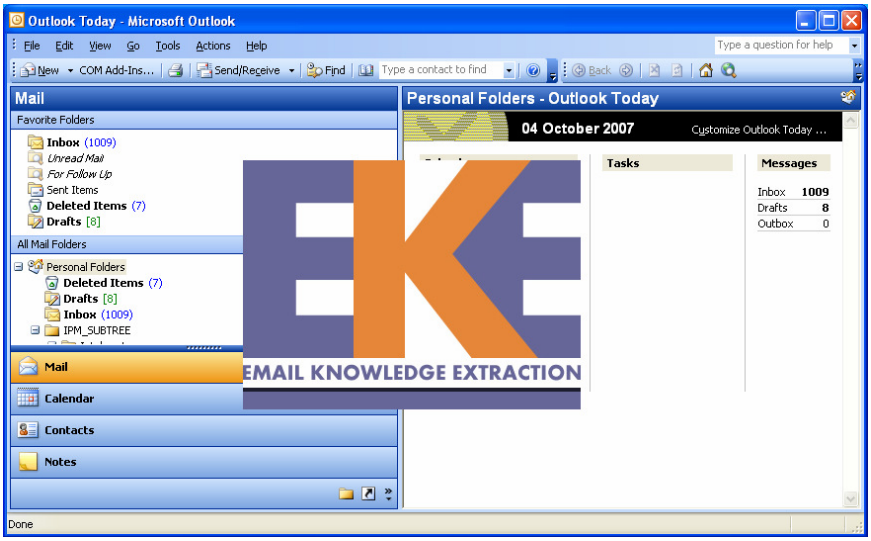


Figure 4.5: Outlook 2003 – Introduction Screen with EKE Splash Screen

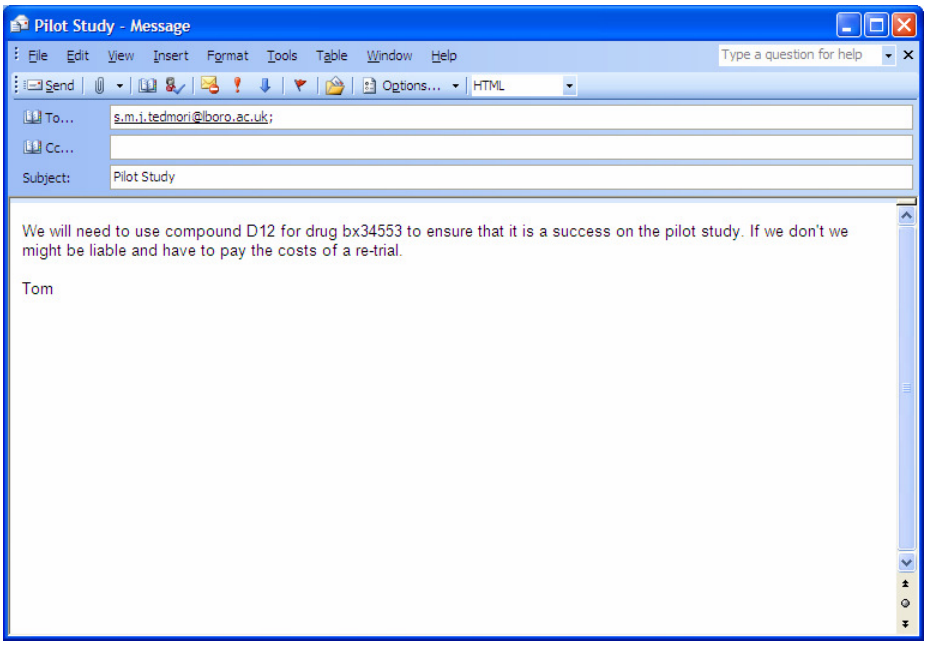


Figure 4.6: New Mail Message Interface

Once the user clicks the "send" button, they are presented with a list of keyphrases extracted from the email, from which they can select the keyphrases that best reflect their knowledge. Figure 4.7 illustrates the form listing the keyphrases.

Figure 4.7: Keyphrase Form

The keyphrase form employs the categorisation established in section 4.2.1.3, to label the radio buttons and to allow users to categorise their skill levels. The keyphrases are displayed as separate rows to assist the user in easily identifying the keyphrases.

4.3.1.2 Keyphrase Extractor Web Service

The KEWS was developed to extract keyphrases from the information received from the EKE add-in (i.e. the email body). The KEWS is responsible for filtering the extracted keyphrases and checking that the keyphrases have not been previously extracted and stored for the sender. The keyphrases that have not been previously extracted are then sent back to the EKE add-in. Those which have already been identified by the system and categorised by the user, will have their frequency field in the database incremented by one.

Software

The keyphrase extractor was built using the Natural Language Toolkit (NLTK), which is a suite of open source libraries and programs for symbolic and statistical natural language processing (NLP) for python. A version of it, NLTK-Lite, is used for identifying keyphrases from the email body. NLTK is mainly be used to identify the parts of speech (POS) of the individual words in the email body and to chunck the phrases, as will be discussed. Figure 4.8 shows an overview of the extraction algorithm. A detailed explanation of the algorithm along with an example demonstrating its operation are available in Paper 3, section 3, Appendix C.

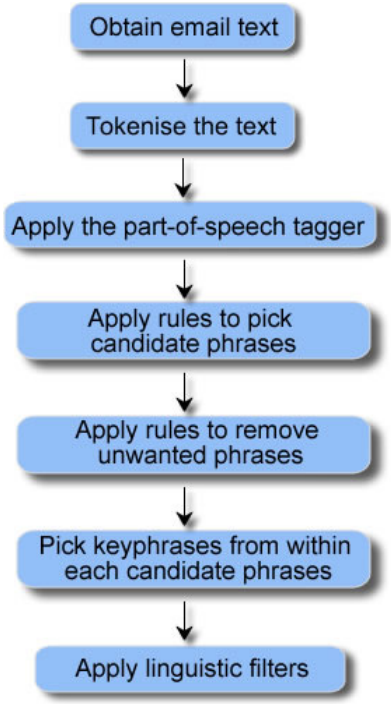


Figure 4.8: Stages of the Extraction Process

The algorithm is fully implemented and embedded in the KEWS. To summarise, the extraction algorithm has two stages. The first stage involves a training process in which a model for POS tagging was created. The second stage involves an extraction process in which keyphrases are extracted from email messages using the created model.

The input to the system is a single email message. Once obtained, the email text is split into tokens using regular expression rules. The tokens are then tagged by their parts of speech using the POS model created. Rules are then applied to select candidate keyphrases by grouping all occurrences of specific sequences of tags together. A rule is a sequence of grammatical tags that is most likely to contain words that make up a keyphrase. These rules were manually set by the author. After the sequences of tags are grouped together, rules are applied to remove a subset of irrelavent phrases. Keyphrases are then selected from the identified candidate keyphrases. Finally, the system uses linguistic filtering to extract more important keyphrases, resulting in a set of lines, each a sequence of tokens containing at least one letter.

The NLTK is a suite of open source Python modules and as such to enable the Apache server with the python programming language, a Mod_python module was added to the Apache installation. This allows python scripts to be executed on the web server.

4.3.1.3 EKE Database Server

The EKE Database Server was designed as the backend of the Expert Search Engine to store all the extracted and approved keyphrases from users' emails. The database was designed to optimise these activities, as discussed below.

Software Design

The database was built using MySQL (Structured Query Language), a relational database, and administered using phpMyAdmin. Table 4.3 describes the function of the database tables.

Table 4.3: The EKE Database Table Descriptions and Functions

Table Name	Table Description
Users	Represents the users of the developed system. It contains information related to the users' contact details. The "Users" table has a relationship with the "Keyphrases" table.
Keyphrases	Stores the keyphrases extracted from the users' email messages. The details stored include the frequency of the keyphrase, level of expertise regarding the keyphrase, and the keyphrase insertion date.

A common visual representation of a database model is an entity relationship diagram. Figure 4.9 depicts the EKE database entity relationship diagram.

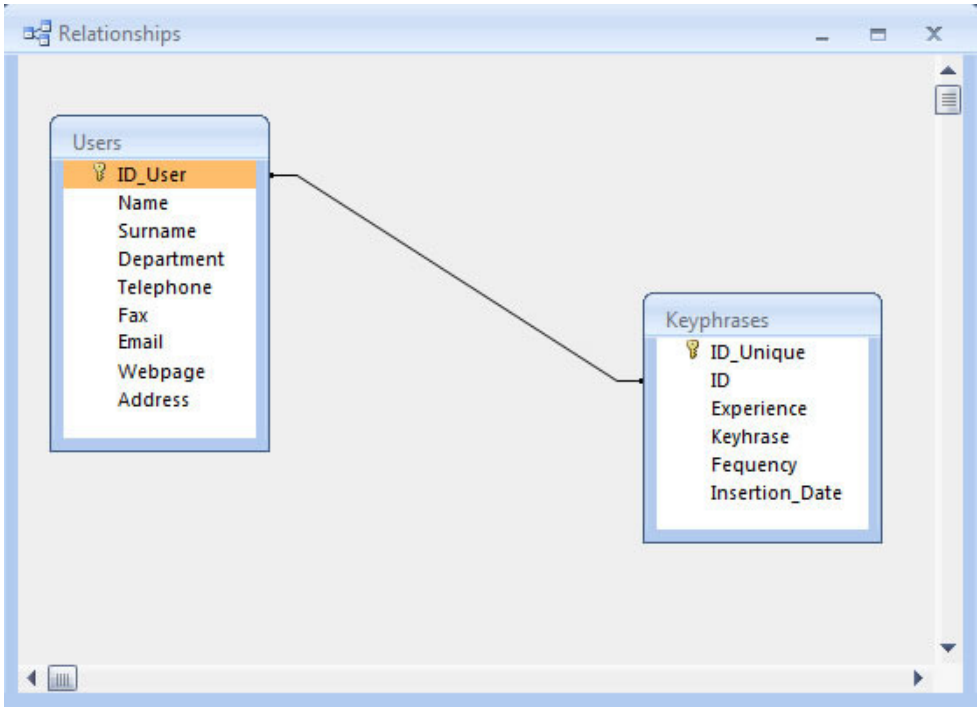


Figure 4.9: EKE Entity Relationship Diagram

4.3.1.4 Expert Search Engine

The Expert Search Engine was developed to enable users to search for other users with knowledge in areas relevant to their query. The engine queries the previously described EKE database and returns a list of experts and their contact details ranked by their suitability to answer the user's query.

Software Design

The search engine was created using PHP (PHP: Hypertext Preprocessor), a server side scripting language. The PHP scripts uses SQL commands to query the expert database. The results are displayed to the user in a friendly manner as shown in the following interface design section. It accepts logical operations "AND" and "OR" in the search query. The results returned are ordered according to how frequently the system has identified the queried phrase from the expert's email (i.e. User "X" who has emailed

the phrase "eclipse" fifty times will appear higher in the returned list of experts than user "Y" who has only emailed the phrase "eclipse" ten times).

Interface Design

A search interface that enables expert seekers to locate people with relevant knowledge has been designed to be clear, simple, and straightforward to use (see Figure 4.10). The expert seeker can enter their query in the search box provided and click the "Search" button. The interface provides users with access to help information.

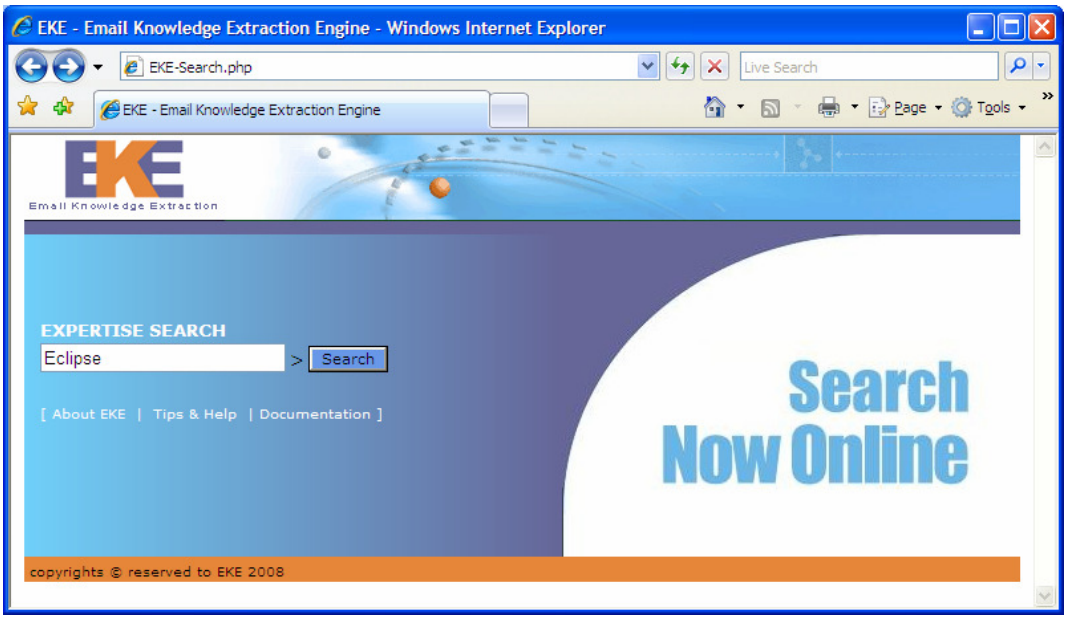


Figure 4.10: EKE Search Engine Interface

Once the expert seeker clicks "search", they are presented with the list of experts whose expertise matches the searched knowledge area. To obtain further information about the returned experts (e.g. contact details, areas of knowledge) and to aid the expert seeker in their selection process, a link is provided to help them choose the right person to contact (see Figure 4.11).

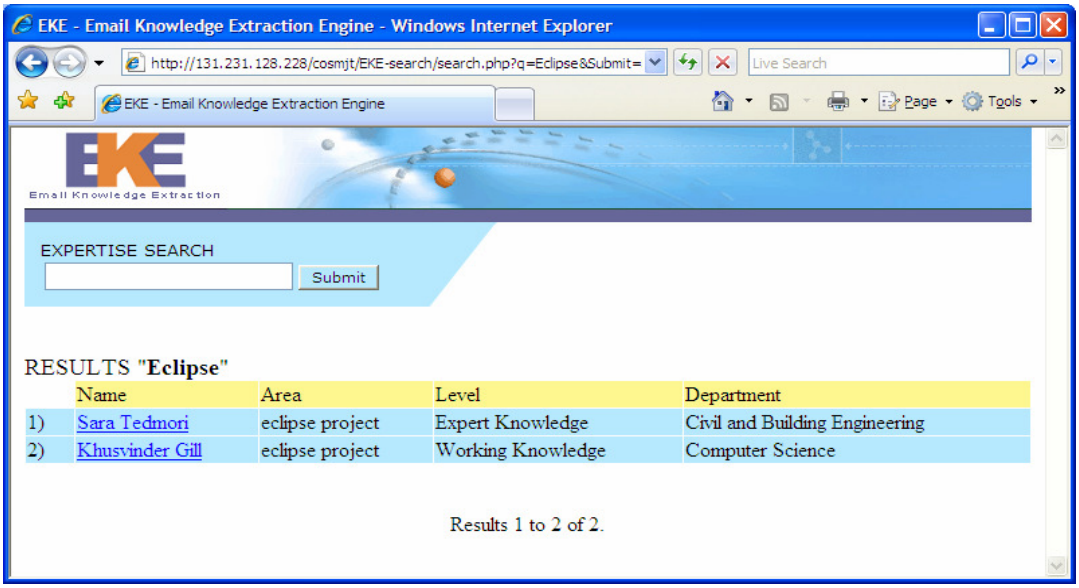


Figure 4.11: EKE Search Engine Return Results

Clicking on a contact link brings up a new window with this information (see Figure 4.12).

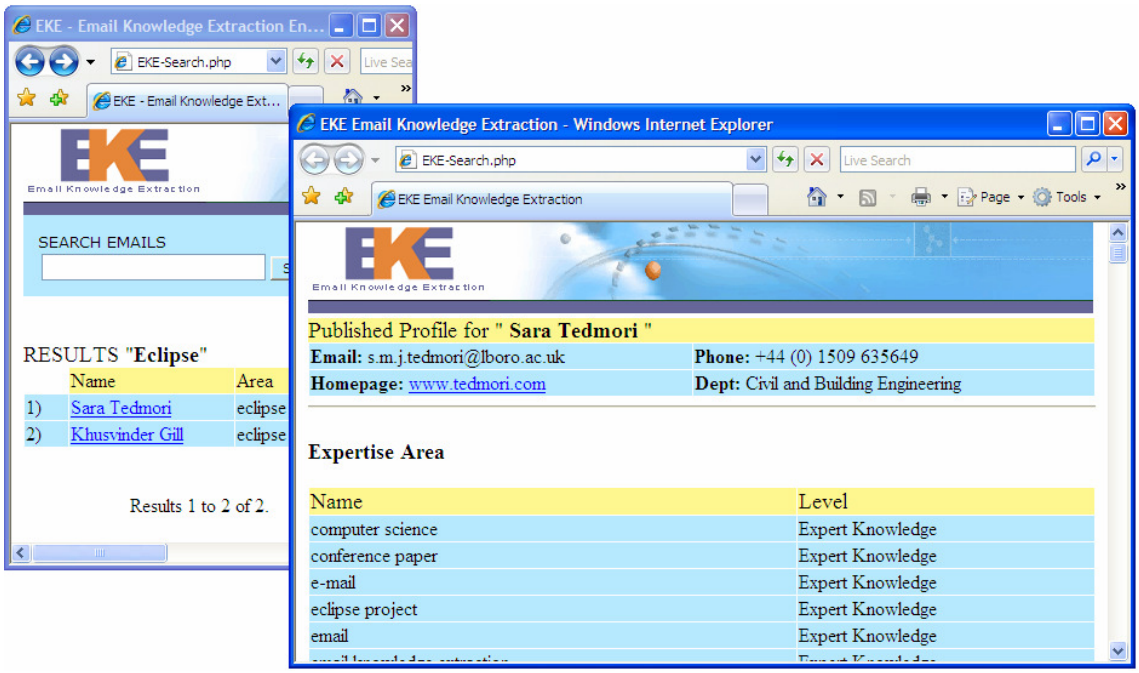


Figure 4.12: EKE Search Engine User Profile View

4.3.2 UNIT TESTING

During this stage, the keyphrase extraction unit was evaluated twice. During the first evaluation, the keyphrase extractor unit developed during Phase 2 (see section 4.3) was tested and compared with other extraction systems. Paper 3, section 4 describes the first evaluation. The experiments undertaken were based on three email collections. The evaluation approach utilised the standard Information Retrieval metrics of precision and recall to show how well generated phrases match phrases, which are considered to be "relevant". For the purposes of this evaluation, the authors of the emails highlighted the "relevant" phrases from the body of the email. When the authors of the emails (e.g. the email sender) were not available, the author (e.g. the author of this research) had to manually assign keyphrases to the emails. These keypharses were then used as the "Gold Standard".

The task involved processing an email message from which a list of keyphrases (a sequence of one or more words appearing in the body of the email) with no duplicates was generated. The f-measure, a widely used performance measure in information retrieval, was employed and is defined as:

$$f-measure = \frac{2 \times precision \times recall}{precision + recall}$$

where *precision* is the estimate of the probability that if a given system outputs a phrase as a keyphrase, then it is truly a keyphrase and *recall* is an estimate of the probability that, if a given phrase is a keyphrase, then a given system will output it as a keypharse.

Table 4.4 shows the precision, recall, and f-measure results. The highest precision (59.6), recall (63.1), and f-measure (61.3) were achieved on the smallest sample (19

messages). The performance measurements obtained were higher than those from other algorithms (Hulth, 2003; Turney, 1997). However, the system requires refining.

Table 4.4 Precision, Recall and F-measure Values from First Unit Evaluation

Corpus Name	Precision	Recall	f-measure
Corpus 1	53.3	57.6	55.4
Corpus 2	59.6	63.1	61.3
Corpus 3	41.7	48.3	44.8

During the second unit evaluation, the keyphrase extractor developed during Phase 3 (see section 4.3) was evaluated. During Phase 3 as described earlier, *WordNet,* a well recognised linguistic tool, was used at the filtering level to help decide whether a phrase is a keyphrase or not. This time the evaluation was conducted to see whether an improvement in the results can be attained, and if so of what magnitude. However, it was found that the improvement in the extracted keyphrases was not significant enough to justify the performance decrease in processing time. The experiments conducted employed the same three email collections used in the previous evaluation. Table 4.5 shows the evaluation results obtained from the two approaches.

Table 4.5: A Comparison between the First and Second Unit Evaluations

Corpus Name	Precision	Recall	f-measure
Corpus 1	53.3	57.6	55.4
with Wordnet	62.5	57.9	60.1
Corpus 2	59.6	63.1	61.3
with Wordnet	70.1	73.6	71.8
Corpus 3	41.7	48.3	44.8
with Wordnet	46.4	49.9	48.1

4.3.3 SYSTEM TESTING

During this stage, the system components (EKE add-in, Keyphrase Extractor Web Service, EKE Database Server, and the Expert Search Engine) were combined and tested as a group. These components were implemented on the development machine, and test cases (dummy emails) were constructed to check the correct interaction of the components. The dummy emails contained known keyphrases, some of which pre-existed in the EKE database. The output of the Keyphrase Extraction Web Service was checked to ensure that the expected outcome had been obtained. Furthermore, the results returned to the user were examined to ensure that existing keyphrases had been removed from those displayed to the user, and that the instance of the keyphrase in the EKE database (i.e. the frequency field) had incremented by one.

During integration testing, it was found that the add-in only operated for the user who installed the add-in. It is important for the system to run on a personal computer with multiple windows user accounts. After further investigation, it was found that the EKE

add-in registry keys (required for EKE add-in to run) were missing for users who are not logged on during installation. To overcome this problem, a program was designed by the author to add these registry keys. The windows "active setup" (a special part of the windows registry which contains locations of programs which have to be run once for all users) was used to run this software and install the registry keys for all computer users.

4.4 STAGE 3: SYSTEM EVALUATION (TASKS 8 –9)

As stated in Chapter 3, this stage involved evaluating the system that was developed during Stage 2 against the requirements identified in Stage 1.

4.4.1 EKE SYSTEM EVALUATION

The evaluation of the developed system was an integral part of this research. As described in section 4.4.1, an evaluation was undertaken with potential end users. The evaluation provided good feedback on the end-users perceptions of the tool in terms of the enablers and the barriers to its adoption. To obtain feedback from potential end users on functionality, robustness, clarity, and ease of use in a working environment, a study was conducted in the Information Science Department at Loughborough University. EKE was installed on 10 members of staff's machines for a six week period. On the installation date, participants were given a walkthrough of the system and were asked to complete a brief pre-study questionnaire designed to elicit their first impressions. After the six week study, the software was uninstalled from the participants' computers. Participants were then provided with the list of key areas identified from their email communications. Participants were asked to assess how well these terms reflected their areas of expertise. This was undertaken as part of a post-study questionnaire addressing areas related to EKE's performance, usability, and handling of socio-ethical challenges. Results from this evaluation are summarised in Chapter 5 and documented in Paper 5, Appendix E.

4.4.2 EKE CONCEPT EVALUATION

The organisational setting in which EKE should be used, made it necessary to establish proof of concept of the system through a demonstration of the implemented prototype to an industrial audience. The evaluation was undertaken for the following reasons: Firstly, it was important to establish end-users' perspectives on usability and functionality of the EKE tool and on socio-ethical challenges raised by its adoption. Secondly, the project was initiated in an industrial context at AstraZeneca whose main interest was establishing the potential benefit(s)/barrier(s) of incorporating an expert locator in their knowledge management initiatives. A critical success factor for the proposed EKE system was therefore potential end-user buy-in. The focus group method was selected to achieve these goals. Focus groups facilitate in-depth understanding of potential end-users views (Krueger & Casey, 2000). The focus group method was adopted as an evaluation method for understanding the end-users' perception of EKE. A working hands-on demonstration during the focus groups would have been the best option for the evaluation. However, due to the many challenges that might arise, when attempting to test the system on the organisational network (i.e. most organisations would not permit their networks to be utilised for this purpose), the next best option for

the EKE concept evaluation was a presentation based hands-off demonstration. The results from this evaluation are further discussed in Chapter 5.

EKE will be used in a social-organisational context. It therefore makes sense to provide industrial participants with the opportunity to supply feedback regarding the EKE tool. The focus group technique facilitated the circulation of creative ideas through the process of interaction as it stimulated discussion and allowed further querying of responses.

The accruement of organisations to participate in the focus groups proved to be a difficult endeavour. Ultimately, it was possible to conduct the study at two organisations. It was felt that this would overcome single group bias, and provide the necessary feedback. The key objective of the evaluation was to assess the potential level of acceptance of end-users of the developed system. During the organisation selection phase of the EKE concept evaluation study, an invitation and product brief were emailed to potential organisations. Those that accepted the offer, received an information pack including the agenda of the focus group.

Group A comprised five participants from SAP, one of the largest European software enterprises whilst Group B comprised seven participants from the UBS (United Bank of Switzerland) one of the leading financial firms. As mentioned in section 3.5.4, Group A was conducted over the telephone while Group B was conducted using the technique of focus group interviewing.

In both focus groups, the sessions started with a fifteen minute presentation which provided the participants with background information about the project. The participants were shown a video of EKE in operation. A group discussion then followed 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.

In particular, participants were asked to complete a SWOT (Strengths, weaknesses, opportunities, and threats) analysis of the tool by writing down the internal and external factors facilitating/hampering its adoption. Participants were then asked to discuss amongst their groups the factors identified. Following this, the groups were asked to conduct a brainstorming session to identify any improvements which they felt could be made to EKE. A record of these discussions was kept. The focus group ended with asking the participant to fill in a questionnaire (see Appendix F, Focus Group Questionnaire).

4.5 SOCIO-ETHICAL CHALLANGES (TASK 10)

The review of the literature revealed key socio-ethical challenges engendered by the proposed expert locator system, EKE. These challenges are addressed in Paper 4, Appendix D and are summarised in Table 4.6. The post-study questionnaire and the questionnaires distributed at the end of the two focus group sessions included a section investigating these challenges. Results from these investigations are summarised in Chapter 5.

Table 4.6: Key Challenges Involved in the Implementation and Use of EKE

Ethical Challenges	Description
Employee Rights	The use of an email locator system such as EKE may threaten employee rights, such as privacy. Thus, it is important to take into account employee rights when implementing EKE.
Privacy and Monitoring	EKE involves monitoring as employees email messages are scanned for expert keyphrase content, which workers may object to. Obtaining informed consent of employees to email scanning is one possible solution to the monitoring problem. However, it is not possible to ensure that the data will be used for the intended purposes as consented to, therefore raising the possibility of loss of privacy. EKE partially addresses this sensitive area by enabling employees to select which keyphrases they would like to share.
Motivational Issues for Knowledge Sharing	Employee motivation to share knowledge is a complex issue and by no means assured. Providing recognition and incentives are key motivators in sharing knowledge, however, they risk users providing misleading feedback to gain additional benefits.
Relationships	Individuals tend to seek information from people they trust, thus employees might feel uncomfortable seeking expertise from certain experts returned by EKE. In addition, in cases where asymmetric relationships are detected between the knowledge seeker and the knowledge holder, the knowledge holder may feel unduly used and the seeker may feel guilty about the one way relationship. Hence, this may result in decreased system usage.
Expert or Non-expert Classification	EKE may never classify some people as experts on certain topics, thus making them feel inferior or unappreciated. Alternatively, some people may resent being classified as experts due to the anticipated higher demand on their time. Moreover, others may be inaccurately classified as experts.
Misleading Email Use	Experts might attempt to escape detection by avoiding the use of technical terminology or by using personalised email accounts. Alternatively, novices may attempt to be classified as experts by using technical terminology and confirming themselves as experts when prompted.
Expert Headhunting	The expertise database contains critical information about an organisation and its employees, highlighting the organisations capabilities and competencies at a given point in time. The expertise database should therefore be kept confidential and inaccessible by competitors or others who might misuse the information.

4.6 SUMMARY

This chapter has discussed the research undertaken to meet the aim and objectives of the EngD project. It has highlighted the main stages undertaken including concept development, system building, and system evaluation. The key findings of the research (including the results from the EKE concept evaluation and EKE system evaluation) are presented in Chapter 5.

5 FINDINGS AND IMPLICATIONS

5.1 INTRODUCTION

This chapter presents the key research findings of the EngD project. To avoid duplication, the findings drawn from the concept development stage detailed in Chapter 4, section 4.2 are summarised below. This is followed by a summary of results from both the EKE concept and EKE system evaluations.

5.2 CONCEPT DEVELOPMENT STAGE FINDINGS

This research has explored the subject of locating an expert through examining email content and found that:

- Several studies exist in the literature highlighting that people remain the most valuable sources of information (see section 4.2.1.1).
- Results from the questionnaire which targeted employees at Leicestershire Constabulary (see section 4.2.1.3) showed that people have difficulties identifying sources of new information. It was found that 75% of respondents "sometimes/frequently" have difficulties identifying resources for new information. Moreover, it was felt that time could be saved if people know where to look for the information required. 59% of the respondents believed that a minimum of 21% of their time could be saved if they knew where to look for the information. For a more detailed analysis see Paper 2, Appendix B.
- The literature review has shown that there are several problems associated with the previous and current approaches to expert finding that still remain to be addressed (see Paper 1, Appendix A).
- Results from the online questionnaire which targeted academics from Loughborough University and employees at an IT firm in the US have shown that it is possible to profile users' knowledge based on the emails they send (see section 4.2.1.3). It was found that 59% of the respondents' emails in the academic sector and 73% of the respondents' emails in the industrial sector contained keyphrases that could be used to profile their knowledge. The results were published Tedmori et al. (2006).
- Results from the questionnaire which targeted employees at Leicestershire Constabulary show that more than half of the respondents (52%) believed that EKE would benefit their organisation. Moreover, if their organisation was to use EKE, 89% were willing to contribute by making their knowledge areas public.

One of the outputs of the EngD project is a tool that exploits email for the purpose of finding experts. The next section provides the findings of the evaluations of EKE.

5.3 EVALUATION OF THE EKE SYSTEM

This section provides a brief overview of the EKE system evaluation study conducted in the Information Science Department at Loughborough University. The study consisted of a pre-study questionnaire, 6 weeks study period, and post-study questionnaire. A total

of 10 participants took part in this study. They all had several years of work experience at Loughborough University, averaging 10.6 years. All the participants use email on a daily basis for knowledge related activities, with half sending more than 10 knowledge related emails a day.

5.3.1 PARTICIPANTS' FIRST IMPRESSIONS ABOUT THE SYSTEM

On the day EKE was installed on each of the participants' machines, they were given an overview of the tool and asked to provide their initial thoughts. From the comments made, there were indications that the EKE concept was regarded as being a good, useful and interesting idea. It was suggested that the information captured by EKE could be used to enable personal ratings for the purpose of personal development and that EKE's ability to integrate with existing technologies (i.e. Outllook) would encourage its adoption and use.

In relation to the clarity of the keyphrase categorisation form, 50% of the participants thought the form was very clear, 30% thought the form was somewhat clear, 20% expressed neutral views, and no one felt the form was not clear. However, opinion was divided between those who thought the form could be improved (50%) and those who thought it could not (50%). A few suggestions for improvement were made. Among the suggestions was the ability to turn the form off, the addition of a "No Knowledge" category, and the provision of links to the knowledge categories on the form to help users make their selection.

5.3.2 END USER EVALUATION OF EKE

After the six week trial period of using EKE, a follow up questionnaire was distributed. The analysis of the results obtained from this questionnaire are summarised in this section. For a more detailed review of the study, please see Paper 5, Appendix E.

The responses shown below were grouped according to whether the participant was positive, negative, or neutral in their response. This was done in order to make it easier for the reader to gauge whether the majority of respondents provided a positive or negative answer. The positive responses attained indicated that:

- EKE's concept is easy to grasp (90%).
- EKE is easy (100%) and intuitive (80%) to use.
- EKE does not adversely affect the ability to send emails (100%).
- EKE's keyphrase categorisation form layout is clear (80%).
- EKE's knowledge categories are appropriate for their purpose (70%).
- The task of extracting keyphrases from email is worth the effort (100%).
- It is not desirable for email attachment to be included in the analysis (70%).

Although the majority of the feedback from participants was positive, the responses indicated that certain aspects of EKE need further consideration and could be improved. Some of the keyphrases extracted by EKE were seen as irrelevant and hence, the keyphrase extractor should be refined to identify more specific terms and reject more common terms. The participants' views on this issue were very subjective. For example,

some participants were not pleased that the system picks up module codes and geographical information (e.g. names of places), others were quite happy that it does.

EKE examines the whole email chain, from which a long list of keyphrases is often extracted even in the case of short emails. Moreover, the process of sending an email is slowed down, and as such, it would be desirable if the tool could be configured to run if certain conditions are satisfied (i.e. email is business related). More general improvements were suggested, including:

- Replacing the "BK" category with an "Interest" category: Temporary interests are time bounded. A lot of knowledge is temporal and tied to a particular context. It was indicated that users tend to regard themselves as knowledgeable in a particular context, and not in a generic sense.
- Introducing a finer level of granularity by increasing the range of knowledge categories: It was agreed that the use of more knowledge categories would be helpful. However, this may slow down the categorisation process.

Participants were shown a list of the keyphrases that had been extracted from the emails they sent over the six week study period. They were asked to rate for each of the knowledge categories, how well these keyphrases represented their knowledge. The findings showed that people are better at identifying the areas they perceive themselves as experts, as opposed to the areas in which they have basic or working knowledge. This can be attributed to a number of factors such as email is a better source of keyphrases which represent expert knowledge areas; the extraction algorithm is better at identifying keyphrases representing knowledge areas employees are experts in; participants are better at identifying the areas which they deem themselves as experts in; and/or the group of academics, who composed the sample, are experts in their particular fields, hence they send more specialised emails.

Each participant was shown a keyphrase list. The list was divided under the three knowledge categories, "BK", "WK" and "EK". Under each category, a list of ten random keyphrases was selected, where available, from the corresponding category in the user's profile. The participants were asked to indicate whether or not each of the keyphrases reflected their area and level of knowledge. The results showed that the keyphrases extracted under "WK" and "EK" represented the respective users' knowledge more accurately than those classified under "BK". This is interesting because the users classified the keyphrases themselves. This could be attributed to the factors described earlier.

5.3.3 SOCIO-ETHICAL CHALLENGES

In relation to the impact on employee privacy, half of the participants (50%) felt that the use of EKE has little or no impact. The other respondents thought that the system has some impact (40%) on employee privacy. This can be attributed to the participants' perception that any personal information shared across organisational boundaries affects privacy. Another reason is related to the keyphrase form which gives the sender the impression that big brother is watching. Only 10% felt that the use of EKE has a moderate impact on employee privacy and expressed that the scale of impact depends on how EKE will be used.

Half the participants (50%) agreed that Loughborough University's privacy policies provide sufficient protection against the potential privacy problems that might result from the use of EKE**.** The other half (50%) expressed neutral views due to a lack of knowledge about the available legislation or due to their perception that if the organisation wants to know what you are doing, they can always find out. It was felt by the 78% of the respondents (one person did not answer this question) that the current legislation in the UK offers sufficient protection against the potential privacy problems that might result from the use of EKE. As such, the majority (80%) felt comfortable with EKE analysing their emails.

EKE enables users to select what they would like to share with the rest of the organisation. Opinion was divided on the level of privacy protection this provides. Respondents felt that it is the employee's responsibility to choose the appropriate keyphrases. Others said they are happy as long as the keyphrases can be removed from the database. On the other hand, others expressed that there will always be organisational boundary issues to contend with.

All respondents (100%) were willing to share knowledge when identified by EKE as experts. Some stated that they are willing to share knowledge, however being identified as an expert does not imply willingness to share knowledge with all those who contact them. Others linked this to the nature of information that might be shared. For example, if sharing such information contradicts preserving intellectual property rights then they are reluctant to share it. Some expressed a higher degree of readiness to share knowledge in areas in which they regarded themselves as experts, rather than in areas in which they have basic knowledge.

The majority of participants (70%) felt that the provision of recognition and incentives is to some extent a key motivator to use EKE and share knowledge. Some said that this is dependent on the individual; what may be an incentive for one individual does not necessarily represent an incentive for another. Others regarded the use of EKE as sufficient in itself as it may lead others to contact them, enhancing therefore the prospectus of networking. It was also highlighted that to encourage participation, some incentive (e.g. appraisal, research performance) may be more important to provide at the early stages of adoption rather than at later stages.

Participants were generally comfortable with seeking information from experts they do not know. They clarified this issue by saying that care needs to be taken when approaching those experts. Others said that this depends on who the experts are and how well they are prepared to answer their queries. From those who took a neutral stance, it was evident that they found contacting someone they don't know a bit difficult, but sometimes necessary. For example, if time is pressing, then they would directly contact people they don't know. However, if more time was available, they preferred in the first instance to either contact someone they know or to seek an introduction from someone they are acquainted with.

All respondents expressed willingness to reply to enquiries generated through EKE from co-workers. A link was established between the number of enquiries they are willing to reply to and the nature of the enquiry (e.g. how complex the query is, how much time it requires, who is asking) and/or the circumstances of the respondent (e.g. time of request). For example, the less complex the enquiries are, the more enquiries they are willing to reply to.

Discomfort was expressed with the possibility that EKE might incorrectly specify a user as an expert on a particular topic. It was stated that such a situation could be misleading as it raises questions regarding the validity of the system. Finally, all respondents reported that EKE had minimal or no negative effect on their email usage.

EKE was seen as interesting because it allows users to see the areas they have been focusing on and the other areas they have been neglecting. EKE at times can be distracting (especially when the list returned is long or doesn't make sense). Categorising the terms extracted can at times be difficult. Some thought it was very hard to rank themselves in comparison to others (i.e. should they be ranking themselves in comparison to other people in their department, or in comparison to other people in the organisation). Providing education and training was seen as integral, in terms of how to make the most out of the tool and how it would benefit them.

5.4 EKE CONCEPT EVALUATION FINDINGS

This section provides the overall findings obtained from the discussions of the two focus groups. As described in section 4.4.2, 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. Group A was conducted over the telephone at SAP and comprised five participants. Group B was conducted using the technique of focus group interviewing at UBS and comprised seven participants. For a thorough review of Group A and Group B, see Appendix F, Focus Groups Discussions and Questionnaire Analysis. A questionnaire was used at the end of the focus group sessions to capture the participants' views on socio-ethical challenges (see Appendix F, Focus Group Questionnaire).

5.4.1 FOCUS GROUPS FINDINGS

The groups felt that EKE is well designed, easy to use, and would greatly aid employees in finding other employees with the relevant knowledge, 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 because it 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.

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 necessarily 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. From the author's point of view, the added complexity, efforts, and expenses required to create and maintain

such a list are excessive for large organisations, to which such tools offer the most benefits. Moreover, creating and maintaining domain specific ontologies is difficult, because it requires a deep understanding of the industry under consideration. In addition, ontologies can not be easily adapted for reuse in a different domain.

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, through the adoption of an ontology based approach. The second suggestion was that users should be able to select when and how often 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 the facilitation of better organisational communication networks. 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. Both groups, however, had different reasons for their concerns. 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 the knowledge areas of the meeting attendants.

5.4.2 FOCUS GROUP SOCIO-ETHICAL QUESTIONNAIRE

In total, 12 questionnaires were distributed. Of these, 5 questionnaires were distributed to Group A and 7 questionnaires were distributed to Group B. Tables of complete findings are provided in Appendix F, Focus Group Discussions and Questionnaire

Analysis. The response rate for both groups was 100%. Group A and Group B participants' years of work experience at their current organisations averaged 2 years and 2.5 years respectively.

Email Usage: Group A participants were more frequent email users than Group B. This can be attributed to Group B's use of alternative communication mediums (i.e. instant messaging), as inferred from the focus group discussion.

Perceptions of EKE: All participants in Group A rated the use of EKE as easy. The majority of the respondents (80%) felt that the concept of EKE was easy to grasp. When asked about the extent of their agreement/disagreement with the statement, "I would feel comfortable using this tool", 60% expressed their agreement. The tool was found to be intuitive by 80% of the participants. When asked about the appropriateness of EKE's knowledge categories, the majority (60%) opted for the neutral option. Again, the majority (80%) opted for the neutral choice when asked about clarity of the keyphrase categorisation form.

The majority of participants in Group B indicated that the tool was easy to use and felt that the EKE concept was easy to grasp. None of the participants were uncomfortable with using the tool with 100% indicating they found the tool to be intuitive to learn. 86% found the layout of the keyphrase categorisation form clear and 71% found the knowledge categories appropriate.

All participants from both groups stated that the tool would be beneficial to them when searching for experts. Group B participants provided further clarifications and attributed this to:

"Finding the right person can often hold up other work, this tool should help reduce this time wasted".

"The tool has provided a foundation for the claims of people i.e. people marked as experts by the system would have foundations for calling themselves experts".

"The tool's integration into existing work practices will help remind people to keep their expertise profiles up-to-date compared to other tools".

"The tool provides a single point where I can find all the information I need".

Socio-ethical Challenges: This section summarises the socio-ethical challenges linked with the adoption of EKE. The structure of this section is composed of several subsections, each representing a socio-ethical aspect. Each subsection starts with Group A analysis followed by Group B analysis, and where necessary followed by further discussions.

In relation to the impact on employee privacy, all participants in Group A and Group B acknowledged the existence of the threat to some degree. Group B attributed this to three factors:

- **Improper use of the tool:** Participants felt the tool may be used for purposes other than those they had consented to such as monitoring of their activities.
- **Unauthorised Access:** Some of the participants felt they would not like certain information to be made available to the whole organisation. Instead, they said that they would like certain information to be shared only amongst their immediate colleagues.

- **Email usage:** It was felt by some that the use of email in itself may cause concern to certain individuals as it is a personal form of communication and as such less personal sources of information are better to use for knowledge profiling.

The participants who expressed moderate concern over privacy (14%), related their concern to EKE scanning emails irrespective of their nature (i.e. business or private email). However, none of the participants felt that EKE threatens privacy to a great extent.

All participants in Group A agreed that the current organisational privacy policies provide them with sufficient protection. Some felt that the organisational policies in Germany provide very strong data protection. Although others were not completely aware of such policies, they still expressed faith that they are being protected. When asked about the current governmental legislation, the majority of participants didn't respond to the question. They stated in the clarification section that they were unable to answer the question due to a lack of awareness of the existing legislation.

The majority of participants in Group B felt that their current organisational policy provided them with sufficient protection against privacy issues that might arise from the use of EKE. The reasons given for the high degree of confidence in organisational policy included a general feeling that EKE's integration into existing work practices meant that the existing policies would cover most aspects. The comments included.

"We already provide our skill level manually. So, I believe the existing policies should be okay for the EKE tool".

"Email is already sent around other systems and I don't see this system as any different".

The participants in Group B also expressed faith in UK legislation with regards to protecting their privacy rights. The majority of participants agreed that they had confidence in legislation to protect their rights. Most felt that knowledge sharing is good and that the keyphrase extraction process is anonymous enough to comply with the legislation.

It is worth noting that Group A's and Group B's belief in the organisational and national legislation to protect them against any privacy issues that may arise from use of EKE, indicates that few changes have to be made, if any, to existing practices. It can be inferred that there is a significant and widespread lack of awareness of both organisational and national privacy legislations.

Participants were asked to express how comfortable they would be with EKE analysing their emails. The vast majority were happy. However, both groups stated that they would be even more comfortable if only business related emails were analysed.

Participants were asked to indicate the extent to which they think the ability to select the keyphrases they wished to share with the organisation helps overcome privacy issues. In Group A, the majority indicated that such ability helps protect privacy "to some extent". When asked for clarification, respondents highlighted that it is a good idea to be able to select the keyphrases they want to share. Others thought that there was still some threat to privacy because currently all the emails are analysed. Another suggestion was made to limit the range of keyphrases returned by introducing ontologies, which would help reduce the impact of the threat to privacy. It was also felt that storing any information

centrally in one location introduces a privacy threat, leading to 25% indicating that the keyphrase selection protected privacy to a little extent.

A wide selection of opinions emerged from Group B. The dominant opinion amongst participants was that the keyphrase selection feature does to some extent help overcome certain privacy issues. Supportive arguments that they gave included, "*it is good that you can filter out the words you are not comfortable with*" and "*since you can opt for no data to be sent this seems to cover privacy concerns*". Some participants felt that this feature did not provide them with sufficient protection of their privacy, as certain keyphrases are too confidential to be scanned.

A crucial factor in the success of EKE is the participants' willingness to share knowledge with others when requested. All participants in Group A expressed willingness to share knowledge. In addition, the majority said that in a given week they would be happy to reply to more than 10 queries. There was a general acknowledgement among the participants that sharing information and knowledge is good practice because it reaps benefits for both parties involved in the information sharing process.

Similarly, participants in Group B were all happy to share their knowledge with others. However, the extent to which they could provide help is dependent on their workload, availability and the confidentiality of the information sought. In addition, the participants expressed the common view that they would be happy to respond to most enquires; however, as discussed previously, this would be dependent on their work load and the amount of time needed to deal with the specific query. It was suggested that one method to aid searchers and knowledge holders would be to include an availability setting, with the returned results showing those employees who are currently too busy to answer queries. The willingness of groups' respondents to share knowledge via EKE, which is one the most crucial aspects of EKE's success, is encouraging.

Participants were asked to specify the extent to which they think that the provision of recognition and incentives is a key motivator to use EKE and share knowledge. In Group A, the majority felt that such provision has some impact on their motivation to use the system and suggested that the benefits, whether individual or collective, need to be made explicit if the organisation is to reap more benefits from the system.

Similarly, in Group B, the majority felt that the provision of recognition and incentives would to some extent encourage the use of EKE. The participants comments in this regard included "*can be useful*", "*people do care for what they get back in return for whatever they do*" and "*it would motivate everyone to contribute*". Responses obtained from both groups, if realised in the real world, would suggest that there is a strong relationship between the successful adoption of EKE and providing incentives.

The EKE system provides a web based search engine that allows users to search for colleagues who may possess the knowledge needed to provide them with assistance. The search engine's return results may include employees with which the user is not personally acquainted. Both groups indicated that they would have little hesitation in contacting employees with whom they are not familiar, with some even indicating that they would enjoy the new social interaction and the possibility of networking. The groups' responses show that participants have little hesitation in contacting employees returned by the search that they do not know. From the author's point of view, not returning unknown employees would reduce the usefulness of EKE, especially in large

organisations, where the majority of users can be widely distributed across the world and unknown to each other.

EKE's primary objective is to categorise users knowledge. As with all tools, there is the possibility that the tool may incorrectly classify an individual as an expert. The majority (80%) were neutral in their attitudes towards such misclassification. They suggested that generally people after some time, will no longer be experts on a particular topic. Hence, the participants were keen to have the EKE tool extended so that they are able to modify their profiles. The rest of the participants (20%) expressed discomfort by stating that EKE should classify users very carefully, in order to avoid misclassification.

A significant percentage of participants in Group B stated that the possibility of misclassification would make them feel uncomfortable. Some participants still felt that this should not be an issue by stating that "*if someone is contacted by someone incorrectly, one should not mind it*". However, as previously mentioned, a significant proportion felt at unease with this situation and as such, this issue requires particular attention for successful adoption of the tool. The participants expressed that their fears could be overcome by extending EKE to permit users to modify their profiles.

The participants were asked to state the extent to which they felt the tool would negatively affect their email usage. The majority of the participants in Group A felt that it would have no effect as they felt that EKE is seamlessly integrated into email, their main communication medium.

The majority of the participants in Group B felt that the tool would have little or no effect, stating that "*the form seems simple and quick so shouldn't cause any massive delays*". Some felt that EKE may affect their email usage to some extent and commented that the tool may slow down a little the process of sending an email. They further commented that any tool that adds extra steps during busy periods might be annoying; however this could be overcome by allowing users to disable the tool at certain times.

Comments: From the comments section, participants expressed a general satisfaction with the EKE concept and tool with such comments as:

"The tool is of good use and can be used effectively for sharing knowledge".

"An interesting system, which would be extremely helpful in classifying users as experts of particular technologies/ business areas".

"I feel I have developed a good understanding behind the basic concept of the EKE tool, and feel it is a new and innovative idea".

"There are other tools on the market; however the sophistication and intelligence of this tool give it an edge".

Certain suggestions for improvements were made, these included:

"Providing an option to the employee to select their profile areas and then screening extracted keyphrases against these areas and only displaying those keyphrases within these areas." This will reduce intrusiveness and the number of keyphrases returned to the user

"Improving the keyphrase extraction".

"It would be nice to get a better understanding of the system by a hands on trial where we could use EKE in the daily working environment".

5.5 SUMMARY

This chapter has focused on the evaluation of EKE. From the number of different end-user evaluations, it can be concluded that the usefulness of email, as a source for knowledge profiling, has been established. Email based expertise finding systems such as EKE were found to provide a solution that large organisations can adopt into their existing systems and work practices. Moreover, it has been shown that users are more willing to accept a system that fits into everyday work practices. The research findings suggest that once users are involved with these systems, they would be happy to contribute and to share knowledge with others, when requested.

6 CONCLUSIONS AND RECOMMENDATIONS

6.1 INTRODUCTION

This chapter discusses the implications of the research on both the industrial sponsor and the wider industry. It provides a critical evaluation of the research and offers recommendations for further work. The chapter presents the overall conclusions of the research findings.

6.2 THE IMPLICATIONS/IMPACT ON THE SPONSOR

AstraZeneca's main objective for sponsoring this research was to establish the potential of email content for expert finding within organisations. This was done with the intention of reducing the time their employees waste looking for information. The EngD project has made a positive contribution to the sponsoring company. The sponsors have learned that email content is a rich source of information that can be used to accurately profile their employees' knowledge. They have found that users are willing to accept systems, such as EKE, that fit into everyday work practices. The research findings suggest that users are happy to share knowledge and contribute to such systems, thus aiding in increasing employees' productivity and in turn the organisation's profits.

Additionally, several other applications for EKE have emerged, which AstraZeneca can make use of, including: 1- Personal development - where individual employees can monitor their own profiles to identify areas for improvement. 2- Effective recruitment - where AstraZeneca can use EKE's database to monitor their employees' expertise, find areas of knowledge gaps, and as a result, recruit more effectively.

EKE consists of several components, one of which is the keyphrase extractor. This component can be used to profile knowledge using alternative information sources, such as published documents. Furthermore, the keyphrase extractor can be applied to other applications such as indexing, categorising, and summarising pharmaceutical documents.

6.3 THE IMPLICATIONS/IMPACT ON THE WIDER INDUSTRY

The wider industry would benefit from the implementation of EKE in ways similar to those of the industrial sponsor. Additionally, exploiting the knowledge that lies dormant in organisational resources has recently been a subject of interest for large organisations, government institutions, and academia. Organisations are reluctant to instantaneously adopt new systems. This can be attributed to the significant investments they make in their existing systems and exacerbated by the difficulty of obtaining universal buy-in across the whole organisation. Consequently, considerable efforts have been made to incorporate knowledge management into existing systems and work practices, with the aim of overcoming these barriers. EKE has been engineered to integrate with existing email systems and as such, it goes a long way in overcoming these barriers.

The research findings indicate that the practical application of EKE has a broad acceptance in the community and offers significant benefits. EKE can be integrated into the email system of any knowledge intensive organisation. Consequently, organisations

have shown enthusiasm to investigate the potential of adopting such a system. EKE has been demonstrated and accepted in academia (Loughborough University) and industry (large finance and software organisations).

People will always require information, and thus they will always need tools to assist them in efficiently and effectively locating the relevant sources of knowledge. The evaluation conducted at the end of the system development stage revealed EKE's potential for obtaining high user acceptance. Potential employees have shown a great deal of interest and support in the knowledge sharing concept. They have indicated a willingness to both contribute and share knowledge. The use of expert locators result in a better utilisation of time spent by employees looking for information, and as such organisations should appreciate the net gains of such approaches.

6.4 CRITICAL EVALUATION OF THE RESEARCH

The aim of this research is to investigate the capture of knowledge within emails for the purpose of locating expertise within organisations. This is a very challenging task considering the complexity of capturing knowledge from email content, coupled with the limited time frame available for the EngD project. Six major limitations have been identified which have given rise to further research.

- **Questionnaires, focus groups, and EKE study:** All survey and evaluation methods employed in this research require a sample that is large enough to obtain results representative of the target population. However, this was not possible due to the wide scope of the subject domain, coupled with financial and time constraints.
- **Benchmarking:** This has been difficult primarily due to two reasons. Firstly, organisations are not willing to share data on the performance of existing expert locator systems. Secondly, most expert locators are bespoke systems and as such are not available for researchers to evaluate.
- **Software Compatibility:** The existing implementation of EKE is designed to be interoperable with Microsoft Outlook 2003. Due to time constraints, it was not possible to expand the interoperability of EKE to other email clients.
- **Extraction Techniques:** There undoubtedly remain a number of existing techniques which have not been explored, and which can be the subject of further research.
- **Participation:** Recruitment of participants posed a significant challenge for the software evaluation. Organisations were conservative in terms of their willingness to try an untested system on their networks.
- **Sponsoring Company:** The industrial sponsor AstraZeneca was unable to provide testing facilities for the project as their email system is outsourced to a third party organisation.

6.5 RECOMMENDATIONS FOR INDUSTRY/FURTHER RESEARCH

In spite of the extensive research conducted over the past four years, there remain a number of issues that require further research. Some organisations do not have an

appreciation of the vast amount of knowledge that resides dormant in places such as email and of the benefits they can gain from utilising this knowledge. Awareness should be raised regarding the benefits of easily finding people with expertise. This would enable employers and employees to realise that expert locators are gradually becoming indispensable to companies seeking to take advantage of the under-utilised knowledge resources.

There is an obvious need for more focused information dissemination on non-technical aspects of agent technology. A significant amount of the available literature on expert locators is focussed on technical issues such as system architecture, with very little attention given to issues related to the development and deployment of expert locators. Early adopters of expert locators should be encouraged to provide benchmarking and case studies of their experiences.

As stated previously, other recommendations include:

- Wider evaluations involving a larger participant base;
- Increased interoperability of the EKE tool with other email clients;
- Experimentation with other extraction techniques, such as social networks;
- Conducting cost-benefit analysis to highlight the viability of EKE to organisations; and
- Addressing further issues raised by end user evaluation.

6.6 OVERALL CONCLUSIONS

The aim of this research study, as has been stated in Chapter 1, was to "explore the potential of using email messages to determine and locate expertise within organisations". In order to achieve this aim, five specific objectives were defined:

- Objective 1: To review literature on approaches to expert finding;
- Objective 2: To determine the potential value of email in supporting the process of expert finding;
- Objective 3: To develop the expert locator system, EKE;
- Objective 4: To evaluate the use of EKE; and
- Objective 5: To identify enablers and barriers to the use of EKE as an expert locating system.

The research achieved all the above objectives, as detailed throughout the thesis. Below are the main contributions of the research (for a thorough review see section 2.9). The research has:

- Identified areas to consider when designing systems similar to EKE;
- Identified that email contains valid expertise indicators;
- Developed, implemented, and tested an expert locator system which exploits email in a real world environment;
- Highlighted that the majority of people recognise the benefits of the EKE concept; and

- Explored socio-ethical challenges, leading to an enhanced understanding of end users' perceptions.

The review of the contributions shows that the research objectives have been satisfied. The first stage of the research involved a literature review to get an in-depth understanding of expert locators (Objective 1). A domain analysis of the literature identified shortcomings in the research subject and further aspects that needed to be addressed. This stage also involved an online questionnaire that investigated whether it was possible to profile individuals' knowledge from the emails they send (Objective 2). The second stage involved building EKE (Objective 3) that lead to Stage 3 which involved trailing and evaluating EKE (Objective 4). An additional stage investigated the socio-ethical challenges associated with the adoption of EKE by organisations (Objective 5).

In brief, this research has shown that email is a rich source for knowledge profiling, which offers many advantages over existing expert locators. The integration between EKE and organisations' existing email systems reduces implementation costs, and lowers users' resistance to change by integrating into their daily work practices. The user centric approach to knowledge elicitation helps overcome some of the users' privacy concerns and encourages knowledge sharing. Moreover, networking and greater collaboration between colleagues is encouraged through the use of EKE, leading to the potential for increased productivity. Better education and training in regards to the benefits offered by EKE, will help provide the greatest returns for organisations and users. These factors when coupled with the many additional applications of EKE highlight that the adoption of EKE will greatly benefit both academia and industry.

REFERENCES

1. Alavi, M. & Leidner, D.E. (1999). "Knowledge Management Systems: Issues, Challenges, and Benefits", *Communications of the Association for Information Systems,* vol. 1, no. 2, pp. 1-37.

2. Allen, T. (1977). *Managing the Flow of Technology,* MIT Press, Cambridge, MA.

3. AskMe . Available: http://www.askmecorp.com/ [2005, 09/16].

4. AstraZeneca , *About us*. Available: http://www.astrazeneca.com/article/11148.aspx [2007, 12/20].

5. Avison, D.E. & Fitzgerald, G. (2003). *Information Systems Development: Methodologies, Techniques and Tools,* 3rd edn, McGraw Hill, Maidenhead.

6. Awad, E.M. & Ghaziri, H.M. (2004). *Knowledge Management,* Pearson Education, Upper Saddle River, N.J.

7. Balog, K. & de Rijke, M. (2007). "Determining Expert Profiles (With an Application to Expert Finding)", *International Joint Conference on Artificial Intelligence*, pp. 2657.

8. Bannon, L.J. (1986). "Helping Users Help Each Other" in *User Centered System Design: New Perspectives on Human-Computer Interaction*, eds. D.A. Norman & S.W. Draper, Erlbaum, Hillsdale, NJ, pp. 399-410.

9. Barker, K. & Cornacchia, N. (2000). "Using Noun Phrase Heads to Extract Document Keyphrases", *The Thirteenth Canadian Conference on Artificial Intelligence (LNAI 1822)*Montreal, Canada, pp. 40.

10. Becerra-Fernandez, I. (2006). "Searching for Experts on the Web: A Review of Contemporary Expertise Locator Systems", *ACM Transactions on Internet Technology,* vol. 6, no. 4, pp. 333-355.

11. Becerra-Fernandez, I. (2000). "The Role of Artificial Intelligence Technologies in the Implementation of People-Finder Knowledge Management Systems", *Knowledge-Based Systems,* vol. 13, no. 5, pp. 315-320.

12. Beijerse, R.P.U. (1999). "Questions in Knowledge Management: Defining and Conceptualising a Phenomenon", *Journal of Knowledge Management,* vol. 3, no. 2, pp. 94-110.

13. Bertolotti, L. & Calzarossa, M.C. (2001). "Models of Mail Server Workloads", *Performance Evaluation,* vol. 46, no. 2-3, pp. 65-76.

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

15. Blau, P.M. (1986). *Exchange and Power in Social Life,* Transaction Publishers, New Brunswick, NJ.

16. Bontis, N., Fearon, M. & Hishon, M. (2003). "The e-Flow Audit: An Evaluation of Knowledge Flow Within and Outside a High-tech Firm", *Journal of Knowledge Management,* vol. 7, no. 1, pp. 6-19.

17. Bryman, A. & Bell, E. (2003). *Business Research Methods,* Oxford University Press, New York.

18. Burstein, F. (2002). "System Development in Information Systems Research" in *Research Methods for Students, Academics and Professionals: Information Management and Systems*, ed. K. Williamson, 2nd edn, Centre for Information Studies, Charles Sturt University, Wagga Wagga, NSW, pp. 147-158.

19. Burt, R.S. (1992). *Structural Holes: The Social Structure of Competition.* Harvard University Press, Cambridge.

20. 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.

21. Carroll, J. (1997). *Human-Computer Interaction: Psychology as a Science of Design*.

22. Chen, K. (1999). *Automatic Identification of Subjects for Textual Documents in Digital Libraries*, Los Alamos National Laboratory, Los Alamos.

23. CMS 2005, *Selecting a Development Approach*, Centers for Medicare and Medicaid services, Office of Information Services.

24. 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.

25. Cohen, D. & Prusak, L. (2001). *In Good Company: How Social Capital Makes Organizations Work,* Harvard Business School Press, Cambridge, MA.

26. Corporate Smarts . Available: http://www.corporatesmarts.com/ [2005, 09/16].

27. Cross, R. (2000). "More than an Answer: How Seeking Information Through People Facilitates Knowledge Creation and Use", Toronto, Canada.

28. Cross, R. & Baird, L. (2000). "Technology is not Enough: Improving Performance by Building Organizational Memory", *Sloan Management Review,* vol. 41, no. 3, pp. 41-54.

29. Cross, R., Parker, A. & Rice, R.E. (2001). "Information Seeking in Social Context: Structural Influences and Receipt of Information Benefits", *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,* vol. 31, no. 4, pp. 438-448.

30. Cross, R. & Sproull, L. (2004). "More Than an Answer: Information Relationships for Actionable Knowledge", *Organization Science,* vol. 15, no. 4, pp. 446-462.

31. 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.

32. Davenport, T.H. & Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know,* Harvard Business School Press, Boston, MA.

33. Davis, G.B. & Olson, M.H. (1985). *Management Information Systems: Conceptual Foundations, Structure, and Development,* 2nd edn, McGraw-Hill, Inc, New York, NY, USA.

34. Derrington, S. *, Getting Control of the Storage Environment*. Available: http://www.it-observer.com/articles/1249 [2007, 11/22].

35. Desouza, K.C. (2003). "Facilitating Tacit Knowledge Exchange", *Communications of the ACM,* vol. 46, no. 6, pp. 85-88.

36. Dix, A., Finlay, J.E., Abowd, G.D. & Beale, R. (2003). *Human-Computer Interaction,* 3rd edn, Prentice Hall.

37. Ducheneaut, N. & Belloti, V. (2003). "Ceci N'est Pas un Objet? Talking about Objects in E-mail", *Human-Computer Interaction,* vol. 18, pp. 85-110.

38. Duffy, N. (1999). "Benchmarking Knowledge Strategy" in *Leveraging Knowledge for Business Performance 1999: Knowledge In Action*, N. Duffy, A. Jooste, and L. Whittaker Eds edn, WITS Business School, Johannesburg.

39. Erickson, B.H. (1988). "The Relational Basis of Attitudes." in *Social Structures: A Network Approach:*, Barry Wellman and S. D. Berkowitz (eds.), edn, Cambridge University Press., New York:, pp. 99-121.

40. Feldman, S. (2004). *The High Cost of not Finding Information*, KM World Magazine.

41. Fellows, R. & Liu, A. (2003). *Research Methods for Construction,* 2nd ed. edn, Blackwell Science Limited, Oxford.

42. Fink, A. & Kosecoff, J. (1998). *How to Conduct Surveys: A Step by Step Guide,* 2nd edn, Sage Publications, Inc, Beverly Hills, CA.

43. 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.

44. Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C. & Nevill-Manning, C.G. (1999). "Domain-Specific Keyphrase Extraction", *The Sixteenth International Joint*

*Conference on Artificial Intelligence*Morgan Kaufmann Publishers, San Francisco, CA, pp. 668.

45. Garcia, I. *, The Good in E-mail (or why Email is Still the Most Adopted Collaboration Tool).* Available: http://blog.centraldesktop.com/comments.php?y=06&m=04&entry=entry060403-214628 [2007, 11/22].

46. Gibson, R. (1997). *Rethinking the Future: Rethinking Business, Principles, Competition, Control & Complexity, Leadership, Markets, and the World,* Nicholas Brealey, London.

47. Greenbaum, T.L. (1998). *The Handbook of Focus Group Research,* 2nd ed. edn, Sage, London.

48. Hasan, H. (2004). "Information Systems Development as a Research Method", *Australasian Journal of Information Systems,* vol. 11, no. 1, pp. 4-13.

49. Havens, C. & Knapp, E. (1999). "Easing into Knowledge Management", *Strategy Leadership,* vol. 27, no. 2, pp. 4-10.

50. Hayes-Roth, F., Waterman, D.A. & Lenat, D.B. (1983). "An Overview of Expert Systems", *Building Expert Systems*Addison-Wesley, Reading, MA, pp. 3.

51. Hickens, M. (1999). "Xerox Shares Its Knowledge", *Management Review,* vol. 88, no. 8.

52. Hiltz, S.R. (1985). *Online Communities: A Case Study of the Office of the Future,* Ablex Publishing Corp, Norwood, NJ.

53. Hulth, A. (2003). "Improved Automatic Keyword Extraction Given More Linguistic Knowledge", *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*Association for Computational Linguistics, Morristown, NJ, USA, pp. 216.

54. Humphrey, W.S. (1989). *Managing the Software Process,* Addison-Wesley, Boston, MA, USA.

55. Ibert, J., Baumard, P., Donada, C. & Xuereb, J. (2001). "Data Collection and Managing the Data Source" in *Doing Management Research: A Comprehensive Guide*, 1st edn, Sage Publications, London, pp. 172-195.

56. Jackson, T. & Burgess, A. (2003). "Capturing and Managing Email Knowledge.", *Business Innovation in the Knowledge Economy - Abstracts from the IBM & Stratford-Upon-Avon Conference*, eds. J. Abbot, L. Martin, R. Palmer, M. Stone & L.T. Wright, , pp. 28.

57. Jashapara, A. (2004). *Knowledge Management: An Integrated Approach,* Prentice Hall, Harlow.

58. Jones, S. & Paynter, G. (1999). "Topic-based Browsing within a Digital Library Using Keyphrases", *ACM Conference on Digital Libraries*ACM Press, Berkeley, CA, pp. 114-121.

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

60. Kautz, H., Selman, B. & Shah, M. (1997). "Referral Web: Combining Social Networks and Collaborative Filtering", *Communications of the ACM,* vol. 40, no. 3, pp. 63-65.

61. Kautz, H., Selman, B. & Milewski, A. (1996). "Agent Amplified Communication", *Thirteenth National Conference on Artificial Intelligence*The MIT Press, Menlo Park, CA, pp. 3.

62. Kendall, K.E. & Kendall, J.E. (1988). *Systems Analysis and Design,* 1st edn, Prentice Hall, Upper Saddle River, NJ, USA.

63. Klimt, B. & Yang, Y. (2004). "Introducing the Enron Corpus", *First Conference on Email and Anti-spam*.

64. Kraut, R.E. & Streeter, L.A. (1995). "Coordination in Software Development", *Communications of the ACM,* vol. 38, no. 3, pp. 69-81.

65. Krueger, R. & Casey, M.A. (2000). *Focus Groups: A Practical Guide for Applied Research,* 3rd edn, Sage, Thousand Oaks, CA.

66. Krueger, R.A. (1994). *Focus Groups: A Practical Guide for Applied Research,* 2nd edn, Sage Publicattions, Newbury Park, CA.

67. Krueger, R.A. & Casey, M.A. *, Telephone Focus Groups*. Available: http://www.tc.umn.edu/~rkrueger/focus_tfg.html [2008, 04/23].

68. 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* AAAI Press, Stanford, CA, pp. 110.

69. Lang, J.C. (2001). "Managing in Knowledge-based Competition", *Journal of Organizational Change Management,* vol. 14, no. 6, pp. 539-553.

70. Lang, K.N., Auld, R. & Lang, T. (1982). "The Goals and Methods of Computer Users", *International Journal of Man-Machine Studies,* vol. 17, no. 4, pp. 375-399.

71. Lave, J. & Wenger, E. (1991). *Situated Learning : Legitimate Peripheral Participation,* Cambridge University Press, U.K.

72. Lichtenstein, S. & Swatman, P. (2003). “Email and Knowledge Management”, *Proceedings of 7th Pacific-Asia Conference on Information Systems*, Adelaide, Australia.

73. Lichtenstein, S. (2004). "Knowledge Development and Creation in Email", *The 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 8*IEEE Computer Society, Washington, DC, USA, pp. 245.

74. Lyon, K.L. (2000). "Using Patterns to Capture Tacit Knowledge and Enhance Knowledge Transfer in Virtual Teams" in *Knowledge Management and Virtual Organisations*, Y. Malhotra edn, Idea Group Publishing, Hershey, PA, pp. 124-143.

75. Maltzahn, C. (1995). "Community Help: Discovering Tools and Locating Experts in a Dynamic Environment", *CHI '95: Conference Companion on Human Factors in Computing Systems*ACM, New York, NY, USA, pp. 260.

76. March, J.G. & Simon, H.A. (1958). *Organizations,* Wiley, New York.

77. Markus, M.L. (2001). "Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success", *Journal of Management Information Systems,* vol. 18, no. 1, pp. 57-93.

78. 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.

79. Mattox, D., Maybury, M. & Morey, D. (1999). "Enterprise Expert and Knowledge Discovery", *The 8th International Conference on Human-Computer Interaction*Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA, pp. 303.

80. McDonald, D.W. & Ackerman, M.S. (1998). "Just Talk to Me: A Field Study of Expertise Location", *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*ACM, New York, pp. 315.

81. Mintzberg, H. (1973). *The Nature of Managerial Work,* Harper & Row, New York, NY.

82. Morgan, A.K. & Drury, V.B. (2003). "Legitimising the Subjectivity of Human Reality Through Qualitative Research Method", *The Qualitative Report,* vol. 8, no. 1, pp. 70-80.

83. Morgan, L. (1998). "Practical Strategies for Combining Qualitative and Quantitative Methods: Application for Health Research", *Qualitative Health Research,* vol. 8, no. 3, pp. 362-376.

84. Myers, M.D. (1997). "Qualitative Research in Information Systems", *MIS Quarterly,* vol. 21, no. 2, pp. 241-242.

85. Newell, S., Scarbrough, H., Robertson, M. & Swan, J. (2002). *Managing Knowledge Work,* Palgrave Global Publishing.

86. Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-Creating Company : How Japanese Companies Create the Dynamics of Innovation,* Oxford University Press, New York.

87. Olle, T.W., Hagelstein, J., MacDonald, I.G., Rollands, C., Sol, H.G., Van Assche, F.J.M. & Verrijn-Stuart, A.A. (1991). *Information Systems Methodologies; A Framework for Understanding,* 2nd edn, Addison-Wesley, Boston, MA, USA.

88. Pelz, D.C. & Andrews, F.M. (1966). *Scientists in Organizations: Productive Climates for Research and Development,* Wiley, New York.

89. Pew Internet & American Life Project , *Pew Internet & American Life Project Surveys, March 2000-April 2006*. Available: http://www.pewinternet.org/trends/Internet_Activities_1.11.07.htm [2007, 11/27].

90. 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*, eds. H.S. Nwana & D.T. Ndumu, London, UK, pp. 453.

91. Polanyi, M. (1966). *Tacit Dimension,* Doubleday & Co, Garden City, NY.

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

93. Remenyi, D. & Williams, B. (1995). "Some Aspects of Methodology for Research in Information Systems", *Journal of Information Technology,* vol. 10, no. 3, pp. 191-201.

94. Ryle, G. (1949). *The Concept of Mind,* Hutcheson, London.

95. Schön, D.A. (1993). "Generative Metaphor: A Perspective on Problem-setting in Social Policy" in *Metaphor and Thought*, ed. A. Ortony, 2nd edn, Cambridge University Press, Cambridge, pp. 137-163.

96. 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.

97. 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.

98. Sharp, H., Rogers, Y. & Preece, J. (2007). *Interaction Design: Beyond Human Computer Interaction,* Wiley.

99. Sihn, W. & 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*, pp. 279.

100. Stenmark, D. (2001). "Leveraging Tacit Organisational Knowledge", *Journal of Management Information Systems,* vol. 17, no. 3, pp. 9-24.

101. Stewart, K.A., Baskerville, R., Storey, V.C., Senn, J.A., Raven, A. & Long, C. (2000). "Confronting the Assumptions Underlying the Management of Knowledge: An Agenda for Understanding and Investigating Knowledge Management", *The Data Base for Advances in Information Systems,* vol. 31, no. 4, pp. 41-53.

102. Stewart, T.A. (1997). *Intellectual Capital: The New Wealth of Organizations,* Doubleday, New York, NY, USA.

103. 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.

104. Swaak, J., de Jong, T. & van Joolingen, W.R. (2004). "The Effects of Discovery Learning and Expository Instruction on the Acquisition of Definitional and Intuitive Knowledge", *Journal of COmputer Assisted Learning,* vol. 20, no. 4, pp. 225-234.

105. Tacit Software . Available: http://www.tacit.com/home.asp [2005, 09/16].

106. Tedmori, S., Jackson, T.W., Bouchlaghem, N.M. & Nagaraju, R. (2006). "Expertise Profiling: Is Email Used to Generate, Organise, Share, or Leverage Knowledge", *Proceedings of the 11th International Conference on Computing in Civil and Building Engineering*, ed. H. Rivard, ACSE, , pp. 179.

107. 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.

108. Tiwana, A. (2000). *The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System,* Prentice Hall PTR, Upper Saddle River, NJ, USA.

109. Turney, P. 1997, *Extraction of Keyphrases from Text: Evaluation of Four Algorithms*, National Research Council, Institute for Information Technology.

110. Turney, P.D. (2000). "Learning Algorithms for Keyphrase Extraction", *Information Retrieval,* vol. 2, no. 4, pp. 303-336.

111. Tzoukermann, E., Muresan, S. & Klavans, J.L. (2001). "GIST-IT: Summarizing Email Using Linguistic Knowledge and Machine Learning", *Proceedings of the workshop on Human Language Technology and Knowledge Management*Association for Computational Linguistics, Morristown, NJ, USA.

112. Vivacqua, A.S. (1999). "Agents for Expertise Location", *Spring Symposium Workshop on Intelligent Agents in Cyberspace*AAAI, , pp. 9.

113. Vroom, R.W., Kuiper, C. & Wassink, M. (2004). "Knowledge Search Problems and Strategies used by Design Engineers", *The 4th International Conference on Advanced Engineering Design*, ed. L. Smrcek, Glasgow, UK, pp. 1.

114. Walsh, J.P. (1995). "Managerial and Organizational Cognition: Notes from a Trip down Memory Lane.", *Organizational Science,* vol. 6, no. 3, pp. 280-321.

115. Weick, K.E. (1995). *Sensemaking in Organisations,* Sage, London.

116. Weick, K.E. (1979). *The Social Psychology of Organising,* 2nd edn, McGraw-Hill, New York.

117. Whittaker, S., Bellotti, V. & Moody, P. (2005). "Introduction to this Special Issue on Revisiting and Reinventing E-mail", *Human-Computer Interaction,* vol. 20, no. 1st & 2nd, pp. 1-9.

118. Witten, I.H. & Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations,* Morgan Kaufmann.

119. Worth, D.J. & Greenough, C. 2005, *A Survey of Available Tools for Developing Quality Software Using Fortran 95*, CCLRC, Rutherford Appleton Laboratoy.

120. 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.

121. Yu, B. & Singh, M.P. (2003). "Searching Social Networks", *The 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems*ACM Press, Melbourne, Australia, pp. 65.

122. Zack, M.H. (1999a). "Managing Codified Knowledge", *Sloan Management Review,* vol. 40, no. 4, pp. 45-58.

123. Zack, M. (1999b). "Developing a Knowledge Strategy", *California Management Review,* vol. 41, no. 3, pp. 125-145.

124. 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.

125. Zmud, R.W. (1998). "Conducting and Publishing Practice-driven Research", *IFIP WG8.2 and WG8.6 Joint Working Conference on Information Systems: Current Issues and Future Changes*, pp. 21.

APPENDIX A PAPER 1

Tedmori, S.M., Jackson, T.W., Bouchlaghem, D., Adelmann, H. and Nagaraju, R. (2006), Building a Tool for Expertise Discovery, Proceedings of the *17th Information Resources Management Association International Conference* (IRMA 2006); Washington D.C., USA

BUILDING A TOOL FOR EXPERTISE DISCOVERY

Sara Tedmori[1], Thomas Jackson[2], Dino Bouchlaghem[3], Holger Adelmann[4], Rama Nagaraju[5]

[1] Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, UK, LE11 3TU, Tel: +44 (0) 1509 635649, Fax: +44 (0) 1509 223982, Email: S.M.J.Tedmori@lboro.ac.uk

[2] Department of Computer Science, Loughborough University, Loughborough, Leicestershire, UK, LE11 3TU, Tel: +44 (0) 1509 228231, Fax: +44 (0) 1509 211586, Email: T.W.Jackson@lboro.ac.uk

[3] Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, UK, Tel: +44 (0) 1509 223775, Fax: +44 (0) 1509 223982, Email: N.M.Bouchlaghem@lboro.ac.uk

[4] AstraZeneca R&D Charnwood, Loughborough, Leicestershire, UK, LE11 5RH, Tel: +44 (0) 1509 222681, Fax: +44 (0) 1509 635722, Email: holger.adelmann@astrazeneca.com

5 Eureka Moment Software, 9284 Haden Ln, West Chester OH 45069, USA Email: 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 EXERTISE 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 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, 2004) 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.

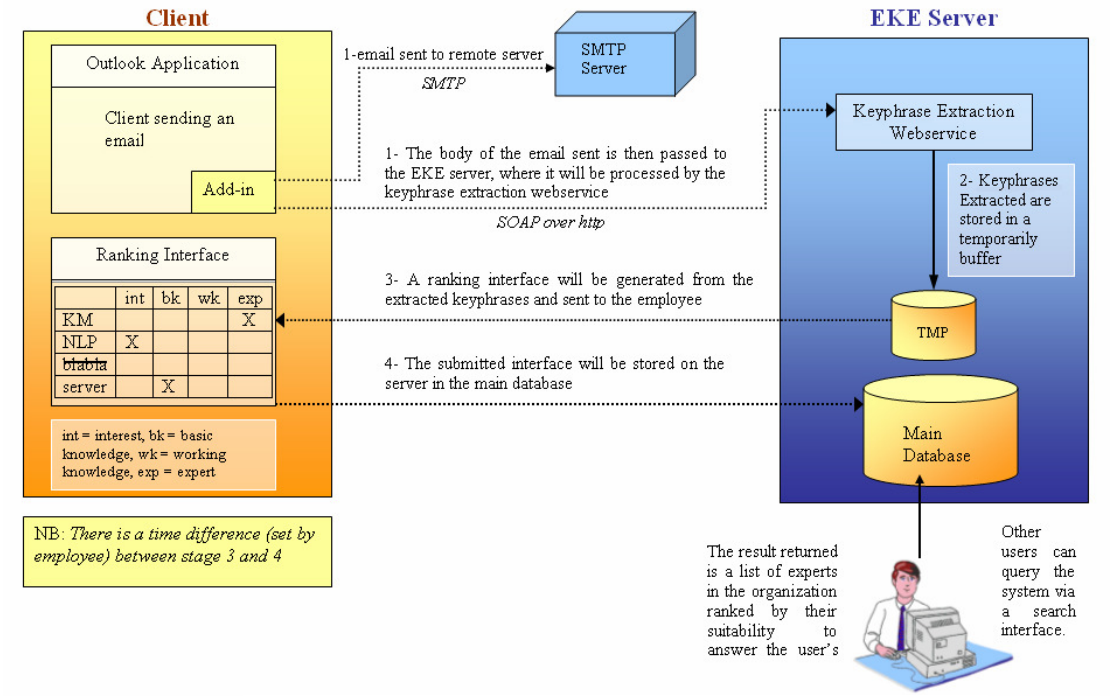


Figure 1. EKE Generic Architecture

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.

6. 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. & Arrango, G. (1991), *Domain Analysis and Software Systems Modeling*, IEEE Computer Society Press, Los Alamitos, CA.

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.

APPENDIX B PAPER 2

Tedmori, S., Jackson, T.W., Bouchlaghem, D., and NewCombe, M. (2007), Information Seeking and Sharing Behaviour of a UK Police Force, Proceeding of the *8th European Conference on Knowledge Management* (ECKM2007), Barcelona, Spain.

Information Seeking and Sharing Behaviour of a UK Police Force

Sara Tedmori[1], Dr. Thomas W. Jackson[2], Chief Inspector Mark Newcombe[3], Prof. Dino Bouchlaghem[1]
[1] Centre for Innovative Construction Engineering, Loughborough University, Loughborough, Leicestershire, UK
[2] Research School of Informatics, Loughborough University, Loughborough, Leicestershire, UK
[3] Leicestershire Police, Leicester, Leicestershire, UK
S.M.J.Tedmori@lboro.ac.uk
T.W.Jackson@lboro.ac.uk
mark.newcombe@leicestershire.pnn.police.uk
N.M.Bouchlaghem@lboro.ac.uk

Abstract: We have reached an age of information overload. It is also an age of information empowerment, an age where people are regularly bombarded with information. People have access to far more information than they can possibly handle. Information plays a vital role in people's lives, as they are constantly challenged to locate the right information that they need in order to make decisions and to complete their tasks. Unfortunately, people often have difficulties in locating relevant information. Early studies on information seeking behaviour show that people searching for information prefer asking other people for advice than searching through a manual. The issue becomes then a matter of searching for the right person. This has led to 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. This study aims to undertake a baseline review of how UK police force employees work and in turn provide a better understanding of how to develop IT systems that will support employees in their daily activities. It documents the analysis of a questionnaire survey that looks at how individuals at Leicestershire Constabulary seek information and how they share information once it's located, in order to determine if an expert locator system would work at their organisation. Results show that officers have difficulties when searching for information. The results give an estimate of the amount of searching time that officers think could be saved by officers if they know where to look for the relevant information and the reasons behind the time being wasted. Moreover, results show that email remains to be the most intensely utilised communication medium, used to help generate, organise, share, or leverage knowledge within the organisation. Although specific websites and online databases were the first sources to be consulted by most officers when searching for information, officers frequently query other peers for references.. The overall results suggest that embracing the concept of an expertise locator at Leicestershire Constabulary could lead to positive outcomes.

Keywords: Information seeking, information sharing, information behaviour, information technology, expertise discovery

1. Introduction

We are living in an information-flooded society, a society characterised by a high level of information intensity. Workers depend on access to information in order to perform their work. They try to obtain the information that they need by using conventional tools of information technology (IT) such as search engines in which information is stored, transmitted and processed. Cutting through the clutter and selecting from the information sources returned in order to get to the right content can be both daunting and time consuming. In order to perform better, workers are increasingly relying on

gaining additional knowledge and experience from other people or sources. Thus, it is essential for them to know where to seek advice and how to access additional expertise from other professionals whilst doing their job. A common way for employees to acquire the needed expertise is to locate internal experts willing to share their specialised knowledge (Ackerman et at 2003). Employees ask others they know for referrals or recommendations, following pointers until the right person is found (Campbell et al 2003). The speed of locating the right expert is important to both the individual worker and the company. When employees gain rapid access to experts, organisational performance can increase (Dooley et al 2002). Therefore, many organisations try to provide employees with help in relation to timely identification and location of expertise as a key part of organisational knowledge management. Expertise locating systems, also known as expertise recommender systems, have emerged for that purpose.

The study presented in this paper, and which was conducted through a questionnaire survey targeting individuals at Leicestershire Constabulary, provides a useful snapshot of the force's views, experience and practical usage in relation to various aspects of information searching and sharing. Moreover, it sheds light on barriers to sharing knowledge, key issues of concern that waste the respondents' time while searching for necessary information, and the respondents views on the authors' latest research.

2. Information Seeking

Information seeking can be described as a reaction to the recognition of an information need (Case 2007). For example, Tom Wilson (2000) has said that information seeking is "the purposive seeking for information as a consequence of a need to satisfy some goal".

Information is increasing, as is the difficulty of finding the required information. Approximately 70% of business professionals responding to a Delphi (2002) survey agree that finding information is difficult. The study also reported that most employees working at large enterprise organisations spend more than two hours a day searching for information necessary to complete their jobs (Delphi, 2002). This calculates to 25% or more of an 8 hour working day. This result is consistent with the results obtained from the IDC study reporting that information workers spend 15% to 35% of their work time just searching for information (Feldman and Sherman 2003). Furthermore, many other studies have reported similar findings. For example, Davenport and Prusak (1997) report that managers spend 17% of their time (6 weeks a year) searching for information.

Outsell's (2001) survey (a survey of US information users carried out in 2001 which surveyed 6,300 people across 20 different industries) found that employees were spending an average of 8 hours a week looking for and using external information content. In 2005, Outsell conducted another study to measure how information users have changed their behaviours over the four years (Outsell, 2005). Comparing the new research with results from 2001, the study found that knowledge workers now spend 11 hours per week searching for information, versus eight in 2001. In addition, the time they spend analysing versus gathering information has changed. In 2005 professionals seemed to spend most of their time (53%) seeking out information. In 2001 however, professionals spent 56% of their time analysing and applying what they had found. The time wasted searching for information results in a significant organisational productivity

cost. The International Data Corp. (IDC) estimates that an enterprise with 1,000 knowledge workers loses a minimum of $6 million a year in time spent searching for and not finding the information needed for knowledge workers to complete their tasks.

Even with the vast array of existing information management tools (such as search engines, document management systems, databases, and the web), knowledge workers are still finding it difficult to locate the needed information to perform their jobs, resulting in an increase in time and money wasted. The concept of an expert locator system emerged as an attempt to contribute towards overcoming these difficulties. Expert locator systems connect people to people rather than people to information. The Email Knowledge Extraction (EKE), a tool the authors are developing, is an example of such a system. It tries to uncover "who knows what" in an organisation by using email content as evidence of expertise.

3. Study Methodology

The data needed for the study was collected through an online questionnaire survey that was completed by individuals at Leicestershire Constabulary during March/April 2007. A questionnaire was used because it enabled the authors to gather anonymous information in a relatively easy manner.

3.1 Questionnaire Structure and Content

The questionnaire consisted of 31 questions divided into 6 sections. A brief introduction about the research, along with instructions on how to complete the questionnaire were presented at the beginning of the questionnaire. No deadline was set for the respondents to complete the questionnaire. Respondents were informed that the completion of the questionnaire was expected to take 10-15 minutes and were assured that all responses would be treated with strictest of confidence. A successful pilot study for the questionnaire was undertaken.

The first part of the questionnaire asked respondents to provide general information about themselves (e.g. gender and number of working years). The second part of the questionnaire asked the respondents about their searching experience. This included questions relating to the number of hours they spend searching for information or advice related to their work, and the amount of time which could be saved if they knew where to look for the information or advice. It also included questions focusing on the respondents' frequency of use of various sources when searching for information, ease of finding information, how often they consult their colleagues for information, and how often they know who to contact when they need information.

The third part seeks to obtain the extent of the respondents' agreement or disagreement with statements relating to their information sharing experience. This included asking the respondents about the most important barrier that hinders them from sharing knowledge. In part four, respondents were asked questions regarding their usage of different types of mediums, with a particular focus on emails.

Questions in part five try to obtain the respondents' views regarding the Email Knowledge Extraction application (EKE) that the authors are developing. The sixth part of the questionnaire allowed the respondents to add any comments they would like to make.

3.2 Questionnaire Sampling, Distribution, and Response

In this study, convenience sampling was used. Such type of sampling may not be optimal for making generalisations (Bryman and Bell 2003), however it was seen appropriate to adopt for the purposes of the study and for accessibility reasons. The study does not aim to make generalisations about employees' information seeking and sharing behaviours. Rather, it aims to obtain better insights of employees' information behaviour and in turn improve understanding of how to develop IT systems that will support employees in their daily activities. With regards to accessibility, contacts developed previously by the authors have facilitated accessing Leicestershire Constabulary to conduct the study.

The questionnaire was sent out to 150 employees working at Leicestershire Constabulary. Two weeks later, a follow up email reminder was sent. Finally, by end of April, 44 responses were received. This presented a response rate of 29%. In relation to the audience, the questionnaire was sent to all Chief Superintendents, Superintendents, Chief Inspectors as well as a police staff head of business.

4. Analysis of the Survey Data

The quantitative data from the survey was analysed using statistical software package SPSS. The analysis carried out was based on obtaining frequencies and conducting non-parametric statistical tests. The first type of analysis (frequencies) was used to obtain percentages of scores in each category. The second type of analysis (non-parametric tests) was used to explore the relationship between two or more categorical variables (particularly this involved conducting chi-square test and fisher's exact test). Conducting the chi-square yielded invalid results due to small sample size. Categories had to therefore be collapsed into two categories in order to provide an appropriate basis for performing the Fisher's exact test (a two-sided probability less than .05 was considered to be statistically significant)**.**

4.1 General Information about the Respondents

Table 1 also shows the general characteristics of the sample that responded; 68.2% of the respondents were male and 31.8% were female, with 84.1% of the respondents have been working at Leicestershire Constabulary for six years or more. This indicates that the respondents involved have generally a significant working experience. Most of the respondents (95.4%) have been working within their current job role for five years or less.

Table 1: General Information

Gender	**Male**			**Female**	
	68.2%			31.8%	
Working Years at Leicestershire Constabulary	**Under 1 year**	**1-5 years**	**6-10 years**	**11-20 years**	**Over 20 years**
	2.3%	13.6%	11.4%	29.5%	43.2%
Working Years within current job role	**Under 1 year**	**1-5 years**	**6-10 years**	**11-20 years**	**Over 20 years**
	31.8%	63.6%	2.3%	2.3%	0%

4.2 Respondents' Searching Experience

Regarding searching experience, results show that 45.5% of respondents spend between one and five hours per week searching for information or advice relevant to their work (See Table 2), a figure lower than IDC's (Feldman and Sherman 2003) and Delphi's (2002) research. Table 2 also shows that 59.1% of the respondents believe that a minimum of 21% of their time could be saved if they knew where to look for this information and that only 18.2% believe that no time is wasted as a direct result of not knowing about information that was available within their organisations.

Table 2: Searching Experience

Number of hours spent per week searching for information/ advice	**1 hour or less**	**1-5 hours**	**6-10 hours**	**11-20 hours**	**over 20 hours**
	18.2%	45.5%	20.5%	11.4%	4.5%
Time saved if you knew where to look for this information/ advice	**0-20%**	**21-40%**	**41-60%**	**61-80%**	**81-100%**
	40.9%	27.3%	13.6%	15.9%	2.3%
Days wasted within the past year as a result of not knowing about available information	**None**	**Up to 5 days**	**6-10 days**	**11-20 days**	**Over 20 days**
	18.2%	52.3%	15.9%	6.8%	6.8%
Frequency of finding the information needed	**Never**	**Rarely**	**Sometimes**	**Frequently**	**Always**
	0%	0%	38.6%	59.1%	2.3%
Frequency of having difficulties identifying resources for new information	**Never**	**Rarely**	**Sometimes**	**Frequently**	**Always**
	4.5%	20.5%	54.5%	20.5%	0%
Frequency of consulting colleagues for information before conducting a search	**Never**	**Rarely**	**Sometimes**	**Frequently**	**Always**
	0%	9.1%	43.2%	47.7%	0%
Frequency of knowing who to contact when you need information	**Never**	**Rarely**	**Sometimes**	**Frequently**	**Always**
	0%	4.5%	54.5%	38.6%	2.3%
Rank of the task of locating information within another department	**Very Easy**	**Easy**	**About Average**	**Difficult**	**Very Difficult**
	0%	11.4%	61.4%	27.3%	0%

Moreover, a significant association was noted between the time the employees report that they spend searching for information relevant to work and the time the employees report that they waste due to not knowing about information available within the company (p=.001, Fisher's exact test). Table 2 also shows the frequency of finding the needed information when searching. The analysis reveals that only a small proportion of respondents (2.3%) always find the information they need while 97.7% of them sometimes/frequently find the information they seek. Moreover, the majority of the respondents (75%), when searching for information that is new to them, sometimes/frequently have difficulties identifying resources for new information. This shows us that almost all officers are unable to always find the information they need to

perform their jobs. Accordingly, a large proportion (90.9%) admitted that they sometimes/frequently consult their colleagues before conducting a search. This concurs with other studies showing that employees prefer asking others for expertise rather than searching documents or electronic knowledge repositories (EKR) (Swaak et al 2004) Moreover, results illustrate that a large percentage of the respondents (93.1%) sometimes/frequently know who to contact when they need information. Analysis of results showed that 88.7% of the respondents do not think that the task of locating information within another department is easy (61.4% think of this task as "about average" and 27.3% think of this task as "difficult").

In order to investigate the time wasted by employees as a direct result of not knowing about information that was available within the company, respondents were asked to provide examples to help better understand the issue. The examples that the respondents provided could be classified into the following categories: Information overload, problems with the search facility, inappropriate titling/labelling, constant change within the force, unpublished material, folder organisation issues, repetition and replication of work, and problems with policies and protocols.

Furthermore, in relation to the sources of information that are difficult to locate, there seemed to be a general consensus among respondents that these include policy and procedures, in addition to persons with expert knowledge. Some respondents stated that it is often difficult to locate persons with expert knowledge of a particular subject.

Table 3: Frequency of use of sources when searching for information

Order of frequency of use when searching for information or advice						
Ranking	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**
Search Engines	15.9%	6.8%	6.8%	70.5%	0%	0%
References from peers	11.4%	43.2%	18.2%	15.9%	2.3%	9.1%
Online databases	20.5%	11.4%	6.8%	11.4%	13.6%	36.4%
Intranet …	15.9%	18.2%	25%	0%	9.1%	31.8%
Mailing lists	13.6%	15.9%	36.4%	2.3%	11.4%	20.5%
Specific Websites	22.7%	4.5%	6.8%	0%	63.6%	2.3%

In relation to the frequency of sources utilised when searching for information (see Table 3), specific websites and online databases were ranked as the most frequently utilised sources by 22.7% and 20.5% of the respondents respectively. It can be noticed that a considerable percentage of the respondents (43.2%) indicated that references from peers was the second most frequently utilised source. Such results could be explained by the respondents searching behaviour. They start searching for the information using the traditional information retrieval tools (e.g. databases). If they fail to locate the required information, they then revert to gaining the required information from their peers.

4.3 Respondents' Information Sharing Attitudes

Table 4 summarises the respondents' information sharing attitudes. It is notable that the majority of respondents (77.3%) widely share the information which they consider

useful or interesting. This result is consistent with the finding which shows that 72.7% of respondents disagreed with the statement, "I only share on a 'need to know basis' or if I am told to do so". This gives an indication of the Leicestershire Constabulary's positive information sharing culture.

A majority of the respondents (90.9%) believe that the knowledge they have is of value to the Force. On the other hand, only half of these (45.5%) believe that the Force fully utilises their knowledge.

Table 4: Information Sharing Experience

I widely share the information which I consider useful/ interesting	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	15.9%	61.4%	13.6%	9.1%	0%
I only share on a 'need to know basis' or if I am told to do so	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	0%	15.9%	11.4%	65.9%	6.8%
The knowledge I have is of value to the Force	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	22.7%	68.2%	9.1%	0%	0%
The Force fully utilises my knowledge	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	2.3%	43.2%	29.5%	25.0%	0%

In order to investigate barriers to sharing knowledge, respondents were asked to select the most important barrier that hinders them from sharing knowledge. They were given six options from which they had to select the most important one. The options they were provided with were:

- 'If I share knowledge I will lose some of my power' (not selected by any of the respondents),
- 'I am insecure about the value of my knowledge' (selected by 9.1% of the respondents),
- 'I lack trust in my colleagues' (selected by 2.3% of the respondents),
- 'I am afraid of negative consequences' (selected by 2.3% of the respondents),
- 'If others do not share, why should I?' (selected by 2.3% of the respondents), and
- 'Other' (selected by 84.1% of the respondents).

If they opt for the 'Other' option, they have to enter in their own words what stops them from sharing knowledge. The majority of the respondents (84.1%) chose the 'Other' option. The comments they provided were grouped into the following categories given their responses: time constraints, sensitivity of position/ information, information overload, lack of systems and mechanisms to sharing knowledge, lack of knowledge among people regarding possessing that expertise, individuals not asked to share, lack of gratitude and feedback, lack of confidence about the value of the knowledge

possessed, unutilised shared information, and lack of knowledge regarding the needs of others.

4.4 Respondent's usage of different types of mediums

It was important to understand how employees use different media to generate, organise, share, or leverage knowledge within the organisation, because it gives an indication on whether email is actually used for those purposes. To address this question, responses to the questionnaire instrument of frequency of use of different mediums were examined (See Table 5).

Table 5: Individual's usage of different types of mediums

Frequency of use of various media to help generate, organise, share, or build upon the knowledge with the organization						
	Daily	**Weekly**	**Monthly**	**6 monthly**	**Yearly**	**Never**
Face-to-face	88.6%	9.1%	0%	0%	2.3%	0%
Telephone	79.5%	20.5%	0%	0%	0%	0%
Email	93.2%	6.8%	0%	0%	0%	0%
Online	2.3%	9.1%	9.1%	4.5%	0%	75%
Memos	11.4%	43.2%	25%	6.8%	0%	13.6%
Intranet	40.9%	13.6%	31.8%	4.5%	4.5%	4.5%
Chatting over lunch/ tea breaks	50%	20.5%	11.4%	2.3%	4.5%	11.4%

Results show that email is the most frequently used medium; it is used by 93.2% of respondents on a daily basis. This is followed by "face-to-face" which is used daily (88.6%) and then the "telephone" which is used daily by 79.5% of the respondents. Such results seem consistent with previous findings by the authors in which results showed that on a daily basis, email was the most frequently used medium, followed by the face-to-face medium and then by the telephone medium (Tedmori et al 2006). Nevertheless, when enquired about the communication medium that one could not do without (Table 6), half of the respondents selected "face-to-face". Email was selected as the second most important medium (favoured by 27.3%). The medium "Telephone" remains the third most favourite medium (ranked by 18.2%).

Table 6: The medium an individual can not do without

The medium that you can not do without	**Face-to-face meetings**	**Telephone**	**Email**	**Online**	**Intranet**	**Chatting over lunch/ tea breaks**
	50%	18.2%	27.3%	0%	2.3%	2.3%

Respondents were asked to indicate what they mostly use email for. The choices were: to ask questions, to answer questions, or both equally. This intended to uncover whether the respondents tend to share knowledge, seek knowledge or use it for both purposes. The findings indicate that 68.2% of employees selected the third option (i.e. both equally), while 6.8% and 25% of employees selected the 'to ask questions' option and

the 'to answer questions' option respectively. Table 7 summarises the number of knowledge related emails sent and received per day.

Table 7: Number of knowledge related emails sent and received per day

Number of knowledge related emails sent a day	**0-5 emails**	**6-10 emails**	**11-20 emails**	**21-40 emails**	**More than 40**
	50%	34.1%	11.4%	2.3%	2.3%
Number of knowledge related emails received a day	**0-5 emails**	**6-10 emails**	**11-20 emails**	**21-40 emails**	**More than 40**
	43.2%	31.8%	13.6%	6.8%	4.5%

4.5 Respondents' views regarding EKE

Results show that more than half of the respondents (52.2%) believe that EKE would benefit their organisation (See Table 8). The promising response was that if their organisation was using EKE, 88.7% are willing to make their interest/knowledge/expert areas public. 9.1% expressed neutral views and only 2.3% indicated that they were not willing to publicise their knowledge areas. Table 8 also shows the number of enquires that respondents are willing to deal with per week. These enquires that could possibly be generated by EKE are not part of an employee's normal workload. They are additional enquiries that the employee is willing to reply to. The result could indicate that employees realise the importance of sharing knowledge to improve their working day by helping their colleagues. The results reported in table 8 indicate that the officers realise the benefits of the system and demonstrate the potential for it to be used.

Table 8: Respondents views regarding EKE

The force would benefit from using this software	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	4.5%	47.7%	34.1%	9.1%	4.5%
I would be willing to make my interest/ knowledge/ expert areas public	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	20.5%	68.2%	9.1%	2.3%	0%
Number of enquires you would be willing to reply to per week	**None**	**1-3**	**4-7**	**8-10**	**More than 10**
	0%	15.9%	34.1%	9.1%	40.9%

Table 9 shows the mediums the respondents prefer to be contacted through by their colleagues. Comparing these results with the results shown in table 5, it can be noted that email, the most frequently used medium, is not the respondents' preferable way of being contacted by co-workers for help. Results show that officers at Leicestershire Constabulary prefer "face-to-face" communication. Telephone and email were ranked as the second and third most favorite mediums respectively. This is also consistent with previous research reported in Tedmori et al (2006).

Table 9: Preferable way of being contacted

Preferable way of being contacted by co-workers for help						
	1st	**2nd**	**3rd**	**4th**	**5th**	**6th**
Face-to-face	47.7%	27.3%	25%	0%	0%	0%
Telephone	20.5%	50.0%	20.5%	0%	4.5%	4.5%
Email	20.5%	13.6%	40.9%	0%	18.2%	2.3%
Online	6.8%	4.5%	11.4%	2.3%	43.2%	31.8%
Memos	0%	4.5%	2.3%	11.4%	29.5%	52.3%
Intranet	0%	0%	0%	86.4%	4.5%	9.1%

5. Concluding Discussion

This paper presents a snapshot of a UK Police Force's information seeking and sharing behaviour. Results in Table 2 show that 59.1% of the respondents believe that a minimum of 21% of their time could be saved if they knew where to look for necessary information. Analysis of results show that the amount of time wasted can be attributed to information overload, problems with the search facility, inappropriate titling/ labelling, constant change within the force, lack of publication, folder organisation issues, repetition and replication of work, and problems with policies and protocols. Furthermore, there seemed to be a general agreement among respondents that the sources of information that were difficult to locate included policy and procedures, in addition to persons with expert knowledge.

As shown in Table 3, specific websites and online databases were ranked as the most frequently utilised sources when searching for information. This result may be an indication of how EKE might help employees search for the right expert. Results presented in Table 2 show that when searching for information, the minority of respondents (2.3%) always find the information they need, while the majority of them (75%), sometimes/ frequently have difficulties identifying resources for new information. This could explain why a large proportion, 90.9% sometimes/ frequently consult their colleagues before conducting a search.

Fisher's exact revealed a significant relationship between the frequency of knowing who to contact when you need information and the number of years working at Leicestershire Constabulary (P = .022). The higher the number of years an employee has worked at the constabulary, the higher the probability of knowing who to contact. The people who found it difficult to locate expert knowledge are mostly likely to be new to the organisation.

This was reinforced by one of the participants who said, *"As a new person to the Force I believe that this system would be very useful. I often am unaware of who I need to ask in relation to specific topics, having this would aid this and also cut out time asking the wrong person."*

In summary, the survey results showed the amount of searching time that could be saved by officers if they know where to look for the relevant information and the reasons behind the time being wasted. The overall results suggest that embracing the

concept of an expertise locator at Leicestershire Constabulary could lead to positive outcomes as the majority of officers realise the benefits of the system and are willing to use it. This finding will be further explored by the deployment of EKE at Leicestershire Constabulary.

6. Acknowledgement
The authors wish to thank all individuals at Leicestershire Constabulary in the UK for their co-operation.

References

Ackerman, M., Pipek, V. and Wulf, V. (Eds.) (2003) *Sharing Expertise: Beyond Knowledge* Management, Cambridge, MA: The MIT Press.

Bryman, A. and Bell, E. (2003) *Business Research Methods* Oxford University Press

Campbell, C.S., Maglio, P.P., Cozzi, A. and Dom, B. (2003) "Expertise Identification Using Email Communications", in *Proceedings of CIKM*, Louisiana.

Case, D.O. (2007) *Looking for information: a survey of research on information seeking, needs, and behavior.* 2nd ed. Amsterdam: Academic Press.

Davenport, T.H. and Prusak, L. (1997) Information Ecology Mastering the Information and Knowledge Environment, Oxford University Press, New York, NY.

Delphi Group. (2002) "Taxonomy and Content Classification", Market Mileston Report.

Dooley, K.J., Corman, S.R. and McPhee, R.D. (2002) "A knowledge directory for identifying experts and areas of expertise", *Human Systems Management*, Vol. 21, No. 4, pp. 217-228.

Feldman, S. and Sherman, C. (2003) "The high cost of not finding information", IDC Report #29127.

Outsell (2001), "Outsell, Inc. Releases Findings Analyzing Content Habits, Preferences and Budgets of More than 6,300 Information End Users", press release, available at: http://findarticles.com/p/articles/mi_m0EIN/is_2001_May_3/ai_74003361 (accessed 16 July 2007).

Outsell (2005), "HotTopics: 2001 vs. 2005: Research Study Reveals Dramatic Changes Among Information Consumers", available at:
http://www.outsellinc.com/store/products/219

Swaak, J., de Jong, T., and van Joolingen, W. R. (2004) "The effects of discovery learning and expository instruction on the acquisition of definitional and intuitive knowledge", *Journal of Computer Assisted Learning*, Vol. 20, No. 4, pp. 225- 234.

Tedmori, S.M., Jackson, T.W., Bouchlaghem, N.M., Adelmann, H. and Nagaraju, R. (2006) "Expertise Profiling: Is Email Used to Generate, Organise, Share or Leverage Knowledge", Proceedings of the *11th Int'l Conf. on Computing in Civil and Building Engineering (ICCCBE-XI)*, Montreal, Canada.

Wilson, T.D. (2000) Human information behaviour, *Informing Science*, Vol. 3 No. 2, pp. 49-55.

APPENDIX C PAPER 3

Tedmori S, Jackson TW and Bouchlaghem D (2006). Locating knowledge sources through keyphrase extraction. *Knowledge and Process Management,* 13, pp. 100-107.

Locating Knowledge Sources through Keyphrase Extraction

Sara Tedmori[1], Thomas W. Jackson[2] and Dino Bouchlaghem[3]

[1]Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, UK
[2] Research School of Informatics, Loughborough University, UK

[3]Department of Civil and Building Engineering, Loughborough University, UK

Abstract:

There are a large number of tasks for which keyphrases can be useful. Manually identifying keyphrases can be a tedious and time consuming process that requires expertise, but if automated could save time and aid in creating metadata that could be used to locate knowledge sources. In this paper, the authors present an automated process for keyphrase extraction from e-mail messages. The process enables users to find other people who might hold the knowledge they require from information communicated via the e-mail system. The effectiveness of the extraction system is tested and compared against other extraction systems and the overall value of extracting information from e-mail explored.

1. INTRODUCTION

Keywords and keyphrases are useful for a variety of purposes (throughout this paper, the authors use the latter to subsume the former). They can be used to summarise, index, label, categorise, cluster, highlight, browse, and search information (Turney, 2003). They can be used in many text-mining and knowledge management related applications. The great majority of documents come without keyphrases, and manually assigning keyphrases is a tedious process that requires knowledge of the subject matter (Witten et al., 1999). Numerous techniques have been proposed to automatically extract keypharses from documents. However, these techniques mainly focus on extracting keyphrases from journal articles. Many other types of documents would also benefit from having keyphrases, including web pages, e-mail messages, news reports, magazine articles, and business papers (Turney, 2003).

A relatively new area of research is trying to extract keyphrases from e-mail messages to aid in determining who knows what within an organisation (Jackson and Tedmori, 2004). The keyphrases that are extracted should give some sort of indication of skills and experience exchanged in e-mails. Such keyphrases ought to disclose skills such as technical expertise, management skills, industry knowledge, education and training, work experience, professional background, knowledge in subject areas, etcetera. However, extracting keyphrases that describe the individual's expertise from an e-mail body poses an immense challenge. E-mails are freestyle text, not always syntactically well formed, domain independent, of variable length, and on multiple topics (Tzoukermann et al. 2001). Commercial systems for expert identification using e-mails include: *Tacit's ActiveNet* (Tacit, 2005), *AskMe Enterprise* (Ask Me, 2005) and *Corporate Smarts' Intelligent Directory* (Corporate Smarts, 2006). Figure 1 shows how such systems can be used to analyse e-mails to identify individuals or groups that have specific expertise. When an e-mail is sent (step 1 in Figure 1), keyphrases are extracted

(step 2 in Figure 1). The extracted keyphrases are then sent back to the user (step 3 in Figure 1) and placed into an expertise profile that the user can edit (step 4 in Figure 1). The expertise profile contains information about 'who knows what' within the organisation. This information is then distilled into a searchable database (step 5 in Figure 1) which users can query to find relevant people. Not all systems perform steps 3 and 4 and in this particular case these steps are specific to the system the authors are developing. Users are provided with an interface to rank their knowledge in the extracted keyphrases. With regards to similar extraction systems and how they work, most of the system information is only available in the form of white papers serving as a marketing tool to promote an organisations product and point of view which potentially could be biased.

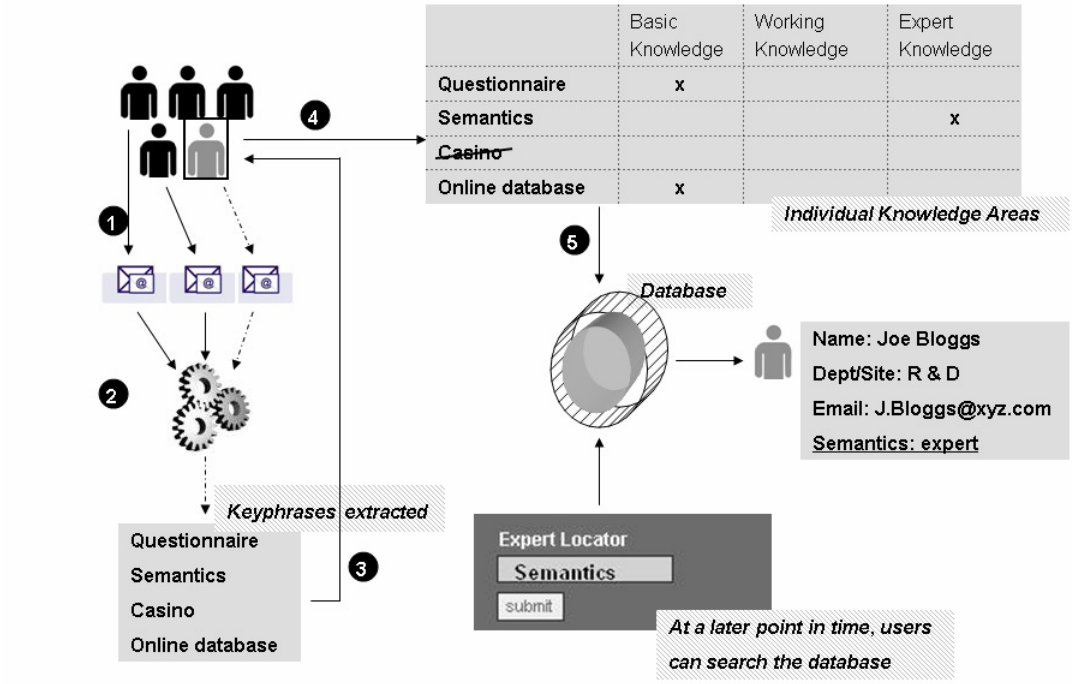


Figure 1 System overview

In this paper, the authors review current keyphrase extraction techniques and present an automatic e-mail message keyphrase extractor that will extract keyphrases and convey them to the user, by combining machine learning techniques and linguistics. The paper evaluates the proposed technique and concludes with a discussion of the proposed technique, including suggestions for future research.

2. EXTRACTION TECHNIQUES

Numerous papers explore the task of producing a document summary by extracting key sentences from the document (Brandow et al., 1995; Edmundson, 1969; Jang and Myaeng, 1997; Johnson et al., 1993; Kupiec et al., 1995; Luhn, 1958; Marsh et al., 1984; Paice, 1990; Paice and Jones, 1993; Salton et al., 1994; Tzoukermann, 2001). While similar to keyphrase extraction, Turney (1999) argues that document summarisation is more difficult than keyphrase extraction. The end result is a set of

sentences that often lack cohesion because anaphoric references are not resolved (Johnson et al., 1993; Brandow et al., 1995). Anaphors are pronouns (e.g. 'it','they'), definite noun phrases (e.g. 'the car'), and demonstratives (e.g. 'this', 'these') that refer to previously discussed concepts. Turney (1999) continues by saying that it may be impossible or very difficult for the reader of the summary to determine the referents of the anaphors. Johnson et al. (1993) tried to automatically resolve anaphors, however this resulted in overly long summaries. The problem of resolving anaphors does not arise in keyphrase extraction tasks, because anaphors are not keyphrases. Moreover, unlike a list of sentences, a list of keyphrases has no structure; a list of keyphrases can be randomly permuted without significant consequences. (Turney, 1999). There have been a number of techniques proposed for extracting keyphrases from text (Barker and Cornacchia, 2000; Frank et al., 1999; Krulwich and Burkey, 1996; Turney, 1999). Some of these techniques are domain specific while others are domain independent. Domain dependent techniques use machine learning, and require a collection of documents with keyphrases already attached for training purposes. Furthermore, these techniques (both domain dependent and domain independent) typically have some kind of connection to linguistics and/or use pure statistical methods. A number of applications have been developed using these techniques and their merits and pitfalls are discussed in the following paragraphs in order to determine the most effective way of extracting keyphrases from e-mail.

Peter Turney (1999) devised *GenEx*, a hybrid genetic algorithm for keyphrase extraction. GenEx has two components: *Genitor,* a genetic algorithm, and *Extractor*, a keyphrase extraction algorithm. After stopword removal, candidate keyphrases (unigrams, bigrams, and trigrams) from the input document are scored based on a number of parameters. These parameters include frequency of the stemmed words in the phrase, length of the phrase, position of the phrases, etcetera. To maximise the performance on the training data, the *Genitor* genetic algorithm tunes the parameters of *Extractor*. *Genitor* is no longer needed after the training process. When the optimal set of parameters are found, *Extractor* can extract the best set of keyphrases, that is the one that has the most matches to the known keyphrase set in the training document set (Turney, 1999).

KEA has been developed by Frank et al. (1999). *KEA* is based on the naïve Bayes machine learning method. *KEA* uses a simpler set of features than Turney's *GenEx* algorithm. The two feature values that *KEA* calculates for each candidate keyphrase are the *TFxIDF*, a measure of a phrases frequency in a document compared to its rarity in general use; and first occurrence, which is the distance into the document of the phases first appearance. The machine-learning scheme first builds a prediction model using training documents where the author's keyphrases are known, and then uses the model to find keyphrases in new documents. *KEA* chooses candidate keyphrases using lexical methods, calculates feature values for each candidate, and uses machine learning to predict which candidates are good keyphrases. The length of candidate phrases is limited to three. Frank et al. (1999) evaluate the *KEA* algorithm in relation to *GenEx* algorithm. The experiments show that *KEA*'s performance is statistically equivalent to *GenEx*. Initially, *KEA* was used by the authors (Jackson and Tedmori, 2004) to extract keyphrases from electronic messages. However, after testing the system it was apparent that the keywords extracted by *KEA* were inappropriate for the task of extracting keyphrases from e-mail messages. As a result, GenEx was deemed inappropriate for the task at hand and alternatives had to be explored.

Peter Turney (2003) argues that a limitation of previous automatic keyword extraction algorithms is that the output keyphrases are at times incoherent. That is, that the majority of the extracted keywords may fit well together however, there will be a minority of outliers with no semantic relation to the majority or each other and he continues to argue that discarding this minority might improve the quality of the machine-extracted keyphrases**.** He suggests a different approach which is to use the degree of statistical association among candidate keyphrases as evidence that they may be semantically related, and thus avoiding them tends to improve the quality of the extracted keywords. These coherence features are not domain specific, and his experiments show that their use improves the quality of extracted keywords even when the testing domain is different than the training domain (Turney, 2003). Hulth (2003) pinpoints two common drawbacks *of GenEx* and *KEA* algorithms. The first drawback is that the number of words in a keyphrase is limited to three knowing that in the training data 9.1% of the manually assigned keywords consist of four or more words. The second drawback is that the user must state how many keywords to extract from each document (Hulth, 2003).

Common to these systems is the approach of extracting keyphrases from text as a supervised learning task (Turney, 1999; Frank et al., 1999). These systems require a separate training document set with keyphrases already assigned in order to function properly. E-mail messages with pre-assigned keyphrases, to be used as a training set, are difficult to obtain. Moreover, these systems are intended for larger electronically stored documents such as journal articles, novels, and newspaper articles and not for e-mails which are considerably shorter.

A common approach to extracting keyphrases when no machine learning is involved is by means of parts-of-speech (POS) patterns. Barker and Cornacchia (2000) describe a system where noun phrases are chosen from a document as keyphrases. The system first skims the input document for base noun phrases (non-recursive structure consisting of a head noun and zero or more premodifying adjectives and/or nouns), then it uses the length of the phrase, the frequency of its use and the frequency of its head noun to assign scores to noun phrases, and finally it filters some noise from the set of top scoring keyphrases. Barker and Cornacchia (2000) reported that there was no change in the performance of the system in comparison to the trained *Extractor* system in experiments involving human judges (Barker and Cornacchia, 2000).

Hulth (2003) reports that keyword extraction from abstracts can be achieved by using simple statistical measures as well as syntactic information from the documents as input to the machine-learning algorithm. His experimental results show that extracting noun phrase chunks gives better precision than n-grams (sequence of 1…N words), and by adding POS tag(s) assigned to the term so that all words or sequences of words matching any of a set of POS are extracted and a dramatic improvement is achieved. By using phrases, the length of the potential words is not restricted, rather potential terms are treated as units. When inspecting manually assigned keywords, the vast majority turn out to be nouns or noun phrases (Hulth, 2003).

Tzoukermann et al. (2001) present *GIST-IT*, a system for automatic extraction of salient information from e-mail messages, for the purpose of providing an informative, generic, ‘at-a-glance’ summary. *GIST-IT* follows a process similar to *KEA* in that a set of candidate noun phrases are built up and assigned features that are then used to decide

on the keyphrases. *GIST-IT* offers two significant improvements on *KEA*. Firstly, *GIST-IT* is intended for single e-mail messages, and the training for feature selection takes place largely on an e-mail corpus. This implies that *GIST-IT* is much more specific to e-mail keyphrase extraction than *KEA*. *GIST-IT* uses some linguistic filtering which include: removing unimportant modifiers (i.e. most, more, etc), removing common words, and removing empty nouns (i.e. lot, group, set).

Common to these systems (Tzoukermann et al., 2001; Hulth, 2003) is the extraction of noun phrases from text. However, the downside is that in spite of the filtering employed, many of the extracted keyphrases are common words that are likely to occur in numerous e-mails in a whole range of contexts. Therefore, it is important to distinguish between more general nouns and those that are more likely to form keyphrases. In the following section, the authors present an approach for keyphrase identification from e-mail text, which is purely based on the grammatical POS tags that surround these phrases.

3. KEYPHRASE EXTRACTION FROM E-MAIL MESSAGES

This section describes a keyphrase extraction algorithm for e-mail text. The algorithm is fully implemented and embedded in E-mail Knowledge Extraction (*EKE)*, an agent developed by the authors that enables its users to find other people who hold the required knowledge of a specific domain. The extraction algorithm has two stages. The first stage involves training in which a model for POS tagging is created. The second stage involves extraction in which keyphrases are extracted from e-mail messages using the created model. Figure 2 shows the basic overview of the extraction stage. The input to the system is a single e-mail message.

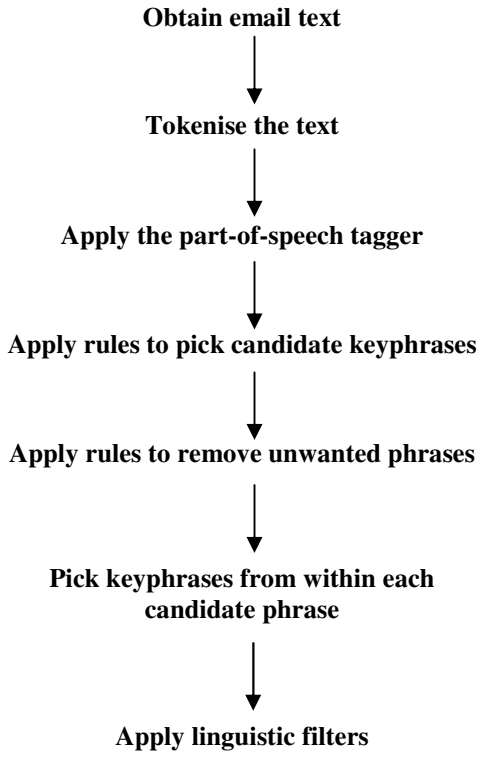


Figure 2 Stages of the extraction process

After the e-mail text is obtained, the text is split into tokens using regular expression rules. In order to discover patterns in text, the next step is to tag the words in the e-mail message by their parts of speech. The Brill rule-based tagger is used to assign the most likely single part of speech tag (noun, verb, adjective, etc.) to each word in the e-mail. Brill tagging is a type of transformational-based learning. It is a supervised learning method since it needs annotated training data. It compiles a list of transformational correction rules. This tagger works by automatically recognising and remedying its weakness, thereby incrementally increasing its performance. The Brown corpus was used as the annotated training document set. The Brown corpus consists of 500 texts, each consisting of just over 2,000 words. In total, it contains 1,014,312 words sampled from 15 text categories. The result of the supervised learning is a prediction model that will be used to tag new text.

Following POS text tagging, rules are applied to select candidate keyphrases by grouping all occurrences of specific sequences of tags together. A rule is a sequence of grammatical tags that is most likely to contain words that make up a keyphrase. These rules were manually set by the authors by manually identifying keyphrases from an e-mail sample consisting of 50 e-mails and looking at the grammatical properties that surround these phrases. After the sequences of tags are grouped together, rules are applied to remove a subset of phrases that are not relevant. Keyphrases are then selected from the identified candidate keyphrases. Finally, the system uses linguistic filtering to extract more important keyphrases. The result is a set of lines, each a sequence of tokens containing at least one letter. Table 1 shows a working example of an e-mail sent through the keyphrase extraction system based on the stages of the extraction process shown in Figure 2. The primary advantage of this technique is that it is domain and genre independent.

Table 1 A Working Example

A working example of an e-mail sent through the keyphrase extraction system
>>> Obtain e-mail text Hi Dany, I've had some experience with online surveys. I usually use html to design the survey and php to process the html and store the results in a database. I know there are alternative languages that you can use, but php is easy to learn and you can find a lot of material on the web. I recommend you start with designing your survey in html! Sara
>>> Tokenise the text <hi>, <dany>, <,>, <i've>, <had>, <some>, <experience>, <with>, <online>, <surveys>, <.>, <i>, <usually> and so on….
>>> Apply POS Tagger <hi/NN>, <dany/NN>, <,/,>, <i've/NN>, <had/hvd>, <some/dti>, <experience/nn>, <with/in>, <online/NN>, <surveys/nns>, <./.>, <i/nn>, <usually/rb> and so on….
>>> Apply rules to pick candidate keyphrases and to remove unwanted phrases (S: <hi/NN> <dany/NN> <,/,> <i've/NN> <had/hvd> <some/dti> <experience/nn> <with/in> (Key phrase: <online/NN> <surveys/nns>) <./.> <i/nn> <usually/rb> and so on….
>>> Pick Keyphrases from within each candidate phrase <online/NN><surveys/nns>, <php/NN>, <html/NN>, <database/NN>, <php/NN>, <html/NN> and so on….
>>>Apply linguistic filters <online/NN><surveys/nns>, <html/NN>, <database/NN>, <php/NN> and so on….
For the complete set of tags used in the Brown corpus please refer to http://www.comp.leeds.ac.uk/amalgam/tagsets/brown.html

4. EVALUATION AND RESULTS:

In this section, the authors firstly describe the test corpus used to measure the performance of the keyphrase extraction process. The authors then describe the evaluation criteria that will be used to measure the performance of the keyphrase extraction application.

4.1 Test corpus

The experiments in this report are based on three e-mail collections. The authors refer to the e-mail collection as the *sample* and to each individual e-mail as the *sampling unit*. For each *sampling unit*, there is a target set of keyphrases that have been generated by hand.

Table 2 below details the three corpuses used. The *sampling units* are collected from subjects from different backgrounds (people with English as their first language and people who can communicate in English, but is not their first language). All subjects belong to the age group 24-60.

Corpus Name	**Description**	**Size**
Corpus 1	E-mails form various academic domains	45
Corpus 2	Employee E From Company XYZ	19
Corpus 3	Enron	50

Table 2 Details of the 3 e-mail collections

All the *sampling units* were outgoing mail. The authors believe that *sampling units* are representative of typical messages that are sent out in institutional and corporate environments. The *sampling units* of the *sample*, Corpus 1, were collated from various academic disciplines (computer science, information science, building and construction engineering). The *sampling units* of the second *sample,* Corpus 2, are specific to one employee from a large supplier of total office solutions in the UK & Ireland, which for confidentiality reasons in is referred to as Employee E from Company XYZ. The *sampling units* of the final sample, Corpus 3, are collated from the Enron e-mail dataset, which is freely available on the net.

4.2 Evaluation Approach

There are two basic approaches to evaluating automatically generated keyphrases (Jones and Paynter, 1999). The first adopts the standard Information Retrieval metrics of precision and recall to reflect how well generated phrases match phrases, which are considered to be 'relevant'. Author phrases are usually used as the set of relevant phrases, or the 'Gold Standard'. Author phrases stand for the list of phrases usually found at the beginning of many articles such as academic journals.

The second approach is to gather subjective keyphrase assessments from human readers. Previous studies involving human phrase assessment (Barker and Cornacchia,

2000; Chen, 1999; Turney, 2000) follow essentially the same methodology. Subjects are provided with a document and a phrase list and asked to assess in some way the relevance of the individual phrases (or of sets of phrases) to the given document. A study by Jones and Paynter (1999), shows that authors do provide good quality keyphrases and thus can be used as the 'Gold Standard' against which other keyphrases can be compared.

The work in this paper adopts the first approach. However, the authors of the e-mails need to highlight the phrases that they think are relevant. The authors of the e-mails need to highlight keyphrases that appear in the body of the e-mail text. Keyphrases consisting of more than one word should be in the same order as in the body of the e-mail text. At occasions, when the authors of e-mails were not accessible, the authors of this paper had to manually assign keyphrases to the e-mails. These keypharses were then used as the 'Gold Standard'.

The task is to take an e-mail message as input and automatically generate a list (containing no duplicate keyphrases) of keyphrases as output. The output keyphrases always appear somewhere in the body of the input e-mail document. The performance measure is based on the number of matches between the machine-generated phrases and the human generated phrases. In the following subsections, the authors will define what matching keyphrases means and how the performance measure is calculated from the number of matches.

4.2.1 Criteria for Matching Phrases

A match occurs, if for example an author suggests the keyphrase 'wordnet relation' and a keyphrase generation algorithm suggests the keyphrase 'wordnet relations'. Yet, if the author suggests 'wordnet relation' and the algorithm suggests 'relation', this is not counted as a match, since there are many different kinds of 'relations'. However, if the authors suggest 'wordnet' and the algorithm suggests 'wordnet relations', this is counted as a match because the algorithm is specifying the term. To summarise, a human selected keyphrase matches a machine-generated keyphrase when they either correspond to the same sequence of stems or when the machine generated keyphrase makes the human selected phrase more specific.

4.2.2 The Performance Measure

Researchers in information retrieval commonly use *precision* and *recall* to evaluate the performance of the returned results (e.g. search results returned). In the keyphrase extraction context, *precision* is the estimate of the probability that if a given system outputs a phrase as a keyphrase, then it is really a keyphrase. *Recall* is an estimate of the probability that, if a phrase is a keyphrase, then a given system will output it as a keypharse. However, there is a well-known trade-off between *precision* and *recall*. One can be optimised at the expense of the other (Turney, 1999). For example, if it is guessed that all phrases are keyphrases, then recall is 100%, but precision will be close to 0%. On the other hand, if one relevant keyphrase is guessed as the only keyphrase then precision might be 100%, but recall would be close to 0%. What is required is a performance measure that yields a high score only when precision and recall are balanced. A measure that is widely used in information retrieval is the f-measure, defined as

$$f-measure = \frac{2 \times precision \times recall}{precision + recall} \qquad \text{(Formula 1)}$$

4.3 Results

In Table 3, precision, recall, and the f-measure results are shown. The highest precision (59.6), recall (63.1), and f-measure (61.3) were achieved on the smallest sample (19 messages). Since only three sets were evaluated, one cannot determine the coloration between size of the sample and performance of the extractor.

Table 3 Shows the precision, recall and f-measure values for each of the collections.

Corpus Name	Precision	Recall	f-measure
Corpus 1	53.3	57.6	55.4
Corpus 2	**59.6**	**63.1**	**61.3**
Corpus 3	41.7	48.3	44.8

Turney (1997) evaluates four keyphrase extraction algorithms using 311 e-mail messages collected from 6 employees, and in which 75% of each employee's messages was used for training and 25% (approximately 78 messages) was used for testing. His evaluation approach is similar to the authors of this paper and the highest f-measure reported was that of the NRC, the extractor component of GenEx, which uses supervised learning from examples. The f-measure reported is 22.5, which is, as expected, significantly less than the f-measures shown in Table 3. Moreover, Hulth (2003) reports results from three different term selection approaches. The highest f-measure reported was 33.9 from the n-gram approach with POS tags assigned to the terms as features. All unigrams, bigrams, and trigrams were extracted, after which a stop list was used where all terms beginning or ending with a stopword were removed. Again, the result reported is less than the authors lowest f-measure (44.76). The system Hulth (2003) reports, limits itself by limiting the number of tokens in the keyphrases to three.

5. CONCLUSION

In this paper, the authors presented a process for keyphrase extraction from e-mail messages. The method uses machine learning to tag new text by its part of speech, then extracts keyphrases purely based on POS tags that surround these phrases. The system was evaluated using three samples. The highest f-measure obtained was 61.3. If comparing with other reported performance measurements from other algorithms, the f-measure obtained by the authors is higher. The f-measure results detailed in this paper are higher than previously reported findings and the keyphrases extracted have provided an effective means of determining who knows what within an organisation. However, the efficiency of the system still requires refining as the end user still has to delete a large number of irrelevant keyphrases (noise) that do not depict their expertise. Therefore, future research should be conducted into exploring ways to improve the process detailed in this paper in order to obtain higher performance measurements.

6. REFERENCES

AskMe. 2005. http://www.askmecorp.com [16 December 2005].

Barker K, Cornacchia N. 2000. Using noun phrase heads to extract document keyphrases. In *Proceedings of the Thirteenth Canadian Conference on Artificial Intelligence* (LNAI 1822), Montreal, Canada, 40-52.

Brandow R, Mitze K, Rau LR. 1995. The automatic condensation of electronic publications by sentence selection. *Information Processing and Management*, **31** (5): 675-685.

Chen KH. 1999. Automatic Identification of Subjects for Textual Documents in Digital Libraries. *Los Alamos National Laboratory*, Los Alamos, NM, USA.

Corporate smarts. 2006. http://www.corporatesmarts.com [22 January 2006]

Edmundson HP. 1969. New methods in automatic extracting. *Journal of the Association for Computing Machinery*. **16** (2): 264-285.

Frank E, Paynter GW, Witten IH, Gutwin C, and Nevill-Manning CG. 1999. Domain-specific keyphrase extraction, In *Proceedings of Sixteenth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers, San Francisco, CA, pp. 668-673
http://www.nzdl.org/Kea/Frank-et-al-1999-IJCAI.pdf

Hulth A. 2003. Improved Automatic Keyword Extraction Given More Linguistic Knowledge. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'03)*. Sapporo.

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

Jang DH, Myaeng SH. 1997. Development of a document summarization system for effective information services. *RIAO 97 Conference Proceedings: Computer-Assisted Information Searching on Internet*; 101-111. Montreal, Canada.

Johnson FC, Paice CD, Black WJ, Neal AP. 1993. The application of linguistic processing to automatic abstract generation. *Journal of Document and Text Management* **1**, 215-241.

Jones S, Paynter G. 1999. Topic-based browsing within a digital library using keyphrases. *Proceedings of the fourth ACM conference on Digital libraries*; 114-121, Berkeley, California, United States

Krulwich B, Burkey C. 1996. Learning user information interests through the extraction of semantically significant phrases. *In AAAI 1996 Spring Symposium on Machine Learning in Information Access.*

Kupiec J, Pedersen J, Chen F. 1995. A trainable document summarizer. *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 68-73, New York: ACM.

Luhn HP. 1958. The automatic creation of literature abstracts. *I.B.M. Journal of Research and Development*, **2** (2), 159-165.

Marsh E, Hamburger H, Grishman R. 1984. A production rule system for message summarization. *In AAAI-84, Proceedings of the American Association for Artificial Intelligence*, pp. 243-246. AAAI Press/MIT Press: Cambridge, MA.

Paice CD. 1990. Constructing literature abstracts by computer: Techniques and prospects. *Information Processing and Management*, **26** (1): 171-186.

Paice CD, Jones PA. 1993. The identification of important concepts in highly structured technical papers. *SIGIR-93: Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 69-78, ACM: New York

Salton G, Allan J, Buckley C, Singhal A. 1994. Automatic analysis, theme generation, and summarization of machine-readable texts. *Science*, **264**: 1421-1426.

Tacit. 2005. http://www.tacit.com [10 December 2005].

Turney PD. 1997. Extraction of Keyphrases from Text: Evaluation of Four Algorithms, National Research Council, Institute for Information Technology, *Technical Report ERB-1051*. (NRC #41550)

Turney P. 1999. Learning to Extract Keyphrases from Text. *Technical Report ERB-1057*, Institute for Information Technology, National Research Council of Canada.

Turney PD. 2000. Learning Algorithms for Keyphrase Extraction. *Information Retrieval* **2**(4); 303-336.

Turney P. 2003. Coherent Keyphrase Extraction via Web Mining. *In Proceedings Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, 434-439, Acapulco, Mexico

Tzoukermann E, Muresan S, Klavans JL. 2001. GIST-IT: Summarizing email using linguistic knowledge and machine learning. In *Proceeding of the HLT and KM Workshop*, EACL/ACL.

Witten IH, Paynter GW, Frank E, Gutwin C, Nevill-Manning CG. 1999. KEA: Practical automatic keyphrase extraction. In *Proceedings of DL '99*, pp. 254-256. (Poster presentation.) http://www.nzdl.org/Kea/Frank-et-al-1999-IJCAI.pdf

APPENDIX D PAPER 4

Tedmori, S., Lichtenstein, S., and Jackson, T.W. (2007), Expertise Location Using Keyphrases in Electronic Mail: Socio-ethical Challenges, Proceedings of the *Australian Conference for Knowledge Management & Intelligent Decision Support* (ACKMIDS 2007), Melbourne, Australia.

Expertise Location Using Keyphrases in Electronic Mail: Socio-ethical Challenges

Sara Tedmori

Loughborough University

Loughborough
Leicestershire
LE11 3TU

UK

Email: S.M.Tedmori@lboro.ac.uk

[*]Sharman Lichtenstein

Deakin University

221 Burwood Highway

Burwood, VIC, 3125

Australia

Email: sharman.lichtenstein@deakin.edu.au

Thomas Jackson

Loughborough University

Loughborough
Leicestershire
LE11 3TU

UK

Email: T.W.Jackson@lboro.ac.uk

[*] Corresponding author

Abstract

Employees frequently lack sufficient expertise to perform their jobs effectively. This paper describes a recently developed organisational expertise locator system based on keyphrase identification from e-mail messages and follow-up feedback from message senders. The paper provides an analysis of the key socio-ethical challenges involved in the implementation and use of this system. Findings include a set of complex socio-ethical challenges, and managerial and theoretical implications. The paper highlights the potential sensitivities of employees with respect to their potential identification by the system as topical experts. It also highlights the potential for employee misuse of e-mail expertise locator systems. Such systems must be carefully managed to reduce the risks involved. The paper highlights the need for knowledge management (KM) system designers to carefully consider socio-ethical issues when designing and implementing organisational KM solutions.

Keywords

Expertise location, knowledge management, electronic mail, socio-ethical

Introduction

Knowledge workers in contemporary organisations increasingly require additional expertise to perform their jobs effectively. In recent years sustained strategic and economic restructuring in companies has displaced important organisational expertise and obfuscated its internal location (Hamel & Prahalad, 1994). A popular strategy for an employee to obtain needed knowledge involves seeking an internal expert willing to share specialised knowledge (Ackerman et al., 2003). Speedy expertise location is also important to companies. When employees gain rapid access to experts, organisational performance may increase (Dooley et al., 2002). As a result, many modern organisations endeavour to assist employees with the timely identification and location of expertise, often within a framework of organisational knowledge management (KM).

In recent years several types of expertise location approaches have been developed and implemented. First, electronic expert directories ("Yellow Pages") offer direct expert links (Dooley et al., 2002). Second, social networks provide a complex social structure for the development of social capital and the connection of novices and experts (Cross et al., 2001). A third approach centres on the use of artificial intelligence techniques to analyse stored knowledge, seeking to identify expertise and experts (c.f. Balog & de Rijke, 2007; Maybury et al, 2002; Yimam-Seid & Kobsa, 2003). However so far, existing approaches to expertise location have not proven effective at locating expertise for a variety of reasons.

A new approach to expertise location, based on keyphrase identification in electronic mail (e-mail) messages (Tedmori et al., 2006a), is considered in this paper. The approach has been implemented in several UK organisations. The keyphrase identification technique used is based on the extraction of phrases that contain important information, from the text of an email message. The proposed approach addresses some of the key weaknesses in existing approaches which rely mainly on expert contributions and dynamic updating of profiles (Pipek et al., 2003). As expertise location approaches may have socio-ethical implications (Braun & Schmidt, 2007), it is important to consider such implications for the e-mail keyphrase identification approach. For example, there may be significant employee privacy issues in the e-mail keyphrase

identification approach, as employee expertise is disclosed to the entire organisation when this approach is used.

This paper describes an expertise location system based on keyphrase identification from e-mail messages, and explores the socio-ethical challenges involved. The remainder of the paper is set out as follows. The paper first provides a review of the value of e-mail as a KM tool and resource. Second, the paper reviews prior keyphrase-based approaches for identifying, categorising and locating expertise from e-mail messages. Third, the research design is outlined. Fourth, the paper describes a new approach to keyphrase-based expertise location from analysing e-mail messages and obtaining message sender feedback. The approach offers significant value compared with existing approaches. Fifth, the paper analyses the key socio-ethical challenges involved and discusses possible solutions. Finally, the paper reflects on the main findings, identifies key theoretical and practical implications, and proposes several future research directions.

E-mail, Knowledge Work and Expertise Location

This section reviews the role of e-mail in knowledge work, focusing on its potential value for expertise location.

E-mail is an important knowledge medium, well-used by organisational knowledge workers worldwide (Doubleclick, 2005). In December 2006, ninety-one percent of U.S. internet users used e-mail daily, equal to performing searches as the most frequent daily internet activity (Pew, 2007). It is well regarded as an essential collaboration tool (Garcia, 2006). Highlighting its organisational use, a recent study shows that e-mail messages were estimated to involve an average of 15.8 Megabytes of archive storage per corporate end-user per day (Derrington, 2006).

However considering e-mail's popularity and ubiquity, there is surprisingly little research on the value that e-mail provides to organisational KM. E-mail enables greater knowledge work than possible in earlier technological eras (Jackson and Burgess, 2003; Whittaker *et al*., 2005). It enables knowledge creation (Ducheneaut & Bellotti, 2003), knowledge sharing and knowledge flow (Bontis et al., 2003). According to Lichtenstein and Swatman (2003), employees are motivated to use e-mail for knowledge work for reasons including:

- E-mail messages attract worker attention;
- E-mail is well integrated with everyday work;
- E-mail discourse provides a context for sense-making about ideas, projects and other types of business knowledge;
- E-mail enables the referencing of work objects (such as digital documents), and provides a history via quoted messages;
- E-mail's personalised messages are appealing, meaningful and easily understood;
- E-mail encourages commitment and accountability by automatically documenting e-mail exchanges;

- E-mail is collected in inboxes and organisational archives, e-mail represents valuable individual, collective and organisational memories that may be tapped later;
- E-mail discourse facilitates the resolution of multiple conflicting perspectives which can stimulate an idea for a new or improved process, product or service.

E-mail provides several important opportunities for expertise-finding. Knowledge in e-mail can: (1) be accessed and reused directly (e.g. Swaak et al., 2004) or (2) serve indirectly as a pointer to an expert (Balog & deRijke, 2007; Campbell et al., 2003)). A recognised definition of an expert is someone who possesses specialised skills and knowledge derived from training and experience (Shanteau & Stewart, 1992).

Traditionally, e-mail clients are designed for the reuse of personal knowledge archives. For example, folders are popular structures for organising e-mail messages for ease of knowledge reuse. This organisation was highlighted by a recent early study of Enron's publicly available e-mail archive (Klimt & Yang, 2004). Indeed, employees often search personal e-mail archives seeking knowledge, in preference to searching electronic knowledge repositories (Swaak et al., 2004), raising questions about the effectiveness of current electronic knowledge repositories. Of interest to this paper, the same study also found that employees prefer to find an expert ("knowledge source") to help them with their knowledge-based concerns, rather than searching directly for the needed knowledge.

Identifying Keyphrases in E-mail for Expertise location

This section reviews several prior approaches to identifying keyphrases in e-mail for expertise location, and highlights their deficiencies. In practice very few organisations exploit their e-mail message content and very few software applications enable such exploitation. Current commercial systems for expert identification by e-mail include: Tacit's ActiveNet (Tacit, 2005), AskMe Enterprise (Ask Me, 2005) and Corporate Smarts' Intelligent Directory (Corporate Smarts, 2006). Tacit Software (formally known as Tacit Knowledge Systems), is a firm with a product that purports to transform enterprise e-mail into a shared knowledge resource. The company's Knowledge-email product ("ActiveNet", known previously as "KnowledgeMail") scans and organises messages according to user-defined profiles, key concepts and phrases, and virtual community recognition (Willen, 2003). When a user enters a request for expertise, the system displays the available (expert) employees' contact details in order of closest match of area of expertise to the user query. The strength of the match is determined by the frequency, intensity, and history of the topic within the person's expertise profile. ActiveNet does not enlist knowledge within electronic mail messages to make decisions about expertise, but adopts a more subjective approach by soliciting expertise from users themselves, thus allowing employees to nominate themselves as experts if so desired.

Expertise location is approached in two contrasting ways. It is either viewed as an expert finding task (c.f. Campbell et al, 2003) or an expert profiling task (c.f. Balog and de Rijke, 2007). Expert finding addresses the task of finding the right person with the appropriate skills and knowledge, as summarised by the question: "Who are the experts on topic X?" By contrast, expert profiling addresses the task of discovering, collecting,

and producing information regarding the skills and knowledge of an individual, as summarised by the question: "What does person Y know?"

Campbell and colleagues (2003) proposed a system for e-mail expertise extraction (abbreviated as e^3) based on exploiting e-mail content and communication patterns. e^3 locates all e-mail messages on a topic and builds an expertise graph by analysing the e-mail messages exchanged between every sender and recipient pair for the topic correspondence. The expertise graph can be analysed by employing a modified version of the Hyperlink-Induced Topic Search (HITS) algorithm to identify experts. However, the size of the networks studied is very small and does not reflect the characteristics of social networks in practice (Zhang et al, 2007).

Balog and de Rijke (2007) propose an approach for creating topical profiles that are descriptors of employee work areas. Such profiles are valuable for assisting an employee in locating the most relevant expert. When an employee searches for an expert, the information returned is richer than merely a list of names and contact details. Context and evidence are employed to help employees select the most relevant expert.

The main underlying technical challenge in utilising e-mail content for expert identification is the extraction of keyphrases that provide a good indication of the sender's skills and experience. Such keyphrases 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 on. To date, current systems in the marketplace have failed to achieve this objective, mainly due to the technical difficulty in identifying keyphrases that represent an e-mail sender's knowledge, as e-mails are freestyle text, not always syntactically well formed, domain independent, of variable length, and based on multiple topics (Tzoukermann et al., 2001). Moreover, knowledge is not necessarily represented in one message, but in a thread (or several threads).

Several methods have been proposed for the automatic extraction of keyphrases (Barker & Cornacchia, 2000; Frank et al, 1999; Krulwich & Burkey, 1996; Turney, 2000]. The two main techniques are domain dependent and domain independent. Domain dependent techniques employ machine learning and require a collection of documents with keyphrases already attached, for training purposes. Furthermore, the techniques (both domain dependent and domain independent) are related to linguistics and/or use pure statistical methods. A number of applications have been developed using such techniques. A discussion of existing approaches, together with their merits and pitfalls, is provided in (Tedmori et al., 2006b).

There are many weaknesses with current approaches to automatic keyphrase identification, several of which are discussed here to illustrate the issues. First, the extraction of noun phrases from a passage of text is common to all such approaches (Tzourkman et al., 2001; Hulth, 2003). However, a disadvantage of the noun extraction approach is that, despite the application of filters, many extracted keyphrases are common words likely to occur in numerous e-mails in many contexts. Therefore it is important to distinguish between more general nouns and nouns more likely to comprise keyphrases. Second, Hulth (2003) pinpoints two common drawbacks with existing algorithms. The first drawback is that the number of words in a keyphrase is limited to three. The second drawback is that the user must state the number of keywords to extract from each document (Hulth, 2003).

This paper describes a new approach which addresses the limitations in prior approaches, and discusses the socio-ethical challenges involved, an important topic very much neglected in earlier research. However before doing so, the research design for the project is overviewed.

Research Design

This section outlines the research design for the study, which adopted the systems development methodology (Burstein & Gregor, 1999). The study comprised four stages. In Stage One a comprehensive literature review of expertise location systems (reported in detail in Tedmori et al., 2006a) and key-phrase extraction techniques (Tedmori et al., 2006b) was conducted. This review provided a strong theoretical background, helped to identify how current approaches to expertise locator systems using key-phrase extraction are deficient, and suggested possible improvements. However the review did not explore the socio-ethical issues involved in such approaches, as such issues only emerged later in the research cycle.

In Stage Two, the expertise locator tool – EKE – was developed, and an evaluation of the keyphrase extraction engine was conducted (Tedmori et al., 2006b). Frequent development cycles and project meetings helped identify user requirements and obtain feedback which was later analysed. The evaluation of the EKE system was an integral task in the development process. Evaluation of the key-phrase extraction process was essential to assessing its performance. The process was evaluated for the effectiveness of extracted keyphrases using a recognised performance measure (Tedmori et al., 2006b).

In Stage Three, the complete system (including the key-phrase extraction engine) was piloted by technologically competent end-users who were frequent and experienced e-mail users, selected from within the school of Informatics at Loughborough 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 obtained was either resolved or marked for further action.

In *Stage Four* (*the stage reported in this paper*) the key socio-ethical challenges involved in the implementation and use of expert locator systems were identified from a literature review of relevant KM, information systems and ethics literatures. Key socio-ethical issues as they might apply to the developed tool were identified and synthesised. Later in this paper, the key socio-ethical issues are discussed and addressed.

Expertise Location from E-mail – EKE

In this section, we describe a new approach to keyphrase identification from e-mail, recently implemented in several companies in the UK. An E-mail Knowledge Extraction (EKE) agent, developed by Tedmori et al. (2006a), aims to enable end-users ("users") to locate employees ("experts") who may possess the knowledge that the user seeks. Figure 1 provides an overview of how the EKE system analyses e-mail messages to identify individuals who have the experience, knowledge and skills to help resolve a problem, find a solution or accelerate task accomplishment. Once an e-mail is sent by a user (step 1 in Figure 1), that e-mail is intercepted and the body of the e-mail message is captured. Keyphrases are then extracted by EKE's keyphrase extraction engine (step 2 in Figure 1). Following that step, the extracted keyphrases are displayed to the sender for her approval (step 3 in Figure 1).

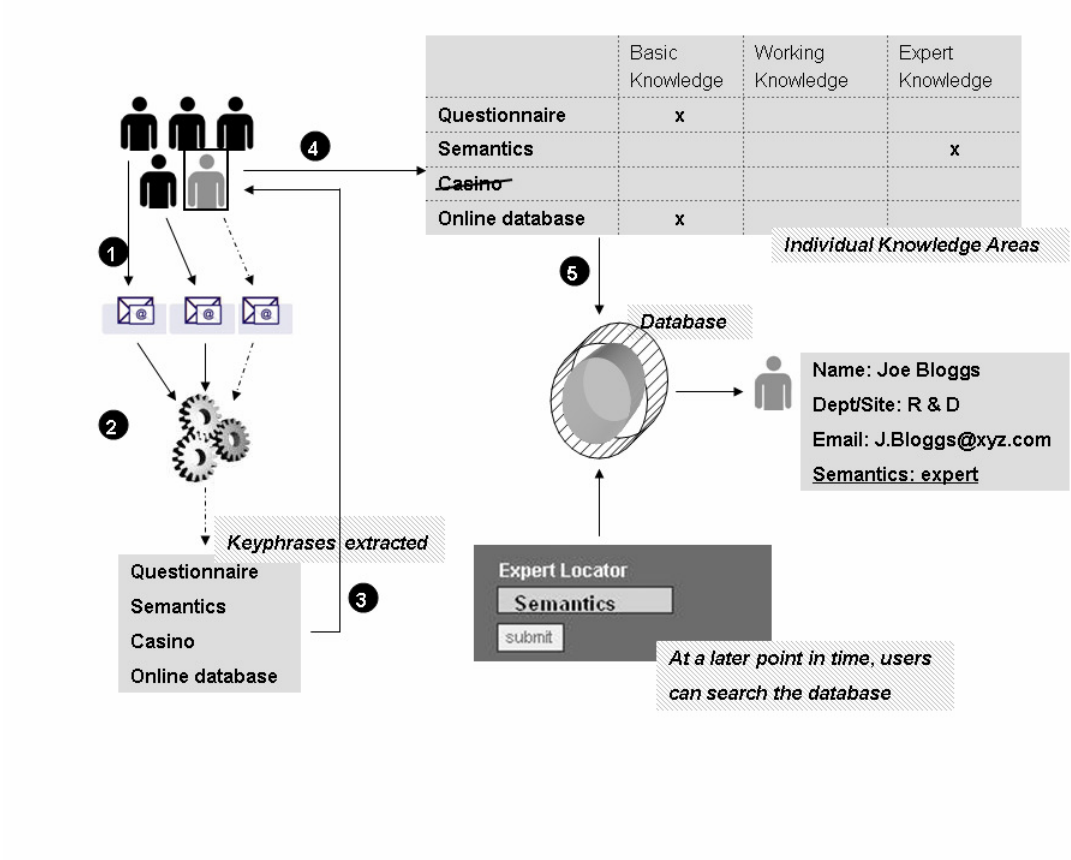


Figure 1 – Overview of the E-mail Knowledge Extraction Agent (Tedmori et al., 2006a)

The user is responsible for ranking the extracted keyphrases, using a scale of 1 - 3, to denote her expertise in that field (basic knowledge, working knowledge, or expert knowledge) and publish her profile for others to review later. The keyphrases are stored in an expertise profile which the user can access and edit (step 4 in Figure 1). The expertise profile therefore contains explicit knowledge about 'who knows what' within the organisation. This knowledge is then distilled into information in a searchable database (step 5 in Figure 1) which users can query to find relevant experts. Not all comparable existing systems perform steps 3 and 4, however the advantage of the EKE approach over those systems that do not perform steps 3 and 4 is that the system becomes more accurate by asking the end-user to rank the keyphrases extracted. Most of the current systems attempt to rank experts automatically - without human intervention – producing a system with a very low success rate of correctly identified expertise level. This is a key advantage of the EKE system over other approaches.

The above approach has been trialled in practice and shown to be effective in correctly identifying experts in the firm (Tedmori et al,, 2006a). However there are important socio-ethical questions that have yet to be addressed, which are explored in this paper.

Socio-ethical Challenges for E-mail Expertise-locator System

This section identifies the key socio-ethical challenges engendered by the proposed expertise locator system (EKE) by reviewing relevant literature.

Employee rights

The use of an e-mail expert-locator system threatens employee rights such as privacy, as will be discussed in the next section. Many studies have shown that employee performance and organisational citizenship are affected by perceived organisational justice (Cohen-Charash & Spector, 2001). Clearly it is important that employee rights are respected in the implementation and management of the e-mail expertise locator system. Employer and employee rights often differ in internet use policy (Lichtenstein & Swatman, 1997) and e-mail policy (Kelleher & Hall, 2005) and thus there may be a conflict of rights. A key employee right is the right to privacy and absence of monitoring, discussed next.

Privacy and Monitoring

Information privacy addresses the legitimate collection, use and disclosure of personal information, as well as "the claims of individuals that data about themselves should generally not be available to other individuals and organizations, and that, where data is possessed by another party, the individual must be able to exercise a substantial degree of control over that data and its use" (Clarke, 1999).

According to this definition the e-mail expertise locator system suggests a loss of privacy as data will be made available to certain individuals in the organisation and used for purposes to which the user has not consented. The approach also involves monitoring, as employee e-mail messages are being scanned and filtered for expert keyphrase content. Workers may object if they know their e-mail messages are being subjected to a form of organisational monitoring, even if only for the purpose of identifying expertise.

Arguments have been made against monitoring in an internet context. Martin and Freeman (2003) highlighted issues of liability, privacy, security, creativity, paternalism and social control arguments. Interestingly, Urbaczewski and Jessup (2002) discovered that workers monitored electronically for feedback purposes were more satisfied with monitoring than workers monitored for reasons of management control. In addition the researchers found that high-performing, highly motivated workers were more accepting of monitoring than other worker categories. According to a review of current studies on this topic, many groups accept electronic monitoring (Stahl *et al*., 2005).

Obtaining the informed consent of employees to e-mail scanning in the proposed way is a possible solution to the monitoring dilemma. However there is a need for companies to provide employees with greater negotiating power to protect employee privacy from increasing workplace surveillance (Palm, 2004; Stahl, 2005) even when informed consent is obtained. Stahl (2005) suggests that the role of managers, when addressing surveillance, might be as participants or moderators of the discourse. Recent studies have highlighted a role for e-mail committees, training, policies, and sustained awareness when e-mail monitoring is pursued as policy (Duane & Finnigan, 2005). Clearly there are some precedents in the e-mail management literature for possible managerial/policy alleviation of monitoring concerns. The proposed system (EKE) addresses this sensitive area by enabling employees to select the keyphrases they would like to share with the rest of the organisation.

Motivational issues for knowledge sharing

It is assumed by EKE that experts will be willing to share expertise with others when requested. However, it is well known in the KM literature that employee motivation to share knowledge is a complex issue and by no means assured. KM literature suggests that providing recognition and incentives are key motivators for employees to share knowledge (McDermott and O'Dell, 2001; Newell et al, 2002). At one organisation where EKE will be implemented, managers are considering the introduction of "knowledge miles". Each employee will be allocated a number of knowledge miles and when a colleague shares useful knowledge, the knowledge recipient can transfer knowledge miles to the colleague. At the end of each year knowledge miles can be exchanged for real air miles. Other reported barriers to knowledge sharing may be relevant to the use of EKE, such as geographic distance between expert and novice leading to a reliance on electronic communications of the missing knowledge. Such communication can be complex and is fraught with opportunities for misunderstandings.

Relationships

Schmidt and Braun (2007) pose an interesting challenge. What if employees seeking expertise do not feel comfortable learning from certain individuals, even though those individuals may be classified as experts? According to Schmidt and Brown, individuals may seek learning from others based on existing trusted relationships where there is established reliability and other positive characteristics favourable to a positive learning experience.

Such considerations raise an interesting issue for e-mail expertise locator systems. Knowledge acquisition from non-experts is a feature of social computing, or Web 2.0, which relies on social networks. Social networks can be leveraged for information retrieval (Kirsch et al. 2006). Perhaps social network analysis techniques such as that of Zhang and Ackerman (2005) could be cross-referenced with the e-mail expertise-locator tool to address this issue. Alternatively, when an employee locates an expert by means of EKE, she could be connected with the expert via a path in the corporation's social network, thereby smoothing the initial contact via leveraging established relationships.

Another relationship-oriented issue is where asymmetric relationships are detected between knowledge seeker and expert. Such a relationship occurs when one person is always the expert and the other is always the novice. An imbalance of this type can lead an employee to avoid seeking knowledge from the system. Perhaps the integration of the social network approach with the expertise locator tool can be designed to limit such situations. If a frequently used expert-novice pairing is returned by the system in response to a worker enquiry, a different combination could be offered where the "expert" has "working knowledge" only, in order to save face for the employee seeking help.

To be or not to be – an Expert?

With the EKE system, some people may never be classified as an expert on any topic and may therefore feel inferior, unappreciated or marginalised. On the other hand some people may resent being classified as experts due to the anticipated higher demand on their time. Still others may be inaccurately classified by keyphrases, particularly as, according to a recent survey, twenty-three percent of messages sent by e-mail are

personal (Radicati, 2005). In addition, according to the same survey, over sixty percent of employees use personal e-mail services to send business e-mail, thereby significantly reducing the likelihood that the corporate e-mail archive is a wide and rich representation of internal expertise. Clearly, this issue requires further investigation, however it is clear the e-mail expertise locator tool faces considerable challenges on these points.

Will the E-mail Knowledge Extractor Change E-mail Use?

It is possible that, with the knowledge that keyphrases are considered significant in terms of whether one is identified as expert, employees will change their e-mail use patterns. Experts who seek to escape detection in order to avoid the work obligations of sharing knowledge with other workers may avoid using technical terms, or use personal e-mail accounts for highly specialised subjects. Novices may attempt to be classified as experts by using specialised terms in messages, and confirming themselves as experts when questioned. The knowledge base resulting from the e-mail analysis may therefore need to be reviewed periodically by a knowledgeable group in the company, who can detect the absence of an expert, or the misclassification of an expert as novice.

Expert Headhunting

Underpinning EKE is an expertise database containing critical information about an organisation and its employees. The database may highlight an organisation's capacity to undertake specific jobs at a particular point in time, based on the stored expertise. In addition other organisations could head-hunt employees from the organisation by consulting expert profiles. The expertise database should therefore be kept confidential and inaccessible by competitors or others who might misuse the information.

Conclusion

This paper described a recently developed KM system that uses keyphrase identification of expertise from organisational e-mail message content. It also discussed the key socio-ethical challenges involved in such a system's implementation and use. Several potential solutions to the identified challenges were proposed.

The theoretical implications from the paper are several. First, a set of socio-ethical challenges for the implementation of an effective e-mail expertise-locator system adds to existing theory in the area of expertise location systems. Second, important insights are offered into the socio-ethical issues involved.

At a practical level, the paper suggests a need for software developers to extend the EKE tool to cater for the socio-ethical challenges involved. Corporate e-mail policy should also address the challenges by structuring the use of the tool according to ethical precepts. According to recent reports, 50 - 76 percent of US organisations have e-mail usage and content policies (AMA, 2006; Radicati, 2005). While the proportion of organisations with e-mail policies may differ in other countries, e-mail policy improvement is suggested. E-mail policies should reflect a negotiated consensus between employer and employee needs in acceptable e-mail use (c.f. APC, 2006). For example, the Australian privacy guidelines on Workplace E-mail and Web-browsing state: "The Privacy Commissioner encourages organisations to develop *in consultation with staff* [our italics] a clear privacy policy in relation to staff use of computer networks, particularly with regard to the use of e-mail and the Internet" (APC, 2006).

Extending this concept to the e-mail expertise locator context, employees and managers should develop e-mail policies based on negotiated rights.

In conclusion this paper highlights the need for KM system designers to carefully consider *socio-ethical issues* when designing and implementing KM solutions for organisations.

References

Ackerman, M., Pipek, V. & Wulf, V. (Eds.) (2003) *Sharing Expertise: Beyond Knowledge Management, Cambridge*, MA: The MIT Press.

AMA (2006) Workplace E-Mail, Instant Messaging & Blog Survey**,** American Management Association (AMA) and The ePolicy Institute, US.

Balog, K. & de Rijke, M. (2007) Determining expert profiles (with an application to expert finding), in *Proceedings of International Joint Conferences on Artificial Intelligence 2007*, pp. 2657-2662.

Barker K, & Cornacchia N. (2000) Using noun phrase heads to extract document keyphrases. In *Proceedings of the Thirteenth Canadian Conference on Artificial Intelligence* (LNAI 1822), Montreal, Canada, pp. 40–52.

Braun, S. & Schmidt, A. (2007) Expert Finding as Informal Learning Support: Competency-Aware and Socially-Aware, in *Proceedings of 1st International Expert Finder Workshop*, Berlin, January 16, 2007.

Bontis, N., Fearon, M. & Hishon, M. (2003) The e-flow audit: an evaluation of knowledge flow within and outside a high-tech firm. *Journal of Knowledge Management, 7*(1), 6-19.

Burstein, F. & Gregor, S. (1999) The systems development or engineering approach to research in information systems: an action research perspective, in *Proceedings of the 10th Australasian Conference on Information Systems*, B. Hope and P. Yoong, eds., Victoria University of Wellington, New Zealand, pp. 122-134.

Campbell, C.S., Maglio, P.P., Cozzi, A. & Dom, B. (2003) Expertise Identification Using Email Communications, in *Proceedings of CIKM '03*, Louisiana.

Clarke, R. (1999) "Internet Privacy Concerns Confirm the Case for Regulation", *Communications of the ACM*, *42*(2), February, 60-67.

Cohen-Charash, Y., & Spector, P. E. (2001) The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes, 86*, pp. 278-321.

Cross, E., Parker, A., Prusak, L. & Borgatti, S.P. (2001) Knowing What we Know: Supporting knowledge creation and sharing in social networks, Organizational Dynamics, 30(2), 100-120.

Derrington, S. (2006) Getting Control of the Storage Environment, IT-Observer, October 11, http://www.it-observer.com/articles/1249/gaining_control_storage_environment/

Dooley, K.J., Corman, S.R. & McPhee, R.D. (2002) A knowledge directory for identifying experts and areas of expertise, Human Systems Management, 21(4), 217-228.

Doubleclick (2005) Doubleclick's Sixth Annual Consumer Email Study, Doubleclick, Retrieved on 14 March 2007 from: http://www.doubleclick.com/us/about_doubleclick/press_releases/default.asp?p=534

Duane, A. & Finnigan, P. (2005) Monitoring and Controlling E-mail Systems: A Cross-Case Analysis, *International Journal of E-Business Research*, 1(4), 24-40.

Ducheneaut, N. & Bellotti, V. (2003) Ceci n'est pas un objet? Talking about objects in e-mail. *Human-Computer Interaction, 18*, 85-110.

Frank E, Paynter GW, Witten IH, Gutwin C, Nevill-Manning CG. Domain-specific keyphrase extraction. In *Proceedings of Sixteenth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers, San Francisco, CA, 1999, pp. 668–673 http://www.nzdl.org/Kea/Frank-et-al-1999-IJCAI.pdf

Garcia, I. (2006) The Good in E-mail (or why Email is Still the Most Adopted Collaboration Tool), Central Desktop, http://blog.centraldesktop.com/index.php?entry=entry060403-214628

Hamel, G. & Prahalad, C. K. (1994) *Competing for the Future. Breakthrough Strategies for Seizing Control of Your Industry and Creating Markets for Tomorrow*. Boston, MA. Harvard Business School Press.

Hulth A. (2003) Improved Automatic Keyword Extraction Given More Linguistic Knowledge. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'03)*. Sapporo.

Jackson, T.W. and Burgess, A., "Capturing and Managing Email Knowledge", *Business Innovation in the Knowledge Economy* - Abstracts from the IBM & Stratford-Upon-Avon Conference, Abbot, J., Martin, L., Palmer, R., Stone, M. and Wright, L.T. (eds), De Montfort University, Leicester, UK, Business Innovation in the Knowledge Economy, Warwick & Stratford-Upon-Avon, UK, June 2003, pp. 28-29, ISBN: 1857213563.

Jackson, T.W. and Tedmori, S., "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, May 2004, pp. 463-466, ISBN: 1-59140-261-1.

Kelleher, Z. & Hall, H. (2005) Response to risk, *Business Information Review, 22*(1), 46-52.

Kirsch, Gnasa & Cremers (2006) European Conference on Information Retrieval (ECIR 2006), http://sites.inka.de/moebius/docs/ecir2006-paper.pdf

Klimt, B. & Yang, Y. (2004) Introducing the Enron Corpus, in *Proceedings of First Conference on Email and Anti-spam*, url: http://www.ceas.cc/papers-2004/168.pdf

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

Lichtenstein, S. & Swatman, P. (2003). "Email and Knowledge Management", *Proceedings of 7th Pacific-Asia Conference on Information Systems*, Adelaide, Australia.

Martin, K. & Freeman, R.E. (2003) Some problems with employee monitoring, *Journal of Business Ethics, 43*(4), p. 353.

Mathiassen, L., (1998) Reflective Systems Development, *Scandinavian Journal of Information Systems*,10, (1&2), pp.67-117.

Maybury, M., D'Amore, R. & House, D. (2002) Awareness of Organizational Expertise, *International Journal of Human-Computer Interaction, 14*(2), 199-217.

McDermott, R., O'Dell, C. (2001). Overcoming cultural barriers to sharing knowledge, Journal of Knowledge Management, 5(1), 76-85.

Newell, S., Robertson, M., Scarbrough, H., and Swan, J. (2002). Managing knowledge work, Palgrave, Basingstoke, Hampshire.

Palm, E. (2004) An Ethical Questioning of Work Place Surveillance - strengthening employees' negotiating power, *Ethicomp Proceedings* 2004, Syros, Greece.

Pew (2007) Internet Activities, Pew Internet & American Life Project, http://www.pewinternet.org/trends/Internet_Activities_1.11.07.htm

Pipek, V., Hinrichs, J. & Wulf, V. (2002) Sharing Expertise: Challenges for Technical Support, in Sharing Expertise: Beyond Knowledge Management (Ackerman, M., Pipek, V. & Wulf, V. Eds.), Cambridge, MA: The MIT Press.

Radicati (2005) Corporate Email User Habits, White Paper, The Radicati Group Inc and Mirapoint, http://www.mirapoint.com/pdfs/whitepapers/EmailUserHabits_Radicati.pdf

Shanteau, J. & Stewart, T.R. (1992) Why study expert decision making? Some historical perspectives and comments, Organizational Behavior and Human Decision Processes, *53* (2), 95-106.

Schmidt, A. & Braun, S. (2006) Context-aware Workplace Learning Support: Concept, Experiences and Remaining Challenges, in *Proceedings of First European Conference on Technology-enhanced Learning (ECTEL'06)*.

Stahl, B.C. (2005) Responsible Management of Information Systems: an Oxymoron? Responsible Leadership track, in *Proceedings of EURAM 2005*.

Swaak, J., T. de Jong, and W.R. van Joolingen (2004) "The effects of discovery learning and expository instruction on the acquisition of definitional and intuitive knowledge", *Journal of Computer Assisted Learning*, Vol 20, No.4, pp. 225- 234.

Tedmori, S., Jackson, T.W., Bouchlaghem, D., Holger, H. and Nagaraju, R., "Building a Tool for Expertise Discovery", *Emerging Trends and Challenges in Information Technology Management*, 2, Khosrow-Pour, M. (ed), Idea group Publishing, *Information Resources Management Association International Conference*, Washington, DC, May 2006a, pp 1053-1056, ISSN 1 59904 019 0

Tedmori, S., Jackson, T.W. and Bouchlaghem, D., "Locating Knowledge Sources Through Keyphrase Extraction", Knowledge and Process Management, 13(2), 1st April 2006b, 100-107, ISSN 10924604

Turney, P (2000) Learning Algorithms for Keyphrase Extraction. *Information Retrieval,* **2**, 303-336

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

Urbaczewski, A. and Jessup, L.M. (2002) Does Electronic Monitoring of Employee Internet Usage Work?, *Communications of the ACM*, 45, 1, 80-83.

Whittaker, S., Bellotti, V. & Moody, P. (2005) Introduction to this special issue on revisiting and reinventing e-mail, *Human-Computer Interaction*, *20* (1-2), 1-9.

Willen, C. (2003), "Mining email uncovers knowledge resources but could be intrusive", http://www.intelligentkm.com/feature/010308/editpage.shtml, 22 July 2003

Yimam-Seid, D. & Kobsa, A. (2003) Expert finding systems for organizations: problem and domain analysis and the DEMOIR approach, *Journal of Organizational Computing and Electronic Commerce, 13*(1), 1-24.

Zhang, J. and Ackerman, M. (2005) Searching for Expertise in Social Networks: A Simulation of Potential Strategies, in *Proceedings of GROUP 2005*, Sanibel Island, Florida.

Zhang, J., Ackerman, M. S., and Adamic, L. (2007). Expertise Networks in Online Communities: Structure and Algorithms, in *Proceedings of 16th International World Wide Web Conference (WWW2007)*, Banff, Canada.

APPENDIX E PAPER 5 (TO BE SUBMITTED)

Tedmori, S., Jackson, T.W., Bouchlaghem, D (2008), Evaluation of an Email Knowledge Extraction System.

Evaluation of an Email Knowledge Extraction System

Sara Tedmori[1], Dr. Thomas W. Jackson[2], Prof. Dino Bouchlaghem[1]
[1] Centre for Innovative Construction Engineering, Loughborough University, Loughborough, Leicestershire, UK
[2] Research School of Informatics, Loughborough University, Loughborough, Leicestershire, UK
S.M.J.Tedmori@lboro.ac.uk
T.W.Jackson@lboro.ac.uk
N.M.Bouchlaghem@lboro.ac.uk

Abstract
This paper reviews an approach to locating knowledge holders within organisations through the use of a well established communication medium, email. The approach has been used to develop the Email Knowledge Extraction (EKE) tool. EKE was then evaluated at a leading academic institution in the UK. This study represents the first effort to validate the viability of the email medium as a source of knowledge profiling data, to be used for finding employees who possess the required knowledge. The socio-ethical challenges associated with EKE's adoption are also explored.

Keywords
Expertise Locator, Email, Knowledge Extraction, System Evaluation, Socio-ethical

1. INTRODUCTION
In this highly competitive global economy it is important to provide employees with the right tools to capture and share knowledge. This allows employees to resolve problems, leading to increased creativity and higher levels of innovation. A recent development addressing this requirement has been the introduction of expertise finding tools that provide users with convenient access to individuals with particular skills. The aim of such tools is to categorise users' knowledge, by utilising information not generally exploited by knowledge workers in a way that can be later queried by knowledge seekers.

Everyday employees encounter problems that require resolving. The tendency is to seek help from colleagues they know. This often involves following referrals to other people until the right person is located (Campbell et al., 2003). As organisations grow and become more geographically dispersed, the difficulty of finding where and with whom the knowledge resides increases. This is when organisations recognise that they need to enhance their employees' expertise finding behaviour by developing tools that facilitate collaboration. Expert finders have been developed for this purpose. An expert finder can be viewed as a repository that holds pointers to knowledge holders. Employees at various locations can query this repository in search for individuals capable of providing them with the required assistance. 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 (Shanteau & Stewart, 1992).

The aim of this work is to present and evaluate EKE, the expert finding tool developed by the authors as part of an ongoing research project. The paper outlines EKE, as well as the underlying concept and evaluation findings of the tool.

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. 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, expert seekers often require assistance with complex queries which require the combined experience of multiple experts.

There have been numerous attempts by researchers in both academia and industry to semi-automate/automate the process of finding the right expert. These systems can be differentiated depending on the information source(s) they exploit for expertise recognition including: self disclosure, emails, knowledge repositories, bulletin boards, program code, web pages, and other documents. There have been numerous attempts to create tools which utilise email as their source of information including *Tacit's ActiveNet* (Tacit Software, 2007), AskMe Enterprise (AskMe) and Corporate Smarts' Intelligent Directory (Corporate Smarts). 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. It 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 (Lichtenstein, 2004).

Expert locator systems have been implemented in a variety of organisational domains (Maybury, 2007). Successful deployment of such systems is dependent on many factors including: user involvement, clear purpose, measured usage and benefit, ease of use, incremental deployment, appropriate privacy, incentives for use, and effective training (Maybury, 2007). The authors' literature review has shown that there is 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.

3. EKE: A SYSTEM FOR LOCATING EXPERTISE FROM EMAILS

This section outlines the approach developed to identify individual competencies from email messages. EKE reported in detail in Tedmori et al. (2006), aims to allow users to locate others who may possess specialised knowledge.

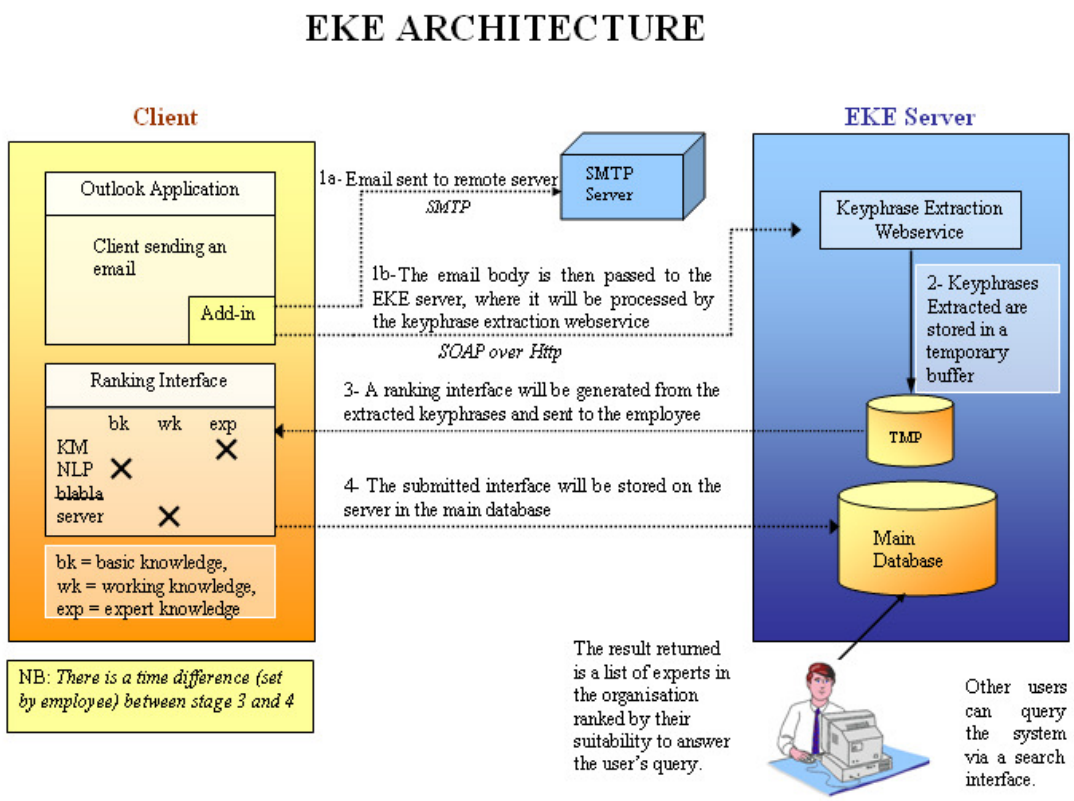


Figure 1: EKE Generic Architecture

Figure 1 illustrates EKE's architecture. The system comprises of client 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 on a keyphrase categorisation form for verification (i.e. whether or not the term returned is indicative of their skills) and categorisation (i.e. their subjective level of knowledge in the returned proficiency area). Once the user verifies that the keyphrases are relevant, they categorise them under one of four knowledge categories. The knowledge categories available to the user are: Basic Knowledge (BK), Working Knowledge (WK), Expert Knowledge (EK), and non applicable (N/A). Once categorised, the keyphrases are stored in the respective user's profile under the appropriate knowledge categories, in a repository which can be queried by all users.

From a technical viewpoint, identifying keyphrases suggestive of sender skills and experiences represents the main challenge. A detailed overview of EKE's keyphrase extraction engine is provided in Tedmori et al. (2006). 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.

4. RESEARCH DESIGN

The research adopted a systems development methodology which recognises three main stages: concept development, system building, and system evaluation (Burstein, 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 Tedmori et al. (2006). A domain analysis identified shortcomings in the research subject as well as further aspects that need to be addressed (also reported in detail in Tedmori et al. (2006)). This 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 (Tedmori et al. (2006). The development process was iterative and evolutionary. Individual programme units were developed, tested, and integrated in order to obtain the final system. Evaluation of the keyphrase extraction engine was essential to assess its performance. The engine was evaluated for the effectiveness of extracted keyphrases using a recognised performance measure (Tedmori et al., 2006b).

During the system evaluation stage *(the stage reported in this paper)*, the system availability was expanded to target potential users. The evaluation consisted of focus group sessions (discussing these groups is beyond the scope of this paper) and a system evaluation study. A detailed description of the EKE system evaluation is provided in section 5.

5. Evaluation of the EKE System

A study was conducted at the Information Science Department at Loughborough University with the aim of obtaining feedback from potential end users on functionality, robustness, clarity, and ease of use in a working environment. EKE was installed on 10 personal computers belonging to members of staff for a period of six weeks. Participants were given a walkthrough of the system and asked to complete a pre-study questionnaire. After the study period, EKE was uninstalled from the participants' personal computers. Participants were then asked to fill in a post-study questionnaire tackling areas related to EKE's performance, usability, and handling of socio-ethical challenges. As part of this questionnaire, participants were provided with a list of their key proficiencies, identified from their email communications. Participants were asked to judge how well the key proficiencies identified reflect their areas of knowledge. Results from this study provide valuable feedback on the end-users' perceptions of the EKE tool with regards to the enablers and the barriers to its adoption.

5.1 General information about the respondents

Table 1 shows the general characteristics of the participants of the study. The participants had several years of work experience at Loughborough University. While 20% of participants had two years or less experience, the majority's experience was significant, with 80% ranging between 5 and 25 years and 50% ranging between 10 and 25 years experience. The participants' experience averaged 10.6 years.

Table 1: Demographic Information

Gender	**Male**			**Female**	
	60 %			40 %	
Age Group in years	**Under 25**	**25 - 34**	**35 - 44**	**45 - 54**	**Over 55**
	0 %	10 %	10 %	70 %	10 %

All participants (100%) indicated that they use email on a daily basis to help generate, organise, share, and leverage knowledge. Table 2 summarises the number of knowledge-related emails that participants send per day. According to the table, half of

the participants (50%) send up to 10 emails a day, while the other half (50%) send a minimum of 11 emails a day.

Table 2: Email Use

Number of knowledge related emails sent a day	0-5	6-10	11-20	21-40	More than 40
	20 %	30 %	40 %	0 %	10 %

5.2 Participants' first impressions of the system

The participants were given a brief overview of the EKE tool and were asked to provide their initial thoughts in relation to the EKE concept. From the comments made, there are indications that the EKE concept is regarded as a good, useful, and interesting idea. It was suggested that the information captured by EKE could be used to enable personal ratings for the purpose of personal development. It was felt that EKE's ability to integrate with existing technologies (i.e. Outllook) would encourage its adoption and use.

Table 3: Clarity of Keyphrase Categorisation Form

Clarity of keyphrase categorisation form	Not Clear	Not Too Clear	Neutral	Somewhat Clear	Very Clear
	0 %	0 %	20 %	30 %	50 %

Using a scale of 1 (not clear) to 5 (very clear), participants were asked to rate how clear they felt that keyphrase categorisation form was. Table 3 summarises the findings. Half of the participants (50%) thought the form was very clear and no one felt the form was not clear. However, opinion was divided between those who thought the form could be improved (50%) and those who thought it could not (50%). The following suggestions for improvements were made:

"It would be nice to be able to turn the form off".

"To be able to differentiate between phrases I have no knowledge in and those which are non-applicable (N/A)".

"One improvement which can be made it to provide hot links to the knowledge categories displayed on the form, to help me make my selection".

5.3 End User Evaluation of EKE

Table 4 shows the results of the system evaluation. The responses attained indicate the simplicity of using the system. EKE was rated as very easy to use by 40% and easy to use by 60%. The majority (90%) felt that the concept of EKE was easy to grasp by the majority.

All participants felt that their ability to send emails was not adversely affected by the use of EKE. The tool was found intuitive to use by 80% of the participants, with only 10% expressing difficulty in using and understanding the tool.

Table 4: System Evaluation

I would rate the use of the EKE software as.	**Very Easy**	**Easy**	**Neutral**	**Difficult**	**Very Difficult**
	40%	60%	0%	0%	0%
The concept of the EKE software was _____ to grasp.	**Very Easy**	**Easy**	**Neutral**	**Difficult**	**Very Difficult**
	30%	60%	0%	10%	0%
I can still effectively send email using this tool.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	10%	90%	0%	0%	0%
I am able to send emails quickly using this tool.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	10%	20%	60%	10%	0%
I feel comfortable using this tool.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	20%	60%	0%	20%	0%
It was intuitive to learn how to use this tool	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	30%	50%	10%	10%	0%
The knowledge categories were appropriate to use.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	20%	50%	10%	20%	0%
The layout of the keyphrase categorisation form is clear.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	30%	50%	10%	10%	0%

EKE uses four categories “BK”, “WK”, “EK”, and “N/A” to allow users to classify their knowledge areas. 70% of the participants agreed that these categories were appropriate for that purpose. In addition, 80% agreed that the layout of the keyphrase categorisation form was clear.

When the participants were asked whether they thought the task of extracting keyphrases from email was worthwhile, 100% were of the opinion that it was. In justification of their choice, the following comments were given:

“There is so much information content in email that is not searchable in any sort of way and is difficult to extract”.

“So much of our daily work and communication takes place via email. It is a lost opportunity not to use its potential”.

When asked whether they would like to see the idea of extracting keyphrases applied to email attachments, the majority (70%) responded by saying that they wouldn’t. Some felt that in the case of a detailed email, a long list of keyphrases was being returned. Applying extraction to attachments implies a longer list would be generated. However, 30% of the respondents thought it might be a good idea and it was suggested by one of the respondents as a possible improvement which might result in the identification of

key knowledge areas that are not identified through using email content alone. In response to whether the tool would benefit users in searching for experts, 77.8% of the respondents agreed that it would.

In response to the question whether there was anything about the tool that they disliked, 66.7% said that there was. The following issues were raised:

"Some of the words extracted made little sense".

"The tool parses through the whole thread instead of only the sender's text". This is not consistent with other participants' feedback who viewed this as a advantage stating that if you don't look at the whole thread, you will loose a lot of the knowledge.

"For a short email, a long list is generated".

"It slightly slows down the process of sending email".

In reference to the first point, that some of the words extracted made little sense, the end-users views were very subjective. For example, some participants were not pleased that the system picks up module codes and geographical information (e.g. names of places), others were quite happy that it does.

The respondents (75%) felt that the tool could be improved and provided the following suggestions:

"Replacing the "BK" category with an "Interest" category": One of the respondents argued that temporary interests are time bounded. A lot of knowledge is temporal and tied to a particular context. Hence, the respondent indicated that they tend to regard themselves as knowing in that particular context, and not in a generic sense.

"Introducing a finer level of granularity by increasing the range of knowledge categories".

"Limiting the frequency of the tool's execution" (i.e. business and not personal email)

"Improving the accuracy and recognition of keyphrases".

Participants were shown a list of the keyphrases that had been extracted from the emails they sent. They were asked to rate for each of the knowledge categories, how well these keyphrases gave an accurate reflection of their knowledge. Table 5 provides a summary of their responses.

Table 5: Evaluation of Extracted Keyphrases

	Poor	**Fair**	**Satisfactory**	**Good**	**Excellent**
Basic Knowledge	20%	20%	10%	40%	10%
Working Knowledge	0%	20%	30%	40%	10%
Expert Knowledge	11.1%	11.1%	22.2%	11.1%	44.4%
Overall	10%	20%	20%	40%	10%

The findings above show that people are better at identifying the areas they perceive themselves as experts in as opposed to the areas they have basic or working knowledge in. As shown, 44.4% said that the keyphrases categorised under "EK" gave an excellent representation of their knowledge. Only 10% said that the keyphrases classified under "BK" or "WK" gave an excellent representation. This can be attributed to a number of factors such as email is a better source of keyphrases which represent expert areas; the extraction algorithm is better at identifying keyphrases representing knowledge areas employees are experts in; participants are better at identifying the areas which they deem themselves as experts in; and/or the group of academics, who composed the sample, are experts in their particular fields, thus send more specialised emails. Each participant was shown a keyphrase list. The list was divided under the three knowledge categories, "BK", "WK" and "EK". Under each category, a list of ten random keyphrases was selected, where available, from the corresponding category in the user's profile.

Table 6: Evaluation of Extracted Keyphrases

	Total Keyphrases	% Yes	% No
Basic Knowledge	96	58%	42%
Working Knowledge	94	94%	6%
Expert Knowledge	73	86%	14%

The results summarised in Table 6 show that the keyphrases extracted under "WK" and "EK" represented the respective users' knowledge more accurately than those classified under "BK". This is interesting because users' classified the keyphrases themselves. The reason could be attributed the four factors described earlier.

When given the choice, 90% of participants said that they preferred to extract keyphrases from email using EKE rather than manually. The majority of participants who favoured EKE attributed this primarily to the speed of the system and its ease of use. Another reason mentioned was the tool's good coverage of main email subject categories.

5.4 Socio-Ethical Challenges

Findings of socio-ethical challenges are summarised in Table 7. In relation to the impact on employee privacy, 50% felt that the use of EKE has little or no impact. They stated that it should not be a threat as long as employees consent to using it. Some added that because of what is stored in the database is optional, it has little impact.

Some respondents (40%) thought that the system has to some extent an impact on employee privacy. This can be attributed to the participants' perception that any personal information shared across organisational boundaries affects privacy. Another reason is related to the keyphrase categorisation form which gives the sender the impression that big brother is watching. Only 10% felt that the use of EKE has a moderate impact on employee privacy and expressed that the scale of the impact depends on how EKE will be used. No one thought that EKE has a great impact on privacy.

Table 7: Socio-ethical Challenges

Loughborough University, Department of Information Science

The use of EKE threatens employee privacy.				
To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
10%	40%	40%	10%	0%
Organisational privacy policies provide sufficient protection.				
Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
0%	50%	50%	0%	0%
The governmental legislation provide sufficient protection. (One respondent did not answer this question)				
Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
11.1%	66.7%	11.1%	11.1%	0%
How comfortable do you feel having the emails analysed by EKE for the purpose of identifying areas of expertise?				
Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
50%	30%	20%	0%	0%
EKE enables employees to select keyphrases they would like to share with others. To what extent do you think this addresses privacy issues?				
To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
20%	10%	20%	10%	40%
To what extent are you willing to share your expertise with others when requested				
To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
0%	0%	20%	20%	60%
Extent to which you think that providing recognition and incentives is a key motivator to use EKE and share knowledge				
To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
20%	10%	70%	0%	0%
How comfortable do you feel seeking information from experts you do NOT know?				
Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
0%	50%	50%	0%	0%
How many enquiries willing to reply to in a week?				
None	**1 – 3**	**4 - 7**	**8 – 10**	**More than 10**
0%	62.5%	12.5%	0 %	25%
How comfortable do you feel about EKE incorrectly classifying a user as an expert on a particular topic?				
Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
0%	10%	30%	50%	10%
To what extent do you think EKE will negatively affect you email usage?				
To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
60%	40%	0%	0%	0%

50% of participants agreed that Loughborough University's privacy policies provide them with sufficient protection. Some accepted the right of employers to have the ability to monitor communications when appropriate. Others, simply had faith that they were protected. 50% expressed neutral views due to a lack of knowledge about the available legislation, some of whom thought it was a fallacy to think that you are protected. If your organisation wants to know what you are doing, they can always find out.

78% of the respondents (one respondent didn't answer this question) agreed with the statement that the current legislation in the UK offers them sufficient protection against the potential privacy problems that might result from the use of EKE. Some thought that it is the responsibility of the employee to limit their use of employer's communication facilities to legal use only. Others, simply had faith that they were protected. Only 11.1% disagreed with the statement and used an example from personal experience relating to the loss of child benefit data by the HM Customs and Revenue (BBC, 2007).

The majority (80%) expressed their comfort with EKE analysing their emails for the purpose of identifying their areas of expertise. Those who provided clarification, attributed this to trusting the system (as with online banking). Others linked their comfort to the knowledge sharing objective of EKE and the people utilising it.

EKE enables employees to select the keyphrases they would like to share with the rest of the organisation. 50% of respondents thought that this feature addresses privacy issues. When asked for clarification, respondents felt that "*it is the employee's responsibility to choose the appropriate keyphrases. If you object to using it, then you don't have to use it*". Some felt that the categorisation of knowledge areas addresses the privacy concerns. Others said they are happy as long as the keyphrases can be removed from the database. 30% of respondents thought that this addresses privacy issues to little or no extent. One participant thought that the task of extracting keyphrases to identify employees' competencies could be achieved by utilising the emails' subject lines instead of scanning the email body. From the authors' point of view, this is not possible due to the size, availability, and lack of knowledge in subject lines.

No one expressed reluctance to share expertise with others when requested. On the other hand, the vast majority (80%) expressed their willingness to share when identified as experts. Some stated that they are willing to share expertise, however being identified as an expert does not imply willingness to share knowledge with all who contact them. Others linked this to the nature of information that might be shared. For example, if sharing such information contradicts preserving intellectual property rights then they are reluctant to share it. Some expressed a higher degree of readiness to share knowledge in areas in which they regarded themselves as experts rather than in areas in which they have basic knowledge. 20% indicated their willingness to share their expertise to some extent claiming that this is dependent on the circumstances of the request.

20% felt that the provision of recognition and incentives does not motivate employees to use the system and share knowledge. 10% felt that such provision does motivate employees, but only to a little extent, because recognition/motivation is not required for a system that is capable of automatically identifying metadata. Metadata provides added value sufficient for one to participate. On the other hand, 70% felt that such kind of

provision does to some degree motivate employees. Some said that the importance of recognition and incentives as key motivators is dependent on the individual, what may be an incentive for one individual does not necessarily represent an incentive for another. Some regarded the use of EKE as sufficient in itself as it may lead others to contact them, enhancing therefore the prospect of networking. Others highlighted the issue that in order to encourage participation, some incentive (e.g. appraisal, research performance) may be more important to provide at the early stages of adoption rather than at later stages.

Participants were asked to indicate how comfortable they feel seeking information from experts they do not know. Half of the participants (50%) expressed their comfort in doing so while the other half (50%) opted for a neutral position. Some of those who indicated that they were comfortable clarified by saying that care needs to be taken when approaching those experts. Others said that this depends on who the experts are and how well they are prepared when answering their queries. From the clarifications provided by the respondents who took a neutral stance, it was evident that they found contacting someone they don't know a bit difficult, but sometimes necessary. For example, if time is pressing, then they would directly contact people they don't know. However, if more time was available, they preferred in the first instance to either contact someone they know or to seek an introduction from someone they are acquainted with. One of the respondents noted that this depends on how much they trust the accuracy of EKE in returning relevant experts.

All respondents expressed willingness to reply to enquiries generated through EKE from co-workers. Those who were willing to reply to more than 10 enquiries established a link between the number of enquiries they are willing to reply to and the nature of the enquiry (e.g. how complex the query is, how much time it requires, who is asking) and/or the circumstances of the respondent (e.g. time of request). For example, the less complex the enquiries are, the more enquiries they are willing to reply to.

60% expressed discomfort with the possibility that EKE might incorrectly specify a user as an expert on a particular topic. It was stated that such a situation could be misleading as it raises questions regarding the validity of the system. 30% expressed neutral position. They highlighted the importance of clarifying at the outset the possibility of such a problem taking place. 10% felt comfortable stating that they always have the option of either taking the advice from the wrongly classified expert or leaving it.

Generally all respondents reported that EKE had minimal or no negative effect on their email usage. The only concern raised was related to the slight delay of sending emails. However, it was reported that such a delay is not significant.

6. CONCLUSION

In this paper, the authors reviewed and evaluated EKE, a tool for keyphrase extraction from email messages. The evaluation showed that EKE was found to be useful, interesting, easy and intuitive to use. It did not adversely affect users' ability to send emails. In particular, the evaluation highlighted that people are better at identifying the areas they perceive themselves as experts as opposed to areas in which they have basic or working knowledge in.

With regards to the socio-ethical challenges, there were no significant privacy concerns, users were happy with EKE analysing their emails. However, there was a general lack of knowledge concerning organisational and governmental privacy policies. It was found that EKE's users would be willing to share and reply to enquiries when identified as experts. Nonetheless, the provision of recognition and incentives was seen as a possible motivator to encourage participation and continuous use. Generally users expressed that they would be comfortable with seeking information from experts they did not know. Though, they would be uncomfortable if EKE incorrectly classified users as experts on particular topics.

It was suggested that EKE could be improved by increasing its flexibility and providing more configuration options, where for example users could specify the types of emails analysed. Adoption could be enhanced by educating and training users on how to make the most of EKE and how it would benefit them. In general, extracting keyphrases from email was considered as a worthwhile activity, with the possible extension to other applications added benefit of extension to other applications.

EKE was evaluated by ten users. Further work using larger sample sizes in a wider variety of environments is required to further evaluate EKE.

Reference:

AskMe . Available: http://www.askmecorp.com/ [2005, 09/16].

BBC *, Data lost by Revenue and Customs*. Available: http://news.bbc.co.uk/1/hi/uk/7103911.stm [2007, May/Sunday].

Burstein, F. (2002). "System Development in Information Systems Research" in *Research Methods for Students, Academics and Professionals: Information Management and Systems*, ed. K. Williamson, 2nd edn, Centre for Information Studies, Charles Sturt University, Wagga Wagga, NSW, pp. 147-158.

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.

Corporate Smarts . Available: http://www.corporatesmarts.com/ [2005, 09/16].

Lichtenstein, S. (2004). "Knowledge Development and Creation in Email", *The 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 8*IEEE Computer Society, Washington, DC, USA, pp. 245.

Maybury, M.T. (2007). "Discovering Distributed Expertise", *Regarding the "Intelligence" in Distributed Intelligent Systems*MITRE, Washington D.C.

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.

Tacit Software . Available: http://www.tacit.com/home.asp [2005, 09/16].

Tedmori, S., Jackson, T.W., Bouchlaghem, N.M., Adelmann, H. & Nagaraju, R. (2006). "Building a Tool for Expertise Discovery", Idea Group Publishing, , pp. 1053.

Tedmori, S., Jackson, T.W. & Bouchlaghem, D. (2006). "Locating Knowledge Sources Through Keyphrase Extraction", *Knowledge and Process Management; Chichester,* vol. 13, no. 2, pp. 100.

APPENDIX F SUPPORTING DOCUMENTS

I. Leicestershire Constabulary Questionnaire

II. DIS Pre-Study Questionnaire

III. DIS Post-Study Questionnaire

IV. Focus Groups Questionnaire

V. Focus Group Discussion and Questionnaire Analysis

I. LEICESTERSHIRE CONSTABULARY QUESTIONNAIRE

Information Seeking and Searching Questionnaire

Once started the survey can not be saved and completed at a later date and therefore must be fully completed and submitted once commenced. You will need approximately 10-15 minutes to complete this questionnaire.

The results will enable us to better understand how employees work and in turn enable us to develop IT systems of the future that will support you in your daily activities. Your input will feed into the world leading research being undertaken at Loughborough University.

If you would like a copy of the results, please contact Sara at s.m.j.tedmori@lboro.ac.uk, and I will email you a copy as soon as it is available.

All responses to this questionnaire will be treated in the strictest of confidence. The information you provide will only be used for this research project. Individuals will not be identified in any results produced.

* indicates required field entry

Section 1: General

*** 1. Gender**

- ❍ Female
- ❍ Male

***2. How many years have you been working at Leicestershire Constabulary?**
Choose only one of the following

- ❍ Under 1 year
- ❍ 1 - 5 years
- ❍ 6 - 10 years
- ❍ 11 - 20 years
- ❍ Over 20 years

***3. How long have you been working within your current job role?**
Choose only one of the following

- ❍ Under 1 year
- ❍ 1 - 5 years
- ❍ 6 - 10 years
- ❍ 11 - 20 years
- ❍ Over 20 years

*4. Please indicate your department by selecting the appropriate code from the following list.
Choose only one of the following

Please choose..

*5. Please indicate your staff grade or rank depending on whether you are a Police Staff, a Police Officer, or a Special Officer.
Choose only one of the following

Please choose..

Section 2: Searching Experience

*6. In a normal week, on average, how many hours do you spend searching for information/advice relevant to your work?
Choose only one of the following

- ❍ 1 hour or less
- ❍ 1 - 5 hours
- ❍ 6 - 10 hours
- ❍ 11 - 20 hours
- ❍ Over 20 hours

*7. How much of this time do you think could be saved if you knew where to look for this information/advice?
Choose only one of the following

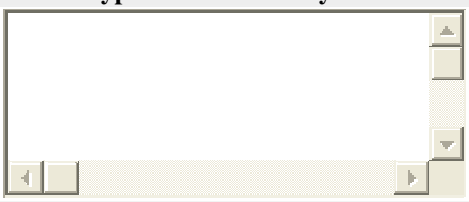
- ❍ 0 - 20%
- ❍ 21 - 40%
- ❍ 41 - 60%
- ❍ 61- 80%
- ❍ 81 - 100%

*8. Within the past year, how many days do you think you have wasted as a direct result of not knowing about information that was available within the company (for example, recreating documents)?
Choose only one of the following

- ❍ None
- ❍ Up to 5 days
- ❍ 6 - 10 days
- ❍ 11 - 20 days
- ❍ Over 20 days

9. In relation to question 8, please state any examples you might have to help us understand the issue.

10. Please state what type of information you find difficult to locate.



***11. Please state in order of frequency of use what sources you utilise when searching for information/advice relevant to your work.**

Click on an item in the list on the left, starting with your highest ranking item, moving through to your lowest ranking item.

Your Choices:	Your Ranking:
Search Engines(Google, Yahoo. MSN Search) References from peers Online databases Intranet (Local Netw ork) Mailing lists Specific w ebsites	1: 2: 3: 4: 5: 6:

Click on the scissors next to each item on the right to remove the last entry in your ranked list

***12. When you search, how often do you find the information you need?**

Choose only one of the following

- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Frequently
- ○ Always

***13. When searching for information that is new to you, how often do you have difficulties identifying resources for new information?**

Choose only one of the following

- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Frequently
- ○ Always

***14. How often do you consult your colleagues for information to aid your query before conducting a search?**

Choose only one of the following

- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Frequently
- ○ Always

***15. How often do you know who to contact when you need information?**
Choose only one of the following

- ❍ Never
- ❍ Rarely
- ❍ Sometimes
- ❍ Frequently
- ❍ Always

***16. Please rank the task of locating information within another department.**
Choose only one of the following

- ❍ Very Easy
- ❍ Easy
- ❍ About Average
- ❍ Difficult
- ❍ Very Difficult

Section 3: Information Sharing Experience

Knowledge is the awareness and understanding of facts, truths, or information gained in the form of experience or learning (Source: Wikipedia).

***17. I widely share information which I consider useful or interesting.**
Choose only one of the following

- ❍ Strongly Agree
- ❍ Agree
- ❍ Neutral
- ❍ Disagree
- ❍ Strongly Disagree

***18. I *only* share on a 'need to know basis' or if I am told to do so.**
Choose only one of the following

- ❍ Strongly Agree
- ❍ Agree
- ❍ Neutral
- ❍ Disagree
- ❍ Strongly Disagree

***19. The knowledge I have is of value to the Force.**
Choose only one of the following

- ❍ Strongly Agree
- ❍ Agree
- ❍ Neutral
- ❍ Disagree
- ❍ Strongly Disagree

***20. The Force fully utilises my knowledge.**
Choose only one of the following

- ❍ Strongly Agree
- ❍ Agree
- ❍ Neutral
- ❍ Disagree
- ❍ Strongly Disagree

***21. Please indicate the most important barrier that hinders you from sharing knowledge.**
Choose only one of the following

Please choose one of the following:

- ❍ If I share knowledge I will lose some of my power
- ❍ I am insecure about the value of my knowledge
- ❍ I lack trust in my colleagues
- ❍ I am afraid of negative consequences
- ❍ If others do not share, why should I?
- ❍ Other (Please enter what stops you from sharing information in the comment box provided)

Please enter your comment here:

Section 4: Individual's usage of different types of mediums

Knowledge is the awareness and understanding of facts, truths, or information gained in the form of experience or learning (Source: Wikipedia).

***22. How often do you utilise the mediums listed below to help generate, organise, share, or build upon the knowledge within your organisation?**
For each medium, please select the answer that best describes your extent of use

	Daily	Weekly	Monthly	6 Monthly	Yearly	Never
Face-to-face	☐	☐	☐	☐	☐	☐
Telephone	☐	☐	☐	☐	☐	☐
Email	☐	☐	☐	☐	☐	☐
Online (e.g MSN messenger)	☐	☐	☐	☐	☐	☐
Memos	☐	☐	☐	☐	☐	☐
Intranet (Local network)	☐	☐	☐	☐	☐	☐
Chatting over lunch/tea breaks	☐	☐	☐	☐	☐	☐

***23. Out of the communication mediums listed below, which one could you NOT do without?**

Choose only one of the following

- ❍ Face-to-face meetings
- ❍ Telephone
- ❍ Email
- ❍ Online (e.g MSN Messenger)
- ❍ Intranet (Local network within the Force)
- ❍ Chatting over lunch/ tea breaks

***24. What do you mostly use email for?**

Choose only one of the following

- ❍ To ask questions
- ❍ To answer questions
- ❍ Both equally

***25. On average, how many knowledge-related emails do you send a day?**

Note: A knowledge-related email is one that engages in the creation, identification, collection, organisation, sharing, adaptation and use of knowledge.

Choose only one of the following

- ❍ 0 - 5 emails
- ❍ 6 - 10 emails
- ❍ 11 - 20 emails
- ❍ 21 - 40 emails
- ❍ More than 40 emails

***26. On average, how many knowledge related emails do you receive a day?**

Note: A knowledge-related email is one that engages in the creation, identification, collection, organisation, sharing, adaptation and use of knowledge.

Choose only one of the following

- ❍ 0 - 5 emails
- ❍ 6 - 10 emails
- ❍ 11 - 20 emails
- ❍ 21 - 40 emails
- ❍ more than 40 emails

Section 5: Your views on Loughborough University's latest research

locate the experts inside the organisation.

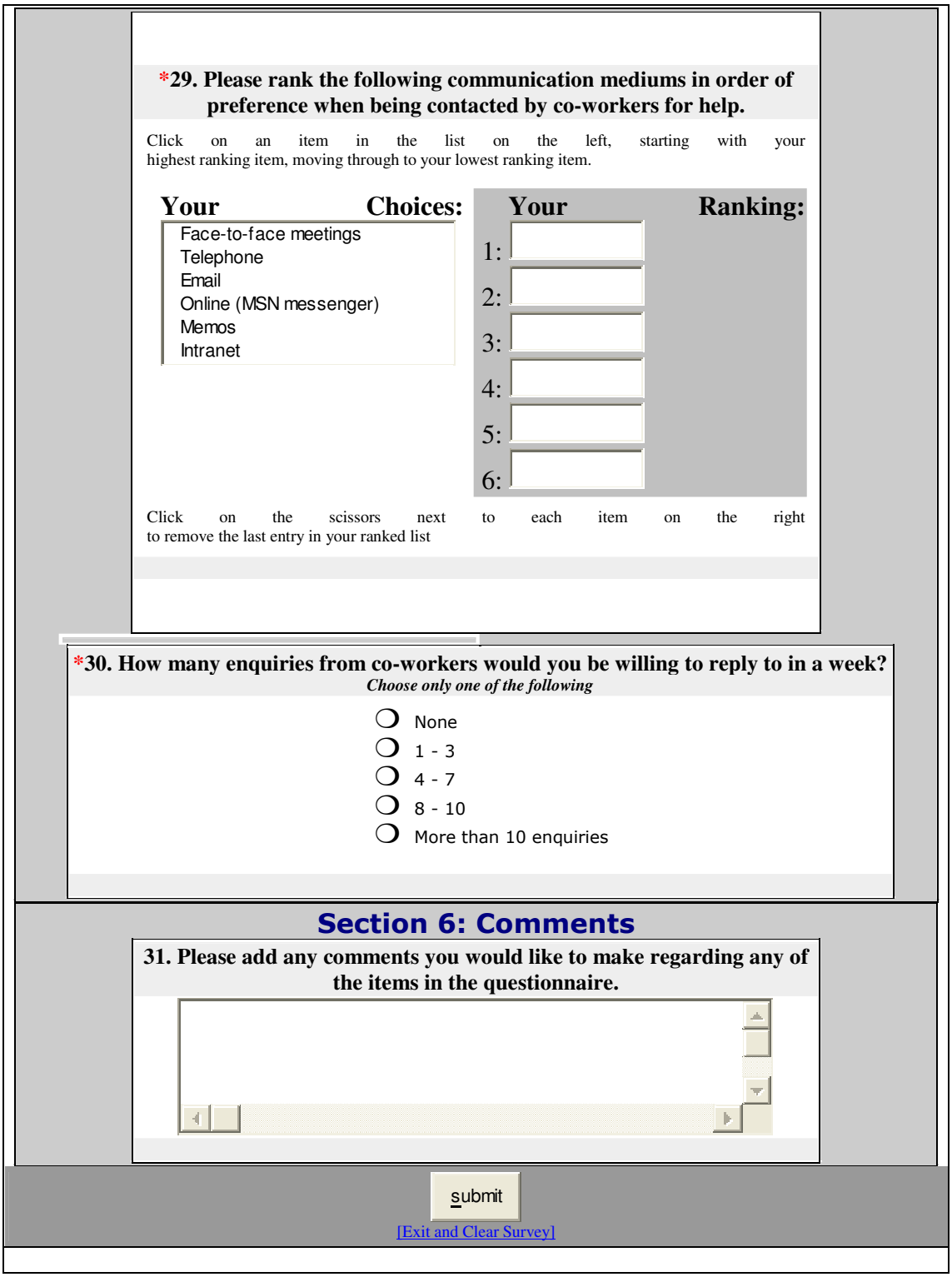
How it works? EKE extracts keyphrases from the body of emails you send. For ethical and privacy reasons, you have the option of authorising whether or not you want the keyphrases extracted from your emails to be published in a database. For example, suppose EKE managed to extract the term "Fraud" from an email you have just sent, EKE will provide you with the options of either publishing and ranking your skill level as being either basic, general, or expert in that area or of discarding that keyphrase. Being a knowledgeable person in that area, you decide to publish the keyphrase "Fraud" and indicate your skill level in it as 'expert knowledge'. A search facility is provided, where a user can enter his query. The return result is a list of individuals in the organisation with knowledge in the area the user is querying. In this way, other employees when conducting a search on "Fraud" can know that you are an expert in that area and can as a result contact you.

***27. The Force would benefit from using this software.**
Choose only one of the following

- ❍ Strongly Agree
- ❍ Agree
- ❍ Neutral
- ❍ Disagree
- ❍ Strongly Disagree

***28. I would be willing to make my interest/ knowledge/ expert areas public to help other employees within the Force.**
Choose only one of the following

- ❍ Strongly Agree
- ❍ Agree
- ❍ Neutral
- ❍ Disagree
- ❍ Strongly Disagree



*29. Please rank the following communication mediums in order of preference when being contacted by co-workers for help.

Click on an item in the list on the left, starting with your highest ranking item, moving through to your lowest ranking item.

Your Choices:

- Face-to-face meetings
- Telephone
- Email
- Online (MSN messenger)
- Memos
- Intranet

Your Ranking:

1:
2:
3:
4:
5:
6:

Click on the scissors next to each item on the right to remove the last entry in your ranked list

*30. How many enquiries from co-workers would you be willing to reply to in a week?

Choose only one of the following

- None
- 1 - 3
- 4 - 7
- 8 - 10
- More than 10 enquiries

Section 6: Comments

31. Please add any comments you would like to make regarding any of the items in the questionnaire.

submit

[Exit and Clear Survey]

II. DIS PRE-STUDY QUESTIONNAIRE

Pre-Study Questionnaire

EKE: A Tool for Finding the Right Experts

Dear Respondent,

All responses to this questionnaire will be treated in ***strict confidence***. The information provided will only be used for the research project purposes. Individuals will not be identified in any results produced. Thank you very much for your time and effort.

Part 1: General Information

Please tick the appropriate box.

1.1 Gender

☐ Male ☐ Female

1.2 Age Group

☐ Under 25 year ☐ 25 – 34 ☐ 35 – 44 ☐ 45 - 54 ☐ 55 and over

1.3 How long have you been working at your current organisation? ……..……

1.4 How frequently do you utilise email to help generate, organise, share and leverage knowledge* within your organisation?

** Knowledge is the awareness and understanding of facts, truths or information gained in the form of experience or learning*

☐ Daily
☐ Weekly
☐ Monthly
☐ 6 Monthly
☐ Yearly
☐ Never

1.5 On average, how many knowledge-related* emails do you send a day?

** A knowledge-related email is one that engages in the creation, identification, collection, organisation, sharing, or adaptation of knowledge*

☐ 0 – 5 emails
☐ 6 – 10 emails
☐ 11 – 20 emails
☐ 21 – 40 emails
☐ More than 40

Part 2: Your first impressions about the system

Email Knowledge Extraction (EKE) is a tool that mines the knowledge contained in employees' emails. EKE automatically discovers interest areas by picking out keyphrases from employees' e-mail messages. For ethical and privacy reasons, each employee has the option of authorising whether or not they want their extracted keyphrases to be published. I will demonstrate how the system operates and ask you a few questions as we go along.

2.1 Is there anything you would like me to clarify or are you happy to proceed?

……..

……..

2.2 What do you think about the concept of EKE?

……..

……..

2.3 On a scale of 1 to 5, how clear do you think this keyphrase ranking form is (5 being very clear, 1 being not clear)? ……..

2.4 Is there any way that you think this keyphrase ranking form can be improved?

☐ Yes ☐ No

If yes, in what way?

……..

……..

Part 3: Comments

Please add any comments you would like to make regarding any aspect of this study..

Thank you very much for your time and effort

III. DIS POST-STUDY QUESTIONNAIRE

Post-Study Questionnaire

EKE: A Tool for Finding the Right Experts

Dear Participant,

Thank you for participating in the Email Knowledge Extraction study (EKE). At this stage, it is important to obtain your feedback about your experience of using the EKE tool, through completing this questionnaire. All responses will be treated in ***strict confidence***. The information provided will only be used for the purposes of the research project. Individuals will not be identified in any results produced. Thank you very much for your time and effort.

Part 1: System Evaluation

Please tick the appropriate box.

1.1 Overall, I would rate the use of the EKE software as ……..

☐ Very Easy ☐ Easy ☐ Neutral ☐ Difficult ☐ Very Difficult

1.2 The concept of the EKE software was …….. to grasp.

☐ Very Easy ☐ Easy ☐ Neutral ☐ Difficult ☐ Very Difficult

1.3 I can still effectively send emails using this tool.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Very Disagree

1.4 I am able to send emails quickly using this tool.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

1.5 I feel comfortable using this tool.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

1.6 It was intuitive to learn how to use this tool.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

1

1.7 The categories "Basic Knowledge", "Working Knowledge", "Expert Knowledge" and "N/A" were appropriate to use for the purpose of rating my level of knowledge in the extracted keyphrases.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

1.8 The layout of the keyphrase ranking form is clear.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

1.9 Do you think that, in general, the idea of extracting keyphrases from emails is worthwhile?

☐ Yes ☐ No

Why? ……..

…….

1.10 Would you like to see the idea of extracting keyphrases applied to email attachments?

☐ Yes ☐ No

1.11 Do you think that this tool would benefit you in searching for experts?

☐ Yes ☐ No

1.12 Do you think that this tool can be improved in anyway?

☐ Yes ☐ No

If 'Yes', how? ……...……..

…….

1.13 Is there anything about this tool that you did dislike?

☐ Yes ☐ No

If 'Yes', what? ……...……..

…….

1.14 Overall, I am satisfied with this software.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

2

Part 2: Evaluation of Extracted Keyphrases

Please look at the list of keyphrases extracted from your emails and categorised by yourself under the categories "Basic Knowledge", "Working Knowledge", and "Expert Knowledge" during the study.

2.1 For each of the expertise levels shown below, please tick how well these keyphrases give an accurate reflection of your work and areas of expertise.

a- Basic Knowledge

- ☐ Poor
- ☐ Fair
- ☐ Satisfactory
- ☐ Good
- ☐ Excellent

b- Working Knowledge

- ☐ Poor
- ☐ Fair
- ☐ Satisfactory
- ☐ Good
- ☐ Excellent

c- Expert Knowledge

- ☐ Poor
- ☐ Fair
- ☐ Satisfactory
- ☐ Good
- ☐ Excellent

d- Overall

- ☐ Poor
- ☐ Fair
- ☐ Satisfactory
- ☐ Good
- ☐ Excellent

3

2.2 Ten randomly selected keyphrases from each of your three knowledge categories are listed below. Please indicate whether or not each of these keyphrases truly reflects your area and level of expertise (i.e. "Basic Knowledge", "Working Knowledge", and "Expert Knowledge").

	Keyphrases Extracted		
Basic Knowledge	1-	☐ Yes	☐ No
	2-	☐ Yes	☐ No
	3-	☐ Yes	☐ No
	4-	☐ Yes	☐ No
	5-	☐ Yes	☐ No
	6-	☐ Yes	☐ No
	7-	☐ Yes	☐ No
	8-	☐ Yes	☐ No
	9-	☐ Yes	☐ No
	10-	☐ Yes	☐ No
Working Knowledge	1-	☐ Yes	☐ No
	2-	☐ Yes	☐ No
	3-	☐ Yes	☐ No
	4-	☐ Yes	☐ No
	5-	☐ Yes	☐ No
	6-	☐ Yes	☐ No
	7-	☐ Yes	☐ No
	8-	☐ Yes	☐ No
	9-	☐ Yes	☐ No
	10-	☐ Yes	☐ No
Expert Knowledge	1-	☐ Yes	☐ No
	2-	☐ Yes	☐ No
	3-	☐ Yes	☐ No
	4-	☐ Yes	☐ No
	5-	☐ Yes	☐ No
	6-	☐ Yes	☐ No
	7-	☐ Yes	☐ No
	8-	☐ Yes	☐ No
	9-	☐ Yes	☐ No
	10-	☐ Yes	☐ No

4

2.3 Given the choice of extracting keyphrases from email either manually or using EKE, which one would you choose and why?

☐ Manually ☐ Using EKE

Why? ……..

…….

Part 3: Socio-ethical Challenges

3.1 To what extent do you think that the use of the EKE threatens employee privacy?

☐ To no extent
☐ To little extent
☐ To some extent
☐ To a moderate extent
☐ To a great extent

Please clarify: ……

……

3.2 I think that the current organisational privacy polices provide me with sufficient protection against the potential privacy problems that might result from the use of EKE.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

Please clarify: …….

…….

3.3 I feel that the current legislation in the UK offers me sufficient protection against the potential privacy problems that might result from the use of EKE.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

Please clarify: …….

…….

5

3.4 How comfortable do you feel about having the emails you send analysed by EKE for the purpose of identifying your areas of expertise?

- ☐ Extremely comfortable
- ☐ Comfortable
- ☐ Neutral
- ☐ Uncomfortable
- ☐ Extremely uncomfortable

Please clarify: ……

……

3.5 EKE enables employees to select the keyphrases they would like to share with the rest of the organisation. To what extent do you think this addresses privacy issues?

- ☐ To no extent
- ☐ To little extent
- ☐ To some extent
- ☐ To a moderate extent
- ☐ To a great extent

Please clarify: ……

……

3.6 Based on the user's search query, EKE identifies experts in the organisation who might be able to provide the user with answers to their query. When you are identified via EKE as an expert on a particular subject, to what extent are you willing to share your expertise with others when requested?

- ☐ To no extent
- ☐ To little extent
- ☐ To some extent
- ☐ To a moderate extent
- ☐ To a great extent

Please clarify: ……

……

6

3.7 To what extent do you think that providing recognition and incentives is a key motivator for employees to use EKE and share knowledge?

- ☐ To no extent
- ☐ To little extent
- ☐ To some extent
- ☐ To a moderate extent
- ☐ To a great extent

Please clarify: ……

……

3.8 The search facility of EKE returns a list of experts whom you may/may not know. How comfortable do you feel seeking information from experts you do NOT know?

- ☐ Extremely comfortable
- ☐ Comfortable
- ☐ Neutral
- ☐ Uncomfortable
- ☐ Extremely uncomfortable

Please clarify: ……

……

3.9 How many enquiries generated through EKE from co-workers would you be willing to reply to in a week?

- ☐ None
- ☐ 1 – 3 enquiries
- ☐ 4 – 7 enquiries
- ☐ 8 – 10 enquiries
- ☐ More than 10 enquiries

Please clarify: ……

……

3.10 The EKE system may incorrectly classify a user as an expert in a particular topic. How comfortable do you feel about that?

- ☐ Extremely comfortable
- ☐ Comfortable
- ☐ Neutral
- ☐ Uncomfortable
- ☐ Extremely comfortable

Please clarify: ……

……

3.11 To what extent do you think that EKE will negatively affect your email usage?

- ☐ To little or no extent
- ☐ To some extent
- ☐ To a moderate extent
- ☐ To a great extent
- ☐ To a very great extent

Please clarify: ……

……

Part 4: Comments

Pease add any comments you would like to make regarding any aspect of this study.

Thank you very much for your time and effort

8

IV. FOCUS GROUPS QUESTIONNAIRE

Questionnaire

EKE: A Tool for Finding the Right Experts

Dear Participant,

Thank you for participating in the Email Knowledge Extraction study (EKE). All responses to this questionnaire will be treated in ***strict confidence***. The information provided will only be used for the purposes of the research project. Individuals will not be identified in any results produced. Thank you very much for your time and effort.

Part 1: General Information

Please tick the appropriate box.

1.1 Gender

☐ Male ☐ Female

1.2 Age Group

☐ Under 25 year ☐ 25 – 34 ☐ 35 – 44 ☐ 45 - 54 ☐ 55 and over

1.3 How long have you been working at your current organisation? ……..……

1.4 How frequently do you utilise email to help generate, organise, share and leverage knowledge* within your organisation?

** Knowledge is the awareness and understanding of facts, truths or information gained in the form of experience or learning*

☐ Daily
☐ Weekly
☐ Monthly
☐ 6 Monthly
☐ Yearly
☐ Never

1.5 On average, how many knowledge-related* emails do you send a day?

** A knowledge-related email is one that engages in the creation, identification, collection, organisation, sharing, or adaptation of knowledge*

☐ 0 – 5 emails
☐ 6 – 10 emails
☐ 11 – 20 emails
☐ 21 – 40 emails
☐ More than 40

1

Part 2: System Evaluation

2.1 Overall, I would rate the use of the EKE software as ……..

☐ Very Easy ☐ Easy ☐ Neutral ☐ Difficult ☐ Very Difficult

2.2 The concept of the EKE software was …….. to grasp.

☐ Very Easy ☐ Easy ☐ Neutral ☐ Difficult ☐ Very Difficult

2.3 I would feel comfortable using this tool.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

2.4 It was intuitive to learn how to use this tool.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

2.5 The categories "Basic Knowledge", "Working Knowledge", "Expert Knowledge" and "N/A" are appropriate to use for the purpose of rating my level of knowledge in the extracted keyphrases.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

2.6 The layout of the keyphrase ranking form is clear.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

2.7 Do you think that this tool would benefit you in searching for experts?

☐ Yes ☐ No

Why: ……...

……

Part 3: Socio-ethical Challenges

3.1 To what extent do you think that the use of the EKE tool threatens employee privacy?

☐ To no extent
☐ To little extent
☐ To some extent
☐ To a moderate extent
☐ To a great extent

Please clarify: …….

…….

2

3.2 I think that the current organisational privacy polices provide me with sufficient protection against the potential privacy problems that might result from the use of EKE.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

Please clarify: …….

…….

3.3 I feel that the current governmental legislation offers me sufficient protection against the potential privacy problems that might result from the use of EKE.

☐ Strongly Agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly Disagree

Please clarify: …….

…….

3.4 How comfortable do you feel about having the emails you send analysed by EKE for the purpose of identifying your areas of expertise?

☐ Extremely comfortable
☐ Comfortable
☐ Neutral
☐ Uncomfortable
☐ Extremely uncomfortable

Please clarify: …….

…….

3.5 EKE enables employees to select the keyphrases they would like to share with the rest of the organisation. To what extent do you think this addresses privacy issues?

☐ To no extent
☐ To little extent
☐ To some extent
☐ To a moderate extent
☐ To a great extent

Please clarify: …….

…….

3

3.6 Based on the user's search query, EKE identifies experts in the organisation who might be able to provide the user with answers to their query. When you are identified via EKE as an expert on a particular subject, to what extent are you willing to share your expertise with others when requested?

- ☐ To no extent
- ☐ To little extent
- ☐ To some extent
- ☐ To a moderate extent
- ☐ To a great extent

Please clarify: …….

…….

3.7 To what extent do you think that providing recognition and incentives is a key motivator for employees to use EKE and share knowledge?

- ☐ To no extent
- ☐ To little extent
- ☐ To some extent
- ☐ To a moderate extent
- ☐ To a great extent

Please clarify: …….

…….

3.8 The search facility of EKE returns a list of experts whom you may/may not know. How comfortable do you feel seeking information from experts you do NOT know?

- ☐ Extremely comfortable
- ☐ Comfortable
- ☐ Neutral
- ☐ Uncomfortable
- ☐ Extremely uncomfortable

Please clarify: …….

…….

3.9 How many enquiries generated through EKE from co-workers would you be willing to reply to in a week?

- ☐ None
- ☐ 1 – 3 enquiries
- ☐ 4 – 7 enquiries
- ☐ 8 – 10 enquiries
- ☐ More than 10 enquiries

Please clarify: …….

…….

4

3.10 The EKE system may incorrectly classify a user as an expert in a particular topic. How comfortable do you feel about that?

- ☐ Extremely comfortable
- ☐ Comfortable
- ☐ Neutral
- ☐ Uncomfortable
- ☐ Extremely comfortable

Please clarify: …….

…….

3.11 To what extent do you think that EKE will negatively affect your email usage?

- ☐ To no extent
- ☐ To little extent
- ☐ To some extent
- ☐ To a moderate extent
- ☐ To a great extent

Please clarify: …….

…….

Part 4: Comments

Pease add any comments you would like to make regarding any aspect of this study.

Thank you very much for your time and effort

5

V. FOCUS GROUP DISCUSSION AND QUESTIONNAIRE ANALYSIS

Part1: Focus Group Discussion

Strengths of EKE

1- General Perceptions	
Group A	EKE is "*well designed*" and "*easy to use*". This can be attributed to EKE's integration with existing technologies and work practices which means that little extra work is required. "*I don't have to do more than just write an email*", explained one of the participants.
Group B	EKE is "*well designed*" and "*easy to use*", which greatly aids employees to find the right experts. EKE's analysis of emails is executed behind the scenes and as such, "*does not impact the real purpose of sending an email*". The participants' organisation currently employs a web based system where they manually enter their expertise areas. EKE's close integration with email provides a better alternative to the system they currently employ. It is "*definitely far preferable than the system we have, where you have to use a website, which nobody ever uses*". EKE would "*encourage people and give them an opportunity to upgrade their profiles*". And in general, with tools such as EKE, "*people do contribute. At the end of the day, these things count towards their annual appraisals*".

2- User Involvement	
Group A	The level of user involvement in the selection and categorisation of specific keyphrases is appropriate. It was quoted, "*from my point, self categorisation of the level of knowledge is one strength*". This argument was based on the reasoning that self rating leads to more accurate profiles. In addition, this level of user involvement leads to the added benefit of activity recording. "*I think EKE has a very interesting side effect when we talk about activity recording*", added one of the participants.
Group B	EKE does not "*require a list of key terms, instead it relies on natural language processing*" where upon users select whether they wish to "*contribute or not*". This is achieved with "*minimal amount of user involvement to store information*".

3- Technological Capabilities	
Group A	The use of email as a source for expertise profiling helps promote the adoption of EKE in large well established organisations. It helps minimise the investment in new technologies and the training required by employees. It also helps decrease the resistance to change inherent in organisations when adopting new technologies. The benefits of exploiting email can only be realised if the integration of the exploiting tool with the organisation's existing email structure is seamless and hassle free. EKE's "*plugin is easy to install*". One of the main strengths of EKE is the pre-processing of the keyphrases prior to presentation to the user, which includes: 1) The ability to recognise compound words (e.g. "knowledge management" is extracted instead of "knowledge" and "management" separately). 2) The domain independent approach adopted by the system which does not limit the returned keyphrases to a predefined list of knowledge areas. 3) EKE's extraction approach, only promotes users to categorise a keyphrase once, after which the frequency is automatically incremented for the instance of that keyphrase each time it is identified. The knowledge categories "BK", "WK", and "EK" are "*accurate*", "*none confusing*", and "*suitable*" for categorising knowledge areas.
Group B	EKE's close "*integration with existing technologies*" and its potential flexibility to integrate with different email systems helps overcome many hurdles faced by existing expertise locators. EKE's domain independent approach is appropriate, as the retuned keyphrases are not limited to a predefined list of knowledge areas. Another strength is the system's approach of dealing with the extracted keyphrases. Only keyphrases that have not previously been stored in the database for the user will be returned back for approval and categorisation. Keyphrases contained in the email chain are often relevant to the users' areas of interest/ knowledge and as such should be included in the analysis.

Opportunities EKE provides

4- Individual Opportunities	
Group A	EKE will facilitate the task of finding experts faster. This leaves users with more time to focus on accomplishing their tasks, resulting in increased productivity. "*It would enable me to find the experts faster especially in larger organisations*" is an example of one of the comments made.
Group B	EKE will help find experts faster. As such, it will "*aid in solving problems very quickly*". This may lead to an increase in productivity and free up more time for other work related aspects.

5- Organisational Opportunities	
Group A	EKE will help enhance communications/networking within organisations. It is "*easy for networking*" and "*hopefully will help create a more effective communication environment*". This will lead to "*improved knowledge management within the organisation*". EKE will provide employees with greater access to where the knowledge lies.
Group B	EKE provides a "*good opportunity for teams to share information and knowledge*". This can lead to better communications and knowledge management within the organisation.

Weaknesses of EKE

6- Knowledge Categories	
Group A	The inflexibility of the knowledge categories to change once the keyphrase is categorised and stored was seen as problematic. This is because, over a period of time knowledge levels may change (i.e. increase or decrease), while the profiled knowledge levels of the user stay the same. If I have "*basic knowledge for a term and over time my knowledge and usage of the word increases to ten times a day or 20 times a day, I think I am no longer in the basic knowledge level. However, the system does not cater for this*".

7- General Perceptions	
Group A	Some participants argued that EKE is more suitable for medium to large organisation and "*may not be necessary in small organisations*" where employees have a smaller communication network and a greater familiarity of colleagues work activities. The initial phase when EKE is first adopted is a crucial stage in its successful adoption. One participant exclaimed, "*people might hate it in the initial phase, and there will be a big hole, because if people don't use it initially then they will stop using it*". This is where training can play an important part.
Group B	The ability to edit keyphrases once the system has stored them (after the user categorisation phase) was felt to be a desirable feature which the system does not currently offer. However, this additional functionality is easy to implement and could be undertaken in future work. Moreover, prior to the user categorisation stage, a returned keyphrase may be misspelled and it would be desirable to be able to modify the keyphrase.

8- Linguistics	
Group A	From the group discussion, a predominant linguistic weakness emerged. This weakness stems from a common natural language problem of when a keyphrase is extracted out of context of its neighbouring sentences. This can make it impossible or very difficult for the person viewing these keyphrases to determine the context of the keyphrase. The suggested inclusion of ontology generated a lot of discussion. It was thought that through the use of an "*ontology*" and "*statistical significance assignment*" the contextualisation problem could be reduced. However, as discussed before, some felt the ontology free approach EKE currently employs is preferable even with the contextualisation problem. Furthermore, another context related issue was highlighted. This was related to the exclusion by EKE of company terms which look like common English words.
Group B	Some argued that the keyphrases contained in the email chain are often not relevant to the users' areas of interest/knowledge and as such should not be included in the analysis.

9- Impact on Performance	
Group A	From the organisational perspective, employee performance is a key factor behind adopting EKE. It was raised by one of the participants that EKE should have little if not any negative impact on performance.
Group B	EKE minimises the interruptions and the intrusiveness to the user by checking if the keyphrase exists in the database before prompting the user for approval. In spite of this, it was still felt there was "*some intrusiveness*" and an "*added step*" which may have a negative impact on performance. In a large organisation, where tens of thousands of employees may use the system, even if a small number of keyphrases are extracted per user on a daily basis, the "*network congestion*" may increase substantially and so may the "*database size*".

10- Confidentiality	
Group B	Financial organisations have to comply with stringent confidentiality legislations, and as such information can be confidential to one particular group within the organisation. It was suggested that EKE's openness would pose significant adoption challenges in such institutions.

Threats EKE arises

11- Individual	
Group A	EKE may increase users' fear of being watched and of the organisation collating a record of their activities. Creativity and openness of employees often relies on them feeling free to try new things. If they feel they are being monitored and their activities recorded, this "*creativity and openness*" may fall or be lost altogether. EKE's self selection of keyphrases overcomes this concern to a certain degree.
Group B	The primary objective of any employee is to get their work done. Because of the keyphrase categorisation stage, people might get distracted with what they are doing. As such, the level of intrusion should be reduced to the absolute minimum required.

12- Organisational	
Group A	One potential downside which may arise if the user is continuously prompted to categorise their expertise for every email is the reduction of productivity. This may cause a general downturn for the organisation as a whole. Moreover, an external threat is the risk of outsiders gaining access to the data stored by the system. The problem is, "*if someone gets the data out, they can get so much out of it*". "*Outdated knowledge*" poses a potential internal threat to the organisation. This may lead to employees wasting time looking for colleagues who have since moved on or no longer have knowledge on a particular topic.
Group B	One particular concern is how secure the system is in terms of "*what is transmitted*", and "*how secure it is to external attacks*". Of particular concern is the possibility of malicious use, both by the organisation and outsiders. It would be "*a threat to the organisation if an unauthorised person is able to access the database*". "*Poaching staff*" was another concern highlighted by some of the participants.

Improvements that could be made to EKE

13- Keyphrase Extraction	
Group A	As time passes, our knowledge levels fluctuate from basic through to expert and vice versa. EKE as it stands today takes a fixed view of users' knowledge levels. Once classified, the level can not be altered. EKE could be further extended by the adoption of ontology. The opinion on this issue was divided as some were in favor of the ontology free approach.
Group B	When EKE is first setup for a particular user, it was recommended that "*somewhere in the profile the user should be able to say what their areas of expertise are. For example, I should be able to specify whether I am a technical person or business analyst or somebody with financial knowledge, so that EKE will have an idea about my profile and only extract keyphrases relevant to my knowledge area. This will reduce the surplus keyphrases returned*". Additionally, pre-feeding the system with terms specific to the organisation was seen as "*helpful*" in reducing the number of keyphrases returned. An added benefit of pre-feeding is the ability to add new terms when new and novel ideas arise.

14- System Adoption	
Group A	The introduction of EKE should be staged to avoid overwhelming the users and encourage adoption. "*Few emails or email communications should be processed by EKE so that people can become familiar with it, without the fear that everything is processed from the beginning*". Furthermore, to encourage adoption, training should be provided to employees on how the system works, how to use the system, and the benefits incurred from using the system. Additionally, EKE's level of surveillance should be made clear from the outset. Because EKE stores personal data about its users, the legal implications need to be taken into account. This is further exacerbated by large geographically distributed organisations that may be operating in different countries where the laws can differ. "*Research into the legal implications*" was seen as an area for further research.
Group B	Many large established organisations have a pre-existing email infrastructure. For EKE to be successful, it would be desirable if "*it could integrated and customised according to the needs of certain clients*". A user may incorrectly categorise a keyphrase or the categorisation may change naturally over time. Currently, the system does not permit the user profile to be modified. This can be a beneficial extension to the existing system. Because of the nature of work undertaken in certain organisations, where information is private within subgroups of the organisation, it was felt that it would be desirable if EKE's users profiles could have different levels of access (i.e. global profile for the whole organisation and private profile to a subset of the organisation). Regardless of how good any technology is, users can often feel overwhelmed. EKE is no exception and as such, for successful adoption, training should be undertaken before rolling out EKE across the organisation.

15- Configuration	
Group A	The ability to limit the analysis of EKE to business emails only and exclude all private email communications is needed. To reduce the number of keyphrases returned to the user and to increase the accuracy of the tool, only phrases which have been identified multiple times by EKE should be returned to the user for categorisation. It was exclaimed, "*I think if you have a threshold say ten for example, a word has to be identified ten times before the term would be returned*". Alternately, the user could "*decide when*" to categorise the extracted keyphrases.
Group B	EKE could be made more flexible by allowing users to configure the system at anytime by choosing whether they want to contribute, how often they contribute, and when they want to categorise the extracted keyphrases. These points are illustrated in comments such as: "*I think it should be configurable by every user, if he or she wants to contribute or not*" and "*Users should be able to configure when they want this popup. For example, I should be able to configure EKE so that I am able categorise my knowledge areas on Friday evenings, when I am bored and tired or some other time within the week*".

16- Extra Functionality	
Group A	The self selection and categorisation by employees can be easily extended to allow activity recording. Activity recording is a task where employees are asked to record the work they are performing for a set period. The information recorded by EKE can be used for this purpose with little amendment and no added workload.
Group B	Instant messaging is a widely used communication medium within organisations and as such would be an interesting and useful extension to EKE. Additionaly, EKE could be extended to make use of the calendar functionality in Outlook. In outlook you can do calendar appointments for business training sessions and you can invite the attendants to get added as experts.

Part2: Questionnaire Results

Table 1: Demographic Information

Group A	**Gender**	**Male**			**Female**	
		100 %			0 %	
	Age Group (in years)	**Under 25**	**25 - 34**	**35 - 44**	**45 - 54**	**Over 55**
		40 %	40 %	20 %	0 %	0 %

Group B	**Gender**	**Male**			**Female**	
		100 %			0 %	
	Age Group (in years)	**Under 25**	**25 - 34**	**35 - 44**	**45 - 54**	**Over 55**
		28.6 %	71.4 %	0 %	0 %	0 %

Table 2: Frequency of Email Usage

Group A	**Frequency of email usage within the organisation.**	**Daily**	**Weekly**	**Monthly**	**Six Monthly**	**Yearly**	**Never**
		100%	0%	0%	0%	0%	0%

Group B	**Frequency of email usage within the organisation.**	**Daily**	**Weekly**	**Monthly**	**Six Monthly**	**Yearly**	**Never**
		57.1%	28.6%	14.3%	0%	0%	0%

Table 3: Email Use

Group A	**Number of knowledge related emails sent a day**	**0-5**	**6-10**	**11-20**	**21-40**	**More than 40**
		40 %	20 %	0 %	40 %	0 %

Group B	**Number of knowledge related emails sent a day**	**0-5**	**6-10**	**11-20**	**21-40**	**More than 40**
		100 %	0 %	0 %	0 %	0 %

Table 4: System Evaluation

Group A	I would rate the use of the EKE as	**Very Easy**	**Easy**	**Neutral**	**Difficult**	**Very Difficult**
		20%	80%	0%	0%	0%
	The EKE concept was _____ to grasp	**Very Easy**	**Easy**	**Neutral**	**Difficult**	**Very Difficult**
		40%	40%	20%	0%	0%
	I feel comfortable using EKE	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		0%	60%	20%	20%	0%
	EKE was intuitive to learn	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		20%	60%	20%	0%	0%
	The knowledge categories are appropriate	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		20%	0%	60%	20%	0%
	The layout of the keyphrase categorisation form is clear	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		20%	0%	80%	0%	0%

Group B	I would rate the use of the EKE as.	**Very Easy**	**Easy**	**Neutral**	**Difficult**	**Very Difficult**
		14.3%	85.7%	0%	0%	0%
	The EKE concept was _____ to grasp.	**Very Easy**	**Easy**	**Neutral**	**Difficult**	**Very Difficult**
		57.1%	42.9%	0%	0%	0%
	I feel comfortable using EKE.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		14.3%	71.4%	14.3%	0%	0%
	EKE was intuitive to learn.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		28.6%	71.4%	0%	0%	0%
	The knowledge categories are appropriate.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		42.9%	28.6%	28.6%	0%	0%
	The layout of the keyphrase categorisation form is clear.	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
		28.6%	57.1%	14.3%	0%	0%

Table 5: Group A Socio-ethical challenges

Group A	The use of EKE threatens employee privacy.				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	0%	0%	20%	60%	20%
	Organisational privacy policies provide sufficient protection.				
	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	20%	80%	0%	0%	0%
	The governmental legislation provide sufficient protection.				
	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	0%	50%	0%	50%	0%
	How comfortable do you feel having the emails analysed by EKE for the purpose of identifying areas of expertise?				
	Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
	0%	60%	20%	20%	0%
	EKE enables employees to select keyphrases they would like to share with others. To what extent do you think this addresses privacy issues?				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	25%	0%	50%	25%	0%
	To what extent are you willing to share your expertise with others when requested				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	0%	0%	40%	40%	20%
	Extent to which you think that providing recognition and incentives is a key motivator to use EKE and share knowledge				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	20%	0%	40%	20%	20%
	How comfortable do you feel seeking information from experts you do NOT know?				
	Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
	20%	60%	20%	0%	0%
	How many enquiries willing to reply to in a week?				
	None	**1 – 3**	**4 - 7**	**8 – 10**	**More than 10**
	0%	20%	20%	0 %	60%
	How comfortable do you feel about EKE incorrectly classifying a user as an expert on a particular topic				
	Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
	0%	0%	80%	20%	0%
	To what extent do you think EKE will negatively affect you email usage				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	60%	20%	0%	20%	0%

Table 6: Group B Socio-ethical challenges

Group B	The use of EKE threatens employee privacy.				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	0%	28.6%	57.1.%	14.3%	0%
	Organisational privacy policies provide sufficient protection.				
	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	28.6%	42.9%	28.6%	0%	0%
	The governmental legislation provide sufficient protection.				
	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
	14.3%	57.1%	28.6%	0%	0%
	How comfortable do you feel having the emails analysed by EKE for the purpose of identifying areas of expertise?				
	Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
	28.6%	71.4%	0%	0%	0%
	EKE enables employees to select keyphrases they would like to share with others. To what extent do you think this addresses privacy issues?				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	14.3%	14.3%	42.9%	0%	28.6%
	To what extent are you willing to share your expertise with others when requested				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	0%	0%	20%	42.9%	57.1%
	Extent to which you think that providing recognition and incentives is a key motivator to use EKE and share knowledge				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	0%	14.3%	28.6%	14.3%	42.9%
	How comfortable do you feel seeking information from experts you do NOT know?				
	Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
	14.3%	85.7%	0%	0%	0%
	How many enquiries willing to reply to in a week?				
	None	**1 – 3**	**4 - 7**	**8 – 10**	**More than 10**
	0%	28.6%	42.9%	28.6 %	0%
	How comfortable do you feel about EKE incorrectly classifying a user as an expert on a particular topic				
	Extremely comfortable	**Comfortable**	**Neutral**	**Uncomfortable**	**Extremely uncomfortable**
	0%	28.6%	28.6%	42.9%	0%
	To what extent do you think EKE will negatively affect you email usage				
	To no extent	**To little extent**	**To some extent**	**To moderate extent**	**To a great extent**
	28.6%	28.6%	0%	0%	42.9%

APPENDIX G RESEARCH PUBLICATIONS

Jackson, T.W. & **Tedmori, S.** (2004)."Capturing and Managing Electronic Knowledge: The Development of the Email Knowledge Extraction", *The International Resource Management Association Conference*, Idea Group, New Orleans, USA, May 2004, pp. 463-466.

Tedmori, S., Jackson, T.W., Bouchlaghem, D., Adelmann, H. and Nagaraju, R. (2006). "Building a Tool for Expertise Discovery", *Information Resources Management Association Conference*, ed. M. Khosrow-Pour, Idea Group Publishing, Washington, pp.1053.

Tedmori, S., Jackson, T.W., Bouchlaghem, N.M. and Nagaraju, R.(2006), "Expertise Profiling: Is Email Used to Generate, Organise, Share or Leverage Knowledge", *Proceedings of the 11th International Conference on Computing in Civil and Building Engineering*, ed. H. Rivard, Quebec, Canada, pp. 179.

Tedmori, S., Jackson, T.W. and Bouchlaghem, D., "Locating Knowledge Sources Through Keyphrase Extraction", *Knowledge and Process Management*, vol. 13, no. 2, pp. 100-107.

Tedmori, S., Jackson, T.W. and Bouchlaghem, N.M. (2007), "The Email Knowledge Extraction (EKE) System: Discovering the Limitations of Natural Language Processing", *Proceedings of the 15th International Software Quality Management*, Stafford, UK.

Tedmori, S., Jackson, T.W. and Bouchlaghem, N.M. (2007), "Optimising the Email Knowledge Extraction System to Support Knowledge Work", *Proceedings of the 15th European Conference on Information Systems*, St Gallen, Switzerland, pp. 681-691.

Tedmori, S., Jackson, T.W., Bouchlaghem, N.M. and Newcombe, M. (2007), "Information Seeking and Sharing Behaviour of a UK Police Force", *Proceedings of the Eighth European Conference on Knowledge Management*, Barcelona, Spain, pp. 982 - 988.

Tedmori, S., Lichtenstein, S. and Jackson, T.W. (2007), "Expertise Location Using Keyphrases in Electronic Mail: Socio-ethical Challenges", *Proceedings of the Australian Conference for Knowledge Management & Intelligent Decision Support*, Melbourne, Australia.

Lichtenstein, S., **Tedmori, S.,** and Jackson, T.W., “Socio-ethical Issues for Expertise Location from Electronic Mail”, Paper accepted for publication in the International Journal of Knowledge and Learning.

AstraZeneca R&D Charnwood
Bakewell Road
Loughborough
Leicestershire
LE11 5RH

Centre for Innovative and Collaboartive Engineering (CICE)
Loughborough University
Loughborough
Leics, LE11 3TU