IDEA GROUP PUBLISHING

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

ITP4991

Capturing and Managing Electronic Knowledge: The Development of the Email Knowledge Extraction System

Thomas Jackson
Computer Science Department, Loughborough University, UK, t.w.jackson@lboro.ac.uk

Sara Tedmori
Computer Science Department, Loughborough University, UK

ABSTRACT

"Knowledge management" has become the buzzword in the press in recent times and although there has been a strong emphasis on the advantages of knowledge management in organisations, applications developed to take advantage of knowledge within emails are still very limited. This paper details the development and implementation of a graphical online email knowledge-based system called EKE, that captures the knowledge within electronic messages.

1. INTRODUCTION

As Francis Bacon said, "Knowledge is power". The power of knowledge lies in its being a very powerful resource for preserving valuable heritage, for exploring and understanding, for finding and correlating answers, for creating core competences, and for initiating new situations for both individuals and organisations now and in the future. However, the term knowledge brings with it a set of semantic problems since it has many meanings depending upon the speaker and the listener's interpretation of the context within which it is spoken and heard. There is still no consensus on a suitable definition of the term. However, from a business perspective, knowledge provides guidance to humans so that they can formulate decisions to do their work. Knowledge can be categorised in two major types, Tacit and Explicit. Tacit knowledge (know how) is what exists in a person's mind. Explicit knowledge (know that) is what has been codified or documented in an artefact of some type outside the human being (Polanyi 1967; Ryle 1949; Nonaka, 1994).

It is generally recognised that communities perform better when all members adopt certain behaviours, such as sharing their knowledge or making effective use of the knowledge produced by others. Unfortunately, the embracing of such behaviours is not easy for various reasons. For example, some people might not be willing to share their knowledge; others might not have access or know how to use tools that make the sharing of knowledge possible. Organisations might not have educated their employees properly about the benefits of sharing their knowledge (Roda, et. al., 2003).

Electronic mail (email) is a fast, effective and efficient means of communication. The introduction of email dates back to the late 1971 when a computer engineer named Ray Tomlinson sent the first email message. The number of emails sent on an average day was expected to hit 14 billion in 2002 and by 2005, this amount is expected to triple to 35 billion emails sent each day. 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 (Jackson, Dawson and Wilson, 2003).

Email is part of our daily lives and it has achieved a wide spread acceptance for personal communication and become a fundamental prerequisite for doing any kind of business (Bertolotti and Calzarossa, 2001). One of the basic features of electronic messaging in general is that it aids easy access to the needed information which empowers employees to make more informed decision and foster innovative thinking which is essential to the success of any organisation. 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 become documents and records with legal requirements and restrictions, and the data that they contain is not made use of (Abaza, 2003). Unfortunately, the majority of current employee interaction with email is far from efficient and 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 both the sender and receiver and is therefore not fully utilised within an organisation. Both public and private organisations are quickly discovering that they have an obligation to apply the appropriate preservation for the knowledge in emails to make use of the knowledge that reside in electronic messages and an equal obligation towards their employees personal privacy. This is when the need for a solid knowledge management strategy becomes essential and its benefits lie in the ability to try to capture and manage the knowledge in emails and deliver it more accurately and consistently to the person who needs it, for the purposes of exploring and understanding information, finding and correlating answers without sacrificing privacy.

Capturing messaging-based knowledge and managing information for its own sake adds little, if any, value to organisations. The value added comes when knowledge is applied in order to improve, change or develop specific tasks and activities by making this knowledge accessible from other applications, ensuring compliance with legal regulations. As detailed within in this paper, the authors have developed an application that should increase the rate of knowledge sharing between employees by taking advantage of all the information that reside in emails for the benefit of the company as a whole.

2. Turning Email into Knowledge

Knowledge management is an area that has attracted the interest of many academics and organisations. In its broadest sense, knowledge management (KM) is the ability to control knowledge for achieving organisational goals. Organisations are quickly discovering that their power lies in the intellectual capital of their employees (Rubenstein-Montano, et. al., 2001). If organisations succeed in utilizing this intellectual capital effectively, they can increase their intangible assets such as expertise, skills, information and knowledge.

Organisations today are making every effort to improve their products and services while controlling costs. They always try to find better ways to perform complex tasks. However, performing complex tasks requires organisations to know their experts and specialists. Unfortunately, the knowledge is almost always difficult to find and is frequently known only by an expert and not by an electronic information system. The challenge is to acknowledge and discover the organisation's experts and specialists so the company can make good use of their knowledge and make it available to others as well. However, to find the organisational experts is not based just around technology,

but through providing a corporate culture that enables employees to recognise and believe that sharing knowledge and using it wisely will contribute to the organisations success. Employees should be encouraged to learn form each other. For information sharing not only increases productivity and minimises costs, but can also improve the employees' ability to react to varying situations (Bumper Brain Technologies, 2003)

Currently only few organisations exploit their firm's email content assets and only few software applications exist to make this possible. Tacit Knowledge Systems Inc. is a firm with a product that purports to transform enterprise email into a shared knowledge resource. The company's KnowledgeMail product scans and organises messages according to user-defined profiles, key concepts and phrases, and virtual community recognition. (Willen, 2003). KnowledgeMail only allows people in a company to store their public profile of expertise. So when the user types in their request, results which are contact details are shown to the user and are ordered by the strength of the match to the query. However, the strength of the match is determined by the frequency, intensity, and history of the topic within the person's expertise profile. KnowledgeMail is not free and it does not make use of the knowledge within electronic messages, but takes a more subjective approach through soliciting what the users thinks their expertise are. KnowlegdeMail only allows people to add or modify their expertise profile so other people in the organisation can contact them.

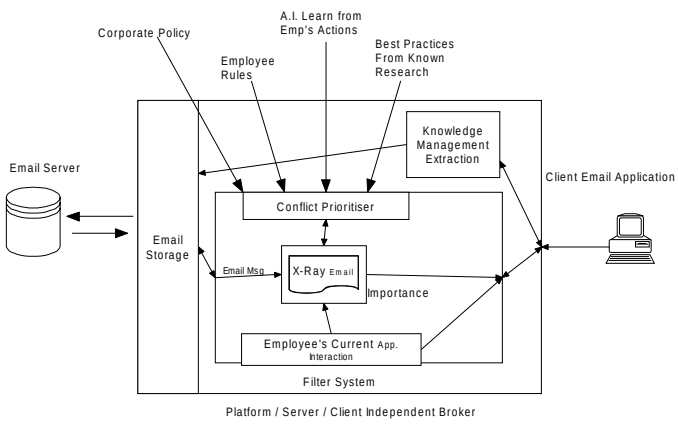
From 2000 to 2001, a lucrative market emerged for KM tools that index, categorize, and archive email, so specific messages can be easily located. KM email tools that exist nowadays are generally divided into two groups. One group categorises emails based on header information like dates. The second group does selective categorisation of only relevant emails based on content in the email body and attachments. To organize and archive, the first group use traditional search-and-retrieval indexing, while the selective email applications usually require other methods like content filtering and natural language selection, which apply various policies and rules to the content in order to organise data by theme (Harney, 2003).

Unlike EKE, email management applications today are most commonly implemented to improve customer services without making use of the knowledge that resides in electronic messages. The applications do not exploit the firm's email content assets. A product like KnowledgeMail relies only on the users' stored expertise profile, what they think their expertises are. However with EKE, the users have to rank keywords extracted from their emails from a scale which denotes their expertise within that field.

3. DEVELOPMENT AND CONSTRUCTION OF EKE

The EKE project is part of a wider research programme at Loughborough University. Currently researchers at Loughborough University are developing an Email Broker system, as shown in Figure 1, to not only manage knowledge, but also to aid employees to become more efficient when dealing with email (Jackson et. al., 2003)

Figure 1 – The Email Brokering System



As a whole, the system will consist of two parts, the Intelligent Email Classification and Notification Management System and the Knowledge Management Extraction (EKE) System. The Email Brokering System has been derived from the author's research into email, where it was found that the majority of employees had their email application set to check for email every five minutes. This meant that the employees could be interrupted by their incoming email every five minutes and the email could range from very important business matters to spam. The author noted that employers could try and minimize the number of interruptions and time wasted by setting the employee's electronic message applications to check for messages every 45 minutes instead of five, but of course employee's can always override this. Some employers do not try using this approach in the first place because they do not want to delay important business emails that need to be dealt with immediately. Thus, the first part of the email brokering system has been dedicated to the development of a sophisticated Intelligent Email Classification and Notification Management System to determine the importance of an email and when the email message should be delivered to increase employee effectiveness and efficiency when working (Jackson and Burgess, 2003).

The second part of the Email Broker System is the Knowledge Management Extraction System. The purpose of the system is to scan emails before they are sent out within an organisation, picking out keywords and then storing them into a database. The top-twenty frequently used keywords will then be ranked every month by the employee to denote their expertise in relation to them. The employee also has the opportunity to rank all the other keywords stored for that month, to ensure the employee's expertise is captured. For ethical and privacy reasons and knowledge accuracy, the sender of the email has the option not to keep any or some of the extracted keywords in the database, as keywords that are not ranked by the employee are removed from the database. Only once the keywords have been ranked will the organisation be allowed to search that part of the database. As well as the keywords, EKE also stores the sender's name, the recipient's name, the date and time of when the email was sent so that when the ranked keywords are accessible in the database they can be accessed by other members of the organisation, who will be able to contact the relevant member(s) of staff who hold the pertinent knowledge.

3.1 Capturing the Email Message in EKE

One of the key elements of EKE is to be able to capture an email message before it is sent to the server so individual keyword extraction profiles can be deployed rather than generic ones that apply to the whole organisation. To capture the email message before it reached the server the authors adopted a software application called E-MailRelay (Walker, 2003). E-MailRelay is a simple store-and-forward message transfer agent (MTA) and proxy server. It runs on Unix-like operating systems (including Linux), and on Windows. When running as a proxy all email messages can be passed through a user-defined program, such as a spam filter, which can drop, re-address or edit messages as they pass through. When running as a store-and-forward MTA incoming messages are stored in a local spool directory, and then forwarded to the next SMTP server on request. As part of EKE, E-MailRelay was set-up as a store-and-forward MTA, to enable the message to be processed and the keywords extracted. When used as a store-and-forward transfer agent E-MailRelay runs in two modes; the storage daemon part, and the forwarding agent. The storage daemon waits for incoming mail and stores anything it receives in a user specified directory. Once the keywords have been extracted from the email they are then moved to the spool directory where the forwarding agent E-MailRelay pulls messages out of the spool directory and passes them on to a remote server.

3.2 Extracting the Keywords in EKE

After E-MailRelay has intercepted the email messages the keywords have to extracted. A pre-written program called KEA was used to extract the keywords from the email messages (Yong, 2003). The KeyPhrase Extractor which is part of the KEA suite is used to extract the keywords from an email message. KeyPhrase Extractor extracts

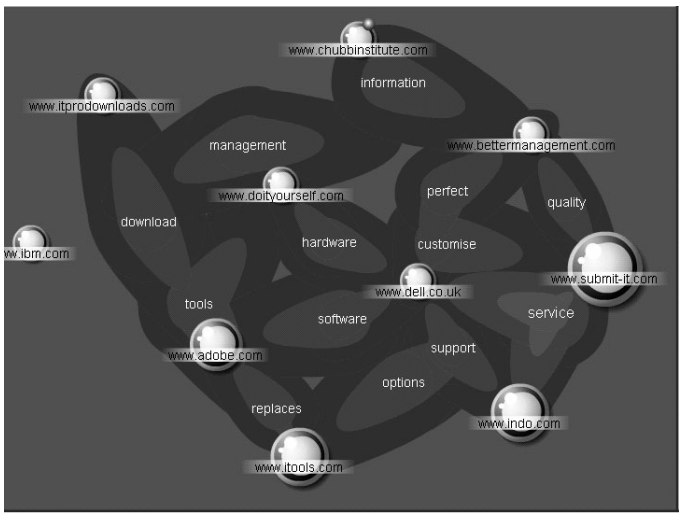
key-phrases from the documents in a given directory. It assumes that the file names for the documents end with ".txt" and then puts extracted key-phrases into corresponding files ending with ".key" (if those are not already present). However, before the keywords from the email can be extracted, user specific keyword extractor models have to be built using the KeaModel Builder. The KeaModel Builder builds a key-phrase extraction model from the user specific documents and predefined user keywords in a given directory. Again like the KeyPhrase Extractor it assumes that the file names for the documents end with ".txt" and it also assumes that files containing corresponding author-assigned key-phrases end with ".key".

As mentioned earlier, at the end of the month a client-side application called the Database Populator, which was developed by the authors is executed. The Database Populator collates all of the extracted keywords over the last month and displays them to the user for their acceptance. The user has to rank the extracted keywords from a scale of one to ten to denote their expertise in that field. The process may appear to be time consuming, but rarely does the whole process of ranking the keywords take more than five-minutes and is considerably less when compared to other methods of capturing the data. Personal interviewing or sending out questionnaires via mail, email, internet, or intranet to determine the staff's areas of expertise can be thought of as traditional methods of collecting information and extracting knowledge into regular knowledge management systems without worrying greatly about privacy issues. However, they also have problems that include speed, cost, ambiguous and biased questions that do not always reflect the real environment. First of all, it is very difficult to produce questionnaires which are unambiguous or contain biased questions. Personal interviews can be time consuming and expensive to arrange and conduct. In addition, the interviewer can bias the results and if the respondent is known to the interviewer they may be reluctant to provide certain information. Furthermore, the response rates for questionnaires can be very low. The above factors have contributed to the authors' favouring the use of an email keyword extractor such as KEA over personal interviewing or sending out questionnaires in order to meet the challenges of accurately determining the fields of expertise within the organisation.

3.3 Information Visualisation

Visual search engines have enormous advantages over classical textual searches in that a visual search easily catches the user's gaze. The user will have all the information in front of their eyes and they can identify relevant results depending on stimulus such as, size, lines and colours. (Dürsteler, 2003). To ensure employees get the maximum from EKE, a graphical user interface search facility was developed for its front end. The search facility enables users to search for keywords to discover who has communicated about a topic and more importantly who within the organisation has knowledge about the topic. Figure 2 shows a screen shot from a web site called Kartoo.com. Kartoo, developed in France by cousins Laurent and Nicolas Baleydier, is a relational search engine interface. Rather than displaying a list of ranked ordered links in the usual monotonous list format, it displays the results of a given search in a graphical, networked format**.** Kartoo uses links that suggest how the results interconnect. Laurent Baleydier and his team spent three years perfecting the technology, and chose Flash as the primary interface mechanism. When a user types in a search query, Kartoo creates an interactive map of the users search results. The map is a blue screen dominated by nodes/circles that are linked to each other by common terms. The terms are related to the users search query and are all connected with a mesh of lines. The user can click on the terms to add them to the query and focus their search and to further restrict the search graphically. More experienced users can further refine their searches with the advanced search interface (URL, near, domain, images). The interface is intuitive, and it makes navigating through the space of a query topic easier than textual methods and speeds up searches. At present the front end for EKE is very primitive when compared to Kartoo, but it is envisaged that a similar front-end will be developed for EKE within the next few months.

Figure 2 – Screenshot from Kartoo.com



4. PRELIMINARY RESULTS & SUMMARY OF EKE

A pilot study for EKE has recently been deployed within a major pharmaceutical organisation within the UK. The early feedback from the employees has been very positive, but the pilot is highlighting that the keyword extraction model either requires further training for individual employees or another method needs to be constructed to ensure all key technical words and phrases are captured. As already mentioned within the paper the front-end to EKE also requires further development to ensure the end-user gets the maximum benefit from the knowledge locating tool.

Within emails, massive amounts of information and knowledge are shared. However, this knowledge is no longer just shared between the sender and receiver. The development of EKE has also opened-up an ethical debate and organisations should be aware of their legal obligation towards their employees' privacy. The development of a front end that enables email senders to confirm keywords extracted from their emails before they appear to the users of the search facility was an absolute necessity and has largely overcome the anxieties that the users first had about EKE. With the increasing use of email in organisations and with the increasing number of electronic messages sent and received, the development of EKE has been a timely one, as it makes better use of the information contained within electronic messages and turns it into a better managed asset from which the whole organisation can benefit.

REFERENCES

Abaza, "Managing Email Content—Challenges and Benefits", http://www.kmworld.com/publications/whitepapers/ECM/abaza.htm, 2 July 2003

Bertolotti Laura and Calzarossa Maria Carla, "Models of mail server workloads", Performance Evaluation ,Volume 46, Issues 2-3 , October 2001, Pages 65-76

Bumper Brain Technologies Inc, 2003, http://www.bumperbrain.biz/services.htm, 22 November 2003

Dürsteler Juan C, "KartOO", http://www.infovis.net/E-zine/2002/num_97.htm, 26 June 2003

Harney John, "With the right tools, email messages can be valuable KM resources", http://www.intelligentkm.com/feature/010831/feat1.shtml, 11 July 2003

Jackson, T.W., Dawson, R.J. and Wilson, D., ''Breaking Bad Habits: The Negative Effect of Email and Instant Messaging on the Workplace'' , *Information Technology & Organisations: Trends, Issues, Challenges & Solutions* , Khosrow-Pour, M. (ed.), Idea Group, Hershey, USA, Information Resources Management Conference , Philadelphia, 2003, pp. 399-401, ISBN: 1-59140-066-X .

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 .

Nonaka, I. (1994) *Organization Science*, **5,** 14-37.

Polanyi, M. (1967). *The Tacit Dimensio*n, Doubleday, New York.

Roda Claudia, Nabeth Albert AngehrnThierry and Razmerita Liana, "Using conversational agents to support the adoption of knowledge sharing practices", Interacting with Computers, Volume 15, Issue 1 , January 2003 , Pages 57-89

Rubenstein-Montano Bonnie, Buchwalter Judah and Liebowitz Jay, "Knowledge management: A U.S. Social Security Administration case study", Government Information Quarterly, Volume 18, Issue 3 , Autumn 2001, Pages 223-253

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

Walker, G., "E-mailrelay", http://emailrelay.sourceforge.net/, 2003

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

Yong, Chien-Chin, "KEA", http://www.cse.unsw.edu.au/~ccyong/keadoc/, 2003