Motivating intelligent email in business: an investigation into current trends for email processing and communication research

Michal Laclavík
Institute of Informatics
Slovak Academy of Sciences
Bratislava, Slovakia
laclavik.ui@savba.sk

Diana Maynard
Dept. of Computer Science
University of Sheffield
Sheffield, UK
diana@dcs.shef.ac.uk

***Abstract*—According to recent surveys, information workers send and receive an average of 133 messages per day[1], and users talk about "living" in email, spending an average of 21 percent of their time on it, as well as reporting general problems with overload. Information created by a business can represent either an asset or a liability, depending largely on how well it is managed. Email is no different in this respect: it can be a highly efficient and useful tool for communication, but only if the information it contains can be managed effectively. One of the main drawbacks of email usage today is its insufficient integration into the collective workspace environment. We believe that by integrating it with other external information (both on the desktop and on distributed servers), one can migrate some of this information to more appropriate storage environments, thereby partly addressing the problem of overload and offering users an integrated access to data and functionality. Currently, there is much research in the area of both personalised and business information management, but very little research that focuses on email as the primary information source, despite its ubiquity. In this paper we survey the current state of the art in email processing and communication research, focusing on the current and potential roles played by email in information management, and commercial and research efforts to integrate a semantic-based approach to email.**

***Keywords*-email research; state of the art; intelligent email**

I. Introduction

With the ever increasing availability of digital information and the enormous volumes of data transmitted electronically every day, information overload is a serious problem for today's businesses. As a result of the growing ubiquity of email, office workers can easily become swamped by electronic data: in 2001, information workers received about 20 email messages a day and sent about 6 messages[2] . While the number of received messages is increasing, sending stays pretty much at the same level [1]. Findings from 2003 also show that 80% of users prefer email[3] for business communication. Similarly, some case studies[4] indicate that Web 2.0 applications often depend on email, so with Facebook, Twitter and even older applications such as Amazon, the number of email communications (notifications, transactions etc.) has increased as a result. In the US, the number of pieces of first-class mail sent has decreased by over 2 billion in the last 10 years as people turn increasingly to electronic methods of communication. While this has enormous potential benefits to efficiency and ultimately, profitability, this information is only useful if it can be stored and managed efficiently. Furthermore, email is rarely a standalone information source, but often contains pointers to further information such as files (e.g., saved attachments), links to items on the web, and references to other resources.

One major documented problem with email is that in addition to the asynchronous communication that it was designed for, it is also used as a conduit for multiple additional functions that it was not designed for, such as alerting, collaboration, archiving and task management. For example, around 18% of messages contain attachments, and users may have huge archives of thousands of stored messages that include attachments that they use as a knowledge repository [2] [3]. It is also well known that people use their email inboxes as an active "todo" list to manage and control workflow, keeping active conversational threads in their inboxes where they will encounter them when they access new messages. In addition, email is used as a conduit for often low value information (messages forwarded or cc-ed as FYIs). Various problems arise from using email for all these functions it was not designed for. Users find it hard to access archival information because a) search in email is often ineffective; and b) lack of integration means that users do not remember whether information such as attachments is in their email archive or has been downloaded to their file system. Finally, up to one third of folders that users create contain 2 or fewer messages [3], which is clearly an inefficient storage method.

[1] http://www.radicati.com/wp/wp-content/uploads/2008/09/hp_whitepaper.pdf

[2] http://www.gallup.com/poll/4711/Almost-All-EMail-Users-Say-Internet-EMail-Made-Lives-Better.aspx

[3] http://www.mariosalexandrou.com/technology-trends/2003/80-percent-of-users-prefer-email.asp

[4] http://www.returnpath.net/blog/2008/07/case-study-web-20-runs-on-emai.php

Studies have also shown that problems arise in using email to manage active workflows. If users receive a large number of incoming messages (including spam and other irrelevant messages), then this makes it hard to monitor active important tasks or collaborations as these "disappear" in the inbox as new irrelevant messages arrive [2]. A final problem is that email can fill up with messages that are ephemeral, or of little current relevance (such as information about meetings that have already taken place). Again these sit in the inbox and distract the user's attention away from more important active tasks [4].

Some research and commercial attempts have been made to address these problems. They are described and discussed in the paper. However, there are various problems with these prior research efforts. First, many of these systems have involved the development of new standalone clients, forcing people to switch to new software and change their work practices. Second, these systems are often not well integrated with users' other systems (e.g. their file system, or corporate databases) which means that information in these systems cannot be exploited or leveraged to allow more effective information processing.

The paper is structured as follows: in the following section, we situate the problem more completely, discussing the role of email in the field of information management and further discussing its troublesome issues. Section III describes some commercial tools which have been developed to deal with some of the problems mentioned. Section IV discusses some of the main research efforts tackling these issues; while Section V puts forward some proposals for the way ahead and outlines some ways in which current systems can be improved.

II. Email Communication Role in Knowledge and Information Management

Recent R&D investments for business Knowledge Management (KM) systems have not been very successful. Some authors claim [5] that if more than 33% of investment goes into technology, this results in a radical change to working culture in the enterprise, and thus KM projects will therefore fail to be successful. It is difficult for KM to succeed if it requires changing working practices by introducing completely new tools for tasks which are usually performed by person to person communication and other social interactions. However, since email is already a widely used and accepted technology, used on a daily basis by knowledge workers, we do not need to introduce completely new tools for these tasks. In order to create a suitable knowledge management and collaboration platform, we can include standard ICT infrastructure, email and Web tools, available in most enterprises, rather than offering new working tools which are costly, hard to install and maintain, and require changes in organisational culture and working practices. Communication channels such as email already have some features typically required from a knowledge management solution, such as universal SMTP protocol, which facilitate user interaction, communication or information sharing as well as being available in all types of organisations.

Email repositories and email activity are valuable assets in any modern, internet-based business organization. Email is the number one application on the internet, and even small companies can generate large volumes of email traffic and fill email repositories with enormous amounts of data, just to accomplish their daily tasks. The following features are common in the use of emails in enterprises and communities of all sizes [6]:

- Every organisation, without exception, will have an email infrastructure before it reaches the stage of developing or adopting any knowledge management solution.
- E-mail communication in a modern organisation is over 78% action-oriented, according to a study [7]. Communication can be considered the foundation for most organisational actions.
- Managers, and knowledge workers of all kinds, interact with their email systems on a daily basis.

By building a solution on top of an existing email communication mechanism, an organisation does not have to change its working practices when such a solution is installed and set up. Users simply receive emails as before, but additional information or knowledge relevant to the knowledge management or collaborative aspects is attached, as appropriate to the email. A successful email-based solution can learn and adopt many ideas from existing systems and prototypes, but has the following requirements:

- to be used within any email client or webmail and thus integrate with the email or extend current email standards in a similar way to multipart messages and mime types;
- to inter-connect existing SME environments such as document repositories, intranet systems, databases or other legacy systems;
- to integrate with widely used general purpose platforms: for example, integrating with platforms for collaboration (such as wikis) and task management (calendars, todo lists and so on).

III. Existing Commercial Tools

In this section we shall discuss some commercial innovative approaches dealing with email communication, considering what they have to offer and/or how they could be further developed.

Gmail[5] a web-based email program developed by Google, has perhaps changed email services and email use more significantly than any other tool. By grouping messages into conversations, offering fast full-text search, providing context sensitive advertisements, offering actions such as

[5]http://gmail.com/

add event to calendar, *track package* or *show address on the map* as well as introducing labels instead of folders and providing gigabytes of storage for email archives, Gmail has advanced the state of email services significantly. However, its capabilities could be extended still further.

Zimbra[6] is a web-based client accessible from any device with search functionality, offering integration with calendars, IM and document authoring, thus providing a collaboration suite. Similar functionality is offered by Microsoft Outlook, Gmail integrated with Google services, or ClearContext. In addition, Zimbra detects objects such as phone numbers or addresses, and allows some actions on these objects, as well as the possibility to build mash-ups for specific needs or applications.

MarkMail[7] is a community-focused searchable message archive service, which allows an organisation with huge volumes of email to leverage the large amounts of collective knowledge accumulated over time through email discussions. Users can find technical information, research historical decision making, spot trends, and locate the subject matter experts for any topic. While it provides extensive search facilities for email, it does not combine information in email with other desktop knowledge or context. MarkMail thus focuses on finding information rather than connecting it with other forms of data. It is important to extend the search capabilities in email, and while MarkMail makes a good attempt at structuring messages, it does not provide novel ways of relevance ranking. We discuss some possible methods for improvement in the next section.

iWantSandy was an email-based tool aimed at helping with the *organisation of tasks* in a person's daily life. Essentially it operated as a reminder service to a person or group of people (family, colleagues etc.), based on the user sending emails to the service containing details of the information to be reminded about, and the service sending reminders by text or email at the appropriate time. While it was a very useful tool in assisting with organisational activity, it did not integrate information from external sources, and operated only on a fairly restricted language as far as the instructions go. The main drawbacks were therefore that it could not simply analyse regular emails, it could not deal with related information not explicitly mentioned, and it operated on a push rather than pull technology (the user had to explicitly inform the service that they wanted to remember something). iWantSandy ceased to operate in December 2008, however, the intellectual property has been acquired by Twitter, so a revised version of the tool may be reincarnated in the future.

The Attent solution from Seriosity[8] is one of the few commercial applications which really attempts to address the information overload problem resulting from the enormous amount of email found in business. It works by prioritising a user's email based on perceived importance. The idea stems from the field of interactive gaming and relies on the sender attaching a number of Serios to their messages, which are essentially units of importance. This enables the recipient to prioritise their emails according to importance levels. While this is very useful in some situations, there are again a number of drawbacks. First, the perceived importance of an email may differ wildly between sender and recipient: the recipient therefore has no real control of importance or of topics of interest to themselves. Second, there is no importance attached to information other than the email itself: serios are attached only to emails themselves. Third, it only deals with importance but offers no possibilities for searching and navigating information, nor of relating information with other emails or with other kinds of information.

Both iWantSandy and Attent operate on a user-input basis: they require the user (either the sender in the case of Attent or the recipient in the case of iWantSandy) to be pro-active in deciding what they consider important or what they want to be notified about. Neither use any form of understanding of information or intelligent analysis. While the purpose of all the last three tools (these two and MarkMail) is to make life easier for the user by saving time and effort in finding, dealing with or remembering important information, they do not aim to exploit many of the existing useful properties of email, such as linguistic features, relational information and related metadata in order to improve knowledge management tasks.

Xobni[9] is a recent Outlook plug-in, which supports extended search capabilities, a better organisation of the inbox, and management of the media and contacts within emails by integrating social networking aspects into the email communication. Xobni offers a variety of information related to the message sender, such as attachments, contact information extracted from signatures or contacts related to the sender. Xobni does not support any intelligent analysis or understanding of the email communication and relies heavily on integration with social networking standards. However, the idea of using social networks within email is quite innovative and can have a high potential in the enterprise context, if social networks within email are semantically integrated with the enterprise environment.

In addition to the efforts discussed, the Postbox[10] email client built on top of Thunderbird offers additional contact details including phone number or contact's picture and status from Facebook. It also provides better search capabilities including attachments and improved conversation views. ClearContext[11] detects and organises events/tasks, contacts and attachments inside Outlook in a similar way to Xobni.

[6] http://www.zimbra.com/
[7] http://markmail.org/
[8] http://www.seriosity.com/products.html
[9] http://www.xobni.com/
[10] http://www.postbox-inc.com/
[11] http://clearcontext.com/

IV. Current Research Effort

While there is much hype on web-based research, email communication has not really been widely addressed by the research community. Most of the research that does exist in this domain has been connected with spam detection. There is also ongoing research on email in the Human Computer Interaction field. In 2004 the series of Conferences on Email and Anti-Spam (CEAS) started. The primary focus is spam detection; however, research related to extracting information and knowledge, its management, social network analysis and topic discovery is also addressed. The publishing of the Enron emails [8] was also a significant driver of research in email communication. The Enron corpus consists of approximately 200,000 messages after cleaning of the original 600,000 messages. Research concerning communication analysis and conversation threads has also been performed on the first annotated email corpora[12], based on a small portion of Enron emails related to the California Energy Crises. The next significant movement in this area was the introduction of the Enterprise track within the series of TREC conferences in 2005. Part of the Enterprise track[13] makes use of a corpus of W3C mailing lists compiled in 2004[14] containing approximately 300,000 email messages. The task was to address email search and expert search, although mainly the latter was addressed. While in 2004 and 2005 the W3C email corpus was used, in 2007 the CSIRO corpus[15] was used where email communication was no longer present, and expert and document search tasks were performed on various kinds of documents, excluding email communication. We believe that email communication in the enterprise or community is the most important missing piece of the puzzle to manage and use corporate knowledge effectively. We maintain also that email communication is an important source of information and knowledge as well as a tool for communication, collaboration and for performing daily work activities widely used but not yet widely addressed by research.

Improving search in email should be also addressed more in research. Some of the TREC results in the Enterprise Track mentioned earlier can probably be applied to this task as a way to improve ranking. In TREC, several advanced algorithms concerning expert search were developed. Other attempts focus on improving email search using social network and information retrieval to allow activity centric search within email [9]. When searching information on the web, one can use well-established ranking algorithms for result ordering, such as PageRank, OPIC or HITS: while email is not the same as hypertext, it does contain multiple distinct parameters such as threads of conversations, date, subjects, a sender and receivers, that can be used by algorithms to better rank search results for emails. The above approaches and results of semantic message understanding can enable the development of appropriate ranking algorithms and thus improve email search. The most effective ordering of results from a search in a personal email archive is the date of the message. If we go to mailing lists or an enterprise archive, where one has to find the answer to knowledge intensive tasks, the rank problem pops up immediately. Search result ranking is then needed to find the most relevant and trusted information.

Social Networks included in email archives are becoming increasingly valuable assets in organizations, enterprises and comunities, though to date they have been little explored. In personal archives, Xobni exploits social networks to help the user manage contacts and attachments, but at the enterprise or community level, social networks can be exploited to improve email search, manage customers and suppliers, prioritise emails or improve inference mechanisms when connected with other detected semantic information from the email. Social networks with email communication have been studied to some extent. For example, communication on the Apache Web Server mailing lists and its relation to CVS activity was studied in [10]. This work also introduces the problem of identifying email users' aliases. Extracting social networks and contact information from email and the Web and combining this information is discussed in [11]. Similarly new email clients (e.g. Postbox) or plugins (Xobni) try to connect email social networks with web social networks like Linkedin or Facebook. The extraction of social networks from large email archives and network transformations using a semantic model is discussed in [12]. Another research effort [13] exploits social networks to identify relations and tests proposed aproaches on the Enron corpus. To conclude, there is much research work done on social networks within web social network applications, but email social networks are a bit different since in the email you can discover the level of interactions (number of messages exchanged, time, relation to content and possibly discovered semantics), and the influence of these differences on better information and knowledge management still needs to be explored.

One of the first attempts to apply semantic web technologies to email was performed by McDowell [14], who tries to resolve problems stemming from one to many communication tasks such as event planning, by communicating semantic web formal data such as RDQL queries in the message. While such an approach is not very user friendly, there is definitely a need to communicate and share data such as events from calendar, tasks or contact details over email in a standardized way.

Efforts to connect knowledge or context-sensitive information with emails have been realised in already discussed Zimbra and Gmail as well as in kMail research prototype

[12] http://bailando.sims.berkeley.edu/enron_email.html
[13] http://www.ins.cwi.nl/projects/trec-ent/wiki/index.php/Main_Page
[14] http://research.microsoft.com/users/nickcr/w3c-summary.html
[15] http://es.csiro.au/cerc/

[6], which integrates email communication with organisational memories or in the Acoma framework[16] [15]. Acoma connects to any email client as a proxy (similarly to antivirus programs) and adds HTML or text attachments including context sensitive hints into an email message. Acoma is being extended within the Commius[17] project focused on email-based interoperability for SMEs.

In addition, the following R&D prototypes have been developed, which are focused on solving problems of email communication to handle various tasks such as task management, information archiving or collaboration aspects: Telenotes, ContactMap, TaskMaster, Snarf, ReMail or Priorities. One major strand of research has been to address task management. One such effort looks at ways to better track existing threads. Threads account for up to half of the user's email, and these can become very complex to track. When email messages accumulate "replies to replies", it is difficult to see which replies relate to which elements of the original messages. This work developed thread visualisations so that users can better detect relations between complex conversations [16], [17]. Gmail also groups threads and deals with multiple messages in one thread as a single item of communication. Other clients allow users to more easily monitor and access active tasks, using machine learning to recognise relations between messages when these are implicit, e.g. messages that appear in different threads may actually be about the same task [18]. Yet more work has extracted social profiles from user interactions with prior emails to determine who are the user's important contacts. The system can learn that the user routinely replies to messages from one person but ignores messages from another. It can then promote messages from people who are important to the user, demoting messages from others [19]. Finally machine learning has been applied to users' email behaviour to determine which kind of messages users reply to quickly and which they tend to ignore [20], and also to assist them with filing messages [21]. Another approach tackling the email overload problem is to stratify conversations. Telenotes integrates email with instant messaging and also Notes backend databases. This allows users to migrate conversations to different email clients - allowing quick question and answer emails to migrate to instant messaging, but longer term focused group discussions to discussion databases [4]. IBM Research has spent nearly a decade studying email within their ReMail[18] [22] project, since Lotus software was one of their important products. ReMail focused mainly on visualization and management of emails in threads, offering email annotations with colour, icons and notes as well as integration with instant messaging. Most of ReMail findings have already been exploited in some way in current email clients or webmails.

One of the most significant attempts to understand email communication has been performed by DERI's Semanta system[19], within the EU Nepomuk project. This applies speech act theory to email communication processes, eventually giving a formal structure and semantics to ad-hoc workflows which are characteristic of email communication. Semanta focuses only on speech act understanding but not on other aspects such as support for business tasks, interoperability, connection to existing infrastructure and involve changes to working practices on certain level. It also does not focus on any approaches to gather the semantics of email other than speech acts. Speech act theory was applied earlier also by V. Carvalho and W. Cohen [23] for "email acts" classification. The Nepomuk project[20], which focused on creating a Social Semantic Desktop, introduces several other workspace integrations with email such as a data wrapper that automatically adds emails to the semantic desktop infrastructure or the Nepomuk Task Management - Kasimir prototype, which also integrated email functionality.

While current research in the field of HCI and applied knowledge management aims to change users' behaviour with respect to email communication, this is only one side of the story. For example, current advice to users involves setting up the email application to display in the inbox the sender, the subject and three lines of the email, so that the recipient can quickly determine if the email requires immediate attention. Clearly, this advice is useless if the first three lines of the email do not contain relevant information to help the user decide what to do with the email. A better solution is therefore to have a mechanism for providing a three-line summary of the email that provides the main points and action types (is the email for information only, what kind of action is required, etc.), as well as labels that are attached automatically to the document providing categorisation information (e.g. what is the main topic of the email, which social circle does it fall into (work/family/friends, which project at work, and so on)). There has been little work to date on email summarisation: while there has been some work on summarising email conversations [24], this is more for the purpose of getting an idea of the main attitudes and ideas represented in emails, rather than a general idea of the topics of the emails themselves and the salient facts, or of grouping similar emails together and summarising the group as a whole. Generating summary from the keywords is discussed in [25]. From the HCI field comes also a few works on task or activity management discussed earlier in this section. There are also recent efforts in managing tasks, activities or workflows in the email (e.g. in Nepomuk), managing flexible processes in small and medium enterprises within email [26], dealing with

[16]http://acoma.sf.net/

[17]http://www.commius.eu/

[18]http://www.research.ibm.com/remail/

[19]http://smile.deri.ie/projects.html#semanticemail

[20]http://nepomuk.semanticdesktop.org/

requests and commitments in email [27] or using machine learning aproaches to provide activity centric views on email communication in the enterprise environment [28].

Some of the latest email research was promoted also in the AAAI-08 conference and its Workshop on Enhanced Messaging (EMAIL-2008). After this event participants created an email research Google discussion group and website[21] with a periodically updated bibliography on email research. We hope that such activities will stimulate community building, promoting and extending email research activities and further improvement of email communication systems, services and tools.

V. The Way Ahead: Implementing a Semantics-based Approach to Email Processing

Many proponents of the Semantic Web seek a universal medium for information exchange based upon XML syntax. This has given rise to such standards as the Resource Description Framework (RDF)[22] and its elaboration in RDF Schema or the Web Ontology Language (OWL)[23]. The predominant use of ontologies to foster semantic interoperability is reflected by the numerous research efforts, and software tool development and support in this area. In response to approaches to ontological modelling, such as those cited above, e.g. RDF, a number of tools for ontology editing, storage, querying and reasoning are now available. These include several semantic frameworks for accessing and manipulating documents in OWL, RDF and RDFS. There are several RDF/RDFS-based reasoners and repositories, such as OWLIM, Sesame, Jena, Joseki, Kowari and 3store.

Automated annotation of Web documents is a key challenge for the realisation of the Semantic Web. Web documents are structured, but this structure is typically understandable only for humans. This is one of the major problems to be addressed by the Semantic Web. Emails, on the other hand, are primarily composed of unstructured text, which is even more resistant to automatic processing by other tools tied to HTML structure such as wrappers. Thus annotation is a crucial step in the transformation of this unstructured information and knowledge, before processes such as search and reasoning can be performed. Manual annotation currently plays an important role in the email management process: most email programs enable the user to label individual emails with tags such as importance rankings, topics and so on. However this process requires the user to first read the email and then to classify it, which is a laborious process. Better by far would be the automation of as much of this process as possible, so that the user automatically gets informed about the nature of the email in his inbox before he has read it, and perhaps gets his email sorted according to topic, priority, etc. Semi- and fully automatic annotation processes are thus required: partly as a standalone task in order to create tags and even highlight keywords and phrases in documents, but also, and perhaps more importantly, in order to pave the way for processes such as clustering of documents along a thread, summarisation of groups of related emails, extraction of important information (such as urgent tasks required of the user, according to the content of the email, e.g. upcoming deadlines for papers, requests for authorisation, request for attendance at a meeting, and so on), and search facilities. Furthermore, it can be used to link emails to related documents which may have no explicit connection. One of the few approaches to semantic email is the Semanta framework developed by DERI, which was described in the previous section. However, while this is certainly a step in the right direction, it is still far from being a complete solution.

[21] http://emailresearch.org/

[22] http://www.w3.org/RDF/

[23] http://www.w3.org/TR/owl-feature/

VI. Conclusions

Email is the number one application on the internet. It has existed for more than 20 years, and for a long time we have been using it at pretty much the same level. As time passes, there have of course been some improvements with services such as handling large email archives, better email clients with various functionalities and fast full text search, but compared with other web tools which are much younger, one can ask why email has been addressed so little in research and development. Although email, wikis, and task management applications are currently poorly integrated, a continuum across these environments would be quite natural. While complex activities and collaboration may require full task management or wiki functionality, tasks or ad-hoc collaboration can be more easily handled by email. All the above mentioned requirements can only really be achieved with semantic understanding of messages.

In this paper, we have described the importance of email in the everyday life of information and knowledge workers, as well as discussing its value, namely not only the contents of email but also its tacit and unexplored interconnection with the community or enterprise business context and environment. We have also discussed existing commercial and research attempts to address various problems stemming from current email communication use, and describing some of the challenges which need to be addressed by email communication research, providing possible research paths to follow.

Acknowledgment

The authors would like to thank to reviewers for useful comments, which helped to improve the paper. This work is partialy supported by projects Commius FP7-213876, APVV DO7RP-0005-08, AIIA APVV-0216-07 and VEGA 2/7098/27.

REFERENCES

[1] A. Lantz, "Does the use of e-mail change over time?" *International Journal of Human-Computer Interaction*, vol. 15, no. 3, pp. 419–431, 2003.

[2] S. Whittaker and C. Sidner, "Email overload: Exploring personal information management of email," in *Proceedings of ACM CHI96 Conference on Human Factors in Computing Systems*, 1996, pp. 276–283.

[3] D. Fisher, A. Brush, E. Gleave, and M. Smith, "Revisiting Whittaker and Sidner's "email overload" ten years later," in *CSCW2006*, New York, 2006.

[4] S. Whittaker, G. Swanson, J. Kucan, and C. Sidner, "Telenotes: managing lightweight interactions in the desktop," *Transactions on Computer Human Interaction*, vol. 4, pp. 137–168, 1997.

[5] T. H. Davenport and L. Prusak, *Working knowledge: how organizations manage what they know*. Boston, MA: Harvard Business Press, 2000.

[6] D. Schwartz and D. Te'eni, "Tying knowledge to action with kmail," *IEEE Knowledge Management*, pp. 33–39, 2000.

[7] D. Te'eni and D. Schwartz, "Contextualization in computer-mediated communication," in *Information SystemsThe Next Generation*, L. Brooks and C. Kimble, Eds. New York: McGraw-Hill, 1999, pp. 327–338.

[8] B. Klimt and Y. Yang, "The Enron Corpus: A New Dataset for Email Classification Research," in *European Conference on Machine Learning*, 2004.

[9] E. Minkov, R. Balasubramanyan, and W. Cohen, "Activity-centric Search in Email," in *Enhanced Messaging Workshop, AAAI*, 2008.

[10] C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan, "Mining email social networks," in *MSR '06: Proceedings of the 2006 international workshop on Mining software repositories*. New York, NY, USA: ACM, 2006, pp. 137–143.

[11] A. Culotta, R. Bekkerman, and A. McCallum, "Extracting social networks and contact information from email and the web," in *CEAS '04: Proceedings of the First Conference on Email and Anti-Spam*, 2004. [Online]. Available: http://www.ceas.cc/papers-2004/176.pdf

[12] M. Laclavik, M. Seleng, M. Ciglan, and L. Hluchy, "Supporting collaboration by large scale email analysis," in *CGW '08: Proceedings of Cracow'08 Grid Workshop*, 2009, pp. 382–387. [Online]. Available: http://laclavik.net/publications/cgw08_laclavik_final.pdf

[13] C. P. Diehl, G. Namata, and L. Getoor, "Relationship identification for social network discovery," in *The AAAI 2008 Workshop on Enhanced Messaging*, 2008.

[14] L. McDowell, O. Etzioni, A. Halevy, and H. Levy, "Semantic email," in *WWW '04: Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM, 2004, pp. 244–254.

[15] M. Laclavik, M. Seleng, and L. Hluchy, "ACoMA: Network Enterprise Interoperability and Collaboration using E-mail Communication," in *Proceedings of eChallenges 2007; Expanding the Knowledge Economy: Issues, Applications, Case Studies*, 2007, pp. 1078–1085.

[16] G. Venolia and C. Neustaedter, "Understanding sequence and reply relationships within email conversations," in *Conference on Computer Human Interaction*, 2003.

[17] M. Wattenberg, S. Rohall, D. Gruen, and B. Kerr, "Email research: Targeting the enterprise," *Human Computer Interaction*, vol. 20, 2005.

[18] V. Bellotti, N. Ducheneaut, M. Howard, and I. Smith, "Taking email to task: The design and evaluation of a task management centered email tool," in *Conference on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, 2003.

[19] S. Whittaker, Q. Jones, B. Nardi, M. Creech, L. Terveen, E. Isaacs, and J. Hainsworth, "ContactMap: Organizing Communication in a Social Desktop," *ACM Transactions on Human Computer Interaction*, 2004.

[20] E. Horvitz, A. Jacobs, and D. Hovel, "Attention-sensitive alerting," in *Conference on Uncertainty and Artificial Intelligence*, 1999, pp. 305–313.

[21] R. Segal and J. Kephart, "MailCat: an intelligent assistant for organizing email," in *Proceedings of Conference on Autonomous Agents*, 1999, pp. 276–282.

[22] S. L. Rohall, D. Gruen, P. Moody, M. Wattenberg, M. Stern, B. Kerr, B. Stachel, K. Dave, R. Armes, and E. Wilcox, "Remail: a reinvented email prototype," in *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, 2004, pp. 791–792.

[23] V. Carvalho and W. Cohen, "On the collective classification of email" speech acts"," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM New York, NY, USA, 2005, pp. 345–352.

[24] G. Carenini, R. Ng, and X. Zhou, "Summarizing Emails with Conversational Cohesion and Subjectivity," in *Proceedings of ACL-08*, 2008, pp. 353–361.

[25] M. Dredze, H. M. Wallach, D. Puller, and F. Pereira, "Generating summary keywords for emails using topics," in *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*. New York, NY, USA: ACM, 2008, pp. 199–206.

[26] T. Burkhart, D. Werth, and P. Loos, "Commius - an email based interoperability solution tailored for smes," *Journal Of Digital Information Management*, vol. 6, no. 6, 2008.

[27] A. Lampert, R. Dale, and C. Paris, "The nature of requests and commitments in email messages," in *The AAAI 2008 Workshop on Enhanced Messaging*, 2008.

[28] N. Kushmerick, T. Lau, M. Dredze, and R. Khoussainov, "Activity-centric email: A machine learning approach," in *AAAI'06: Proceedings of the 2006 American National Conference on Artificial Intelligence*.