Email between Private Use and Organizational Purpose

Uwe V. Riss

SAP Research Karlsruhe

Vincenz-Priessnitz-Str. 1, 76131 Karlsruhe, Germany

uwe.riss@sap.com

ABSTRACT

Emails have become an eminent source of personal and organizational information. They are not only used for personal communication but also for the management of information and the coordination of activities within organizations. Email traffic also exhibits the social networks existing in organizations. However, the central problem, which we still face, is how to tap this rich source appropriately. Main problems in this respect are the personal character of emails (their privacy) and the mainly unstructured character of their contents. Since these two features are essential success factors for the use of email they cannot be simply ignored. Meanwhile there are various approaches to recover the hidden treasure and make the contained information available to information and process management. For example, semantic or mining technologies play a prominent role in this attempt. The paper gives an overview of different strategies to make organizational use of emails, also touching the role of privacy.

Categories and Subject Descriptors

H.4.1 [**Office Automation**]: *Workflow management*; H.4.3 [**Communications Applications**]: Electronic mail; H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work;

Keywords

Email, semantic email, personal information management, social network analysis, task management, privacy.

1. INTRODUCTION

Although email has been designed for the exchange of personal messages, meanwhile its use has extended to personal information management, collaborative activity handling and other purposes [1]. In this way email systems have become resources of versatile and extensive information.

Various research initiatives have started to exploit this so far mainly unused information source. In the course of these initiatives it has become clear that the use of email can serve various purposes and we need better ways to support them. The email corpus consists of individual messages and many approaches first aim at the identification of common features and connections between them. While some research directions such as email for task and process management already started more than 20 years ago, we still observe a growing number of new analytical approaches towards a better integration of email and web technologies or regarding various kinds of social network analysis. In particular in organizations a vast amount of information is hidden in emails and this makes it attractive to use analytical methods to get access to it. However, in order to make such approaches work in an organizational context we must ensure that privacy of people's communication is respected.

In this paper we will discuss which opportunities the recent developments in email analysis and integration bring about and shortly touch the role of privacy in this respect.

2. TASKS AND PROCESSES IN EMAIL

The support of task management via email can regarded as one of the first directions towards a better exploitation of email use. Prominent examples of these attempts are Taskmaster [2] and TaskView [3], which aim at improving the task management functionality for emails. Such task support systems were extended by analytic tools, for instance, making use of "speech act" theory [4]. They help classify emails [5,6] and better identify their task features. The central aim of using speech act theory was to tackle the problem of unstructured email content that has made it difficult to automatically extract proper task descriptions. One well known system that is based on this idea is Semanta [7]. Another approach based on speech acts that aims to detect action requests from emails is that of Lampert et al. [8]. However, Semanta is not the only system that aims at the identification of speech acts but also at the application of semantic technologies to emails. Recently we find various solutions, which go into this direction, embraced by the concept of semantic email.

2.1 Semantic Email

One of the groundbreaking papers fostering the idea of introducing semantic technologies to email was "Semantic Email" by McDowell at al. [9], who suggested a Semantic Web vision of email that aims at the enrichment of emails by semantic descriptions such as RDF queries in order to improve their automatic processing. Semantic web technologies even went beyond this goal and were also used to seamlessly embed email handling in the Nepomuk semantic desktop [10]. In the Nepomuk project[1] even two semantically enabled email approaches came up, besides Semanta [11] also the Email extension of the Nepomuk Semantic Task Management [12]. By this integration, email could not only be used to delegate tasks but also became part of the personal information management on the semantic desktop. Similarly, Khoussainov and Kushmerick have used machine learning for the identification of tasks making use of the relations between messages via semantic analysis [13]. Due to the potential benefits it is not surprising that semantics is also seen as a key aspect of future email systems [14].

2.2 Business Process Support

Via task management email has not only been integrated in personal information management and work coordination but it has also become a means of business process management, in particular in the domain of knowledge work that is characterized by rather

---

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

flexible and hardly standardizable processes. An example of research in this direction is the Commius project[2] where email is tailored to address the needs of SME to provide email-based business processes [15,16]. Here the advantage of email and its lightweight character becomes apparent and allows SME to pass on expensive IT infrastructure. In a similar way the Nepomuk Semantic Task Management has developed a semantic task management framework based on work coordination through email and lightweight process management [17].

3. Expert Finder and Social Network Analysis

Another major strand of email research aims at the analysis of email traffic with the goal to identify expertise [18] and social networks [19,20]. In contrast to the previous approaches, both directions have a strong organizational focus. Moreover, they are related since they aim at the role of persons in an organization. Consequently, both aspects expertise and network identification are analyze in combination [21].

3.1 Social Network Analysis via Email

Analysis of email has been already carried out to identify personal relationships for some time. Whittaker et al. [22] have shown that several email parameters such as frequency, reciprocity, or longevity predict the relevance of email contacts. As well, the mining of email traffic has revealed various communication patterns [23]. A certain focus of such network analysis has been the use of extracted email information for enterprise purposes such as knowledge management due to the huge amount of organizational knowledge contained in email corpora [24]. Finally, researchers have not only looked for plain relationships between people but also tried to extract process knowledge from email traffic [25]. In this attempt network and process analysis have come together.

3.2 Expert Finding via Email

Culotta et al. [19] enriched the social network information from email by additionally extracted keywords that refer to the expertise of those persons involved in the email traffic, however, merely from a technical perspective. Campbell et al. [18] addressed the privacy issues that are related to such analysis and solved it by extracting only personal relations from email while the expertise was obtained from publicly available sources. IBM's SmallBlue [26] takes account of the privacy issues but tries to solve the problem in a different way, making use of the email contents but introducing a set of policies to restrict the selection of the extracted data. Guy et al. [27] point at the difference of the social network and expert analysis of one's own emails (egocentric network analysis) and of the entire organizational email corpus necessary for the analysis of proper social networks. While the latter has to deal with privacy issues, the private analysis is uncontroversial. Organization-wide analysis of email for organizational purposes cannot do without privacy considerations.

3.3 Approaches to Privacy

There are still other possibilities to address the privacy challenge. For instance, public social networks often make more private information visible by way of mutual agreement. Starting from egocentric networks one could use this strategy for the mutual access to these networks based on the participants' permission. In fact, Nardi et al. [28] have pointed to the fact that users spend effort

to establish networks and deliberately share information with others in these. The only precondition is that they keep control over their data. Another possibility is to depersonalize information extracted from emails via aggregation. Hereby the same information is obtained from a larger number of independent sources so that its private character vanishes. However, even this approach is not completely unproblematic. Generally, we observe a growing amount of publicly available personal information, for instance, in public social networks. This trend definitely opens up new opportunities for network and expert analysis – also from emails as one source among others.

4. Conclusions

We have considered two main strands of today's email analysis. On the one hand, we have the processing of email contents using speech act theory and semantic technologies, which aim at a better automatic understanding and more formal and detailed representation of email contents. On the other hand, we have the analysis of entire email corpora that provide insight into the collaboration structures within organizations. Such analysis, however, can conflict with users' privacy concerns that have to be addressed. Moreover, we find that both strands show a tendency to grow together. This is not surprising since both strands concern the support of collaboration, from an individual and an organizational perspective, respectively. Semantic representations of email content go well together with network descriptions, as the idea of the social semantic desktop shows (e.g., in Nepomuk project). Information that can be obtained in this way constitutes a valuable organizational asset so that it is not surprising that enterprises become more and more interested in it. Privacy sets some boundary line to which extent this information can be exploited but we also observe changes in the general attitude towards privacy as well as various approaches to avoid privacy infringements. This indicates that we will see more opportunities for email integration and analysis in the future.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Whittaker, S., and Sidner, C. 1996. Email overload: Exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground* (CHI '96), Michael J. Tauber (Ed.). ACM, New York, NY, USA, 276-283. DOI= http://doi.acm.org/10.1145/238386.238530.

[2] Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. 2003. Taking email to task: the design and evaluation of a task management centered email tool. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (CHI '03). ACM, New York, NY, USA, 345-352. DOI= http://doi.acm.org/10.1145/642611.642672.

[3] Gwizdka, J. 2002. TaskView: design and evaluation of a task-based email interface. In *Proceedings of the 2002 conference of the Centre for Advanced Studies on Collaborative research* (CASCON '02), Darlene A. Stewart and J. Howard Johnson (Eds.). IBM Press.

[4] Searle, J. R. 1969. *Speech acts* Cambridge University Press, Cambridge, UK..

---

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

[5] Cohen, W. W., Carvalho, V. R., and Mitchell, T. M. 2004. Learning to classify email into "speech acts". In Proceedings of the *Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*.

[6] Carvalho V. R., and Cohen, W. W. 2005. 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* (SIGIR '05). ACM, New York, NY, USA, 345-352. DOI= http://doi.acm.org/10.1145/1076034.1076094

[7] Scerri, S., Davis, B., Handschuh, S., and Hauswirth, M. 2009. Semanta – semantic email made easy the semantic web: Research and applications. Vol. 5554 of *Lecture Notes in Computer Science*, chapter 7, pages 36-50. Springer, Berlin, Heidelberg.

[8] Lampert, A., Dale, R., and Paris, C. (2010). Detecting emails containing requests for action. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 984-992, Stroudsburg, PA, USA.

[9] McDowell, L., Etzioni, O., Halevy, A., and Levy, H. 2004. Semantic email. In *Proceedings of the 13th international conference on World Wide Web* (WWW '04). ACM, New York, NY, USA, 244-254. DOI= http://doi.acm.org/10.1145/988672.988706.

[10] Groza, T., Handschuh, S., Möller, K., Minack, E., Jazayeri, M., Mesnage, C., Reif, G., and Gudjónsdóttir, R. 2007. The NEPOMUK project - on the way to the social semantic desktop. *I-Semantics' 07*, pages 201-211, Graz.

[11] Scerri, S., Handschuh, S., and Decker, S. 2008. Semantic email as a communication medium for the social semantic desktop. In Bechhofer, S., Hauswirth, M., Hoffmann, J., and Koubarakis, M., editors, *The Semantic Web: Research and Applications*, vol. 5021 of *Lecture Notes in Computer Science*, chapter 12, pages 124-138. Springer, Berlin, Heidelberg.

[12] Riss, U. V., Jurisch, M., and Kaufman, V. 2009. E-mail in semantic task management. *IEEE Conference on Commerce and Enterprise Computing, Vienna, CEC '09*, pages 468-475, IEEE.

[13] Khoussainov, R. and Kushmerick, N. 2005. Email task management: An iterative relational learning approach. In *Second Conference on Email and Anti-Spam, CEAS 2005, Stanford, CA, USA*.

[14] Mulsant, J., Recource, G., and Vuillemot, R. 2011. Tomorrow's E-Mail: DLM 3.0 project's vision for the future of E-Mail in enterprises. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, volume 3, pages 377-380. IEEE.

[15] Burkhart, T., Werth, D., Balzert, S., and Loos, P. 2009. An Email-Based interoperability approach for SMEs. In *International Conference on Interoperability for Enterprise Software and Applications China, 2009. IESA '09.*, pages 245-251. IEEE.

[16] Burkhart, T., Werth, D., and, Loos, P. 2012. Context-sensitive Business Process Support Based On Emails. *In these proceedings*.

[17] Riss, U. V. 2012. TAPIR: Wiki-based Task and Personal Information Management Supporting Subjective Process Management. *Communications in Computer and Information 284* (forthcoming).

[18] Campbell, C. S., Maglio, P. P., Cozzi, A., and Dom, B. 2003. Expertise identification using email communications. In *Proceedings of the twelfth international conference on Information and knowledge management* (CIKM '03). ACM, New York, NY, USA, 528-531. DOI= http://doi.acm.org/10.1145/956863.956965

[19] Culotta, A., Bekkerman, R., and Mccallum, A. 2004. Extracting social networks and contact information from email and the web. Defense Technical Information Center.

[20] Diehl, C. P., Namata,G., and Getoor, L. 2007. Relationship identification for social network discovery. In *Proceedings of the 22nd national conference on Artificial intelligence - Volume 1* (AAAI'07), Anthony Cohn (Ed.), Vol. 1. AAAI Press 546-552.

[21] Schwartz, M. F., and Wood, D. C. M.. 1993. Discovering shared interests using graph analysis. *Commun. ACM* 36, 8 (August 1993), 78-89. DOI= http://doi.acm.org/10.1145/163381.163402

[22] Whittaker, S., Jones, Q., and Terveen, L. 2002. Contact management: identifying contacts to support long-term communication. *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 216–225. ACM, New York, NY. DOI= http://dx.doi.org/10.1145/587078.587109

[23] Bird, C., Gourley, A., Devanbu, P., Gertz, M., and Swaminathan, A. 2006. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories* (MSR '06). ACM, New York, NY, USA, 137-143. DOI= http://doi.acm.org/10.1145/1137983.1138016.

[24] Laclavik, M., Dlugolinkỳ, Å., Šeleng, M., Kvassay, M., Gatial, E., Balogh, Z., and Hluchý, L. 2011. Email analysis and information extraction for enterprise benefit. *Computing and Informatics*, 30(1):57-87.

[25] van der Aalst, W. M. P., and Nikolov, A. 2009. Mining E-Mail messages: Uncovering interaction patterns and processes using e-mail logs. In Sugumaran, V., editor, Methodological Advancements in Intelligent Information Technologies: Evolutionary Trends, pages 212-234. IGI Global, Hershey, PA.

[26] Ehrlich, K., Lin, C.Y., Griffiths-Fisher, V. 2007. Searching for experts in the enterprise: combining text and social network analysis. In: GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work, pp. 117–126. ACM, New York, NY. DOI= http://dx.doi.org/10.1145/1316624.1316642

[27] Guy, I., Jacovi, M., Meshulam, N., Ronen, I., and Shahar, E. 2008. Public vs. private: comparing public social network information with email. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (CSCW '08). ACM, New York, NY, USA, 393-402. DOI= http://doi.acm.org/10.1145/1460563.1460627 .

[28] Nardi, B.A.,Whittaker, S., and Schwarz, H. 2000. It's not what you know. It's who you know: Work in the Information Age. *First Monday* 5 (5).