Semi-automatic tagging for email

Dolata, Mateusz ; Jeners, Nils ; Prinz, Wolfgang

Abstract: Processing email messages is an essential part of personal information management. It is a complex task enforcing users to develop individual strategies, which are sometimes not in line with functionalities offered by various email clients. This paper ad- dresses these strategies and offers a tag-based solution for organizing mailboxes. The results of the evaluation confirm the desirability of the presented prototype for email structuring and retrieval. The discussion opens perspective on future developments.

Semi-Automatic Tagging for Email

Mateusz Dolata[1], Nils Jeners[2], Wolfgang Prinz[3]
University of Zurich[1], RWTH Aachen University[2], Fraunhofer FIT[3]
dolata@ifi.uzh.ch, nils.jeners@rwth-aachen.de, wolfgang.prinz@fit.fraunhofer.de

Abstract. Processing email messages is an essential part of personal information management. It is a complex task enforcing users to develop individual strategies, which are sometimes not in line with functionalities offered by various email clients. This paper addresses these strategies and offers a tag-based solution for organizing mailboxes. The results of the evaluation confirm the desirability of the presented prototype for email structuring and retrieval. The discussion opens perspective on future developments.

Introduction

Asynchronous, distributed communication plays an important role in daily work practice. *Email* remains the dominating professional communication paradigm, while its importance for informal correspondence has been decreasing. Email remains popular, even though or rather due to its simplicity (Prinz et al., 2009). The protocol is straightforward and implements the metaphor of sending a traditional letter. This openness of email is considered a reason for its popularity. According to Bellotti et al. (2003) people use their virtual mailboxes as: a calendar, a to-do repository, an archive, a contact list, and finally a message collector. Such observations lead to the definition of *email overload* (Whittaker and Sidner, 1996).

To cope with such problems, particular email clients extend the simple email metaphor. The focus of email processing moves towards proper structuring and efficient retrieval. The offered mechanisms do not always support the user, but demand his attention for preparatory filtering or good memory of wording for an *ad hoc* search. This paper assumes that users apply a mixture of strategies to facilitate the email processing. Field observations and interviews resulted in specific usage scenarios. They were further used to define and develop a prototype de-

Manuscript accepted for ECSCW '13, Work in Progress Session

scribed below, which enables observing how *semi-automatic tagging of messages leverages efficient email structuring and retrieval*. In particular, the following initial research questions are asked: (1) *what are the advantages of supported tagging for email processing*; (2) *do users accept and find it attractive to use such tags for email processing*. Future tests and refinements of the presented prototype shall yield a coherent requirements catalogue for design of semi-automatic approaches for email processing.

Related work

Email processing is addressed by a vivid discussion in the community. Particular studies range from understanding the role of email for communication till evaluation of practical systems.

Classification of the virtual correspondence was addressed in the past. While focusing on email as a communication channel, Winograd (1986) proposed a model based on Searle's Speech-Act Theory (1969). By convention, two standard email acts emerged: response and forwarding. Based on those message attributes, email clients support threading, without support for classification into speech acts. Other approaches for automatic email classification choose transactional activities (Dredze et al., 2006) or tasks (Bellotti et al., 2003) as a target domain.

While addressing email as a personal information management domain, several studies propose classification of activities. Venolia et al. (2001) suggest a model for email workflow consisting of: flow, triage, task management, archive, and retrieve. Based on a user study, they propose, among others, labels as a way to support users at archiving messages, triage, task management and retrieval. They also suggest automatic elicitation and suggestions regarding labelling. This solution has been partially implemented by Google's GMail™ , so that users are able to apply numerous tags to a simple message, without any further automatic support. Venolia et al. (2001), however, explicitly stress the role of supportive and intuitive UI as highly relevant for labelling. Different labelling approaches (with or without automatic processing) resulted from numerous research project, however did not find their way to the market (cf. Crawford et al., 2002; Segal and Kephart, 2000; Kerr and Wilcox, 2004). Matysiak Szóstek (2011) focuses on elicitation and dependencies between latent user needs. She follows a model consisting of just two activities: organization and retrieval. Message annotation seems to be the most relevant need for organization of virtual correspondence, whereas informative overview and flexible sorting play an important role during retrieval. In general, needs linked to retrieval are perceived as more important than those associated with organization. This sets the feeling of email overload in relation with processing older messages (cf. Dabbish and Kraut, 2006). Matysiak Szóstek (2011) provides numerous design requirements regarding email processing, including automatic maintenance of priority indications and linking between related messages, as well as flexible sorting according to people involved.

As discussed, email is the ultimate system in CSCW. However, email clients did not evolve over the past decades. Recently, productive systems appeared

which change the tradition, e.g. GMail™ , MailPilot, etc. The above review shows that academia is attracted by the topic of email management and provides evaluation results of numerous prototypes and broad studies on email usage. However, a clear and consistent catalogue of design principles for automatic or semi-automatic support of email processing is still missing.

Semi-Automatic Tagging

Development: Given the results of the literature review and insights from observations and interviews in form of usage scenarios, a concept for semi-automatic tagging of messages was developed. Particularly, tagging means to add tags to messages: either manually or automatically. Semi-automatic tagging in our prototype is realized by enabling easy and efficient changes to automatically generated tags. This, also, implies that the tag generator learns from examples.

The system generates tags for a respective message when it arrives. The decisions of the system are understandable and reproducible reflecting the content of the message. Also, the user has the possibility to change the behaviour of the system and adjust it to own needs. Consequently, the system does not only tag incoming messages, but also learns how to tag from the previously labelled messages. The desired functionality along with the insights from preliminary interviews leads to additional technical requirements. First, the program shall provide tags, even when no tags are available in the mailbox, i.e., no training data exists. Second, it shall adapt to user needs. Third, the system shall be robust and fast.

Under consideration of the above requirements, a hybrid solution was chosen to generate tags. Its essence lies in combination of heuristic and *machine learning* (ML) approaches. In particular, the algorithm combines information from linguistically motivated text processing and from a learnable keyword extractor when generating set of tags for a given messages. The heuristics rely on the extraction of nouns and named entities from the text. *Nouns* play an important role in transporting meaning, therefore filling variety of semantic roles in Indo-European languages (cf. Fillmore et al., 2003). The Stanford Part-Of-Speech-Tagger (Toutanova et al., 2003) is used to obtain nouns from the text. *Named entities* (NE) are phrases or words that refer to particular, unique entities (Sundheim, 1995). As they are mostly names of people, places or organization, they are assumed good candidates for message tags. The Stanford NE Recognizer (Finkel et al., 2005) is employed for extraction. In addition, results of learnable *key phrase extractor* from MAUI indexer (Medelyan and Witten, 2008) are heuristically combined with nouns and named entities and form a candidate set. Each candidate is assigned a weight depending on its frequency and character (noun vs. NE vs. key phrase). The weights change with number of tagged messages in the mailbox, such that the role of the machine learnable key phrase extractor grows with the number of available examples. Further processing, such as removal of stopwords and nearly duplicates, leverages the quality of the candidate set. Finally, the top ranked candidates are assigned as labels to the considered message.

User interface plays an extraordinary role in our approach. Not only the purely technical possibility to change a tag, but also the low burden related to this, stand for adjusting the tagging system to ones needs. It leverages the interaction with tags, makes the user more familiar with them, and finally raises the trust in system decisions. This paper addresses only tagging and not the design of email clients in general. Therefore, efforts were made to test the approach in a traditional, very common email client interface. The prototype presented here builds on top of Roundcube (0.7.2.). Figure 1 presents the user interface of the prototype.

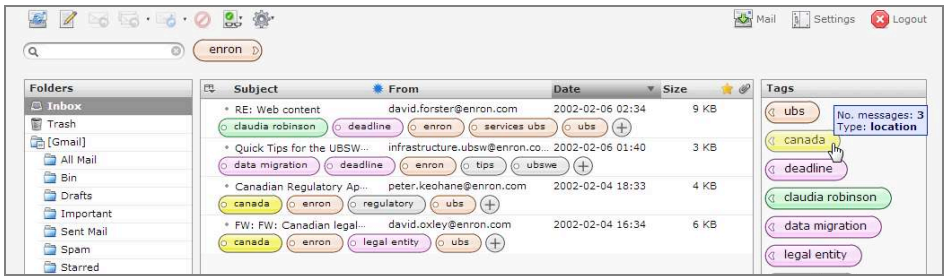

Figure 1. User interface of the prototype showing the toolbar, folders, tags, and messages with given tags.

The most obvious modification is the introduction of a separate frame on the right including all tags used for emails presented in the message list. Labels are ordered according to their frequency in the mailbox. In case the user wants to use tags for retrieval, a single click suffices to filter messages. Figure 1 presents the situation where filtering by tag "enron" was applied already. Choosing additional labels can further specify the search. For instance, if the filter was extended by tag "data migration", only the second message would remain in the view – tags assigned to messages are placed directly below their headers in the message list.

Colours of tags depend on their category (location, topic, time, etc.). Users are of course allowed to adjust them. For automatically generated tags categories are obtained through the NE Recognizer. It suffices to click the tag only once to reach a menu with tag operations, such as: renaming, deleting or category change. Opposite to email clients like GMail™, it is not necessary to define labels first before assigning to a message. Opening the "+" dialogue and providing a name suffices. If the name does not yet exist in the mailbox, a new label will be generated and added to the tag list. Otherwise, the message is assigned the already existing tag.

Evaluation: The evaluation aims at providing answers to the research questions. Since the areas approached by the questions (usability, acceptance and attractiveness) are tightly interwoven, the proposed test observes numerous variables, while giving the possibility to interact with the system and reflect on it.

For evaluating the system, an in-lab experiment with users was conducted. The user was asked to solve two basic tasks testing the usability of the system, such as tagging of two predefined messages, navigational search for a message and summarizing a message given its tags. Between the tasks, short interview was incorporated to collect additional opinions. Finally, data regarding acceptance and attractiveness of the system were collected through UTAUT (Venkatesh et al.,

2003) and AttrakDiff2 (Hassenzahl et al., 2003) questionnaires. All 14 participants, aged 24-59, are frequent email users and merely do not use tagging. Only three participants of the study use it for their main professional mailbox.

The result of the tagging task shows that the tag generator in its original mode makes its predictions with high accuracy measures (0.86 recall, 0.73 precision). The opinions regarding the tagger itself are very positive, but due to the task setting users feel encouraged to change tags. They appreciate the easiness of changing a tag, while seeking faster access to the *remove* command. Indeed, there is a strong tendency towards removal, compared to renaming and adding tags (22%, 5%, 7% respectively). Filtering tests again show vivid user interest and acceptance, even though performance values for tag-based search do not significantly differ from those for query-based search. The average number of clicks, scrolls and typed signs required for finding the desired message is similar with slight tendency towards the tag-based solution (60 vs. 69 operations). Finally, the last assignment yields to the conclusion that tags facilitate message summarization. 10 out of 14 participants can provide full summary and explain the meaning of tags in the context. Three other participants forget to mention one important characteristic. Comparison with other "summarization" paradigms, such as snippets containing first two lines of the message, could provide further insights.

The results of the acceptance and attractiveness questionnaires enable further conclusions on semi-automatic tagging. The UTAUT provides very positive values regarding performance and effort expectancy (5.3 and 6.1 out of 7 respectively). In other words, users anticipate the system to fulfil their needs without requiring much work from them. It is in line with the tendency to assist the user at structuring while providing easy-to-use paradigms. The results of the AttrakDiff2 also confirm the high pragmatic value of the proposed solution (1.3 on a scale ranging from -3 to 3). The general attractiveness reaches the same level, while the hedonic quality is graded 0.8, thus suggesting further improvement regarding, e.g., visual elements and speed, as confirmed in the interviews.

Discussion

This paper shows the drawbacks of the most popular methods for email structuring and retrieval. It aims at launching an intensive research path on semi-automatic support of email processing. It also shows how such a paradigm can be implemented it into daily practice, while extending existing email client with novel functionality. The results of the final evaluation enable observations on positive user's attitude towards the introduced solution, as well as its usability for common email tasks. All this leads to the conclusion, that semi-automatic tagging facilitates easier and efficient structuring and retrieval of messages in the mailbox. Therefore, development of further prototypes, while following the Usability Engineering approach by Rosson and Carroll (2002), will be continued in order to establish a catalogue of relevant and generalizable design principles for semi-automatic email processing.

References

Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. (2003): 'Taking email to task', in Proc. Conf. Human Factors in Computing Systems - CHI, pp. 345.

Crawford, E., Kay, J., and McCreath, E. (2002): 'An intelligent interface for sorting electronic mail', in Proc. Intl. Conf. Intelligent User Interfaces - IUI, pp. 182–183.

Dabbish, L.A. and Kraut, R.E. (2006): 'Email overload at work: an analysis of factors associated with email strain', in Proc. Conf. Comp. Supported Cooperative Work - CSCW, pp. 431–440.

Dredze, M., Lau, T., and Kushmerick, N. (2006): 'Automatically classifying emails into activities', in Proc. Intl. Conf. Intelligent User Interfaces - IUI, pp. 70.

Fillmore, C.J., Johnson, C.J., and Petruk, M.R.L. (2003): 'Background to Framenet', Intl. J. of Lexicography, vol. 16, no. 3, pp. 235–250.

Finkel, J.R., Grenager, T., and Manning, C. (2005): 'Incorporating non-local information into information extraction systems by Gibbs sampling', in Proc. Conf. Assoc. for Comp. Linguistics - ACL, pp. 363–370.

Hassenzahl, M., Burmester, M., and Koller, F. (2003): 'AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität', Mensch & Computer, pp. 187.

Kerr, B. and Wilcox, E. (2004): 'Designing remail: reinventing the email client through innovation and integration', in Extended Abstracts. Conf. Human Factors in Computing Systems - CHI, pp. 837.

Matysiak Szóstek, A. (2011): '`Dealing with My Emails': Latent user needs in email management', Computers in Human Behavior, vol. 27, no. 2, pp. 723–729.

Medelyan, O. and Witten, I.H. (2008): 'Domain-independent automatic keyphrase indexing with small training sets', J. of the American Society for Information Science and Technology, vol. 59, no. 7, pp. 1026–1040.

Prinz, W., Jeners, N., Ruland, R., and Villa, M. (2009): 'Supporting the Change of Cooperation Patterns by Integrated Collaboration Tools', in L.M. Camarinha-Matos et al. (eds.): Leveraging Knowledge for Innovation in Collaborative Networks, pp. 651–658.

Rosson, M.B. and Carroll J.M. (2002): Usability Engineering: Scenario-Based Development of Human-Computer Interaction. Morgan Kaufmann.

Searle, J.R. (1969): Speech acts: An essay in the philosophy of language, Cambridge Univ. Press.

Segal, R. and Kephart, J.O. (2000): 'Incremental Learning in SwiftFile', in Proc. Intl. Conf. Machine Learning - ICML, pp. 863–870.

Sundheim, B.M. (1995): 'Overview of results of the MUC-6 evaluation', in Proc. Message Understanding Conf. - MUC, pp. 13.

Toutanova, K., Klein, D., Manning, C.D., and Singer, Y. (2003): 'Feature-rich part-of-speech tagging with a cyclic dependency network', in Proc. Conf. North American Ch. of the Association for Computational Linguistics on Human Language Technology - NAACL, pp. 173–180.

Venkatesh, V., Morris, M.G., Davis, G.B., and Davis, F.D. (2003): 'User Acceptance of Information Technology: Toward a Unified View', MIS Quarterly, vol. 27, no. 3, pp. 425–478.

Venolia, G.D., Dabbish, L., Cadiz, J.J., and Gupta, A. (2001): Supporting Email Workflow: Technical Report MSR-TR-2001-88, Microsoft Research, Redmond and WA.

Whittaker, S. and Sidner, C. (1996): 'Email overload: exploring personal information management of email', in Proc. Conf. Human Factors in Computing Systems - CHI, pp. 276–283.

Winograd, T. (1986): 'A language/action perspective on the design of cooperative work', in Proc. Conf. Computer-Supported Cooperative Work - CSCW, pp. 203.