

NRC Publications Archive Archives des publications du CNRC

Improving the Usability of Low Precision Text Analysis Systems Through The Use of Alternative Presentation Techniques Crease, Murray

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version.
/ La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

Publisher's version / Version de l'éditeur:

*Proceedings of the 5th Annual Canadian Symposium on Text Analysis
(CaSTA'06), October 12, 2006, Fredericton, New Brunswick, Canada, 2006*

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
<https://nrc-publications.canada.ca/eng/view/object/?id=a22964bf-01d1-4844-bc78-f193a218ed8d>
<https://publications-cnrc.canada.ca/fra/voir/objet/?id=a22964bf-01d1-4844-bc78-f193a218ed8d>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at
PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



National Research
Council Canada

Institute for
Information Technology

Conseil national
de recherches Canada

Institut de technologie
de l'information

NRC-CNRC

*Improving the Usability of Low Precision
Text Analysis Systems Through The Use of
Alternative Presentation Techniques **

Crease, M.
October 2006

* Proceedings of CaSTA06. The 5th Annual Canadian Symposium on Text Analysis. October 12, 2006. Fredericton, N.B.. NRC 48810.

Copyright 2006 by
National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables from this report, provided that the source of such material is fully acknowledged.

Improving the Usability of Low Precision Text Analysis Systems Through The Use of Alternative Presentation Techniques

Murray Crease,

National Research Council Institute for Information Technology

46 Dineen Drive,

Fredericton,

Canada E3B 9W4.

murray.crease@nrc-cnrc.gc.ca

Abstract

Forgetting to attach an attachment to an outgoing e-mail message is at best embarrassing yet the detection of such mistakes is problematic. Existing systems parse a message's body for keywords but typically have limited effectiveness due to their high precision and low recall. An analysis of a small corpus of emails showed that parsing the text of emails- even when supplemented by rules that parsed the high level attributes of the messages – only enabled a high recall at the expense of low precision. A system is described that utilises such rules calibrated for each user to determine the likelihood – or system confidence - of each rule being correct when fired. The form of user alert is modified according to the system confidence in the rule(s) fired allowing the system recall to remain high without affecting the system's usability.

1 INTRODUCTION

Forgetting to attach an attachment to an outgoing e-mail message typically occurs because the sender reaches closure when the message text is complete but before the attachment has been attached. This paper investigates how to detect the non-existence of an attachment and how best to alert users. A survey of existing systems (e.g. <http://www.attachmentglue.com/>) showed that all used a similar approach: parse outgoing messages for certain keywords (e.g. 'attached' and 'enclosed') and if detected display an alert box. This approach has a high precision but the associated low recall minimises its effectiveness. The high precision of these systems is necessary as the alert boxes used would prove annoying if they were regularly wrong. This paper describes a system that employs alternative presentation techniques to allow the recall to be increased without an associated decrease in usability caused by the decrease in precision.

2 DETECTING MISSING ATTACHMENTS

An initial set of rules was generated by analysing messages with attachments. These rules were then iteratively improved by applying them to users' Inboxes with the aim of maximising the system's recall while the precision of individual rules was used to determine system 'confidence' in the rules. An analysis of three users' "Sent Items" mailboxes showed that the individual rules had low recall with varying precision but combined to give the system a high recall and low precision. This paper proposes that such a system can be made usable through the use of alternative feedback mechanisms.

3 ALERTING USERS

Sellen *et al.* described some of the dimensions upon which feedback can be characterised [1]. The dimensions described included:

- *demanding* versus *avoidable*: demanding feedback is hard to avoid whereas users can choose to ignore the latter. An example of demanding feedback is a popup dialog box and of avoidable feedback is a system tray icon.
- *transient* versus *sustained*: sustained feedback persists whereas transient feedback does not

This paper proposes that structured, non-speech audio (or earcons [2]) should be used to either augment or in some cases replace demanding, visual alerts such as alert boxes. It is hoped that the use of this alternative form of feedback, combined with the high overall recall, will improve the usability of this low precision system.

4 IMPLEMENTATION

A prototype of the system was implemented in Java. The first time the user runs the application they are required to calibrate the rules by running them across their ‘Sent Items’ mailbox (Figure 1). This allows the system to determine the effectiveness of the different rules for that particular user. An evaluation is planned to determine what combinations of feedback enable the low precision rules to be deemed usable.

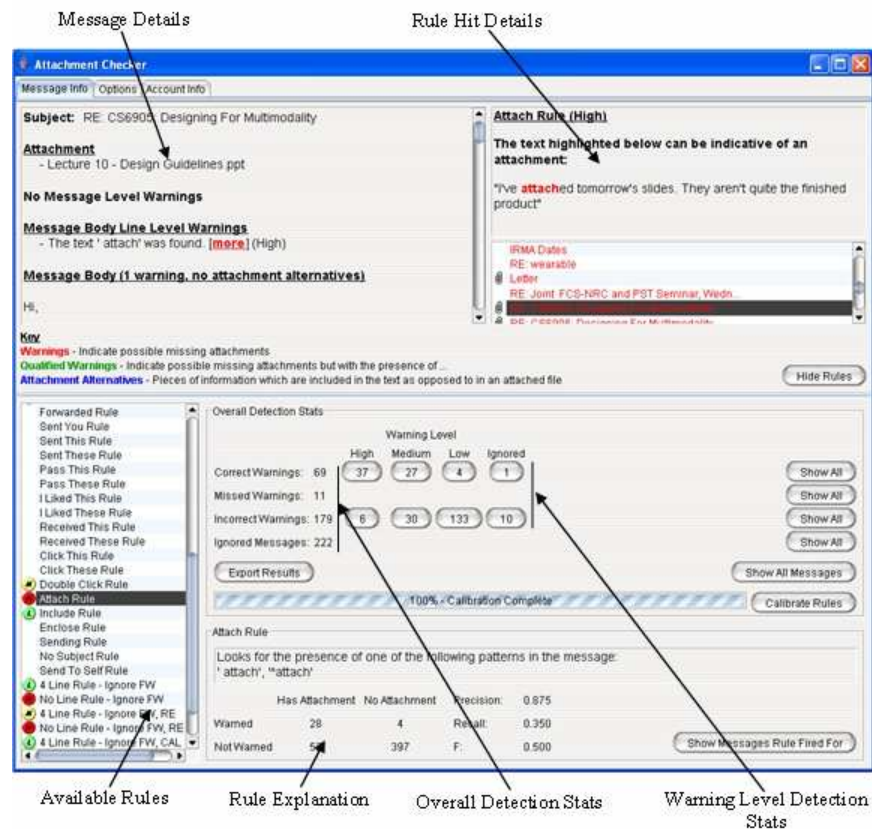


Figure 1: Rule Calibration interface indicating the feedback given to the users regarding the rules. It is hoped that this feedback will make the users trust the system as discussed in [3].

5 CONCLUSIONS

This paper outlines the problems surrounding the detection of emails which should have attachments but do not. An analysis of a small corpus of emails determined that textual parsing alone is insufficient to

detect all missing attachments and even examining higher level details of the message such as message length will not enable 100% recall. A system to detect missing attachments with a high recall but low precision is described. It is hoped that a planned evaluation will show that by employing alternative alerting techniques the low precision of the system does not make it unusable. If successful, this approach could be increase the usability of other low precision text analysis systems.

6 ACKNOWLEDGEMENTS

I would like to thank Anna MacLachlan for her much needed assistance in the field of natural language parsing without which none of this work would have been possible.

REFERENCES

- [1] A. J. Sellen, G. P. Kurtenbach, and W. A. S. Buxton, "The Prevention of Mode Errors Through Sensory Feedback," *Human-Computer Interaction*, vol. 7, pp. 141-164, 1992.
- [2] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," *Human-Computer Interaction*, vol. 4, pp. 11-44, 1989.
- [3] T. Roth-Berghofer, "Explanation and Case-Based Reasoning: Foundational Issues," in *European Conference on Case-Based Reasoning ECCBR 2004*. Madrid, Spain, 2004.