

A NEW EMAIL RETRIEVAL RANKING APPROACH

Samir AbdelRahman¹, Basma Hassan² and Reem Bahgat¹

¹Department of Computer Science, Faculty of Computers and Information, Cairo University, Giza, Egypt

s.abdelrahman@fci-cu.edu.eg , r.bahgat@fci-cu.edu.eg

²Department of Computer Science, Faculty of Computers and Information, Fayoum University, Fayoum, Egypt

bhk00@fayoum.edu.eg

Abstract:

Email Retrieval task has recently taken much attention to help the user retrieve the email(s) related to the submitted query. Up to our knowledge, existing email retrieval ranking approaches sort the retrieved emails based on some heuristic rules, which are either search clues or some predefined user criteria rooted in email fields. Unfortunately, the user usually does not know the effective rule that acquires best ranking related to his query. This paper presents a new email retrieval ranking approach to tackle this problem. It ranks the retrieved emails based on a scoring function that depends on crucial email fields, namely subject, content, and sender. The paper also proposes an architecture to allow every user in a network/group of users to be able, if permissible, to know the most important network senders who are interested in his submitted query words. The experimental evaluation on Enron corpus prove that our approach outperforms known email retrieval ranking approaches..

Keywords : *Email Ranking, Email Fields, Email Threading, Scoring Function, and Email Network Architecture*

Published In:

International journal of computer science & information Technology (IJCSIT) Vol.2, No.5, October 2010
DOI :

References:

- [1] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- [2] Xiang, Y. (2009). Managing Email Overload with an Automatic Nonparametric Clustering System. *The Journal of Supercomputing* , 48 (3), 227 - 242.
- [3] Cselle, G., Albrecht, K., & Wattenhofer, R. (2007). BuzzTrack: Topic Detection and Tracking in Email. *In Proceedings of the 12th international conference on Intelligent user interfaces* (pp. 190 - 197). Honolulu, Hawaii, USA: ACM.
- [4] Erera, S., & Carmel, D. (2008). Conversation Detection in Email Systems. *In Proceedings of the IR research, 30th European conference on Advances in information retrieval* (pp. 498-505). Glasgow, UK: Springer-Verlag.
- [5] Castro, P., & Lopes, A. (2009). Magnet Mail: A Visualization System for Email Information Retrieval. *In Proceedings of the 10th International Symposium on Smart Graphics* (pp. 213-222). Salamanca, Spain: Springer-Verlag.
- [6] Yee, J., Mills, R. F., Peterson, G. L., & Bartczak, S. E. (2009). Automatic Generation of Social Network Data from Electronic-Mail Communication. *In Proceedings of the 10th ICCRTS*. Virginia.
- [7] Enron Email Dataset, <http://www.cs.cmu.edu/~enron/>
- [8] Perkiö, J., Tuulos, V., Buntine, W., & Tirri, H. (2005). Multi-Faceted Information Retrieval System for Large Scale Email Archives. *In Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 557 - 564). France: IEEE Computer Society.
- [9] Lopez, V., Castillo, C., & Codina, J. (2003). *Information Retrieval in Mail Archives*. Technical

Report, Càtedra Telefónica de Producció Multimedia, Universitat Pompeu Fabra.

- [10] Weerkamp, W., Balog, K., & de Rijke, M. (2009). Using Contextual Information To Improve Search In Email Archives. *In Proceedings of the 31st European Conference on Information Retrieval Conference* (pp. 400-411). Toulouse, France: Springer.
- [11] Yeh, J.-Y., & Harnly, A. (2006). Email Thread Reassembly Using Similarity Matching. *In Proceedings of the Third Conference on Email and Anti-Spam (CEAS 2006)*. California, USA.
- [12] Broder, A. Z., Eiron, N., Fontoura, M., Herscovici, M., Lempel, R., Mcpherson, J., et al. (2006). Indexing Shared Content in Information Retrieval Systems. *In Proceedings of the 10th International Conference on Extending Database Technology* (pp. 313-330). Munich, Germany: Springer.
- [13] Shetty, J., & Adibi, J. (2004). *The Enron Email Dataset Database Schema and Brief Statistical Report*. Technical Report, Information Sciences Institute , University of Southern California.
- [14] UC Berkeley Enron Email Analysis, http://bailando.sims.berkeley.edu/enron_email.html
- [15] Berry, M. W., Browne, M., & Signer, B. (2007). 2001 Topic Annotated Enron Email Data Set. Philadelphia: Linguistic Data Consortium.
- [16] Järvelin, K., & Kekäläinen, J. (2002). Cumulated Gain-based Evaluation of IR Techniques. *ACM Transactions on Information Systems (TOIS)* , 20 (4), 422 - 446.