

{tag} International Journal of Computer Applications  
Foundation of Computer Science (FCS), NY, USA

[Volume 144](#)

-  
[Number 6](#)

Year of Publication: 2016

Authors:

G. Girija Rani, M. Indra Sena Reddy

10.5120/ijca2016910399

{bibtex}2016910399.bib{/bibtex}

## Abstract

Emails are parts of everyday life. These messages have become increasingly important and widespread method of communication because of its time speed, where the amount of email messages received per day can range from tens for a regular user to thousands for companies. Everyone is overwhelmed with emails, including relational (structured) and non-relational (semi-structured or non-structured), quite a bit of which is repetitive, stale and of drastically differing quality. This large quantity is confounded. Not just spam messages are thought to be 'garbage', additionally undesirable messages (e.g. advertisements, lottery) individuals' waste a lot of time unknowingly by surfing them. So there is much need to categorization of Emails. Classification can help to meet lawful and administrative necessities for recovering particular data inside of a set time span, and this is frequently the inspiration driving implementing data classification. This paper aims at examining on ways doing supervised and unsupervised grouping of messages as per email content.

## References

1. Bryan Klimt, Yiming Yang. Introducing the Enron Corpus.
2. Ron Bekkerman. Automatic Categorization of Email into Folders: Benchmark Experiments on Enron and SRI Corpora. 2004.
3. “A SURVEY OF TEXT CLASSIFICATION ALGORITHMS” chapter 6 by Charu C. Aggarwal
4. “Emails classification by data mining techniques” by Mohammed A. Naser, Athar H. Mohammed Department of Computer, College of Sciences for Women, University of Babylon.
5. “Data Mining: concepts and Techniques” book by Jiawei Han and Micheline Kamber.
6. Text Classification: The Naïve Bayes algorithm Adapted from Lectures by Prabhakar Raghavan (Yahoo and Stanford) and Christopher Manning (Stanford).
7. [http://en.wikipedia.org/wiki/Naive\\_bayes](http://en.wikipedia.org/wiki/Naive_bayes).
8. “Is Naïve Bayes a Good Classifier for Document Classification?” by S.L. Ting, W.H. Ip, Albert H.C. Tsang, International Journal of Software Engineering and Its Applications Vol. 5, No. 3, July, 2011.
9. “An Improved Naive Bayes Text Classification Algorithm In Chinese Information Processing” by Lingling Yuan1 Jiaozuo, P. R. China, 14-15, August 2010, pp. 267-269 ISBN 978-952-5726-10-7.
10. “Enhanced Classification Accuracy on Naive Bayes Data Mining Models” by Md. Faisal Kabir & Chowdhury Mofizur Rahman, Alamgir Hossain, Keshav Dahal, International Journal of Computer Applications (0975 – 8887) Volume 28– No.3, August 2011.
11. Agarwal, R., Imielinski, T. and Swami, A. (1993) ‘Database Mining: A Performance Perspective’, IEEE: Special issue on Learning and Discovery in Knowledge Based Databases, pp. 914-925.
12. Agarawal, R. and R. Srikant, (1994) ‘Fast algorithms for mining association rules’, Proceedings of the 20th International Conference on Very Large Data Bases, San Francisco, CA., USA., pp: 487-499. [13] Maindonald, J. H. (1999) ‘New approaches to using scientific data statistics, data mining and related technologies in research and research training’ Occasional Paper 98/2, The Graduate School, Australian National University.
13. Quinlan, J. (1986), “Induction of Decision Trees,” Machine Learning, vol. 1, pp.81-106.
14. Berson, A., Smith, S. J. and Thearling, K. (1999) Building Data Mining Applications for CRM McGraw-Hill.

## **Index Terms**

Computer Science

Information Sciences

## **Keywords**

Supervised, unsupervised, classification.