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.
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).

Index Terms

Computer Science

Information Sciences

Keywords

Supervised, unsupervised, classification.