

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Email: An Easy Messaging Service

Sangeeta Maharana¹

Computer Department
SP's Institute Of Knowledge College Of Engineering,
Pimple Jagtap, Pune.

Minal Mohite²

Computer Department
SP's Institute Of Knowledge College Of Engineering,
Pimple Jagtap, Pune.

Pornima Wadekar³

Computer Department
SP's Institute Of Knowledge College Of Engineering,
Pimple Jagtap, Pune.

Abstract: Email system was first developed for ARPANET in 1970. Then later on, it became the widely used messaging service on internet. It faced many milestones in the earlier stage for implementation like timesharing computer, readmail, sendmail, RD, NRD etc. But with time, these milestones were overcome with new technologies. Now emails are used in every sector of life. It is used in industries, companies, organisation, schools and colleges etc. The main techniques involved in email are POP3, IMAP, MAPI, SMTP etc. These are the main protocols used in providing various services of emails. Emails can be clustered to make it more efficient for retrieval of mails. Clustering can be done by various algorithms, clustering merges the documents that share many properties and the time complexity is decreased to logarithmic search. K-means, Suffix tree array and lingo. K-means is a powerful algorithm for formation of clusters, web search results Semantic. Suffix tree array generates clusters with the help of tree formation technique. The distinctive methodology of the SSTC algorithm is that it simultaneously constructs the semantic suffix tree through an on-depth and on-breadth pass by using semantic similarity and string matching. SSTC uses only subject-verb-object classification to generate clusters and readable labels. The algorithm also implements directed pruning to reduce the sub-tree sizes and to separate semantic clusters. Experimental results show that SSTC has better performance than conventional Suffix Tree Clustering (STC). Index Terms semantic search results clustering, semantic clustering, semantic suffix tree clustering, text clustering. Lingo is also a very powerful clustering algorithm. It generates semantic labels. This paper is proposed with an idea to give brief knowledge on what an email is and the various techniques used by the existing email system and the technology which can improve the existing system.

Keywords: Email, SMTP, POP3, IMAP, LINGO

I. INTRODUCTION

Email is an electronic messaging service provider. Email has become one of the driving forces behind connecting businesses to the Internet. It offers fast, economical transfer of messages anywhere in the world. Mankind has always had a compelling desire to communicate. In ancient times this could be verbally or in some form of writing. If remote communication was required (i.e. if the parties were not physically together) then messages had to be physically carried or sent by a messenger. Examples of early forms of remote transmission of messages not requiring a person to actually move between the sender and the receiver would be in 'jungle drum' or 'smoke signal' transmissions. These were somewhat lacking in security and privacy. "The number of worldwide email accounts is expected to increase from an installed base of 3.1 billion in 2011 to nearly 4.1 billion by year-end 2015". Yarow (2011) reported that 1.9 billion emailers sent 107 trillion emails in the first quarter of 2010, on average 294 billion emails per day.

II. RELATED WORK

a) *History of email*

1. **Simple office email system:** Email is required within a company, but not out to the rest of the world. A very simple email system could be installed and maintained, giving interoffice communications:

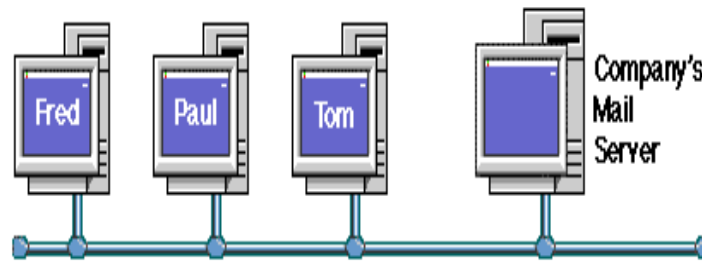


Fig 1. Simple office email system

In the above example, the three workstations are connected to a computer network within a company office. If one user wishes to send email to another user, then the message is simply typed and sent to the mail server, addressed to the recipient using their email name, which would simply be the first name of a user, such as "Mary". Email is required within a company, but not out to the rest of the world. A very simple email system could be installed and maintained, giving interoffice communications.

2. Remote user with access to office email:

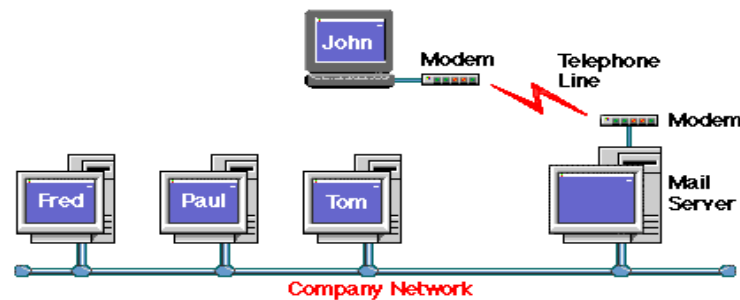


Fig 2. Remote user with access to office email

When John wants to send email to Paul, he types his message within his email client, then, when he wants it to, his email client dials into his company computer network using a modem and telephone line, sends his message to Paul, then collects any waiting messages for him. Once the messages have been collected, the modem disconnects from the phone line, and John can read any messages that were collected.

b) *Existing System*

A number of *users* on a network dialling in to an ISP

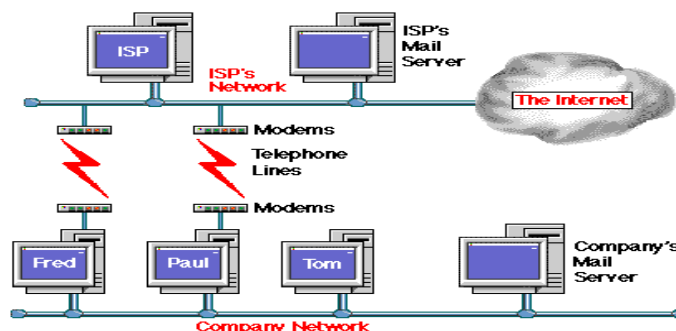


Fig 3. Company network

In this example Fred and Paul have two email addresses: One for internal mail within a company, and one for Internet email. This can sometimes occur if most of the email that a user reads or sends is internal within a company network, yet the user wants access to global Internet email. Each user would have an email account on both the company mail server and the ISP's mail server. Techniques used by Existing system [1]

Port No	Protocols/ Services	Description
25	SMTP SMTP e-mail server	Simple Mail Transfer Protocol core Internet protocol used to transfer from client to server (MUA to MTA) and server to server (MTA to MTA)
110	POP3 POP e-mail server	Post Office Protocol allows clients (MUA's) to retrieve stored e-mail .
143	IMAP IMAP(4) e-mail server	Internet Message Access Protocol provides a means of managing e-mail messages on a remote server and retrieve stored e-mail .
465	SMTSP WSMTP (SSMTP) protocol over TLS SSL	SMTP via SSL encrypted connection (Unofficial)
993	IMAPS SSLencrypted IMAP	IMAP via SSL encrypted connection
995	POP3S SPOP SSL encrypted POP	POP via SSL encrypted connection
587	MSA	Outgoing Mail (Submission)
80	HTTP	Webmail
443	HTTPS	Secure Webmail

Table 1. Commonly used Ports in E-mail Communication

III. PROPOSED SYSTEM

Proposed system gives all the facilities of existing system i.e we can send, receive, forward and can do all the operations of email. The advanced feature is the automatic generation of clusters and labels using lingo algorithm. Existing system also generates labels ,but that is not very effective for use.

Lingo Algorithm[3]

The major steps in this algorithm are:

- 1)Preprocessing
- 2)Frequent Phrase Extraction
- 3)Cluster Label Induction
- 4)Cluster Content Discovery
- 5)Final Cluster Formation

This algorithm then forms clusters by using the steps and organise the mails in the generated labels.Labels are useful in finding the particular mail,and saves time.Indexing,Subphases of LINGO, providing rationale behind the design decisions we had made. We also discuss several implementation and efficiency issues.When designing a web clustering algorithm, special attention must be paid to ensure that both contents and description (labels) of the resulting groups are meaningful to the users.Identities used in E-mail are globally unique and are: mailbox , domain name , message-ID and ENVID . Mailboxes are conceptual entities identified by e-mail address and receive mail. E- mail address has become a common identity identifier on the Internet.

IV. EMAIL ARCHITECTURE

The proposed Email architecture is shown in below.

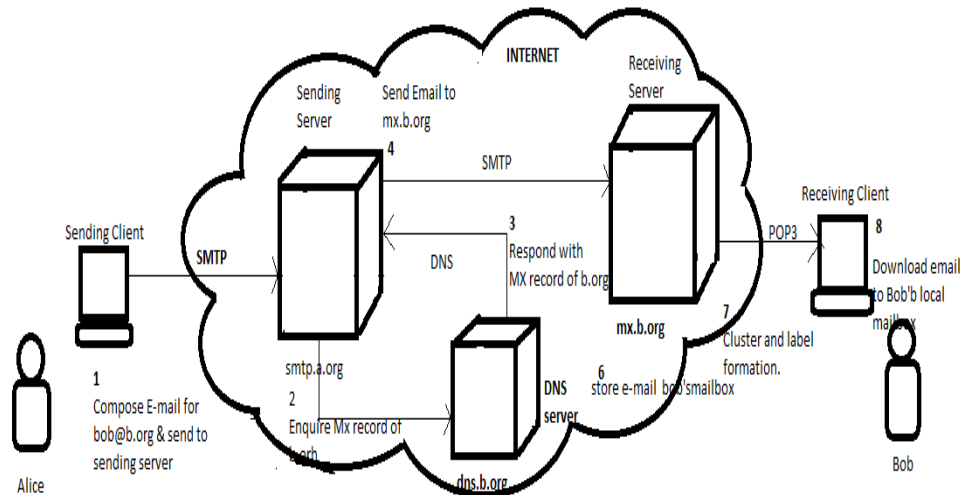


Fig 4. EmailArchitecture

E-mail system is an integration of several hardware and software components, services and protocols which provide interoperability between its users and among the components along the path of transfer.

V. APPLICATION

a) Automatic Email System

The automatic email answering problem can be described as finding approaches to analyze the incoming emails, and then reply them with appropriate answers automatically.

b) Adding Email Social Features

Adding social features to email services is an interesting, yet non-trivial task. It has already attracted lots of attention from large email service providers. According to Delaney and Varal (2007), "the biggest Web email services, including Yahoo Inc., Microsoft Corp. and Time Warner Inc.'s AOL unit, are adding features that allow users to perform such sociable functions as tracking friends".

VI. CONCLUSION

E-mail is a widely used and highly distributed application involving several actors that play different roles. These actors include hardware and software components, services and protocols which provide interoperability between its users and among the components along the path of transfer. It illustrated logical e-mail architecture and underlining various core components, modules and protocols used in the system. It presents the meta-data contained in e-mail message and various techniques used for e-mail forensics. The paper also introduces several clustering algorithm that have functionalities to automatically analyse e-mail and produce labels and clusters.

ACKNOWLEDGEMENT

We hereby are thankful to all the members involved in the completion of paper. The one who guided us at each and every step. First of all, I wish to express my sincere appreciation to Mr. Pratap Singh for enlivening our interest in clustering and for the guidelines and advice he gave us throughout the development of the paper. I would also like to thank Ms. Saba Siraj, project incharge, for her valuable insights and observations. We wish to pay special gratitude to our principal Dr. R.S. Jahagirdar and HOD Prof. Ritesh Thakur. Thanks for their time and effort devoted to evaluation of LINGO. Finally, it is difficult to explain how

grateful I am to my nearest and dearest whose endless patience and personal involvement helped me to complete this challenging venture.

References

1. Email Mining: Tasks, Common Techniques, and Tools Guanting Tang, Jian Pei, and Wo-Shun Luk School of Computing Science, Simon Fraser University, Burnaby BC, CANADA.
2. LUIA FILIPE DA CRUZ NASSIF AND EDUARDO RAUL HRUSCHKA “Document Clustering for Forensic Analysis: An Approach for Improving Computer Inspection.” *IEEE transactions on information forensics and security*, 1, January 2013, vol. 8, no .
3. Decomposition Stanis law Osin´ski, Jerzy Stefanowski, and Dawid Weiss “Lingo: Search Results Clustering Algorithm Based on Singular Value” Institute of Computing Science, Poznan´ University of Technology, ul. Piotrowo 3A, 60–965 Poznan´, Poland, Email: stanislaw.osinski@man.poznan.pl, {jerzy.stefanowski,dawid.weiss}@cs.put.poznan.pl .
4. Peter Hannappel, Reinhold Klapsing, and Gustaf Neumann, “MSEEC—a multi search engine with multiple clustering.” Proceedings of the 99 Information Resources Management Association Conference , May 1999.
5. Zhang Dong “Towards Web Information Clustering”, PhD thesis, Southeast University, Nanjing, China, 2002 .
6. Irmina Mas lowska, “Phrase-Based Hierarchical Clustering of Web Search Results”, In Proceedings of the 25th European Conference on IR Research, ECIR2003, volume 2633 of Lecture Notes in Computer Science, pages 555–562, Pisa, Italy, 2003. Springer.
7. Stanislaw Osinski, Jerzy Stefanowski and Dawid Weiss, “Lingo: Search Results Clustering Algorithm Based on Singular Value Decomposition”, Institute of Computing Science, Pozna´n University of Technology, ul. Piotrowo 3A, 60–965 Pozna´n, Poland .
8. Stanislaw Osinski and Dawid Weiss, Poznan University, “A concept Driven Algorithm For Clustering Search Result”, 2005 IEEE.
9. Claudio Carpinato, Stanislaw Osinski and Dawid Weiss, “A Survey Of Web Clustering Engines”, 2009 ACM.
10. Dawid Weiss and Jerzy Stefanowski. “Web search results clustering in Polish: Experimental evaluation of Carrot “. In Proceedings of the New Trends in Intelligent Information Processing and Web Mining Conference, Zakopane, Poland, 2003.
11. Oren Zamir and Oren Etzioni. Grouper: “a dynamic clustering interface to Websearch results”. *Computer Networks (Amsterdam, Netherlands: 1999)*, 31(11–16):1361–1374, 1999.