

International Journal of Computer Architecture and Mobility (ISSN 2319-9229) Volume 3 -Issue 3, May 2015 Web Based Advanced Email Filter System Using Text Mining

Shantilal Bhayal¹, Lalit Gehlod²

¹Lecturer, ²Assistant Professor

¹CDGI, ²DAVV

Indore, Pin no. 452001

slbhayal@gmail.com, lalit.gehlo@gmail.com

Abstract: The internet has become popular, since it is being used for many purposes. Today internet has brought a globe in a single room. Right from news across the corner of the world, wealth of knowledge to shopping, purchasing the tickets, everything is at finger tips. By using internet a person sitting on any part of world can be contacted easily. Facilities of email have been availed for achieving better communication. Email is now an essential communication tool in business and is also excellent for keeping in touch with family and friends. In this paper, we propose to build a software product to automate the mailing system. This product retrieves, filter, organizes, stores and analyzes the email and their attachments of the recipient's personal email account. Hence, this product is named as "Web based Advanced Email Filter System Using Text Mining".

Keywords- information management; periodic access; mail organizer; email client; text mining; EIA algorithm.

1. INTRODUCTION

In the current scenario, executives and officials are dealing with the busiest schedules at their workplace They are the most prominent internet users around the globe. The main difficulties they face are:

- Maintaining multiple email accounts.
- Accessing the email accounts regularly and organizing them according to the content.
- Manually managing of the emails on the server.
- Need to access email accounts on server again and again to download emails and attachments.

The situation may result

- Delay in work with deadlines (such as bank statements, IT return etc.).
- May not be able to attend the events (personal/official) on time.

Finally, degradation in performance and reputation at both

professional and social front due to not getting the right information at right time.

In this work, we intend to build a software product which automates the mailing system. This product retrieves/downloads the emails automatically from the multiple user accounts and arranges them in the respective preconfigured folders and maintains response status of the email. Also, it extracts the event information (proposed and planned meetings, announced upcoming events, etc.) from

the downloaded emails. Hence, this product is named as Advanced Email Filter Using Text Mining.

2. Background

The work in the field has been evolved in the recent years. Many organizations have worked over it and came up with their products [1]. The products which are related to automate the email system can be classified into three categories depending on their functionality to serve the email. These three categories are 1) Email Notifier, 2) Email Organizer, 3) Mail Delivery Agents.

• Email Notifier

Notifies about the arrival of new emails [2].

• Email Organizer

Organizes email account on the client machine and arranges them.

• Mail Delivery Agent

A mail delivery agent or message delivery agent (MDA) is a computer software component that is responsible for the delivery of e-mail messages to a local recipient's mailbox.

2. System Functionalities

The intent of Personal Mail Organizer is to ease the email related task of executives and officials. This product has been designed to maximize the performance by providing facility to automate the downloading and arrangement of emails on recipient's machine, and hence extraction of event information from organized emails, which would otherwise have to be performed manually.

The product consists of the following basic modules:

A. Notification and Information Management

Application: - Automatically notifies the user about information and manages it accordingly.

General Description: - This module stores information about user in a database which includes login-ids and respective passwords of their email accounts. It retrieves the email messages with their respective attached files, from the server and stores them on the client machine on the basis of the stored predefined keywords which decide the intended folder of the particular email message. It also stores text information defined by the user, using which text mining techniques are to be applied on the content of the email messages. The second major task of this module is to notify the user about the arrival of new email. It also notifies about unavailability of internet connection at threshold value of timer and requests to reset the timer [10].

B. Periodic Access

Application: - Automatically connects to the server.

General Description: - This module checks the internet availability at periodic time intervals. If it gets the connection it accesses the user’s email account and downloads the newly arrived emails. After downloading, it marks them as read on the server. If internet connectivity is not available for three successive time intervals then it doubles its counter and continues the process till the threshold value arrives. If it gets the internet connection before the threshold value then it resets its counter to the default value.

C. File Organization

Application: - Automatically connects to the server.

General Description: - This module analyzes and organizes the downloaded email messages. It applies the constraints and keywords on email messages and arranges them accordingly. It also extracts the desired information from the organized email messages by the application of text mining rules [8] [9].

The interaction of the modules working together is shown in figure 1.

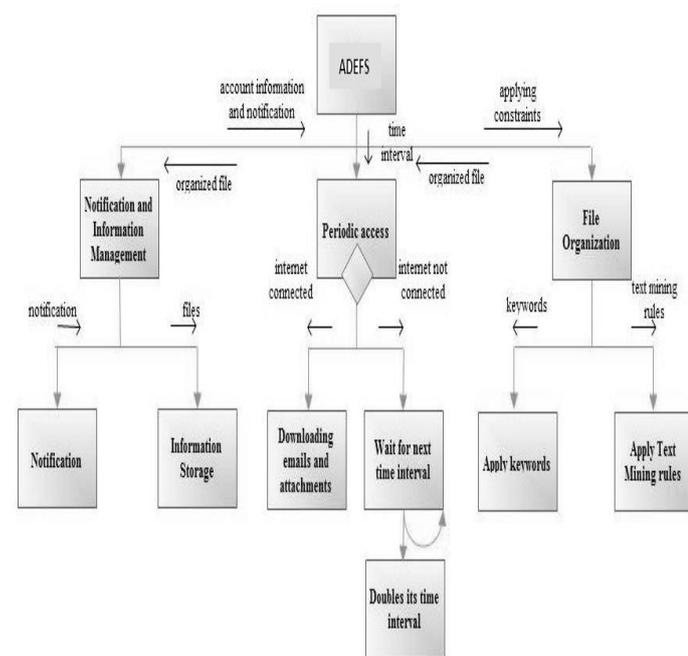


Figure 1. Modular Structure with their Interaction

4. System Framework

The diagrams that describe the preprocessing structure of Personal Mail Organizer and architecture of Text mining system are shown below. The preprocessing structure can be viewed as a pipeline of processes that takes raw email as input, determines whether the email is event related, and, if it is, performs information extraction on it. Each step of the pipeline is discussed in more detail below.

I. Preprocessing Structure

The raw email messages on server are accessed by Advanced Email Filter using Text Mining. Then Advanced Email Filter Using Text Mining performs preprocessing on raw email messages which includes retrieval and downloading of email messages on client’s machine. After preprocessing the downloaded email messages are categorized using application of keywords. These email messages are then organized in intended folders. Then Advanced Email Filter Using Text mining applies its text mining system on meeting related emails to extract scheduling information.

The interaction of the modules working together is shown in figure 1.

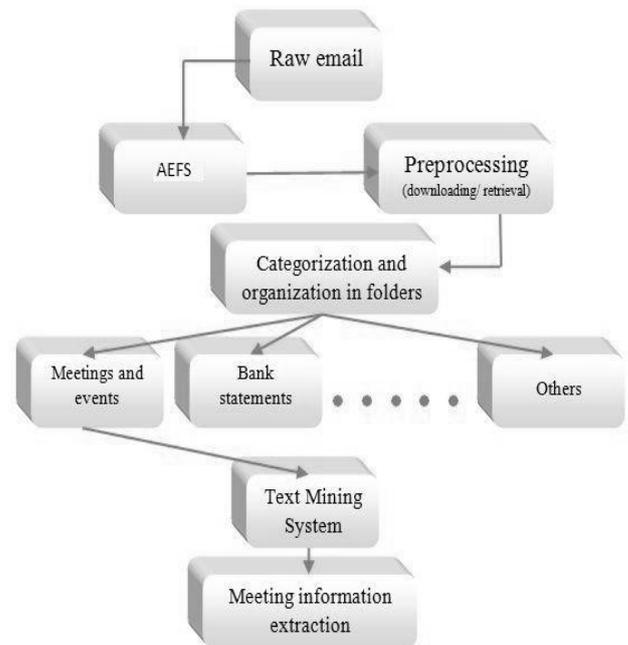


Figure 2. Preprocessing Structure

II. Text Mining System Architecture

The Text mining system architecture has been sub divided into three major phases.

- a) Text preprocessing phase: - In this phase, the downloaded emails are stored in text format from which the meeting related emails are filtered and trimmed.
- b) Rule application phase: - In this phase, Event information extraction (EIE) algorithm is applied on the preprocessed meeting emails.
- c) Visualization phase: - In this phase, the extracted event information can be visualized in text format.

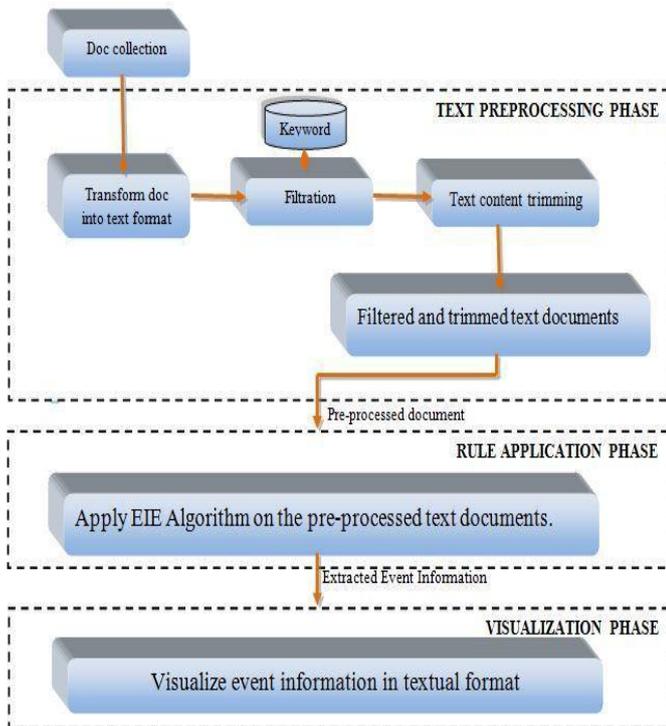


Figure 3. Text Mining System Architecture

5. Proposed Algorithm

In this paper, the EIE algorithm is proposed for extracting scheduling information regarding meetings.

Problem: - There is no specified format for date and time in meeting emails. The problem is to find exact and completely understandable date and time information.

Input: - Meeting emails in text format.

Output: - Date and time information mentioned in the content of input file.

Assumption: - All input files uses English language and numerical to define their content.

The EIE algorithm is as follows: -

- 1) Input the meeting email in text file format.
- 2) Read the contents of the file sequentially.
- 3) Identify the tokens in the content of file and store them in an array temp[] as strings.
- 4) Identify the numeric value in the elements of the array.
 - a) When numeric value is represented by English letter such as one, two, and so on.
 - Replace them with their corresponding numeric representation.
 - b) Check the first character of all the elements of temp[] array as digits.
 - c) Return all the indexes of temp[] where the element has first character as digit.
 - d) Fetch the forward and backward tokens of all the identified indexes.
- 5) Store the extracted tokens into output text file.
6. Other related products

1. Basic Functionalities of Related products

- **Procmail**

Procmail is a mail delivery agent (MDA). It is capable of sorting incoming mail into various directories and filtering out spam messages. It is widely used on Unix-based systems and stable, but no longer maintained [5] [6].

- **Multi email notifier**

Multi email notifier checks multiple email accounts from the same provider, periodically and notifies about the arrival of new email. It also includes the information about the sender, subject and the arrival time of email [2].

- **Outlook Express**

Outlook Express is an email program that allows sending and receiving email messages on client machine. It also allows creating multiple email accounts. One can view emails for all accounts in the same screen. Emails and contacts can be managed by creating folders [3][4].

2. Comparison with Related Products

TABLE I. Comparison of AEMFS with Other Products

S.N	Parameter	Procmail	Multi Email notifier	Outlook Express	ADEFS
1	Operating System Compatibility	Unix based	vista and high ver.	xp and higher version	vista and higher version
2	Category	Mail delivery agent	email notifier	Email client	Email client
3	Notification of New mail	No	Yes	No	Yes
4	Notification of internet unavailability	No	No	No	Yes
5	Downloading of emails	Yes	No	Yes	Yes
6	Organization of emails	Yes	No	Yes	Yes
7	Storing of emails	No	No	Yes	Yes
8	Text mining Application	No	No	No	Yes
9	Event Information Extraction	No	No	No	Yes
10	Automation	Low	Medium	Low	High
11	Periodic Access	No	Yes	Yes	Yes

7. Future Enhancements and Conclusion

The functionality of Personal Mail Organizer can be extended to become compatible with other operating systems and mobile application. The alerts generated by the Personal Mail Organizer can also notify the user on its mobile phone through short message service (sms). It can also include the mobile scheduler to store the extracted event information (date and time). The Personal Mail Organizer can be made self learning software.

The Personal Mail Organizer will be very beneficial to its users, as it provides full automation in retrieval and arrangement of email messages.

8. References

- [1] http://en.wikipedia.org/wiki/Email_client
- [2] <http://www.multiemailnotifier.com/>
- [3] <http://support.microsoft.com/kb/835830>
- [4] http://product.secserr.net/mail/email_express.htm
- [5] <http://www.procmail.org/>
- [6] <http://en.wikipedia.org/wiki/Procmail>
- [7] Mia K. Stern, "Dates and Times in Email Messages" published in ACM digital library, 2004.
- [8] D. S´anchez, M.J. Mart´in-Bautista, I. Blanco, C. Justicia de la Torre, "Text Knowledge Mining: An Alternative to Text Data Mining", published in IEEE, 2008.
- [9] Jan-Peter Kramer, "PIM-Mail: Consolidating Task and Email Management", published in ACM digital library, 2010.
- [10] Santosh Kumar Paul, Madhup Agrawal, Shyam Rajput , Sanjeev Kumar " An Information Retrieval(IR) Techniques for text Mining on web for Unstructured data" published in IJARCSSE, Feb-2014.
- [11] Ms Sonali Gaikwad, Prof Archana Chaugule "Performance Enhancement for Effective Pattern Discovery & Concept Formation in Text Mining" published in IJARCSSE, Sep-2014.
- [12] Deepali Jhanji1 , Priya Garg "Text Mining" published in IJSAE ,Aug-2014.
- [13] Mr. Rahul Patel ,Mr. Gaurav Sharma "A survey on Text mining Techniques." Published in IJECS May-2014.
- [14] Omkar Y. Todkar, Sangarm Z Gawali " Ranking and Searching of Document with New Innovative Method in Text Mining: FirstReview" published in IJAIEM,May-2014
- [15] Jadhav Bhushan G1, Warke Pushkar U2, Kuchekar Shivaji P3, Kadam Nikhil V4 "Searching Research Papers Using Clustering and Text Mining" published in IJETAE, April-2014