Design and Implementation of Email Agent System

Zaid H. Alibadi , Sawsan K. Thamer
Department of Computer Science, Al-Nahrain University.
Loay E. Goerge and
Department of Computer Science, Baghdad University.

Abstract

Email is one of the most useful communication tools over the Internet; Email can be an effective knowledge management tool which conveniently enables fast and accurate communication. On the other side, the increasing volume of email threatens to cause a state of "email overload" at which the volume of messages exceeds individuals' capacity to process them. This paper presents a personal email agent, named PAEA (Personal Assistant Email Agent), which assists the user to send all his/her email messages to their recipients and automatically download the user email message from the email server. The agent is designed to be able to classify the incoming email messages into folders, and to prioritize them so that the user can focus on more important emails first. The agent prioritizes messages according to user profile and his historical reaction. PAEA can instantaneously update his learning from the user behavior to be more effective and adaptive in doing the email sorting task.

Keywords: Email, Email Agent**,** Email Management.

الخلاصة

يعتبر البريد الالكتروني من اهم وسائط الاتصال من خلال شبكة الانترنيت، حيث انه يعتبر اداة فعالة لتنظيم المعلومات كونه وسيلة اتصال سريعه ودقيقه لتبادل المعلومات. ان الاستخدام الواسع للبريد الالكتروني تسبب في زياده كم الرسائل التي يجب على مستخدم البريد معالجتها بسبب اصبحت في بعض الاحيان تتجاوز طاقته.

ان ظاهرة الزيادة في عدد الرسائل الالكترونيه المستقبلة للمستخدم في عمله او حياته اليومية تسببت في خلق مشاكل جديده للمستخدم. حيث تطلب منه بذل جهد اضافي لمعالجة هذه الزيادة في عدد الرسائل، اضافة الى ان اصبحت تشكل تكلفة مادية اضافية لتنظيم وقراءة هذا العدد الكبير بالاضافة الى انها تتطلب الاتصال المستمر بشبكة الانترنيت. وعليه اصبحت هناك حاجة لبناء نظام يساعد المستخدم في عملية تنظيم وقراءة رسائله الالكترونية لكي يتمكن من اكمال اعماله بسهولة.

التطبيق المقترح يساعد المستخدم بارسال رسائله الالكترونية الى مستقبليها اليا وتحميل كل الرسائل الجديدة من بريد المستخدم الى حاسبه الشخصي. اضافة الى ذلك يقوم التطبيق بتنظيم وترتيب الرسائل الالكترونيه بالاعتماد على السجل التاريخي لاهتمامات المستخدم وطبيعة استجابته للرسائل. كما ان التطبيق المقترح ومن خلال الاستخدام المستمر من قبل المستخدم يقوم بتحديث وتطوير عمله من اجل تقديم نتائج اكثر قبولا ومتوافقه مع توقعات المستخدم.

تم اختبار التطبيق من اجل الكشف على ان نتائجه تكون متوافقة مع توقعات المستخدم، كذلك تم اظهار الفوائد المترتبة من استخدام هذا التطبيق حيث انه قلص الزمن المتطلب للاتصال بشبكة الانترنيت. بالاضافة الى تقليص الجهد المتطلب من قبل المستخدم وذلك باظهارالرسائل التي تتعلق باهتمامات المستخدم باعلى القائمة بعد ترتيب الرسائل.

1. INTRODUCTION

The recent phenomena of email-overloading in daily life and business have created new problems to users. It becomes a personal headache for users because they have to process a large number of daily received emails. Also, it is a financial issue because the user checks and read for large amount of email messages needs long online communication connection. Therefore, there is a practical need for developing a software assistant to facilitate the management of personal and organizational emails and to enable users to complete their jobs or tasks smoothly [LiZh09].

Agent systems have been proposed as solutions to the problem of information overload, particularly regarding email and internet searches. Most of the current implementations aiming to ease the burden of dealing with email are text classifiers or keyword extractors, often working as email client plug-ins [AbMc01].

An email client, email reader, or more formally mail user agent (MUA), is a computer program used to manage email. Specifically, the term email client may refer to any agent acting as a client toward an email server, regardless of it being a mail user agent, a relaying server, or a human typing on a terminal. In addition, a web application providing message management, composition, and reception functionality is sometimes considered an email client [Par08].

An intelligent agent is defined as "an agent capable of flexible autonomous action to meet its design objectives". The word "Flexible" in above definition means [WoJe95]:

• Reactivity: intelligent agents perceive and respond in a timely fashion to changes that occur in their environment in order to satisfy their design objectives. The agent's goals and/or assumptions that form the basis for a procedure that is currently executing may be affected by a changed environment and, such case, a different set of actions may be needed to be performed.

• Pro-activeness: reacting to an environment by mapping a stimulus into a set of responses is not enough. Goal directed behavior is needed in intelligent agents. In a changed environment, intelligent agents have to recognize opportunities and take the initiative if they designed to produce meaningful results. The challenge to the agent designer is to integrate effectively goal-directed and reactive behavior.

• Social ability: intelligent agents are capable of interacting with other agents (and possibly humans), through negotiation and/or cooperation, to satisfy their design objectives.

Email provides an example of a rich information management domain, Email is typically short, and it contains a limited amount of structure. The body of an email is usually unstructured text, while the headers provide some tagged information. For example, an agent knows a priori the meaning of a "From" header field it can define the sender; similarly, the "Date" header field should reflect signs about when the email message is received [Res01].

Header fields are necessary for any standards compliant message. Header fields contain information such as where the message came from, where it is going, when it was sent, and more. However, only two header fields are non optional for standards-compliant messages [Los99]:

• From: indicating the originator of the message.

• Date: indicating the origination date of the message.

The "Subject" header field is sometimes problematic. It says something about the contents of the email, but not always. Even for the headers with known content, the utility of their information is limited. Knowing the sender of an email is useful, but often the same sender may discuss different topics with same recipient. So, some form of content understanding is required [Boo98].

A variety of approaches have been taken to address the problem of automating email classification. Some of these systems are described below in approximate chronological order:

• Magi by Payne and Edwards [PaEd97], at 1997, have developed Mail agent interface (Magi) application to work on top of a UNIX mail system. They indicated that depending on the confidence "Magi" it can carry the action out automatically, suggest the action to the user and see if they agree, or make no suggestion at all.

• Bonne [Boo98], at 1998, had introduced "Re:Agent" email tool, "Re:Agent" classifies emails into two categories only, 'work' or 'other'. However, Boone found

that his introduced approach "Re:Agent" can achieve 98% accuracy, while the standard IR approach had 91% accuracy.

- Segal and Kephart [SeKe99], at 1999, introduced "MailCat" system. Their system used TF-IDF approach which computes weighted vectors for each folder based on word frequencies, and then a distance measure is used to estimate the similarity a new message has with each folder. They referred that, when new messages were directory filtered into the most similar folder an error of 20% to 40% resulted.

- Rennie [Ren00], at 2000, had used a naïve Bayes approach for text classification. He called his introduced method "iFile" filter. It works as filter for the EXMH mail client. The system applies stemming and makes use of a stop-list.

- Moreale and Watt [MoWa03], at 2003, have introduced a system that works with several lists, giving users archiving and retrieval assistance through an intuitive and dialectic interface: users can email their query directly to the agent and receive a prompt reply day or night. Alternatively, users can post their query publicly to a forum (monitored by the agent) or run a web-like search over the monitored lists.

- Fawzi [Faw08], at 2008, had introduced "EMFA" system which used machine learning to classify the email messages into two list, Negative list (that contains unwanted messages) and Positive list (that contains the messages that must be forwarded or replied).

The aim of this work is to implement a simple proactive automated system that helps user to automatically manage their email messages according to his\her personal profile. Also, the system offer a number of services to the user (like, give the user the ability to filter-in and download the new email message according to three email messages attributes, auto-reply email message and compose new email message with the ability to attach files).

2. Proposed System

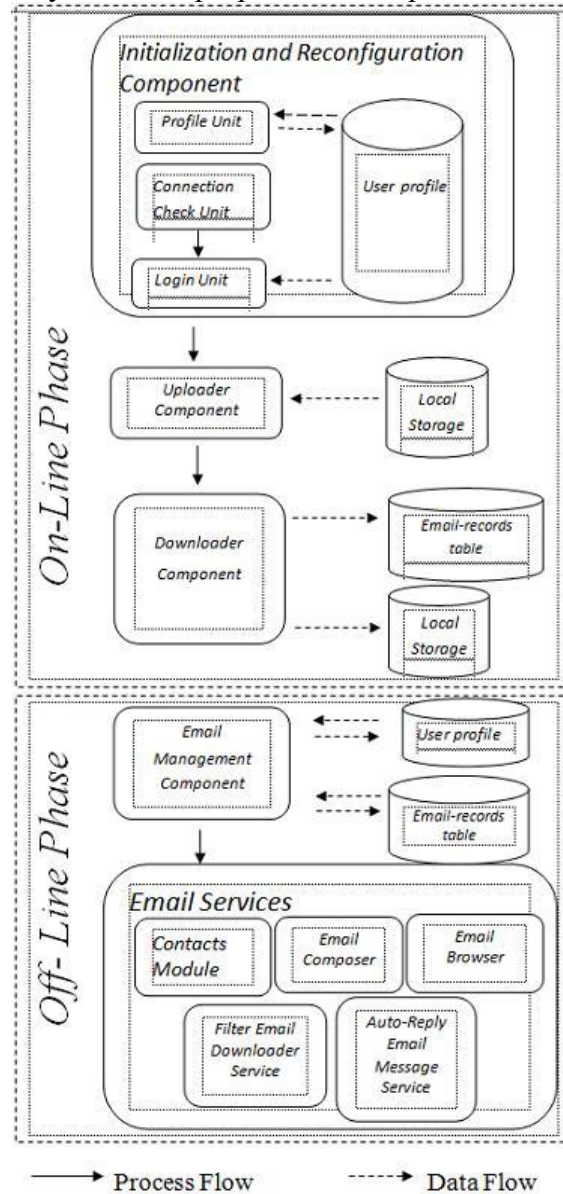Figure (1) present the layout of the proposed and implemented agent system.

Figure (1) PAEA Architecture

PAEA consists of five components: (i) Initialization and Reconfiguration component, (ii) Uploader component, (iii) Downloader component, (iv) Email Management component and (v) Email Services component; as illustrated in figure (1).

The proposed agent system is developed to be reactive and automatic. Its structure consists of two sets of components; the first set of components deal with email account as long as there is an internet connection, while the second set of components offers off-line services (like, email composing, prioritize, archiving and browsing) whether there is an internet connection or not. These two sets are designed to work in an asynchronous and collaborative way.

The PAEA components are:

A. Initialization and Reconfiguration Component (IRC)

The needed predefined information and actions from the user are: (i) to make sure that a successful login process to the user account will be accomplished and (ii) to achieve results that match the user's expectations.

This component consists of three units:

a. Profile Unit (PU): When a new user uses PAEA system, a user profile will be created; the user profile contains the connection's information which should pre-assign by the user to initiate the connection. Also, the user profile contains a brief description about what are the most interesting subjects to his\her which can be used later as keywords in email management process, and this information can be modified by the user whenever he\she wants. The involved user profile information includes:

  i. Login Information: It contains the server name, account username and password; this set of information is a required to enable an access to the user' email account.

 ii. Interest Subjects List: It is a list of "interest subjects" for the user; it could be pre-assigned by him. This list is changeable and could be modified by the user during his continual interaction with the agent system.

iii. Actions Setting: The user can adjust the agent how to react with the uploaded and downloaded email messages (for example, delete the email messages after it is downloaded from the server or not, specifying the folder on local storage where the email messages are saved and if the agent after downloading each new email, will send an auto-replay email message to email's sender).

b. Internet Check Unit (ICU): As a first main stage for all on-line operations a connection with internet must be established to reach the user's email account on an email server; this will let PAEA capable to make the required access to user email account. This unit is continually checks if there is an internet connection to automatically activate Login Unit operation to establish a connection with the user email account.

c. Login Unit (LU): after supplying the user credentials to login method, then by using IMAP the PAEA will automatically choose the available authentication method and log in.

B. Uploader Component (UC): One of the gained benefits due to the usage of PAEA system is "the email messages could be composed in off-line mode (i.e., internet connection is not required)". After the preparation of the email messages, they will be automatically stored in "Outbox" folder with the overhead information; which are required to send the messages later without need to user interception. Later, the agent will automatically send the stored messages to their recipients as soon as an internet connection becomes available.

C. Downloader Component (DC): This function is to download all newly incoming email messages from the user account area on the server to the local storage. The downloaded and saved email messages can be accessed, through the email browser component of the PAEA system. The downloaded email messages are stored into a local storage; this could be useful for reducing the time needed to have on-line with internet media.

D. Email Management Component (EMC): Prioritizing emails according to their personal importance to the user is another assessed function offered by PAEA system. The degree of importance of an email to user could be assigned by a number; this number is considered as the significance weight of the message for certain user.

Calculation of the overall weight of each email message depends the following factors:

a. Address weight: if the email's sender address is saved in user's contacts list, the value of this contact is saved in contact's record; then its weight will be added to the overall weight of the email message.

b. Website weight: in a similar way to that followed with the address weight value. If the email message is received from a registered website in contact table, then the assigned weight to this website will be added to the overall weight of the email message.

c. Subject weight: if the subject's keywords of the email message are related to the interest subject list of the user, a bounce will be added to the overall weight of the email message (in this work the bounce is set 10).

For each unread email message, its overall weight value is calculated, the weight equation is:

$$Mw=Aw+Ww+Sw$$

Where,

Mw: is the overall message weight.

Aw: is a number represents the
        address email's significance (weight) factor to the user.

Ww: is a number represents the website's significance factor to the user.

Sw: is a number represents a bounce that will be added to the email's weight if the subject of the email is related to the interest subject list.

E. Email Services:

A number of services are offered in PAEA; they are either necessary to accomplish agent's work, or to offer extra services for the user. The offered services are:

a. Contacts Module (CM)

The user has the required ability to create its own contacts list. This list holds all email addresses and websites which the user has the interest to exchange emails with them. So, the main function of contact module is to create a table that contains a record for each contact.

b. Email Browser (EB)

PAEA offers a simple Email Browser to present the text content of the email message, and to present the most important information about the email message (i.e., the sender's address and subject's title). Also, through this browser the user can extract the attachments from the email message and save them as separated files into the local storage; this could be done by the "Attach_dwnloder" function.

Another important job of the EB is monitoring the behavior response of the user to the incoming email messages. When user opens the email message, EB increases the weight value of the contact (whether it is person or website) by one unit. If the email message received from unregistered email address (i.e., the email address that is not saved in user's contact list), then PAEA will automatically add this address to user's contacts list, if the user had frequently opened the email messages received from this address. PAEA uses a binary file as a "History file" to save a record of any unknown email address that the user received an email message from it.

When the user read more than three email messages from an unknown email address, it will automatically create a record for this email address and add it to the user's contacts list.

c. Email Composer (EC)

The "MailMessage" class, used in PAEA system, is established to create an instance of "MailMessage" to represent each new email message. PAEA offers Email Composer to set most important attributes in header and body fields of the email

message (like; From, To, Subject, date, MessageID and the content of the email message). Also, the user can attach, up to three, files with each sent email message. Once user completed composing the email message, it will be saved automatically in "Outbox" folder waiting to be sent to its recipient(s) as soon as an internet connection is established.

d. Auto-Reply Email Message Service (AREMS)

PAEA offers the capability of allowing user to compose a default email message, so whenever the agent download a new email message an auto-replay email message are send to the sender. This is a useful tool if the user does not have enough time to handle his email messages, but still want to inform the sender that his message is received and his real replay message will issued later.

e. Filter Email Downloader Services (FEDS)

As mentioned previously, IMAP supports "Search" command. So, the user can search for email messages match a given criteria by the search method. It accepts a variable number of search parameters. Therefore, the user can adjust the downloader to download only the email messages that match a given criteria.

Three types of filters are available to user, and could applied separately his/her new incoming email messages, they are:

1. Specific Contact Filter (SCF): it is used to download only email messages received from a specified contact.

2. Specific Website Filter (SWF): it is used to download only email messages received from a specified website.

3. Specific Time-intervals Filter (STF): it is used to download only email messages received within specific time intervals.

3. System Learning

Beside to autonomous transaction of email messages, PAEA can adjusted by its learning information and, consequently, updated its knowledge. This continual updating of system parameters keeps the system capable to produce rational decisions for each user over the time.

The interaction between the user and the system depends on user behavior (i.e., how the user acts with the email messages that the agent sort and manage them). So, each time the user decided to open the email or not, PAEA will interact with this action as a feedback from the user. Therefore, the reaction of PAEA will be different in case when the previous email message send by some source are already opened or ignored.

• When the user read the email:

The system will automatically increases the contact weight whose receiving messages continually opened. If the user open unread email message, the system will increase the contact weight of its sender by one unit. Also, if the user open more than one email message from unknown email address, the system will automatically add this address to the user's contacts list.

• When the user ignore (doesn't read) the email:

Each time the system download a new set of email messages from the user account on server to local storage, the agent will increase the "seen" field of the email record by one unit to all unread email messages. So, when the agent recalculate the overall weight to user's unread email messages, the EMC will decrease the weight of the old unread emails by an integer value, which is already saved in "seen" field.

After three times of user ignorance for the same email the system will decrease the contact weight of its sender by one unit. Also, the user has the ability to change the contacts weights, all the user's contacts will be list in List-View. The user can make

click on any contact and change its weight to a new value. This will give more flexibility to the work of the agent because at any time the user can update the contacts weight, which is an important factor that the EMC depend on it in the email management process.

4. System Autonomous

Once the user run PAEA system and it is connected to user account; then after 30 minutes PAEA will automatically check:

i. "Outbox" folder to see if the user had composed a new set of email messages, and they are waiting to be sent to their destinations. If there is a waiting message(s), the "Uploader Component" will be activated to send this email message(s).

ii. "Inbox" folder, which exists in user's account area on the mail server, to know whether new email messages have been arrived. Once PAEA find new email messages it will activate the "Download Component" in order to download the new emails and archive them on local storage. After downloading, PAEA will set the "New Email" Flag to True.

The "Email Management Component" will periodically recheck the "New Email" flag to see if there are new email messages downloaded; in case of finding new coming emails then this component will recalculate the overall weight of the user's unread email messages and to rearrange the email messages.

5. System Evaluation

A subject test through a simulated scenario of email flow was applied on the PAEA system to evaluate its work, and to see if its response is rational and match the user expectation. In this simulation we track the steps of the main three components (DC, UC and EMC). For two iterations, we see the status of these components (active, wait or idle) and their effects to the user's email messages. Also, we notice that most of the email messages that related to Mr. X hold interested subjects are in the top of the emails list. During the test we have noticed that the needed internet connection time is minimized (the UC need only 15 seconds to send 5 email messages, their total size is 585 Kbytes; for comparison purpose where Mr. X try to the webmail he took 3 minutes to compose and send these email messages, the DC took 30 seconds to download 4 email messages, their total size is 913 Kbytes. Note that these results mainly depend on the internet connection speed.

Although email is ubiquitous, large and realistic email corpora are rarely available for research purposes due to privacy concerns. So, we use "Alshather Company" staff to test PAEA system and check their evaluation results; a questionnaire was applied on the company staff to check their opinion, and PAEA system get 8.0 (in scale from 1 to 10) average score.

6. Conclusion

We have finished the implementation of all components of PAEA using the programming language Visual Basic.Net under Windows. As next step we have just started an extensive evaluation study with a test group of company staff. Therefore, it is too early to make any final remark about the success of the proposed system. However, in first test sessions we have received quite positive comments and feedback from the users.

References

[AbMc01] S. Abu-Hakima, C. McFarland and J. Meech, "An agent-based system for email highlighting", AGENTS'01 Proceedings of the 5th International Conference on Autonomous Agents, pp. 224-225, Montreal, Quebec, Canada, 2001.

[Boo98] G. Boone, "concept features in Re:agent, an intelligent email agent", in 2nd International Conference on Autonomous Agents, pp. 141-148, MAY 1998.

[Faw08] H. Fawzi, "Design and Implementation of Email Filtering Agent", Master thesis, computer department, college of science, Al-Nahrain University, 2008.

[LiZh09] W. Li, N. Zhong, Y. Yao and J. Liu, "An Operable Email Based Intelligent Personal Assistant", World Wide Web journal 12:125–147, 2009.

[Los99] P. Loshin, "Essential email standards: RFCs and protocols made practical", John Wiley & Sons, Inc, November 1999.

[MoWa03] E. Moreale and S. Watt, "An Agent-Based Approach to Mailing List Knowledge Management", Papers from the AAAI Spring Symposium, pp. 49, Stanford, California, March 24–26, 2003.

[Par08] C. Partridge, "The Technical Development of Internet Email", IEEE Annals of the History of Computing (Berlin: IEEE Computer Society) 30 (2): 3–29. (April-June 2008).

[PaEd97] T. Payne and P. Edwards, "Interface Agents that Learn: An Investigation of Learning Issues in a Mail Agent Interface", Applied Artificial Intelligence, Vol. 11(1), pp.1-32, 1997.

[Res01] P. Resnick, "Internet Message Format", QUALCOMM Incorporated, RFC 2822, IETF, April 2001.

[Ren00] J. Rennie, "ifile: An Application of Machine Learning to E-Mail Filtering", proceedings the KDD-2000 Workshop on Text Mining, Boston, 2000.

[SeKe99] R. Segal and M. Kephart, "Mailcat: An intelligent assistant for organization email", in Proceedings of the 3rd International Conference on Autonomous Agents, pp. 276-282, Seattle, WA, 1999.

[WoJe95] M. Wooldridge and N. Jennings, "Intelligent Agents: Theory and Practice", The Knowledge Engineering Review, 10(2), pp. 115-152, 19995.