A JADE Implementation of Integrated Agent System for E-Mail Coordination (IASEC)

Balakrishnan S
Research Scholar
Bharath University
Chennai

K.L.Shunmuganathan, PhD.
Professor & Head
RMK Engineering College
Chennai

ABSTRACT
With growing usage of computer applications there is growing dependence of business over computer aided services like e-mails. The CRM or E-Mail based marketing companies are becoming pervasive. To run these companies in a way that the customers are well satisfied, there is always a greater need to prioritize the e-mails which has a direct impact on the ROI – Return On Investment. In this paper, a method is proposed to prioritize the unread E-Mails according to the users' interest and priority. This is implemented in this system using agent technology. The Agent mechanism is guided using JADE Middleware and its underlying architecture. The Intelligent agent used here can learn the interest and priority of the user and use this knowledge to arrange the unread mails, according to the users' needs. The Text corpus in the mail would undergo tokenization and each token would be matched with knowledgebase by the agent. The precision of the prioritization may range to approximately 90 percent. By this way, the E-Mails are ranked according to interest and priority of the user.

General Terms
JADE, Integration

Keywords
Intelligent Agent, Tokenization, E-Commerce, Artificial Intelligence

1. INTRODUCTION
E-mail has emerged as the world's most preferred form of communication. Billions of E-mails traverse the globe daily. A person, who relies only on E-mails to send and receive messages, would definitely be getting a large number of E-mails daily. It becomes a difficult task for the user to decide which e-mail to open first and which ones to open later. It is indeed, wise for the email client to sort the emails in the user's inbox, according to their degree of importance. This can be achieved using an Intelligent Agent. The Intelligent Agent will take the responsibility to rank each E-Mail according to its degree of importance, and display the mail according to the rank in the user's inbox.

2. RELATED WORK
Email marketing is a form of direct marketing which uses email as a means of communicating commercial or fund-raising messages to an audience. In its broadest sense, every email sent to a potential or current customer could be considered email marketing. However, the term is usually used to refer to:

- sending email messages with the purpose of enhancing the relationship of a merchant with its current or previous customers, to encourage customer loyalty and repeat business

- sending email messages with the purpose of acquiring new customers or convincing current customers to purchase something immediately

- adding advertisements to email messages sent by other companies to their customers

- sending email messages over the Internet, as email did and does exist outside the Internet e.g., network email.

3. CHALLENGES OF MAIL PRIORITIZATION

- Personalized email prioritization (PEP) is an ordinal regression problem, which is different from conventional text classification where for each category; there are only two levels, true or false. Users may rate their importance from one to five or from not important at all to very important, resulted in ordinal regression problem. Given limited amount of time, users may want to selectively read important emails or may associate actions to certain importance levels.

- The personalized email prioritization entails two main research challenges: (1) the sparse training data and (2) one's own priority definition. First of all, unlike spam filtering, we could not share training data among different users because of privacy issues and different interests. People hesitate to share their very personal labeling information except spam emails. Even though there are users who are willing to share the very personal labeling information, the personal labeling information could not be shared. For instance, the importance of a grant proposal email could be extremely important to the principal investigator but it could be marginally important or not important to the person who is not actively working on the proposal.

- Second, one's own priority definition could lead to diverse way of defining priority. In that case, the assumption of the current state-of-the-art ordinal regression such as Support Vector Ordinal Regression (SVOR) [12] might not be good enough. For instance, regression-based approaches assume one weight vector to model all levels of email priorities from the lowest priority level to the highest priority level, resulted in all decision boundaries to be parallel. Since the email text is very high dimensional space, it is not easy to visualize and check whether regression-based approach assumption will be held or not. Therefore, we have to do any form of empirical evaluation to conform what kinds of approaches are the best.

There are some Customer Relationship Management Tools that come with integrated support for emails. They may discover crucial information for decision making by tracking the e-mail conversations had with the customers. They also enable

- Centralization of all customer information and email correspondence inside CRM
- Association of customer's email conversation with CRM data

For driving such business deeds it is mandatory that the emails are properly prioritized according to the purpose and carry out the workflow guided by these preferences. To enable this functionality is used the Integrated Agent System for Email Coordination (IASEC) in integration with the existing mail system.

4. SYSTEM ARCHITECTURE & RELATED TECHNOLOGY

The following diagram shows the 2 – tier architecture of an Email Prioritizing Agent in Jade. The architecture is explained in detail as below.
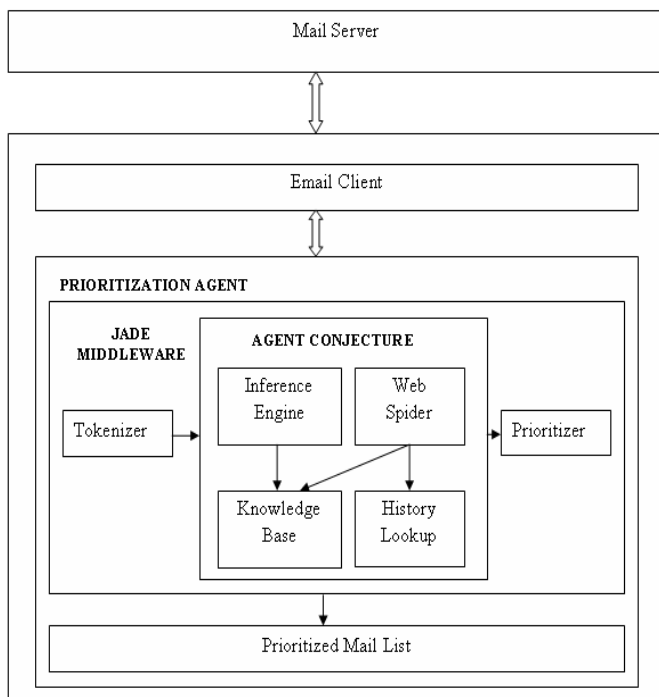

Fig 1: IASEC Architecture

The Agent works in association with the email – client which may in turn be associated to any Mail Server. The Agent takes up the e-mail morphology and tokenizes the entire content. The logic for the inference mechanism is stored in the inference engine which works in association with the Knowledge Base, which stores the list of words along with its weights, which are directly proportional to the interests of the user. The weight of the words in the database is updated by a web spider based upon the interests and activities of the user.

The spider accomplishes the above task by visiting the web pages and parsing it into words following which it performs morphological analysis to update the weight of the words present in the database. It also takes into account the various factors such as the sender of the E-mail, the periodicity of the

mail and the users' interests and the user's web browser history.

The inference Engine would resolve the degree of importance of the mails in the inbox working with the knowledge base. This is summarized as a prioritized list of content on the email client consequently.

The Agent Technology is a rapidly developing area of research. Agent Technology is to shortly commence the development phase of its multi-protocol network. Intelligent Agent is an entity which observes and manipulates an environment (E-mail client in our case) and directs its activity toward achieving its goal. Intelligent Agent will learn and use its knowledge to achieve its goal. An agent program can be defined briefly as a mathematical function of an agent, which maps all the possible sequence of percepts to every action that the agent can perform either to a coefficient or feedback elements or even to a function or constant all of which affects eventual actions.

4.1 Prioritized E-mail coordination

The powerful basis for understanding how agent technology can help people work together more effectively will result from a better understanding of the nature of coordination. The coordination of distributed processing is of great interest to a number of research communities. In particular, Coordination technology is a way of designing and studying innovative computer systems that help people work together in small or large groups. The coordination process must organize the flow of information in such a way that communication between the managers of an agent proves to be most effective.

4.2 JADE

JADE (Java Agent Development Environment) [9] is a software Framework fully implemented in Java language. It is a middleware developed by TILAB that simplifies the development of applications. It simplifies the implementation of multi-agent systems through a middle-ware that complies with the FIPA specifications and through a set of graphical tools that supports the debugging and deployment phases.
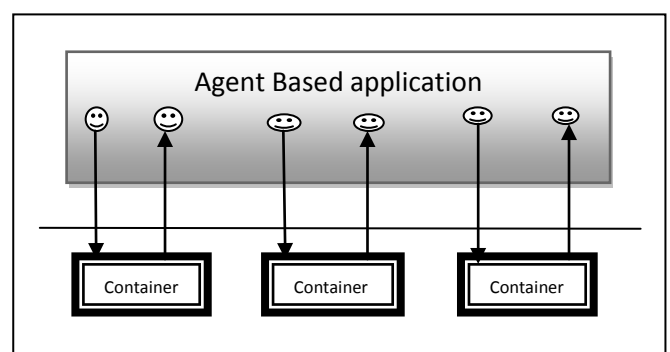

Fig 2: JADE Architecture

5. INTERNAL FUNCTIONALITY OF AGENT

In this section the internal working of the IASEC is explained detail in Figure 3. The system retrieves email from mailbox and prioritizes them using an agent along with the Bots, which are programs capable of performing an automated task over the internet, which in turn uses Web Spider, a kind of Bots which fetches and analyzes the information available in the web pages.
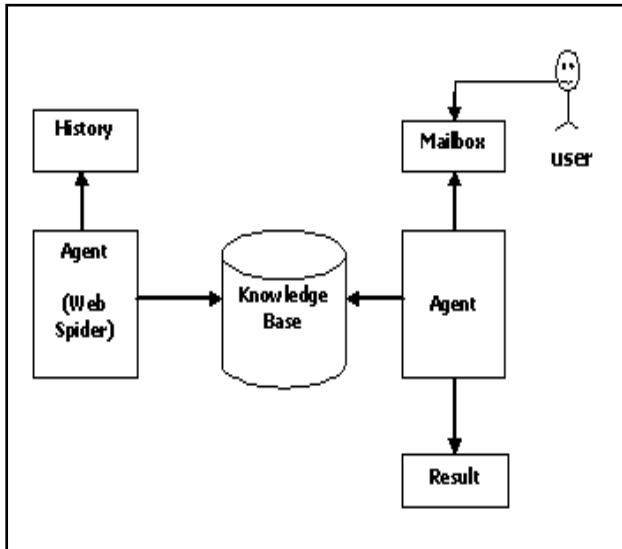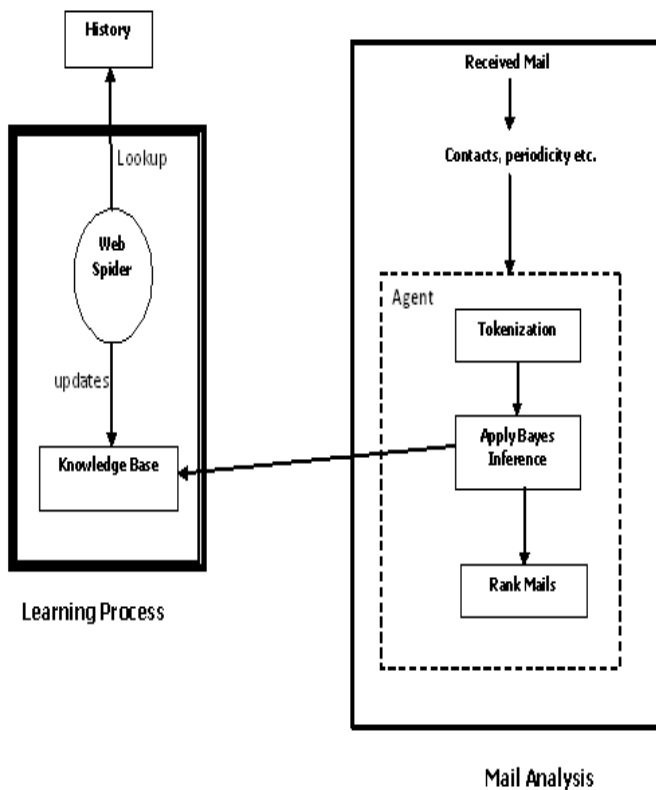
Fig 3: System Layout

Fig 4: Framework

The proposed system consists of the following two parts:

1. The Inference process
2. Mail Analysis

5.1 The Inference process

For the inference process, a database is maintained containing a list of words, which are directly proportional to the interests of the user. During inception, the database contains the words considered universal for any user. The content of the knowledge base are updated by a web spider (Bot) based upon the interests and activities of the user.

5.2 Mail Analysis

In order to declare that a particular E-Mail is important to the user the agent considers some of the factors such as the sender of the E-mail, the periodicity of the mail and the users' interests. The sender and the periodicity can be directly obtained from the mail, whereas the interest of the user can be obtained by studying the user's web browser history.

The agent analyzes the sender and periodicity of the mail in the following way:

5.2.1 Sender

For a given mail the agent checks if the sender of the E-Mail is in the contact list of the user. If the sender is found in the contact list, then preference is given to an E-mail obtained from a known source over an E-Mail from an unknown source. Moreover if the user gets an E-Mail from a person who is in the contact list and is an active member (friend or family member) then, the corresponding E-Mail is provided more preference by the Agent.

5.2.2 Periodicity

The agent analyzes the periodicity of the sender which means that the agent will check if the user gets an E-Mail from a particular contact on a fixed interval of time(as in the case of newsletters). If the agent finds such a kind of mail, the agent awards less priority to that particular E-mail.

5.2.3 Tokenization

Tokenization [1] is the process of isolating individual units called tokens from a Text corpus. A token is considered as a categorized block of text. Tokenization involves the following steps:
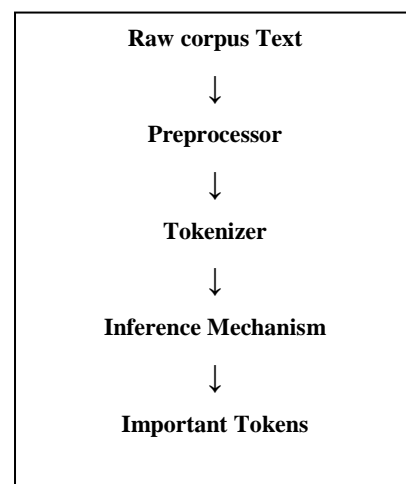
Raw corpus Text

↓

Preprocessor

↓

Tokenizer

↓

Inference Mechanism

↓

Important Tokens

Fig 5: Tokenization process

6. EXPERIMENTED RESULTS

The implementation of IASEC is done using JADE middleware and the following is a screenshot showing the prioritized results,
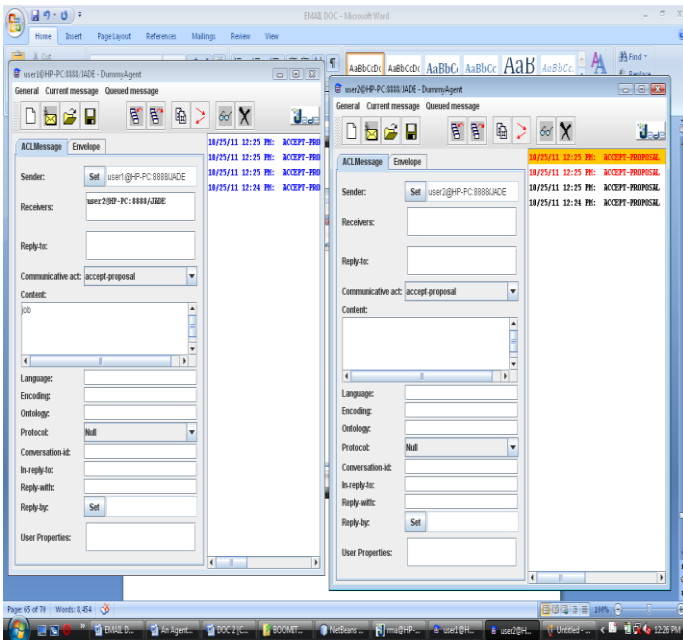

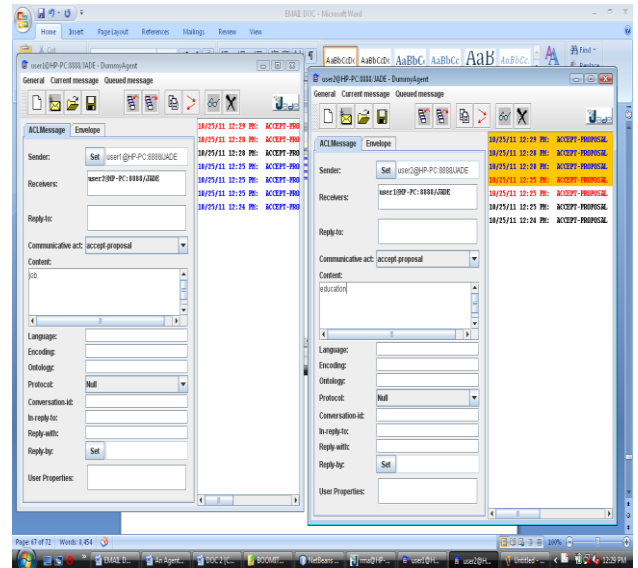**Fig 6: It represents receiving mails from user1 to user2 and priority mails are highlighted**

Fig 7: Prioritized results

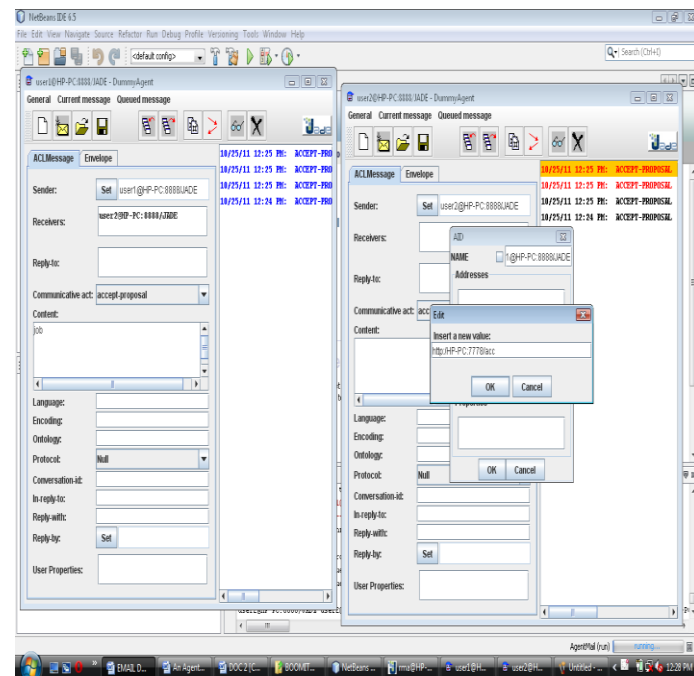**Fig 8: It represents priority mails are highlighted (job, education, sports) etc...**
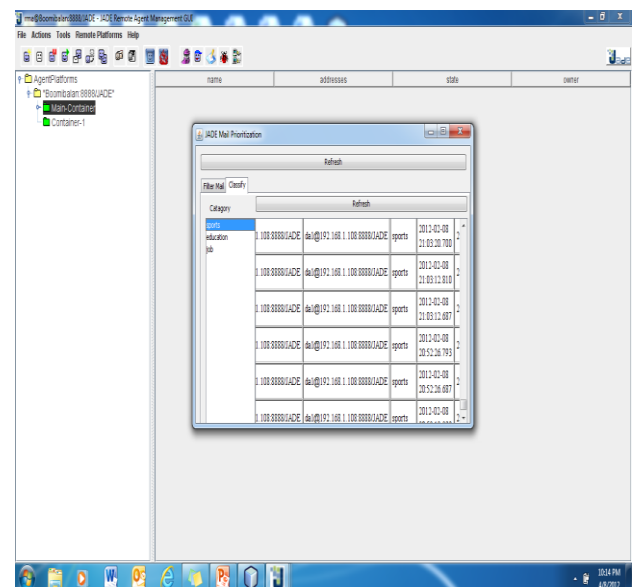

Fig 9: In the Priority view, priority mails are highlighted and displayed.

7. PERFORMANCE ANALYSIS

The system takes into account the performance analyses Table 1.1 provided by [5] and tries to achieve close to 90 percent precision in prioritizing the emails having higher degree of importance.

Table 1. Evaluation of metrics

Evaluation Metric	Education	Sports	Job
Precision	0.5	0.5	0.5
Recall	0.6	0.8	0.4
F-Measure	17.14	17.17	0.16
Accuracy	0.5	0.5	0.5
Error Rate	0.5	0.5	0.5

8. CONCLUSION AND FUTURE ENHANCEMENT

An approach to arrange E-mails according to the priority of the user using an intelligent agent has been presented in this paper. The Intelligent agent can be enhanced with consideration of the attached content also in email by the inference engine.

The alert message (SMS- Short Message Service) can be sent to the user's mobile number, regarding the new high level priority mail in the inbox. This helps the user to get updated of their new prioritized mails, even while the user is offline.

Among the prioritized mails, grouping of several mails under a single subject can be done by using inference engine and checking the keywords in the mail. This helps the user to view the mail as groups based on their content.

9. ACKNOWLEDGMENT

The support, encouragement and the guidance received by the author from Dr.K.L.Shunmuganathan, his research supervisor at Bharath University has been very motivating. The author is thankful to his students Mr.John Royceton, Mr.Jobin George and Mr.Vinothkumar of RGCE for their support.

10. REFERENCES

[1] Yirong Shen and Jing Jiang. "Improving the Performance of Naive Baye's for Text Classification". CS224N Spring 2003

[2] Daphne Koller, Mehran Sahami, Gates Building Computer Science Department Stanford University, "Hierarchically classifying documents using few words", Machine Learning International Workshop, 1997.

[3] Andrew McCallum, Kamal Nigam "Comparison of Event Classification for event models for text classification", AAAI-98 Workshop on Learning for Text Categorization, 1998.

[4] Michael Woolridge, Nicholas R.Jennings "Intelligent Agents: Theory and practice" submitted to Knowledge Engineering Review, 1995.

[5] Mehran Sahami, Susan Dumais, David Heckerman, Eric Horvitz, "A Bayesian Approach to filtering Junk E-mail" ,AAAI-98 Workshop on learning for Text Categorization, 1998.

[6] Stuart J. Russell and Peter Norvig.Artificial Intelligence: A Modern Approach 2003.

[7] Gregory Grefensette, Pasi Tapaninian, "Problems of Tokenization" Rank Xerox Research Center, Grenoble Laboratory, France, 1994.

[8] www.fipa.org

[9] http://jade.tilab.com

[10] www.intelligent-agents.com

AUTHOR'S PROFILE

Mr.S.Balakrishnan is pursuing his Ph.D program at Bharth University (BIHER), Chennai. He completed his M.E degree at Sathyabama University, Chennai in Computer Science & Engineering and his B.E at Manonmanium Sundarnar University, Tirunelveli, and Tamilnadu in Computer Science and Engineering in the year 2000.He has 12 years of experience in teaching Computer Science subjects in various engineering colleges and he is currently working as an Associate Professor & Head of Computer Science & Engineering Department at Rajiv Gandhi College of Engineering, Sriperumbudur, Chennai. He has published more than 10 books.

Dr. K.L. Shanmuganathan B.E M.E. M.S. Ph.D works as the Professor & Head of CSE Department of RMK Engineering College Chennai TamilNadu India. He has more than 18 years of teaching experience and his areas of specializations are Artificial Intelligence Computer Networks and DBMS.