Experience in implementing a Document Delivery Service

Francesco Gennai, Laura Abba, Marina Buzzi

Istituto per le Applicazioni Telematiche - CNR

via Alfieri, 1 56010 Pisa, Italy

E-mail: {F.Gennai, L.Abba, M.Buzzi}@iat.cnr.it

Maria G. Balestri, Silvana Mangiaracina

Biblioteca di Area - CNR

Via Gobetti 101, Bologna, Italy

E-mail: {balestri, mangiaracina}@area.bo.cnr.it

ABSTRACT

In this paper we propose an integration between electronic mail and web services for people such as library operators who need to send large files to Internet users. The proposed solution permits librarians to continue using the e-mail service to send large documents, but at the same time overcomes problems that users can encounter downloading large size files with e-mail agents. The library operator sends the document as an attachment to the destination address, on fly the e-mail server extracts and saves the attachments in a web-server disk file and substitutes them with a new message part that includes the URL pointing to the saved document. The receiver can download these large objects using a user-friendly browser.

KEYWORDS: e-mail, web, MIME, library, Internet document delivery service.

INTRODUCTION

Document Delivery services allows libraries to exchange papers from scientific serials or other types of documents, on the basis of library user requests. The Library of the Research Area in Bologna (BdA) exchanges about 5.000 documents per year with other national libraries. The average length of a document is 10 pages. This high number of requests has to be managed with a very small number of library operators, who often change over time. The main objectives of the biblio-MIME project were:

1. to simplify the Document Delivery activity;
2. to improve the quality of the service, facilitating distribution of large documents;
3. to reduce the time of the Document Delivery process.

To achieve this goal, we decided to start an Internet Document Delivery Service (IDDS). The simplest and most common mechanism to send a document is the electronic mail system. However, the transfer of large files via e-mail can present some limitations in both transport and mailbox access protocols:

Mail servers can impose message size limitations;

- Mailboxes access protocol time-out. In transferring large files, time-out parameters present in the POP (Post Office Protocol [1]) client can interrupt the connection making it impossible to download the message and those which follow it. In this case the user will not be able to download new messages until the server's administrator removes the blocking message.
- The availability of remote and local resources (such as disk space) can influence the behaviour of the IDDS.

Our idea is to combine the e-mail service (for document transmission) with the web service (for document downloading). Downloading documents using a web browser is more flexible and reliable than when using an e-mail client. The user can move between different platforms (where a web client is present) without requiring any client reconfiguration such as that required by e-mail clients.

THE SYSTEM

The Multipurpose Internet Mail Extensions is the keystone of our service. MIME was conceived as an extension to the electronic mail message format (RFC822) to allow for the inclusion of multimedia data in a standard message format. RFCs 2045 [2], 2046 [3], 2047 [4], 2048[5], 2049 [6] contain the basic definitions and descriptions of MIME mechanisms and message formats. An RFC822 message is composed of two parts: the header and the body. The body is a flat ASCII document with line length limitation. MIME redefines the RFC822 body format so as to include also non-textual data and structured body parts. New body formats and content transfer encoding mechanisms are defined by new header fields as the content-type and the content-transfer-encoding. Content-type header field can be used to specify the media type and subtype of data in the body of a message and to specify the canonical form of such data. Content-Transfer-Encoding header fields can be used to specify both the encoding transformation applied to the body and the domain of the result [2]. MIME defines data types: text, image, audio, video, application, message and multipart. For each data type more subtypes can be defined (i.e. image/tiff, image/gif, etc.). A multipart MIME message is a structured message composed of multiple parts containing different data types.

Using a MIME-compliant mail server, we were able to write simple scripts to implement a selective e-mail/web gateway. The system functioning is illustrated in Figure 1:

- The library operator receives one request (by e-mail),

262

locates and retrieves the document in the library, then converts it into electronic format through a standard scanner and e-mails it back to the sender.

- On fly the e-mail server extracts registered document types (.gif, .jpg, .pdf, .tiff) contained in MIME message parts and saves them in a web-server disk file. The extracted part is replaced with a brief service description (instructions such as download viewer applications, copyright rules) and the link pointing to the saved part. The other message parts remain unchanged.

- The resulting multipart message will continue the network trip towards its own final destination.

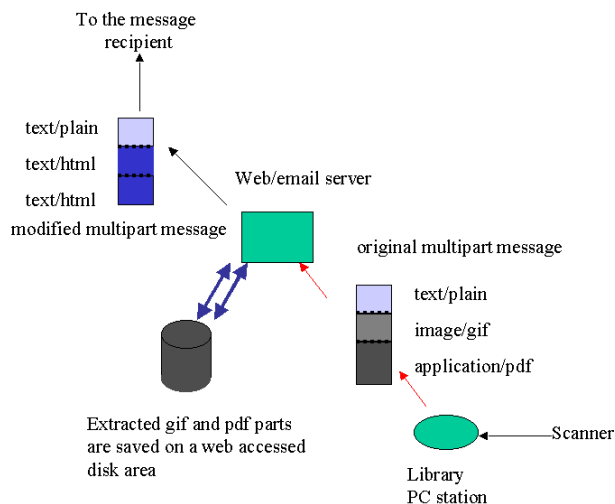- The receiver opens the message, clicks on the URLs and downloads the original files from the web server.

Figure 1: e-mail/web gateway behaviour description

On the same e-mail system we host e-mail services for different CNR domains, we offer a MIME attachment conversion service, e-mail to fax gateway service and recently we began the experimental biblio-MIME service. We have therefore implemented a selective service. In order to have messages processed by the our system, two conditions must be met:

1. the IP address of the client has to be registered in a service authorisation list;

2. the e-mail destination address has to end with the fake top level domain (TLD): **.save** This fake TLD domain (added by the librarian) will be automatically removed by the e-mail server and is necessary to request that the message pass through our service.

The first condition permits only authorised users to use the e-mail/gateway functionality. The second condition is important in order to permit operators to switch between regular e-mail operations or IDDS operations. When the messages match the above conditions the e-mail server pass to our scripts each message part and some variables

including the MIME part type, MIME part subtype, the MIME content-type name parameter (if present). The scripts check if the MIME part type and subtype match one of the registered types (image/tiff, image/jpeg, image/gif, application/pdf or application/octet-stream). The matching part is saved and substituted with a text/html part that includes links to a service description and to the saved part. Content-type and content-transfer-encoding mime headers are changed to reflect the new part content. Parts not matching the above conditions will be restored unchanged in their original positions inside the message body.

Documents are stored in the web directories for a period determined by a system parameter (14 days). All system activities are logged with, and accessible from, protected web pages.

BdA offers the Internet Document Delivery Service free of charge. Libraries using the service agree to respect copyright law.

CONCLUSION

This paper aims to emphasise the capability of the MIME-aware agents that furnish a very strong flexibility in extracting and managing parts of structured data.

Our experimentation has shown numerous advantages:

- Increased efficiency of the service, allowing the library to satisfy all requests received on the same working day;

- The service is user-friendly - training is not required;

- The system offers robustness in delivering and downloading documents, thus increasing the quality of the service.

This proposed system could be advantageously adopted not only in the field of library science but also in any application where automatic filtering and elaboration of message parts is required.

REFERENCES

1. Myers J. and Rose RFC1939. Post Office Protocol v. 3. ftp://ftp.isi.edu/in-notes/rfc1939.txt, May 1996.

2. Freed, Ned and al. RFC 2045. MIME Part One. ftp://ftp.isi.edu/in-notes/rfc2045.txt November 1996.

3. Freed, Ned and al. RFC 2046. MIME Part Two. ftp://ftp.isi.edu/in-notes/rfc2046.txt November 1996.

4. Moore, K. RFC 2047. MIME Part Three. ftp://ftp.isi.edu/in-notes/rfc2047.txt November 1996.

5. Freed, Ned and al. RFC 2048. MIME Part Four. ftp://ftp.isi.edu/in-notes/rfc2048.txt November 1996.

6. Freed, Ned and al. RFC 2049. MIME Part Five. ftp://ftp.isi.edu/in-notes/rfc2049.txt November 1996.