Network Working Group
Request for Comments: 744
NIC: 42857

Joanne Sattley
JZS@CCA

8 January 1978

MARS - A Message Archiving & Retrieval Service

I. Introduction

This document describes a Message Archiving and Retrieval Service (MARS) which has been developed at Computer Corporation of America; it utilizes the Datacomputer, a network database utility developed by CCA for ARPA. [Research and development of a prototype MARS system was supported by the Defense Advanced Research Projects Agency of the Department of Defense, under the ARPA Very Large Databases program, and was monitored by the Office of Naval Research under Contract No. N00014-76-C-0991.]

The Service is available, primarily, to groups for storage of teleconferencing transcripts. Is is also available, upon request, to individual ARPANET correspondents.

There are both 'public' and 'private' messages in the database. Public messages may be retrieved by anyone. The public collection includes the messages of the Header-People [@ MIT-MC] group, and the MsgGroup [@ USC-ISI] proceedings.

Private messages may be retrieved only by the users who have archived them, or anyone whose name appears on the list of message recipients.

Messages archived using MARS are heavily indexed and can be retrieved in a variety of ways, including Boolean combinations of message recipients, message composition date, any text words in the message subject, and text words in the message body. The MARS facilities are integrated very naturally into the existing collection of message-handling tools:

- A message is designated for archiving by sending it to MARS-Filer @ CCA using one of the usual message-mailing tools such as SNDMSG.

- A message is designated for retrieval by sending a request as ordinary mail to MARS-Retriever @ CCA.

The Filer program checks for mail every hour; the Retriever program checks every quarter-hour. The periodicity can be altered to meet demand but the intent is for MARS to operate as a background job and only during extremely low-activity periods.

The next section (II) describes the indexing operation in greater detail, and how to archive and retrieve messages. The last section (III) is an extractable user card.

II. Using MARS

A. Message Indexing

For each message, a vector of parsed tokens is created. The parsed tokens are collected by the message-field in which they occurred -- to be used as "indexes", i.e., values of inverted fields, by the Datacomputer.

The Filer "indexes", essentially without analysis, except for the following:

-- Each distinguishable section of the message is indexed separately; each header line is a separate inversion domain, as is the body of the message.

-- The header lines which contain ARPANET addresses are analyzed in order to index separately on mailbox and host.

-- The date-field is parsed and converted to the standard Tenex internal date/time format, which is better adapted for less-than/greater-than comparisons, as in retrievals which specify a date range.

-- One-character words in both the subject-field and the message-text field arbitrarily discarded.

-- Two-character words in the message-text field are arbitrarily discarded.

-- Hyphenated phrases, i.e., words bound together by hyphens, are retained intact.

-- All message formats which conform to RFC 733 standards are accommodated. The minimum requirements are: a date-field, a from-field, and a blank line between the message-header and message-body.

B. To Archive Messages

There are three modes of filing currently supported by MARS, to wit:

- -- single-message mode, wherein the MARS-Filer mailbox appears in the message as an addressee;

- -- forwarded-message mode, wherein the MARS-Filer mailbox appears as the only primary recipient;

and

- -- batch mode, wherein the mailing envelope is addressed to MARS-Filer and the subject-field contains the keyword "batch".

Until the ARPANET standard for the format of messages is implemented universally, the variability amongst formats is still greater than the Filer can handle as it stands. Nonetheless, a user can successfully file any message in a "foreign" format by forwarding it to the Filer under the aegis of a mail-handling program which does produce good formats. Admittedly, the correct header-field indexing, as described above, will not be done on the enclosed message; but at least, the words in its unreadable header fields will appear as "text" words in the indexing.

In the case of forwarded-message-mode filing, all interesting indexing information is extracted from the message-header of the forwarding envelope prior to discarding it. The name of the archiver, the date and time the message was forwarded, and the subject-line information are recorded. The remainder is handled as though it were a non-forwarded message which had been CC'd to the Filer.

A forwarded message may be 'annotated' by adding text (e.g., notes, comments, keywords) in the forwarding envelope. Annotations are filed and retrieved as part of the archived message.

In the case of batch-mode filing, only the archiver's name and the date and time s/he sent the package are extracted from the mailing envelope. The message-body portion is then treated as a series of individual messages.

C. To Retrieve Messages

Retrievals are initiated by sending a Retrieval Request (which is a specially formatted message) to "MARS-Retriever@CCA". Retrieved messages are mailed back, one at a time, and will appear as new mail in the requester's mailbox.

Retrieval Request messages can be composed using any SNDMSG-type of program, as follows:

. The recipient of the RR message must be MARS-Retriever @ CCA

. Other message header fields are ignored for now

. The message body portion of the RR is used to compose Datalanguage for performing the retrieval. Its format resembles a message header, or selected portions thereof.

The following list defines which field names are recognized, and some notes on their interpretation. The scanning of each field is terminated by a carriage-return.

DATE: The format of the date field is day-month-year. Use of hyphens is optional. This field will cause only those messages composed on the specified date to be retrieved.

AFTER: Use of this field will retrieve messages composed after the specified date.

SINCE: This field is interpreted like the AFTER: field.

BEFORE: Use of this field will retrieve messages composed before the specified date.

UNTIL: This field is interpreted like the BEFORE: field.

FROM: This field is expected to contain a valid mailbox name. The host specification is optional. If more than one name is specified, ORing of the names is implicit. Retrieval based upon host specification alone has not been implemented.

TO: This field is expected to contain one or more valid mailbox names. The host specification is optional. Spaces and commas between the names imply AND.

SUBJECT: Use of this field will retrieve all messages whose indexed subject-field contents match the specified word(s). Spaces and commas imply AND. The use of OR must be explicit.

TEXT: Use of this field will retrieve all messages whose indexed message-body contents match the specified word(s). Spaces and commas imply AND. The use of OR must be explicit.

An interactive TENEX-based program for composing RRs is available; the filename is "RR.SAV". A copy of this program is stored on the Datacomputer, available via DFTP under node COMMON>MARS.

There is also a copy of the program in CCA’s directory at SRI-KA; another in the CCA-ACCAT directory at ISIA.

III. MARS User Card

:: Archiving

Individual Messages

- Include MARS-Filer@CCA on message distribution list
- Forward message to MARS-Filer@CCA [Annotation is optional.]

Batches of Messages

- Incorporate the mail file as the message-body of a single message sent to MARS-Filer@CCA with the clue "BATCH" in its subject-field.

:: Retrieving

Using RR Program

- RR is a TENEX-based interactive program designed to prepare Retrieval Request messages and to mail them to MARS-Retriever@CCA.

Using SNDMSG-Type Program

- Send a message to MARS-Retriever@CCA, specifying the retrieval criteria in the body of the message.

Sample Retrieval Criteria

```
SUBJECT:RFC 733 or RFC733     ; OR must be explicit

TEXT:MARS Project,goals       ; spaces & commas imply AND

DATE: 14 November 1977
SINCE: 1 Nov 77               ; same as AFTER: 1 Nov 77
AFTER: 1  Dec 1977
UNTIL: 15 January 1978        ; same as BEFORE: 15 January 1978
BEFORE: Aug 7 76

FROM:  JZS@CCA         ; host specification is optional
FROM:  Hacker,JZS      ; comma implies OR (in FROM: field only)

TO:  CCA@SRI-KA     ; host specification is optional
TO:  SDD-0:,SDD-1:  ; spaces and commas imply AND
```