

Network Working Group
Request for Comments: 1555
Category: Informational

H. Nussbacher
Israeli Inter-University
Computer Center
Y. Bourvine
Hebrew University
December 1993

Hebrew Character Encoding for Internet Messages

Status of this Memo

This memo provides information for the Internet community. This memo does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Abstract

This document describes the encoding used in electronic mail [RFC822] for transferring Hebrew. The standard devised makes use of MIME [RFC1521] and ISO-8859-8.

Description

All Hebrew text when transferred via e-mail must first be translated into ISO-8859-8, and then encoded using either Quoted-Printable (preferable) or Base64, as defined in MIME.

The following table provides the four most common Hebrew encodings:

Hebrew letter	PC 8-bit Ascii	IBM EBCDIC	PC 7-bit Ascii	ISO 8859-8 8-bit Ascii
-----	-----	-----	-----	-----
alef	128	41	96	224
bet	129	42	97	225
gimel	130	43	98	226
dalet	131	44	99	227
he	132	45	100	228
vav	133	46	101	229
zayin	134	47	102	230
het	135	48	103	231
tet	136	49	104	232
yod	137	51	105	233
kaf sofit	138	52	106	234
kaf	139	53	107	235
lamed	140	54	108	236

mem sofit	141	55	109	237
mem	142	56	110	238
nun sofit	143	57	111	239
nun	144	58	112	240
samekh	145	59	113	241
ayin	146	62	114	242
pe sofit	147	63	115	243
pe	148	64	116	244
tsadi sofit	149	65	117	245
tsadi	150	66	118	246
qof	151	67	119	247
resh	152	68	120	248
shin	153	69	121	249
tav	154	71	122	250

Note: All values are in decimal ASCII except for the EBCDIC column which is in hexadecimal.

ISO 8859-8 8-bit ASCII is also known as IBM Codepage 862.

The default directionality of the text is visual. This means that the Hebrew text is encoded from left to right (even though Hebrew text is entered right to left) and is transmitted from left to right via the standard MIME mechanisms. Other methods to control directionality are supported and are covered in the complementary RFC 1556, "Handling of Bi-directional Texts in MIME".

All discussion regarding Hebrew in email, as well as discussions of Hebrew in other TCP/IP protocols, is discussed in the ilan-h@vm.tau.ac.il list. To subscribe send mail to listserv@vm.tau.ac.il with one line of text as follows:

```
subscribe ilan-h firstname lastname
```

MIME Considerations

Mail that is sent that contains Hebrew must contain the following minimum amount of MIME headers:

```
MIME-Version: 1.0
Content-type: text/plain; charset=ISO-8859-8
Content-transfer-encoding: BASE64 | Quoted-Printable
```

Users should keep their text to within 72 columns so as to allow email quoting via the prefixing of each line with a ">". Users should also realize that not all MIME implementations handle email quoting properly, so quoting email that contains Hebrew text may lead to problems.

In the future, when all email systems implement fully transparent 8-bit email as defined in RFC 1425 and RFC 1426 this standard will become partially obsolete. The "Content-type:" field will still be necessary, as well as directionality (which might be implicit for 8BIT, but is something for future discussion), but the "Content-transfer-encoding" will be altered to use 8BIT rather than Base64 or Quoted-Printable.

Optional

It is recommended, although not required, to support Hebrew encoding in mail headers as specified in RFC 1522. Specifically, the Q-encoding format is to be the default method used for encoding Hebrew in Internet mail headers and not the B-encoding method.

Caveats

Within Israel there are in excess of 40 Listserv lists which will now start using Hebrew for part of their conversations. Normally, Listserv will deliver mail from a distribution list with a "shortened" header, one that does not include the extra MIME headers. This will cause the MIME encoding to be left intact and the user agent decoding software will not be able to interpret the mail. Each user is able to customize how Listserv delivers mail. For lists that contain Hebrew, users should send mail to Listserv with the following command:

```
set listname full
```

where listname is the name of the list which the user wants full, unabridged headers to appear. This will update their private entry and all subsequent mail from that list will be with full RFC822 headers, including MIME headers.

In addition, Listserv usually maintains automatic archives of all postings to a list. These archives, contained in the file "listname LOGymm", do not contain the MIME headers, so all encoding information will be lost. This is a limitation of the Listserv software.

Example

Below is a short example of Quoted-Printable encoded Hebrew email:

```
Date:          Sun, 06 Jun 93 15:25:35 IDT
From:         Hank Nussbacher <HANK@VM.BIU.AC.IL>
Subject:      Sample Hebrew mail
To:          Hank Nussbacher <Hank@BARILVM>,
            Yehavi Bourvine <yehavi@hujivms>
MIME-Version: 1.0
Content-Type: Text/plain; charset=ISO-8859-8
Content-Transfer-Encoding: QUOTED-PRINTABLE
```

The end of this line contains Hebrew .=EC=E0=F8=F9=E9 =F5=
=F8=E0=EE =ED=E5=EC=F9

Hank Nussbacher =F8=EB=E1=F1=E5=
=F0 =F7=F0=E4

Acknowledgements

Many thanks to Rafi Sadowsky and Nathaniel Borenstein for all their help.

References

- [ISO-8859] Information Processing -- 8-bit Single-Byte Coded Graphic Character Sets, Part 8: Latin/Hebrew alphabet, ISO 8859-8, 1988.
- [RFC822] Crocker, D., "Standard for the Format of ARPA Internet Text Messages", STD 11, RFC 822, UDEL, August 1982.
- [RFC1425] Klensin, J., Freed N., Rose M., Stefferud E., and D. Crocker, "SMTP Service Extensions", RFC 1425, United Nations University, Innosoft International, Inc., Dover Beach Consulting, Inc., Network Management Associates, Inc., The Branch Office, February 1993.
- [RFC1426] Klensin, J., Freed N., Rose M., Stefferud E., and D. Crocker, "SMTP Service Extension for 8bit-MIME Transport", RFC 1426, United Nations University, Innosoft International, Inc., Dover Beach Consulting, Inc., Network Management Associates, Inc., The Branch Office, February 1993.

- [RFC1521] Borenstein N., and N. Freed, "MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies", Bellcore, Innosoft, September 1993.
- [RFC1522] Moore K., "MIME Part Two: Message Header Extensions for Non-ASCII Text", University of Tennessee, September 1993.

Security Considerations

Security issues are not discussed in this memo.

Authors' Addresses

Hank Nussbacher
Computer Center
Tel Aviv University
Ramat Aviv
Israel

Fax: +972 3 6409118
Phone: +972 3 6408309
EMail: hank@vm.tau.ac.il

Yehavi Bourvine
Computer Center
Hebrew University
Jerusalem
Israel

Phone: +972 2 585684
Fax: +972 2 527349
EMail: yehavi@vms.huji.ac.il