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:

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

```bash
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 LOGyymm", 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=
= = = = = = = = = = = = = = = = = = = = = = = = = =
Hank Nussbacher                                  =F8=EB=E1=F1=E5=
= = = = = = = = = = = = = = = = = = = = = = = = = =

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,


[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

[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.
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