

Greek Character Encoding for Electronic Mail 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.

Overview and Rational

This document describes a standard encoding for electronic mail [RFC822] containing Greek text and provides implementation guidelines. The standard is based on MIME [RFC1521] and the ISO 8859-7 character encoding. Although the implementation of this standard is straightforward several non-standard but "functional" - though unlikely to inter-operate - alternatives are in common use. For this reason we highlight common implementation and mail user agent setup errors.

Description

In order to transfer Greek text via electronic mail the text is first translated into the ISO 8859-7 character set, and then encoded using either the Base64 (preferable for text that is mainly Greek) or the Quoted-Printable (justifiable in cases where some Greek words appear inside predominately Latin text) method, as defined in MIME.

The following table provides most common Greek encodings (see also [RFC1345]):

| 0646 | 37 | M7 | 51 | MC | 23 | 69 | LG | L1 | G7 | GO | GC | 28 | 97 | Description |
|------|----|----|----|----|----|----|----|----|----|----|----|----|----|-------------------------------|
| ---- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | ----- |
| 0386 | ea | a2 | 86 | cd | 71 | 86 | | | | | | | | b6 Capital alpha with acute |
| 0388 | eb | b8 | 8d | ce | 72 | 8d | | | | | | | | b8 Capital epsilon with acute |
| 0389 | ec | b9 | 8f | d7 | 73 | 8f | | | | | | | | b9 Capital eta with acute |
| 038a | ed | ba | 90 | d8 | 75 | 90 | | | | | | | | ba Capital iota with acute |
| 038c | ee | bc | 92 | d9 | 76 | 92 | | | | | | | | bc Capital omicron with acute |
| 038e | ef | be | 95 | da | 77 | 95 | | | | | | | | be Capital upsilon with acute |
| 038f | f0 | bf | 98 | df | 78 | 98 | | | | | | | | bf Capital omega with acute |
| 0390 | | c0 | a1 | fd | | a1 | | | | | | | | c0 Small iota with acute and |

| | | | |
|------|-------------------|----------------|---------------------------------|
| 03be | a5 ee e8 ea ab e8 | 6f 4a 6e 71 ee | Small xi |
| 03bf | a6 ef e9 ef ac e9 | 70 4f 6f 72 ef | Small omicron |
| 03c0 | a7 f0 ea f0 ad ea | 71 50 70 73 f0 | Small pi |
| 03c1 | a8 f1 eb f2 ae eb | 72 52 71 75 f1 | Small rho |
| 03c2 | aa f2 ed f7 af ed | 77 57 72 77 f2 | Small final sigma |
| 03c3 | a9 f3 ec f3 ba ec | 73 53 73 76 f3 | Small sigma |
| 03c4 | ab f4 ee f4 bb ee | 74 54 74 78 f4 | Small tau |
| 03c5 | ac f5 f2 f9 bc f2 | 75 59 75 79 f5 | Small upsilon |
| 03c6 | ad f6 f3 e6 bd f3 | 76 46 76 7a f6 | Small phi |
| 03c7 | ae f7 f4 f8 be f4 | 78 58 77 7b f7 | Small chi |
| 03c8 | af f8 f6 e3 bf f6 | 79 43 78 7c f8 | Small psi |
| 03c9 | e0 f9 fa f6 db fa | 7a 56 79 7d f9 | Small omega |
| 03ca | e4 fa a0 fb b4 a0 | | fa Small iota with diaeresis |
| 03cb | e8 fb fb fc b8 fb | | fb Small upsilon with diaeresis |
| 03cc | e6 fc a2 de b6 a2 | | fc Small omicron with acute |
| 03cd | e7 fd a3 e0 b7 a3 | | fd Small upsilon with acute |
| 03ce | e9 fe fd f1 b9 fd | | fe Small omega with acute |

Note: All values are in hexadecimal.

The column headers refer to the following character sets:

- 0646 The ISO 2DIS 10646 code.
- 37 PC code page 737 also known as 437G. Note that some implementations of this code page do not include capital letters with acute.
- M7 Character set 8859-7 as implemented in Microsoft Windows 3.1, Microsoft Windows 3.11, and Microsoft Windows 95.
- 51 IBM code page 851.
- MC The Greek code page implemented on the Apple Macintosh computers.
- 23 IBM code page 423 (EBCDIC-CP-GR).
- 69 IBM code page 869.
- LG Latin Greek (iso-ir-19).
- L1 Latin Greek 1 (iso-ir-27). This page only contains the Greek capital letters whose glyphs do not exist in the Latin alphabet. The other capital letters are rendered using the equivalent Latin letter (e.g. "Greek capital letter alpha" is rendered as "Latin capital letter A"). When mapping "Latin Greek 1" text to ISO 8859-7 the Latin capital letters should only be transcribed to the equivalent Greek ones if a suitable heuristic determines that the

specific Latin letters are used to represent Greek glyphs.

G7 7 bit Greek (iso-ir-88).

G0 Old 7 bit Greek (iso-ir-18).

GC Greek CCITT (iso-ir-150).

28 Character set ISO 5428:1980 (iso-ir-55).

97 The target character set ISO 8859-7:1987 (ELOT-928) (iso-ir-126).

MIME Headers

A mail message that contains Greek text must contain at least the following MIME headers:

```
MIME-Version: 1.0
Content-type: text/plain; charset=ISO-8859-7
Content-transfer-encoding: BASE64 | Quoted-Printable
```

In the future, when all email systems implement fully transparent 8-bit e-mail as defined in RFC 1425 and RFC 1426 the message body encoding phase described in this standard will be no longer needed. In this case the requisite MIME headers are modified as follows:

```
MIME-Version: 1.0
Content-type: text/plain; charset=ISO-8859-7
Content-transfer-encoding: 8BIT
```

Even when RFC 1425 is used, Q or B encoding will continue to apply to message headers as detailed in the following section.

Optional

It is recommended, although not required, to support Greek encoding in mail headers as specified in RFC 1522. Specifically, the B-encoding format is to be the default method used for encoding Greek text in RFC-822 mail headers, and the Q-encoding format the method to use for the exceptional case of encoding a single Greek word or letter in an otherwise Latin-character-based header.

Example

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

```
Date:          Wed, 31 Jan 96 20:15:03 EET
From:         Diomidis Spinellis <dds@senanet.com>
Subject:      Sample Greek mail
To:          Achilleas Voliotis <achilles@theseas.ntua.gr>
MIME-Version: 1.0
Content-ID:   <Wed_Feb_14_18_49_50_EET_1996_0@senanet>
Content-Type: Text/plain; charset=ISO-8859-7
Content-Transfer-Encoding: Base64
```

```
yuHr5+zd8eEsCgrU7yDl6+vn7enq/CDh6/bc4uf07yDh8O/05evl3/Th6SDh8PwgMjYg4/Hc
7Ozh9OEuCG==
```

Discussion

It is possible [RFC1428] (and unfortunately common practice) to set up an arrangement of mail user and transfer agents that allow end users to communicate with Greek e-mail messages while violating a number of standards. Such arrangements are unlikely to offer wide scale interoperability.

One common error is to arrange the rendering and composition of Greek messages by rigging a mail user agent hosted in an ISO 8859-1 environment to use a presentation font that contains Greek glyphs and a keyboard input method that generates Greek text using those glyphs. The resulting messages begin with header items indicating contents in the ISO 8859-1 character set and include text in a totally different encoding. Unfortunately this "solution" appears to "work" across similar systems and is widely used.

One other error is to tag Greek text generated on Microsoft Windows platforms as ISO 8859-7 without an intermediate translation phase. It is important to note that the character set used by the Microsoft Windows Greek implementations is NOT the same as the ISO 8859-7 representation. First of all, the character set used to represent Greek characters differs slightly from the ISO 8859-7 encoding (this difference was instrumented in order to rectify the appearance of an early version of Microsoft Word for Windows in which the end-of-section symbol clashed with the "Greek capital alpha with acute" glyph). In addition, a number of 8-bit characters available on Greek Windows implementations are not part of the ISO 8859-7 character set.

Note that the ISO 8859-7 encoding is equivalent to the Greek Standards Organisation EL0T-928 encoding.

References

- [ISO-8859] Information Processing -- 8-bit Single-Byte Coded Graphic Character Sets, Part 7: Latin/Greek alphabet, ISO 8859-7, 1987.
- [RFC822] Crocker, D., "Standard for the Format of ARPA Internet Text Messages", STD 11, RFC 822, UDEL, August 1982.
- [RFC1345] Simonsen, K., "Character Mnemonics & Character Sets" RFC 1345, Rationel Almen Planlaegning, June 1992.
- [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.
- [RFC1428] Vaudreuil, G., "Transition of Internet Mail from Just-Send-8 to 8bit-SMTP/MIME", RFC 1428, CNRI, 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.

Author's Address

Diomidis Spinellis
SENA S.A.
Kyprou 27
GR-152 47 Filothei
GREECE

Phone: +30 (1) 6854535
Fax: +30 (1) 6840631
EMail: D.Spinellis@senanet.com

