Network Working Group
Request for Comments: 1947
Category: Informational

D. Spinellis
SENA S.A.

May 1996

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

```
                                                        diaeresis
0391 80 c1 a4 b0 41 a4 61    41 61 41 41 c1 Capital alpha
0392 81 c2 a5 b5 42 a5 62    42 62 42 42 c2 Capital beta
0393 82 c3 a6 a1 43 a6 67 23 43 67 43 44 c3 Capital gamma
0394 83 c4 a7 a2 44 a7 64 40 44 64 44 45 c4 Capital delta
0395 84 c5 a8 b6 45 a8 65    45 65 45 46 c5 Capital epsilon
0396 85 c6 a9 b7 46 a9 7a    46 7a 46 49 c6 Capital zeta
0397 86 c7 aa b8 47 aa 68    47 68 47 4a c7 Capital eta
0398 87 c8 ac a3 48 ac 75 5c 48 75 48 4b c8 Capital theta
0399 88 c9 ad b9 49 ad 69    49 69 49 4c c9 Capital iota
039a 89 ca b5 ba 51 b5 6b    4b 6b 4a 4d ca Capital kappa
039b 8a cb b6 a4 52 b6 6c 5e 4c 6c 4b 4e cb Capital lamda
039c 8b cc b8 bb 53 b7 6d    4d 6d 4c 4f cc Capital mu
039d 8c cd b7 c1 54 b8 6e    4e 6e 4d 50 cd Capital nu
039e 8d ce bd a5 55 bd 6a 21 4f 6a 4e 51 ce Capital xi
039f 8e cf be c3 56 be 6f    50 6f 4f 52 cf Capital omicron
03a0 8f d0 c6 a6 57 c6 70 3f 51 70 50 53 d0 Capital pi
03a1 90 d1 c7 c4 58 c7 72    52 72 51 55 d1 Capital rho
03a3 91 d3 cf aa 59 cf 73 5f 53 73 53 56 d3 Capital sigma
03a4 92 d4 d0 c6 62 d0 74    54 74 54 58 d4 Capital tau
03a5 93 d5 d1 cb 63 d1 79    55 79 55 59 d5 Capital upsilon
03a6 94 d6 d2 bc 64 d2 66 5d 56 66 56 5a d6 Capital phi
03a7 95 d7 d3 cc 65 d3 78    58 78 57 5b d7 Capital chi
03a8 96 d8 d4 be 66 d4 63 3a 59 63 58 5c d8 Capital psi
03a9 97 d9 d5 bf 67 d5 76 5b 5a 76 59 5d d9 Capital omega
03aa    da    ab    91                    da Capital iota with
                                              diaeresis
03ab    db    bd    96                    db Capital upsilon with
                                              diaeresis
03ac e1 dc 9b c0 b1 9b                    dc Small alpha with acute
03ad e2 dd 9d db b2 9d                    dd Small epsilon with acute
03ae e3 de 9e dc b3 9e                    de Small eta with acute
03af e5 df 9f dd b5 9f                    df Small iota with acute
03b0    e0 fc fe    fc                    e0 Small upsilon with acute
                                              and diaeresis
03b1 98 e1 d6 e1 8a d6       61 41 61 61 e1 Small alpha
03b2 99 e2 d7 e2 8b d7       62 42 62 62 e2 Small beta
03b3 9a e3 d8 e7 8c d8       63 47 63 64 e3 Small gamma
03b4 9b e4 dd e4 8d dd       64 44 64 65 e4 Small delta
03b5 9c e5 de e5 8e de       65 45 65 66 e5 Small epsilon
03b6 9d e6 e0 fa 8f e0       66 5a 66 69 e6 Small zeta
03b7 9e e7 e1 e8 9a e1       67 48 67 6a e7 Small eta
03b8 9f e8 e2 f5 9b e2       68 55 68 6b e8 Small theta
03b9 a0 e9 e3 e9 9c e3       69 49 69 6c e9 Small iota
03ba a1 ea e4 eb 9d e4       6b 4b 6a 6d ea Small kappa
03bb a2 eb e5 ec 9e e5       6c 4c 6b 6e eb Small lamda
03bc a3 ec e6 ed 9f e6       6d 4d 6c 6f ec Small mu
03bd a4 ed e7 ee aa e7       6e 4e 6d 70 ed Small nu
```

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).

GO 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 ELOT-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