

Network Working Group
Request for Comments: 1557
Category: Informational

U. Choi
K. Chon
KAIST
H. Park
Solvit Chosun Media
December 1993

Korean 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.

Introduction

This document describes the encoding method being used to represent Korean characters in both header and body part of the Internet mail messages [RFC822]. This encoding method was specified in 1991, and has since then been used. It has now widely being used in Korean IP networks.

This document also describes the name of the encoding method which is to be used in order to match the message header and body format of MIME [MIME1, MIME2].

This document describes only the encoding method for plain text. Other text subtypes, rich text and similar forms of text, are beyond the scope of this document.

Description

It is assumed that the starting code of the message is ASCII. ASCII and Korean characters can be distinguished by use of the shift function. For example, the code SO will alert us that the upcoming bytes will be a Korean character as defined in KSC 5601. To return to ASCII the SI code is used.

Therefore, the escape sequence, shift function and character set used in a message are as follows:

SO	KSC 5601
SI	ASCII
ESC \$) C	Appears once in the beginning of a line before any appearance of SO characters.

The KSC 5601 [KSC5601] character set that includes Hangul, Hanja (Chinese ideographic characters), graphic and foreign characters, etc., is two bytes long for each character.

For more information about Korean character sets please refer to the KSC 5601-1987 document. Also, for more detailed information about the escape sequence and the shift function you can look for the ISO 2022 [ISO2022] document.

Formal Syntax

Where this document in its formal syntax does not agree with the description part, priority should be given to the formal syntax of the document.

The notations used in this section of the document are according to those used in STD 11, RFC 822 [RFC822] with the same meaning.

* (asterisk) has the following meaning :
l*m "anything"

The above means that "anything" has to be used at least l times and at most m times. Default values for l and m are 0 and infinitive, respectively.

body = *e-line *1(designator *(e-line / h-line))

designator = ESC "\$" ")" "C"

e-line = *text CRLF

h-line = *text 1*(segment *text) CRLF

segment = SO 1*(one-of-94 one-of-94 SI
; (Octal, Decimal.)

ESC = <ISO 2022 ESC, escape> ; (33, 27.)

SO = <ASCII SO, shift out> ; (16, 14.)

SI = <ASCII SI, shift in> ; (17, 15.)

SP = <ASCII SP, space> ; (40, 32.)

one-of-94 = <any char in 94-char set> ; (41-176, 33.-126.)

CHAR = <any ASCII character> ; (0-177, 0.-127.)

text = <any CHAR, including bare CR & bare LF, but NOT
 including CRLF, and not including ESC, SI, SO>

MIME and RFC 1522 Considerations

The name to be used for the Hangul encoding scheme in the contents is "ISO-2022-KR". This name when used in MIME message form would be:

Content-Type: text/plain; charset=iso-2022-kr

Since the Hangul encoding is done with 7 bit format in nature, the Content-Transfer-Encoding-header does not need to be used. However, while using the Hangul encoding, current Hangul message softwares does not support Base64 or Quoted-Printable encoding applied on already encoded Hangul messages.

The Hangul encoded in the header part of the message is Korean EUC [EUC-KR]. In the EUC-KR encoding, the bytes with 8th bit set will be recognized as KSC-5601 characters. To use Hangul in the header part, according to the method proposed in RFC 1522, the encoded Hangul are "B" or "Q" encoded. When doing so, the name to be used will be EUC-KR.

Background Information

The Hangul encoding system is based on the ISO 2022 [ISO2022] environment according to its 4/4 announcement. However, the Hangul encoding does not include the announcement's escape sequence.

The KSC 5601 used in this document is, in definition, identical to the KSC 5601-1987, KSC 5601-1989 and KSC 5601-1992's 94x94 octet definition. Therefore, any revision that refers to KSC-5601 after 1992 is to be considered as having the same meaning.

At present, the Hangul encoding system is based on the experience acquired from the former widely used "N-Byte Hangul" among UNIX users. Actually, the encoding method, "N-Byte Hangul", using SO and SI was the encoding method used in SDN before KSC 5601 was made a national standard.

This code is intended to be used for the information interchange of Hangul messages; any other use of the code is not considered appropriate.

References

- [ASCII] American National Standards Institute, "Coded character set -- 7-bit American national standard code for information interchange", ANSI X3.4-1968
- [ISO2022] International Organization for Standardization (ISO), "Information processing -- ISO 7-bit and 8-bit coded character sets -- Code extension techniques", International Standard, 1986, Ref. No. ISO 2022-1986 (E).
- [KSC5601] Korea Industrial Standards Association, "Code for Information Interchange (Hangul and Hanja)," Korean Industrial Standard, 1987, Ref. No. KS C 5601-1987.
- [EUC-KR] Korea Industrial Standards Association, "Hangul Unix Environment," Korean Industrial Standard, 1992, Ref. No. KS C 5861-1992.
- [RFC822] Crocker, D., "Standard for the Format of ARPA Internet Text Messages", STD 11, RFC 822, UDEL, August 1982.
- [MIME1] Borenstein, N., and N. Freed, "MIME (Multipurpose Internet Mail Extensions): Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies", RFC 1521, Bellcore, Innosoft, September 1993.
- [MIME2] Moore, K., "MIME (Multipurpose Internet Mail Extensions) Part Two: Message Header Extensions for Non-ASCII Text", RFC 1522, University of Tennessee, September 1993.

Security Considerations

Security issues are not discussed in this memo.

Acknowledgments

The authors wants to thank all the people who assisted in writing this document. In particular, we thank Erik von der Poel, Felix M. Villarreal, Ienup Sung, Kyoung Namgoong, and Kyuho Kim.

Authors' Addresses

Uhhyung Choi
Korea Advanced Institute of Science and Technology
Department of Computer Science
Taejon, 305-701, Republic of Korea

Phone: +82-42-869-8718
Fax: +82-42-869-3510
EMail: uhhyung@kaist.ac.kr

Kilnam Chon
Korea Advanced Institute of Science and Technology
Department of Computer Science
Taejon, 305-701, Republic of Korea

Phone: +82-42-869-3514
Fax: +82-42-869-3510
EMail: chon@cosmos.kaist.ac.kr

Hyunje Park
Solvit Chosun Media, Inc.
748-16 Yeoksam-Dong, Kangnam-Gu
Seoul, 135-080, Republic of Korea

Phone: +82-2-561-0361
Fax: +82-2-569-4847
EMail: hjpark@dino.media.co.kr