E. Wada
Fujitsu Laboratories
J. Murai
Keio University
K. Yamamoto
IIJ Research Laboratory

Expires in six months

January, 1999

Japanese Character Encoding Scheme for Internet Messages

<draft-yamamoto-charset-iso-2022-jp-02.txt>

Status of this Memo

Abstract

This memo describes the character encoding scheme used in electronic mail, NetNews, and world wide web messages in Japanese networks. It was first specified by and used in JUNET then described in RFC 1468. The name of this character encoding scheme was originally known as 'JUNET code' and is now called 'ISO-2022-JP' when used in the context of MIME.

In ISO-2022-JP text, both one 7-bit byte Latin script (ASCII or JIS X 0201 Latin set) and two 7-bit bytes Kanji, Hiragana, Katakana and some other symbols and characters (JIS X 0208) are employed. Switching the graphic character sets is based on the extension techniques defined in ISO 2022.

This memo eliminates some ambiguities of RFC 1468. However, it NEVER introduces any essential changes against RFC 1468. Since the character encoding scheme is now widely used in Japanese IP communities, backward compatibility is most important in this revision.

This memo revises RFC 1468 on the following points:

1. It is clarified that ISO-2022-JP does NOT conform to ISO 2022.
2. The formal syntax is divided into two new yet compatible rules, namely, ISO-2022-JP decoding syntax and ISO-2022-JP encoding syntax.
3. The bit combinations permitted in JIS X 0208 are explicitly described in the syntax so that the invalid character positions will be excluded in ISO-2022-JP text.
4. Recommended graphic character sets are specified. That is, ASCII is RECOMMENDED rather than JIS X 0201 1976 Latin set. ONLY to embed YEN SIGN and OVER LINE, JIS 0201 Latin set MAY be used. Also, JIS X 0208 1983 (including 1990) is RECOMMENDED rather than JIS X 0208 1978.

1. Introduction

Efforts to exchange Japanese text via electronic mail[MAIL] and NetNews[NETNEWS] messages started in the early days of JUNET[JUNET], consisted of networks arbitrarily connected by UUCP in Japan. There was strong demand for exchanging ASCII[ASCII], JIS X 0201 Latin set[JISX0201-76], and JIS X 0208 graphic character sets.

JIS X 0201 Latin set is identical to ASCII except for REVERSE SOLIDUS (BACKSLASH) and TILDE. REVERSE SOLIDUS and TILDE are replaced by YEN SIGN and OVER LINE, respectively. This set is Japan's national variant of ISO 646[ISO646]. Note that section 4.1.2 of RFC 2046 discourages use of ISO 646 in electronic mail.

The JIS X 0208 graphic character sets consist of Kanji, Hiragana, Katakana and some other symbols and characters. Each character takes up two bytes.

At that time, some implementations of message transfer agents (MTA) had 7-bit restriction in their transport connection. Moreover, European countries used 8-bit text to convey their languages. It was thus crucial to design a 7-bit safe format to carry Japanese text both for robustness against such MTAs and for distinction from European languages.

Popular character encoding schemes(CES) for Japanese graphic character sets were EUC-JP and Shift_JIS. However, they were not chosen for a transfer form of Japanese text because they are 8-bit. To encode ASCII, JIS X 0201, JIS X 0208 1978[JISX0208-78] and 1983[JISX0208-83], JUNET code was designed. This CES switches these four graphic character sets according to the extension techniques defined in ISO 2022[ISO2022], which is described later in this memo.

During JUNET days, some systems erroneously used the escape sequence "ESC (H" for JIS X 0201 1976 due to a typological error in the JIS book. This escape sequence is officially registered for a Swedish graphic character set[ISOREG] and MUST NOT have been used in JUNET code. JUNET had difficulty in eliminating this misusage.

In 1993, the CES was described in RFC 1468[OLDSPEC] and the name "ISO-2022-JP" was given to identify it in the context of MIME[MIME].

The word "encoding" is ambiguous because it is used for both the extension techniques and for conversion of MIME's transport safe. For this reason, throughout this memo, the word "encoding" is always preceded by a prefix word to clarify its semantics and the word "charset"[CHARSET] is introduced to describe ISO-2022-JP CES. "ISO-2022-JP encoding" means creation of ISO-2022-JP text by MIME composers while "ISO-2022-JP decoding" refers to interpretation of ISO-2022-JP text by MIME viewers.

At the time of writing RFC 1468, one more revision for JIS X 0208, namely 1990[JISX0208-90], was available. RFC 1468, however, does NOT suggest using the revision identification sequence "ESC & @" even if two characters added in JIS X 0208 1990 are used.

This choice was practical from the empirical point of view but it was a departure from the ISO 2022 platform. Strictly speaking, ISO-2022-JP does NOT conform to ISO 2022 though the name is associated with ISO 2022.

The purposes of this memo are mainly to clarify ISO-2022-JP CES and to fix a bug of ABNF[ABNF] in RFC 1468, which failed to express null lines. To avoid confusion, NO new escape sequence for JIS X 0208 1990 NOR for JIS X 0212 1990[JISX0212-90] is introduced. Such efforts should be elaborated in other rich charsets.

NOTE: One more revision for JIS X 0208, namely 1997, is available [JISX0208-97]. Since it is not a new graphic character set, most parts of this memo do NOT explicitly talk about it. Likewise, JIS X 0201 1997[JISX0201-97] does NOT appear in the main body of this memo. For more information, see Historical Note.

2. Standard Keywords

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [KEYWORDS].

3. Description

For historical reasons, ISO-2022-JP can contain four graphic character sets, ASCII, JIS X 0201 1976 Latin set, JIS X 0208 1978 and JIS X 0208 1983 (including 1990). A charset should start with and should end in ASCII since many message systems assume so. Moreover, each line should be self-composite so that lines can be safely split. ISO-2022-JP was designed to satisfy these requirements. Its definition is as follows:

ISO-2022-JP text MUST start with ASCII and MUST end with ASCII. Each line of ISO-2022-JP text MUST end with ASCII or JIS X 0201 1976 (i.e. before CRLF). The next line starts in the graphic character set that was switched to before the end of the previous line.

ISO-2022-JP is NOT completely self-composite because a starting graphic character set of a line is not known if the previous line is missing. This is one of the reasons to recommend ASCII, instead of JIS X 0201 1976, in ISO-2022-JP.

An escape sequence is used to specify a start boundary of a graphic character set different form the preceding graphic character set. The following table gives the escape sequences and the graphic character sets used in ISO-2022-JP text. The "ISOREG" is the registration number in ISO's registry[ISOREG].

ESC Seq	Graphic Character Set	ISOREG
ESC (B	ASCII	6
ESC (J	JIS X 0201 1976 (Latin set)	14
ESC $ @	JIS X 0208 1978	42
ESC $ B	JIS X 0208 1983	87

For example, the escape sequence "ESC $ B" indicates that the bytes following this escape sequence are some characters of JIS X 0208 1983. To switch back to ASCII, the escape sequence "ESC (B" is used.

For the control characters in ASCII, [MIME] and [MAIL2] define CR and LF only. Section 4.1.2 of RFC 2046 says that HT and FF have de facto meaning. To make ISO-2022-JP syntactically and semantically larger than US-ASCII, this memo defines the semantics of ESC only. Semantics of other control characters in ASCII is outside the scope. Note that the extension techniques defined here switches graphical character sets, not control character sets. So, ISO-2022-JP CES assumes that %d0-31 is always the control character set in ASCII.

If Japanese text represented within the four graphic character sets is transferred by SMTP[SMTP] or NNTP[NNTP], it is STRONGLY RECOMMENDED to be in ISO-2022-JP form.

Not all implementations distinguish JIS X 0208 1978 and 1983. Moreover, if ISO-2022-JP text including both ASCII and JIS X 0201 1976 AND/OR ISO-2022-JP text including both JIS X 0208 1978 and JIS X 0208 1983 is converted to EUC-JP or Shift_JIS, original ISO-2022-JP text could not be re-produced.

So, to implement best interoperability, message composers SHOULD use ASCII and JIS X 0208 1983 ONLY. However, message composers MAY use JIS X 0201 1976 to embed YEN SIGN and OVER LINE. Message viewers MUST accept all four graphic character sets. Please note that the best strategy for interoperability is "TO BE LIBERAL IN WHAT YOU ACCEPT, AND CONSERVATIVE IN WHAT YOU SEND"[GUIDE].

JIS X 0201 Kana set MUST NOT be used in ISO-2022-JP text. It may erroneously appear in ISO-2022-JP text with the following sequences:

```
<8-bit Kana set char>
SO <7-bit Kana set char> SI
SO <8-bit Kana set char> SI
ESC ( I <7-bit Kana set char> ESC ( B
ESC ) I <8-bit Kana set char>
```

JIS X 0212 1990 MUST NOT be embedded in ISO-2022-JP text. It may erroneously appear following "ESC $ (D" in ISO-2022-JP text. To embed JIS X 0212 1990, other charsets such as ISO-2022-JP-2 [ISO-2022-JP-2] SHOULD be used.

IMPLEMENTATION NOTE: Even if a message composer does not provide any input method for graphic character sets which is not used in ISO-2022-JP, such as JIS X 0201 Kana set and JIS X 0212 1990, there are ways to include them by accident. Typical cases are citing and forwarding a message including JIS X 0201 Kana set and/or JIS X 0212 1990 (but labeled as ISO-2022-JP).

IMPLEMENTATION NOTE ON CITATION: If JIS X 0201 Kana set is cited, it MUST be removed anyway. One possible deletion is to convert the characters to corresponding characters in JIS X 0208 1983. Please note that as of this writing there doesn't exist a charset to contain JIS X 0201 Kana set. When JIS X 0212 1990 is cited, MIME composers MUST take one of the followings: (1) Remove all characters in JIS X 0212 1990 and specify ISO-2022-JP for the charset parameter. (2) Specify other richer charset such as ISO-2022-JP-2.

IMPLEMENTATION NOTE ON FORWARDING: Since this memo talks about ISO-2022-JP text only, forwarding an entire message is outside the scope. But MIME composers are REQUIRED to forward valid messages in the context of MIME. Mismatch between the charset parameter and its content body MUST be eliminated. The situation would be mixed up when the invalid message was digitally signed. But again, procedures to create valid messages are outside the scope of this memo.

4. Formal Syntax

Two sets of formal syntax rules are defined in this memo. One is ISO-2022-JP decoding syntax for message viewers and the other is ISO-2022-JP encoding syntax for message composers. ISO-2022-JP decoding syntax is designed to be redundant to maintain backward compatibility with existing message composers while ISO-2022-JP encoding syntax is designed to be necessary and sufficient. Readers are assumed to be familiar with the notational conventions defined in [ABNF] and [MAIL2].

4.1 Rules for ISO-2022-JP decoding syntax

ISO-2022-JP decoding syntax that embeds ASCII, JIS X 0201 1976, JIS X 0208 1978, JIS X 0208 1983 (including 1990) is defined as follows (see also Figure 1):

```
iso-2022-jp-text-d = *( line-d CRLF ) line-d
    ; When text is terminated by CRLF, the last 'line-d' is empty.

line-d = *single-byte-char-d
         *(*double-byte-segment-d single-byte-segment-d)
    ; line-d MUST be limited to 998 bytes according to [MAIL2].

single-byte-segment-d = single-byte-designator-d *single-byte-char-d

single-byte-designator-d = %d27 %d40 ( %d66 / %d74 )
    ; ESC "(" ( "B" / "J" )

single-byte-char-d = %d0-9 / %d11-12 / %d14-26 / %d28-127
    ; All bit combinations of ASCII code excluding
    ; LF(%d10), CR(%d13), and ESC(%d27).

double-byte-segment-d = double-byte-designator-d *double-byte-char-d

double-byte-designator-d = %d27 %d36 ( %d64 / %d66 )
    ; ESC "$" ( "@" / "B" )

double-byte-char-d = %d33-126.33-126
    ; All bit combinations of 94 x 94 graphic character set
```

NOTE: ISO-2022-JP decoding syntax above describes valid byte combinations for ISO-2022-JP. Other byte patters are invalid. Message viewers MUST handle any combinations of %d0-255 and MUST safely ignore the invalid patterns.

NOTE: The old specs, RFC 822 and RFC 1468, allow bare CR and bare LF. MIME prohibits them. Though [MAIL2] allows message viewers to accept bare CR and bare LF, it prohibits message composers from generating them. This memo conforms this recent trend. Bare CR and bare LF are irrelevantly accepted according to the note above. The ISO-2022-JP encoding syntax defined below prohibits message composers from generating them.

```
                          +---------------------->
                          |         +----------> |
                          |    +--+ |     +--+ | |
             +----------> | +->|1S|->  +->|1C|->->-> exit
             |     +--+ | | |  +--+    |  +--+ | |
    line-d -->  +->|1C|->-> |          <-------+ v
                |  +--+ |   |       +----------> |
                <-------+   |  +--+ |     +--+ | |
                            +->|2S|->  +->|2C|->->
                            |  +--+    |  +--+ | |
                            |          <-------+ |
                            <--------------------+
```

Figure 1: 'line-d' for ISO-2022-JP decoding syntax

NOTE: 1S, 1C, 2S, and 2C stand for 'single-byte-designator-d', 'single-byte-char-d', 'double-byte-designator-d', and 'double-byte-char-d', respectively.

4.2 Rules for ISO-2022-JP encoding syntax

To create ISO-2022-JP text or string, only characters in ASCII and JIS X 0208 1983 (including 1990) SHOULD be used. But if YEN SIGN and/or OVER LINE of JIS X 0201 1976 are to be used, the following rules SHOULD be followed:

(1) It is RECOMMENTED to use YEN SIGN(%d33.111) and OVER LINE(%d33.49) of JIS X 0208 1983 instead of YEN SIGN(%d92) and OVER LINE(%d126) of JIS X 0201 1976, respectively.

(2) If sequences of JIS X 0201 1976 Latin set can be produced, YEN SIGN and/or OVER LINE of the set MAY be used preceded by the designator.

(3) Otherwise, YEN SIGN and OVER LINE of JIS X 0201 1976 SHOULD be converted to REVERSE SOLIDUS and TILDE of ASCII, respectively.

IMPLEMENTATION NOTE: Display of YEN SIGN and OVER LINE of JIS X 0201 1976 is highly dependent on the receiving system. REVERSE SOLIDUS and TILDE of ASCII may be displayed. And again, if ISO-2022-JP text including both ASCII and JIS X 0201 1976 is converted to EUC-JP or Shift_JIS, original ISO-2022-JP text could not be re-produced. This would cause verification failure of digital signature.

The following is ISO-2022-JP encoding syntax to accomplish rule (1) or (3) above (see also Figure 2)

```
iso-2022-jp-text-e = *( line-e CRLF ) line-e
    ; When text is terminated by CRLF, the last 'line-e' is empty.
```

```
   line-e = *single-byte-char-e [*i-segment-e  f-segment-e]
       ; line-e MUST be limited to 998 bytes and SHOULD be limited to
       ; 78 bytes according to [MAIL2].
```

```
    i-segment-e = double-byte-designator-e 1*double-byte-char-e
                  single-byte-designator-e 1*single-byte-char-e

    f-segment-e = double-byte-designator-e 1*double-byte-char-e
                  single-byte-designator-e *single-byte-char-e

    single-byte-designator-e = %d27 %d40 %d66
        ; ESC "(" "B"

    single-byte-char-e = %d1-9 / %d11-12 / %d14-26 / %d28-127
        ; All bit combinations of ASCII code excluding
        ; NUL(%d0), LF(%d10), CR(%d13), and ESC(%d27).

    double-byte-designator-e = %d27 %d36 %d66
        ; ESC "$" "B"

    double-byte-char-e = %d33.33-126 /
        %d34.(33-46 / 58-65 / 74-80 / 92-106 / 114-121 / 126) /
        %d35.(48-57 / 65-90 / 97-122) /
        %d36.33-115 /
        %d37.33-118 /
        %d38.(33-56 / 65-88) /
        %d39.(33-65 / 81-113) /
        %d40.33-64 /
        %d48-78.33-126 /
        %d79.33-83 /
        %d80-115.33-126 /
        %d116.33-38
        ; The valid bit combinations of JIS X 0208 1990.

                          +-------------------------------------->
                          |                         +----------> |
                          |                         |     +--+ | |
             +----------> |                         |  +->|1C|->->-> exit
             |     +--+ | |    +--+    +--+    +--+ |  |  +--+ |
    line-e -->  +->|1C|->->->->|2S|->->|2C|->->|1S|->  <-------+
                |  +--+ |   |  +--+ |  +--+ |  +--+ |     +--+
                <-------+   |       <-------+       +-->->|1C|->->
                            |                          |  +--+ | |
                            |                          <-------+ |
                            <------------------------------------+
```

Figure 2: 'line-e' for ISO-2022-JP encoding syntax

NOTE: 1S, 1C, 2S, and 2C stand for 'single-byte-designator-e', 'single-byte-char-e', 'double-byte-designator-e', and 'double-byte-char-e', respectively.

5. MIME Considerations

This section discusses how to use ISO-2022-JP text or string (i.e. excluding CRLF) in the context of MIME. The scope is limited to the case where messages are transferred by SMTP[SMTP] or

NNTP[NNTP]. Other protocols are outside the scope of this memo.

5.1 The Charset Parameter

To identify the charset described in this memo, "ISO-2022-JP" MUST be used for the charset parameter wherever it is allowed. Please note that the charset parameter is case-insensitive.

For the charset parameter, MIME composers SHOULD choose as small charset as possible. For example, if a content body is ASCII, US-ASCII SHOULD be specified instead of ISO-2022-JP. This rule improves interoperability with other MIME viewers which do not support various charsets.

To the contrary, MIME viewers SHOULD be able to handle text labeled ISO-2022-JP even if its content is ASCII only. Remember the liberal/conservative rule.

5.2 Content Transfer Encoding

ISO-2022-JP text is already in 7-bit form and MIME-encoding, such as Base64 and Quoted-Printable, makes ISO-2022-JP text unreadable for non-MIME viewers. For this reason, MIME composers SHOULD choose "7bit" (case-insensitive) and SHOULD NOT MIME-encode ISO-2022-JP text with Base64 nor Quoted-Printable.

Note that "7bit" means that no MIME-encoding is applied and its syntax is in 7-bit form. MIME defines that "7bit" MIME-encoding is assumed if content transfer encoding is not present. So, this field can be omitted.

In certain situation, Base64 or Quoted-Printable MAY be used. Base64 is RECOMMENDED to MIME-encode ISO-2022-JP text rather than Quoted-Printable. This is because Base64 encoding is considered more suitable for ISO-2022-JP text than Quoted-Printable encoding for the following reasons: (1) Quoted-Printable encoding is designed for text which is mostly ASCII but ISO-2022-JP text is NOT. (2) Quoted-Printable encoding produces various results while Base64 encoding creates a unique output if folding is ignored. This means that it is easier to implement interoperability for Base64 encoding than for Quoted-Printable encoding.

IMPLEMENTATION NOTE: The key word "=" of Quoted-Printable encoding is used in JIS X 0208 (see 'double-byte-char-e'). If only the ASCII portion of ISO-2022-JP text is MIME-encoded with Quoted-Printable, it is likely that such text cannot be MIME-decoded to the original since the "=" characters in JIS X 0208 portion are considered to be the key word of Quoted-Printable. If ISO-2022-JP text needs to be MIME-encoded with Quoted-Printable, the entire text must be MIME-encoded. Some old implementations made this mistake.

MIME viewers MUST be able to handle ISO-2022-JP text even if it is encoded with Base64 or Quoted-Printable. Also, MIME viewers SHOULD be able to handle ISO-2022-JP text even if its MIME-encoding is specified as "8bit".

5.3 Header Extensions

In a header or in MIME content headers where `encoded-word's are allowed, Japanese string (i.e. excluding CRLF) which consists of the four graphic character sets MUST be in the form defined in RFC 2047[MIME]. For "charset", "ISO-2022-JP" (case-insensitive) SHOULD be specified. ISO-2022-JP string MUST end with ASCII (see 'line-e').

Message composers SHOULD use "B" encoding for ISO-2022-JP string while message viewers MUST be able to handle both "B" and "Q" encoding.

IMPLEMENTATION NOTE: It should be noted that these requirements above apply to any fields in MIME content headers. For example, RFC 2047 mechanism MUST be used to represent ISO-2022-JP string when embedded in Content-Description: content header in a part of a multipart.

ISO-2022-JP string MUST NOT be embedded directly in a header NOR in MIME content headers. EUC-JP string and Shift_JIS string SHOULD NOT be inserted even if they are in the form defined in RFC 2047.

The reason why ISO-2022-JP string and "B" encoding is recommended to embed Japanese string in header fields is that RFC 1468 defines so. For historical reasons, not many message viewers support EUC-JP, Shift_JIS, and "Q" encoding,

The "language" extension defined in RFC 2231[PARAMETER] SHOULD NOT be used without any special purposes.

5.4 MIME Parameter Extensions

To embed Japanese string (i.e. excluding CRLF) to a MIME parameter value, the string MUST be in the form defined in RFC 2231. For "charset", "ISO-2022-JP" (case-insensitive) SHOULD be specified. "language" SHOULD be omitted. The entire ISO-2022-JP string MUST end with ASCII.

Both "extended-value" and "extended-other-values" need not to be self-composite since they are to be concatenated first when received. That is, they can contain any portion of ISO-2022-JP string which may or may not end with ASCII.

IMPLEMENTATION NOTE: The MIME encoding defined in RFC 2231 is a variant of "Q" encoding. Since many key words of header (including "%" and "'") may appear in ISO-2022-JP string, they MUST be encoded with the variant. It is a good idea to encode all characters other than [0-9A-Za-z] (%d48-57 / %d65-90 / %d97-122).

6. Requirements for Message Transfer Agents (MTA)

Several MTAs automatically convert any message to a local charset.

This means that it is very likely that a charset other than ISO-2022-JP is converted to the local charset by a routine which assumes that incoming data is ISO-2022-JP text. MTAs SHOULD NOT convert charsets in general. If MTAs need to convert charsets, they MUST do so according to the charset parameter of MIME messages.

Actually, some MTAs convert "ESC (B" into "ESC (J" and "ESC $ B" into "ESC $ @" interchangeably. Gateways SHOULD NOT convert in this way. Such MTAs are the primary cause of verification failure in digital signature.

7. Historical Note

The JUNET code was originally described in the JUNET User's Guide [JUNET].

JIS X 0208, which was originally called JIS C 6226, was published in 1978. It was revised in 1983, 1990 and 1997.

The revision in 1983 was made because the Japanese Government published the new Kanji graphic character set called "Joyo Kanji Hyo (Kanji List for Daily Use)" for use in daily life, in the public documents, in newspapers in 1981. Some of the character shapes were modified or simplified. To harmonize JIS X 0208 1978 with Joyo Kanji, pairs of characters were swapped in the code positions and several characters had their shapes amended. Moreover, since new characters were added to the Kanji set called "Jinmei You Kanji Beppyo (Additional Kanji List for Use with People's Name)", four characters were added to and a lot of special characters were included in JIS X 0208 1983. This graphic character set was treated as a new one and a new escape sequence "ESC $ B" was assigned.

Since the Kanji set for People's names was extended, two more Kanji characters were added at the end of JIS X 0208 in 1990 and the result is called JIS X 0208 1990. The revised graphic character set is indeed backward compatible with the previous set. Leaving the designation sequence unchanged, only the revision identification sequence "ESC & @" was assigned.

To be strict, the designation sequence of JIS X 0208 1990 is "ESC & @ ESC $ B". However, ISO-2022-JP CES uses "ESC $ B" for practical reasons even though the two additional Kanji characters are used.

The purpose of revision in 1997 is to clarify references of each character. No new characters were introduced. As a graphic character set, JIS X 0208 1997 is equivalent to JIS X 0208 1990. The designation sequence is the same.

Therefore, among versions of JIS X 0208, the version of 1978 is of a different graphic character set while the versions of 1983, 1990 and

1997 are regarded as a family.

For more information on the differences, please refer to Appendix J, [GLOBEFISH]. It describes what characters have been added, swapped, simplified, etc, in the revisions.

JIS X 0201, originally called JIS C 6220, was published in 1976. It was revised in 1997 but no change was introduced. They are thus the same graphic character set.

8. Security Considerations

ISO-2022-JP is NOT believed to introduce any new security holes to the Internet.

References

[ABNF] D. Crocker and Paul Overell, "Augmented BNF for Syntax Specifications: ABNF", RFC2234, November 1997.

[ASCII] American National Standards Institute, "Coded character set -- 7-bit American national standard code for information interchange", ANSI X3.4-1986.

[CHARSET] N. Freed and J. Postel, "IANA Charset Registration Procedures", RFC2278, January 1998.

[GLOBEFISH] K. Lunde, "Understanding Japanese Information Processing", O'Reilly & Associates, Inc., 1993.

[GUIDE] G. Scott, "Guide for Internet Standards Writers", RFC2360, June 1998.

[ISO-2022-JP-2] M. Ohta and K. Handa, "ISO-2022-JP-2: Multilingual Extension of ISO-2022-JP", RFC 1554, December 1993.

[ISO2022] International Organization for Standardization (ISO), "Information technology -- Character code structure and extension techniques", International Standard, Ref. No. ISO 2022-1991(E).

[ISO646] International Organization for Standardization (ISO), "Information technology -- ISO 7-bit coded character set for information interchange", International Standard, Ref. No. ISO/IEC 646:1991(E).

[ISOREG] International Organization for Standardization (ISO), "International Register of Coded Character Sets To Be Used With Escape Sequences".

[JISX0201-76] Japanese Industrial Standards Committee, "Code for information interchange", JIS C 6220 1976(aka JIS X 0201 1976).

[JISX0201-97] Japanese Industrial Standards Committee, "7-bit and

8-bit coded character sets for information interchange", JIS X 0201 1997.

[JISX0208-78] Japanese Industrial Standards Committee, "Code of the

Japanese graphic character set for information interchange", JIS C 6226 1978(aka JIS X 0208 1978).

[JISX0208-83] Japanese Industrial Standards Committee, "Code of the Japanese graphic character set for information interchange", JIS C 6226 1983(aka JIS X 0208 1983).

[JISX0208-90] Japanese Industrial Standards Committee, "Code of the Japanese graphic character set for information interchange", JIS X 0208 1990.

[JISX0208-97] Japanese Industrial Standards Committee, "7-bit and 8-bit double byte coded KANJI sets for information interchange", JIS X 0208 1997

[JISX0212-90] Japanese Industrial Standards Committee, "Code of the supplementary Japanese graphic character set for information interchange", JIS X0212 1990.

[JUNET] JUNET Riyou No Tebiki Sakusei Iin Kai (JUNET User's Guide Drafting Committee), "JUNET Riyou No Tebiki (Dai Ippan)" ("JUNET User's Guide (First Edition)"), February 1988.

[KEYWORDS] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.

[MAIL] D. Crocker, "Standard for the Format of ARPA Internet Text Messages", STD 11, RFC 822, August 1982.

[MAIL2] P. Resnick, "Internet Message Format Standard", draft-ietf-drums-msg-fmt-07.txt, January 1999.

[MIME] The primary definition of MIME. "MIME Part 1: Format of Internet Message Bodies", RFC 2045; "MIME Part 2: Media Types", RFC 2046; "MIME Part 3: Message Header Extensions for Non-ASCII Text", RFC 2047; "MIME Part 4: Registration Procedures", RFC 2048; "MIME Part 5: Conformance Criteria and Examples", RFC 2049; November 1996.

[NETNEWS] M. Horton and R. Adams, "Standard for Interchange of USENET Messages", RFC 1036, December 1987.

[NNTP] B. Kantor and P. Lapsley, "Network News Transfer Protocol", RFC 977, February 1986.

[OLDSPEC] J. Murai, M. Crispin, and E. van der Poel, "Japanese Character Encoding for Internet Messages", RFC 1468, June 1993.

[PARAMETER] N. Freed and K. Moore, "MIME Parameter Value and Encoded Word Extensions: Character Sets, Languages, and Continuations", RFC 2231, November 1997.

[SMTP] J. Postel, "Simple Mail Transfer Protocol", RFC 821, August 1982.

Acknowledgements

Many people contributed to RFC 1468. The original authors of RFC 1468 wished to thank in particular Akira Kato, Masahiro Sekiguchi and Ken'ichi Handa.

The authors of this memo would acknowledge the original work of Erik M. van der Poel and Mark Crispin. Our deep gratitude goes to Noritoshi Demizu, Kenichi Handa, Jun'ichiro Ito, Shuhei Kobayashi, Tomohiko Morioka, Makoto Murata, Erik M. van der Poel, Shigeya Suzuki, and Akira Tanaka (in alphabetical order) for their contribution to this memo.

Authors' Addresses

Eiiti WADA
Fujitsu Laboratories, Ltd
4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki 211-8588 JAPAN

Phone: +81-44-754-2608
FAX: +81-44-754-2580
EMail: wada@u-tokyo.ac.jp

Jun MURAI
Keio University
5322 Endo, Fujisawa 252-0816 JAPAN

Phone: +81-466-47-5111
FAX: +81-466-49-1101
EMail: jun@wide.ad.jp

Kazuhiko YAMAMOTO
Research Laboratory, Internet Initiative Japan Inc.
Takebashi Yasuda Bldg., 3-13 Kanda Nishiki-cho Chiyoda-ku, Tokyo
101-0054 JAPAN

Phone: +81-3-5259-6350
FAX: +81-3-5259-6351
EMail: kazu@iijlab.net

Appendix

An example of a MIME text object is as follows:

```
Content-Type: text/plain; charset=iso-2022-jp
Content-Transfer-Encoding: 7bit

ISO-2022-JP text comes here.
```

Another form of the example above is as follows:

```
Content-Type: Text/Plain; Charset=ISO-2022-JP

ISO-2022-JP text comes here.
```

The following is an example of RFC 2047 header encoding:

```
To: jun@wide.ad.jp
Subject: =?iso-2022-jp?B?GyRCJDMkTkZ8S1w4bCQsRkkkYSRsJFAbKEI=?=
 OK
 =?iso-2022-jp?B?GyRCJEckOSEjGyhC?=
From: Kazu Yamamoto (=?iso-2022-jp?B?GyRCOzNLXE9CSScbKEI=?=)
 <kazu@iijlab.net>
```

The following is an example of RFC 2231 parameter encoding:

```
Content-Disposition: attachment;
 filename*=iso-2022-jp''%1B%24B%25U%25%21%25%24%25k%1B%28B
```

Change Log

Differences between 01 and 02

- Clarifies that %d0-31 is always the character set of ASCII.
- ISO-2022-JP is RECOMMENDED for MIME parameter extensions according to suggestions from some vendors.
- The word "Character Encoding Scheme(CES)" is adopted to align to RFC2278.
- Several typos are fixed.
- Message viewers MUST handle any combinations of %d0-255, which was described as %d0-127.
- single-byte-char-e conforms [MAIL2].
- Refers to RFC 2046 to explain one reason to discourage use of JIS X 0201.

Differences between 00 and 01

- Author information was updated.
- References were updated.
- Refers to the globefish book to describes the differences between each revision.
- Explicitly says that ISO-2022-JP is syntactically and semantically larger than US-ASCII.
- line-d and line-d are limited to 998.
- Clarifies that message viewers MUST handle any byte combinations.
- Includes the consideration for bare CR and bare LF.