

- ・ [Java](#)

参考サイト

@ IT > Java Solution > 短期集中連載：Java の文字化け対策 FAQ

■第 1 回 Web ブラウザが文字コードを判定する基準は？

<http://www.atmarkit.co.jp/fjava/rensai3/mojibake01/mojibake01.html>

■第 2 回 JSP で特殊文字が文字化けする場合の対処方法

<http://www.atmarkit.co.jp/fjava/rensai3/mojibake02/mojibake02.html>

■第 3 回 DB や入力フォームで文字化けする場合の対処方法

- ・ <http://www.atmarkit.co.jp/fjava/rensai3/mojibake03/mojibake03.html>

デフォルト・エンコーディングは、Windows 上の JVM では Windows-31J (MS932) となり、UNIX では LANG 変数で設定された文字コード (EUC-JP や UTF-8 など) になる。そのため、例えば Windows 上で開発した成果物を Linux サーバ上で運用するケースなどにおいて、デフォルト・エンコーディングの違いによる文字化けが発生しやすい。

Java アプリケーション開発では、自分が書いたコードがデフォルト・エンコーディングに依存しないように心掛けよう。例えばテキスト・ファイルを読み込む場合は、以下のように FileInputStream クラスと InputStreamReader クラスで実装にすればよいだろう。

```
InputStreamReader isr = new InputStreamReader(  
    new FileInputStream("filename.txt"), "Windows-31J");
```

この方法であれば、InputStreamReader クラスのコンストラクタで文字コードを明示的に記述でき、デフォルト・エンコーディングに依存しない実装となる。

■第 4 回 JSP のインクルードに関連する文字化けを解消

<http://www.atmarkit.co.jp/fjava/rensai3/mojibake04/mojibake04.html>

その他

■ Wikipedia Microsoft コードページ 932

<http://ja.wikipedia.org/wiki/Microsoft%E3%82%B3%E3%83%BC%E3%83%89%E3%83%9A%E3%83%BC%E3%82%B8932>

■ Shift_JIS のエイリアスの変更について

<http://www.ingrid.org/java/i18n/unicode.html>

■文字コードの墓場 - しいしせねっと

<http://siisise.net/charset.html>

検証

よく使うエンコーディング

- ・ [Windows-31J](#)
- ・ [Shift_JIS](#)

- MS932
- euc-jp
- euc-jp-linux

正式名と別名

- EUC-JP
 - eucjis
 - x-eucjp
 - csEUCPkdFmtjapanese
 - eucjp
 - Extended_UNIX_Code_Packed_Format_for_Japanese
 - x-euc-jp
 - euc_jp
- ISO-2022-JP
 - jis
 - jis_encoding
 - csjisencoding
 - csISO2022JP
 - iso2022jp
- JIS_X0201
 - JIS_X0201
 - X0201
 - JIS0201
 - csHalfWidthKatakana
- JIS_X0212-1990
 - jis_x0212-1990
 - iso-ir-159
 - x0212
 - JIS0212
 - csISO159JISX02121990
- Shift_JIS
 - shift-jis
 - x-sjis
 - ms_kanji
 - shift_jis
 - csShiftJIS
 - sjis
 - pck
- UTF-16
 - UTF_16
- UTF-8
 - UTF8
- windows-31j
 - csWindows31J
 - windows-932
 - MS932
- x-JIS0208
 - JIS_C6626-1983
 - JIS0208
 - csISO87JISX0208
 - x0208
 - JIS_X0208-1983
 - iso-ir-87
- x-euc-jp-linux

- euc_jp_linux
- euc-jp-linux
- x-eucJP-Open
 - EUC_JP_Solaris
 - eucJP-open

エンコーディング一覧取得

- ソース

```
import java.nio.charset.Charset;
import java.util.Iterator;
import java.util.SortedMap;

public class ListCharsets {
    public static void main(String[] args) {

        // 全ての Charset を取得
        SortedMap charsets = Charset.availableCharsets();
        Iterator ci = charsets.keySet().iterator();
        while (ci.hasNext()) {
            // それぞれの Charset について
            Charset charset = Charset.forName((String)ci.next());
            // 全ての別名を取得
            String names = charset.toString() + " (";
            Iterator ai = charset.aliases().iterator();
            while (ai.hasNext()) {
                String alias = (String) ai.next();
                names += " " + alias.toString();
            }
            names += ")";
            // 名前と別名を出力
            System.out.println(names);
        }
    }
}
```

■ WindowsXP+jdk1.4.2_08

```
Big5 ( csBig5)
Big5-HKSCS ( big5-hkscs Big5_HKSCS big5hkscs)
EUC-JP ( eucjis x-eucjp csEUCPkdFmtjapanese eucjp Extended_UNIX_Code_Packed_Format_for_Japanese
x-euc-jp euc_jp)
EUC-KR ( ksc5601 5601 ksc5601_1987 ksc_5601 ksc5601-1987 euc_kr ks_c_5601-1987 euckr csEUCKR)
GB18030 ( gb18030-2000)
GBK ( windows-936 CP936)
ISO-2022-JP ( jis jis_encoding csjisencoding csISO2022JP iso2022jp)
ISO-2022-KR ( ISO2022KR csISO2022KR)
ISO-8859-1 ( iso-ir-100 8859_1 ISO_8859-1 ISO8859_1 819 csISOLatin1 IBM-819 ISO_8859-1:1987 latin1
cp819 ISO8859-1 IBM819 ISO_8859_1 l1)
ISO-8859-13 ( iso8859_13)
ISO-8859-15 ( 8859_15 csISOLatin9 IBM923 cp923 923 L9 IBM-923 ISO8859-15 LATIN9 ISO_8859-15 LATINO
csISOLatin9 ISO8859_15_FDIS ISO-8859-15)
ISO-8859-2 ( l2 iso-ir-101 ISO_8859-2:1987 ISO_8859-2 latin2 csISOLatin2 iso8859_2 ISO8859-2)
ISO-8859-3 ( )
ISO-8859-4 ( iso-ir-110 l4 latin4 csISOLatin4 iso8859_4 ISO_8859-4:1988 ISO_8859-4)
ISO-8859-5 ( cyrillic iso8859_5 ISO_8859-5 iso-ir-144 csISOLatinCyrillic)
ISO-8859-6 ( )
ISO-8859-7 ( greek8 ECMA-118 sun_eu_greek EL0T_928 ISO_8859-7:1987 iso-ir-126 ISO_8859-7 iso8859_7
greek csISOLatinGreek)
ISO-8859-8 ( )
ISO-8859-9 ( iso-ir-148 latin5 l5 ISO_8859-9 ISO_8859-9:1989 csISOLatin5 iso8859_9)
JIS_X0201 ( JIS_X0201 X0201 JIS0201 csHalfWidthKatakana)
JIS_X0212-1990 ( jis_x0212-1990 iso-ir-159 x0212 JIS0212 csISO159JISX02121990)
KOI8-R ( koi8 cskoi8r)
Shift_JIS ( shift-jis x-sjis ms_kanji shift_jis csShiftJIS sjis pck)
TIS-620 ( )
US-ASCII ( ISO646-US IBM367 ASCII cp367 ascii7 ANSI_X3.4-1986 iso-ir-6 us 646 iso_646.irv:1983
csASCII ANSI_X3.4-1986 ISO_646.irv:1991)
UTF-16 ( UTF_16)
UTF-16BE ( X-UTF-16BE UTF_16BE ISO-10646-UCS-2)
UTF-16LE ( UTF_16LE X-UTF-16LE)
UTF-8 ( UTF8)
```

```

windows-1250 { cp1250)
windows-1251 { cp1251)
windows-1252 { cp1252)
windows-1253 { cp1253)
windows-1254 { cp1254)
windows-1255 { }
windows-1256 { }
windows-1257 { cp1257)
windows-1258 { }
windows-31j { csWindows31J windows-932 MS932)
x-EUC-CN { gb2312 EUC_CN euccn euc-cn gb2312-80 gb2312-1980)
x-euc-jp-linux { euc_jp_linux euc-jp-linux)
x-EUC-TW { cns11643 euc_tw EUC-TW euctw)
x-eucJP-Open { EUC_JP_Solaris eucJP-open)
x-ISCII91 { iscii ST_SEV 358-88 iso-ir-153 csISO153GOST1976874 ISCII91)
x-JIS0208 { JIS_C6626-1983 JIS0208 csISO87JISX0208 x0208 JIS_X0208-1983 iso-ir-87)
x-Johab { johab ms1361 ksc5601-1992 ksc5601_1992)
x-MS950-HKSCS { MS950_HKSCS)
x-mswin-936 { ms936 ms_936)
x-windows-949 { windows949 ms_949 ms949)
x-windows-950 { windows-950 ms950)

```

文字コード変換

- <http://developer.java.sun.com/developer/technicalArticles/Programming/sprintf/PrintfFormat.java> を取得
- ソース

```

import java.io.UnsupportedEncodingException;

public class Test03 {

    public static void main(String[] args) {
        System.out.println(outputEncodingBytes('A'));
        System.out.println(outputEncodingBytes('1'));
        System.out.println(outputEncodingBytes('あ'));
        System.out.println(outputEncodingBytes('ア'));
        System.out.println(outputEncodingBytes('A'));
        System.out.println(outputEncodingBytes('ー'));
        System.out.println(outputEncodingBytes('~'));
    }

    static String outputEncodingBytes(char c1) {
        String str = c1 + " :";
        try {
            str += outputBytes(c1);
            str += outputBytes(c1, "Shift-JIS");
            str += outputBytes(c1, "MS932");
            str += outputBytes(c1, "euc-jp");
            str += outputBytes(c1, "euc-jp-linux");
        } catch (UnsupportedEncodingException e) {
            e.printStackTrace();
        }
        return str;
    }

    static String outputBytes(char c, String encoding)
        throws UnsupportedEncodingException {
        return " "
            + encoding
            + "("
            + hexBytes(new String("" + c).getBytes(encoding))
            + ")";
    }

    static String outputBytes(char c) {
        return " default(" + hexBytes(new String("" + c).getBytes()) + ")";
    }

    static String hexBytes(byte[] bytes) {
        String output = "";
        for (int count = 0; count < bytes.length; count++) {
            String hex = new PrintfFormat("%2x").sprintf(bytes[count]);
            output += " 0x" + hex.substring(hex.length() - 2, hex.length());
        }
        return output;
    }
}

```

```
}  
}
```

■ WindowsXP + jdk1.4.2_04

```
A : default( 0x41) Shift-JIS( 0x41) MS932( 0x41) euc-jp( 0x41) euc-jp-linux( 0x41)  
1 : default( 0x31) Shift-JIS( 0x31) MS932( 0x31) euc-jp( 0x31) euc-jp-linux( 0x31)  
あ : default( 0x82 0xa0) Shift-JIS( 0x82 0xa0) MS932( 0x82 0xa0) euc-jp( 0xa4 0xa2) euc-jp-linux( 0  
xa4 0xa2)  
ア : default( 0x83 0x41) Shift-JIS( 0x83 0x41) MS932( 0x83 0x41) euc-jp( 0xa5 0xa2) euc-jp-linux( 0  
xa5 0xa2)  
A : default( 0x82 0x60) Shift-JIS( 0x82 0x60) MS932( 0x82 0x60) euc-jp( 0xa3 0xc1) euc-jp-linux( 0  
xa3 0xc1)  
- : default( 0x81 0x7c) Shift-JIS( 0x21 0x29) MS932( 0x81 0x7c) euc-jp( 0x21 0x29) euc-jp-linux( 0  
x21 0x29)  
~ : default( 0x81 0x60) Shift-JIS( 0x21 0x29) MS932( 0x81 0x60) euc-jp( 0x8f 0xa2 0xb7)  
euc-jp-linux( 0x21 0x29)
```

■ WindowsXP + jdk1.4.2_05

```
A : default( 0x41) Shift-JIS( 0x41) MS932( 0x41) euc-jp( 0x41) euc-jp-linux( 0x41)  
1 : default( 0x31) Shift-JIS( 0x31) MS932( 0x31) euc-jp( 0x31) euc-jp-linux( 0x31)  
あ : default( 0x82 0xa0) Shift-JIS( 0x82 0xa0) MS932( 0x82 0xa0) euc-jp( 0xa4 0xa2) euc-jp-linux( 0  
xa4 0xa2)  
ア : default( 0x83 0x41) Shift-JIS( 0x83 0x41) MS932( 0x83 0x41) euc-jp( 0xa5 0xa2) euc-jp-linux( 0  
xa5 0xa2)  
A : default( 0x82 0x60) Shift-JIS( 0x82 0x60) MS932( 0x82 0x60) euc-jp( 0xa3 0xc1) euc-jp-linux( 0  
xa3 0xc1)  
- : default( 0x81 0x7c) Shift-JIS( 0x21 0x29) MS932( 0x81 0x7c) euc-jp( 0x21 0x29) euc-jp-linux( 0  
x21 0x29)  
~ : default( 0x81 0x60) Shift-JIS( 0x21 0x29) MS932( 0x81 0x60) euc-jp( 0x8f 0xa2 0xb7)  
euc-jp-linux( 0x21 0x29)
```

■ WindowsXP + jdk1.4.2_07

```
A : default( 0x41) Shift-JIS( 0x41) MS932( 0x41) euc-jp( 0x41) euc-jp-linux( 0x41)  
1 : default( 0x31) Shift-JIS( 0x31) MS932( 0x31) euc-jp( 0x31) euc-jp-linux( 0x31)  
あ : default( 0x82 0xa0) Shift-JIS( 0x82 0xa0) MS932( 0x82 0xa0) euc-jp( 0xa4 0xa2) euc-jp-linux( 0  
xa4 0xa2)  
ア : default( 0x83 0x41) Shift-JIS( 0x83 0x41) MS932( 0x83 0x41) euc-jp( 0xa5 0xa2) euc-jp-linux( 0  
xa5 0xa2)  
A : default( 0x82 0x60) Shift-JIS( 0x82 0x60) MS932( 0x82 0x60) euc-jp( 0xa3 0xc1) euc-jp-linux( 0  
xa3 0xc1)  
- : default( 0x81 0x7c) Shift-JIS( 0x21 0x29) MS932( 0x81 0x7c) euc-jp( 0x21 0x29) euc-jp-linux( 0  
x21 0x29)  
~ : default( 0x81 0x60) Shift-JIS( 0x21 0x29) MS932( 0x81 0x60) euc-jp( 0x8f 0xa2 0xb7)  
euc-jp-linux( 0x21 0x29)
```

■ Redhat Linux 9 + jdk1.4.1_05

```
A : default( 0x41) Shift-JIS( 0x41) MS932( 0x41) euc-jp( 0x41) euc-jp-linux( 0x41)  
1 : default( 0x31) Shift-JIS( 0x31) MS932( 0x31) euc-jp( 0x31) euc-jp-linux( 0x31)  
あ : default( 0xa4 0xa2) Shift-JIS( 0x82 0xa0) MS932( 0x82 0xa0) euc-jp( 0xa4 0xa2) euc-jp-linux( 0  
xa4 0xa2)  
ア : default( 0xa5 0xa2) Shift-JIS( 0x83 0x41) MS932( 0x83 0x41) euc-jp( 0xa5 0xa2) euc-jp-linux( 0  
xa5 0xa2)  
A : default( 0xa3 0xc1) Shift-JIS( 0x82 0x60) MS932( 0x82 0x60) euc-jp( 0xa3 0xc1) euc-jp-linux( 0  
xa3 0xc1)  
- : default( 0xa1 0xdd) Shift-JIS( 0x81 0x7c) MS932( 0x3f) euc-jp( 0xa1 0xdd) euc-jp-linux( 0xa1 0  
xdd)  
~ : default( 0xa1 0xc1) Shift-JIS( 0x81 0x60) MS932( 0x3f) euc-jp( 0xa1 0xc1) euc-jp-linux( 0xa1 0  
xc1)
```

■ FreeBSD-4.9 + jdk1.4.2-p5

```
A : default( 0x41) Shift-JIS( 0x41) MS932( 0x41) euc-jp( 0x41) euc-jp-linux( 0x41)  
1 : default( 0x31) Shift-JIS( 0x31) MS932( 0x31) euc-jp( 0x31) euc-jp-linux( 0x31)  
あ : default( 0xa4 0xa2) Shift-JIS( 0x82 0xa0) MS932( 0x82 0xa0) euc-jp( 0xa4 0xa2) euc-jp-linux( 0  
xa4 0xa2)
```

```

ア : default( 0xa5 0xa2) Shift-JIS( 0x83 0x41) MS932( 0x83 0x41) euc-jp( 0xa5 0xa2) euc-jp-linux( 0
xa5 0xa2)
A : default( 0xa3 0xc1) Shift-JIS( 0x82 0x60) MS932( 0x82 0x60) euc-jp( 0xa3 0xc1) euc-jp-linux( 0
xa3 0xc1)
- : default( 0xa1 0xdd) Shift-JIS( 0x81 0x7c) MS932( 0x3f) euc-jp( 0xa1 0xdd) euc-jp-linux( 0xa1 0
xdd)
~ : default( 0xa1 0xc1) Shift-JIS( 0x81 0x60) MS932( 0x3f) euc-jp( 0xa1 0xc1) euc-jp-linux( 0xa1 0
xc1)

```

■ RHEL4 + jdk1.4.2_13

```

A : default( 0x41) Shift-JIS( 0x41) MS932( 0x41) euc-jp( 0x41) euc-jp-linux( 0x41)
1 : default( 0x31) Shift-JIS( 0x31) MS932( 0x31) euc-jp( 0x31) euc-jp-linux( 0x31)
あ : default( 0xa4 0xa2) Shift-JIS( 0x82 0xa0) MS932( 0x82 0xa0) euc-jp( 0xa4 0xa2) euc-jp-linux( 0
xa4 0xa2)
ア : default( 0xa5 0xa2) Shift-JIS( 0x83 0x41) MS932( 0x83 0x41) euc-jp( 0xa5 0xa2) euc-jp-linux( 0
xa5 0xa2)
A : default( 0xa3 0xc1) Shift-JIS( 0x82 0x60) MS932( 0x82 0x60) euc-jp( 0xa3 0xc1) euc-jp-linux( 0
xa3 0xc1)
- : default( 0xa1 0xdd) Shift-JIS( 0x81 0x7c) MS932( 0x3f) euc-jp( 0xa1 0xdd) euc-jp-linux( 0xa1 0
xdd)
~ : default( 0xa1 0xc1) Shift-JIS( 0x81 0x60) MS932( 0x3f) euc-jp( 0xa1 0xc1) euc-jp-linux( 0xa1 0
xc1)

```

■ RHEL + jdk1.5.0_11

```

A : default( 0x41) Shift-JIS( 0x41) MS932( 0x41) euc-jp( 0x41) euc-jp-linux( 0x41)
1 : default( 0x31) Shift-JIS( 0x31) MS932( 0x31) euc-jp( 0x31) euc-jp-linux( 0x31)
あ : default( 0xa4 0xa2) Shift-JIS( 0x82 0xa0) MS932( 0x82 0xa0) euc-jp( 0xa4 0xa2) euc-jp-linux( 0
xa4 0xa2)
ア : default( 0xa5 0xa2) Shift-JIS( 0x83 0x41) MS932( 0x83 0x41) euc-jp( 0xa5 0xa2) euc-jp-linux( 0
xa5 0xa2)
A : default( 0xa3 0xc1) Shift-JIS( 0x82 0x60) MS932( 0x82 0x60) euc-jp( 0xa3 0xc1) euc-jp-linux( 0
xa3 0xc1)
- : default( 0xa1 0xdd) Shift-JIS( 0x81 0x7c) MS932( 0x3f) euc-jp( 0xa1 0xdd) euc-jp-linux( 0xa1 0
xdd)
~ : default( 0xa1 0xc1) Shift-JIS( 0x81 0x60) MS932( 0x3f) euc-jp( 0xa1 0xc1) euc-jp-linux( 0xa1 0
xc1)

```