



新潟インターネット研究会
Niigata Internet Society

コンピュータの日本語環境

—特に文字コードについて—

新潟インターネット研究会勉強会 Jun. 06, 2005 鈴木一郎

本日の内容

- ロケールに関する復習
- 文字集合とは
- 文字エンコードとは
- UNICODEの文字集合とエンコード
 - UCSとUTF
- 異体字と包摂

ポイント

- ✓ ロケールとは何か？
- ✓ なぜ国際化が重要なのか？
- ✓ 英語版windowsで一太郎は動くか？
- ✓ UNICODEは漢字を何バイトで表現する？
- ✓ タカシマヤの「タカ」、よしのやの「ヨシ」をワープロで入力するには？

ロケールとは？

- 言語や文化的なルールの集合（セット）
 - 使用言語
 - 通貨の書式
 - 日付の書式
 - 数値の書式
- コンピュータ環境の切り替えはロケールで行なう
 - 複数の国で同じ言語を利用している
 - 同じ国で複数の言語を利用する

ロケールに関連する環境変数

- LC_COLLATE 文字列のソート順
- LC_CTYPE 扱う文字の種類
(大文字小文字の区別など)
- LC_MESSAGE 表示されるメッセージ
- LC_MONETARY 通貨書式
- LC_NUMERIC 数値書式(区切り書式など)
- LC_TYME 時刻と日付の書式
- LC_ALL, LANG すべての環境変数



ロケール名

- <言語コード>_<カントリーコード>.<文字コード>

(例: LC_ALL=ja_JP.eucJP)

言語コード

en	英語
fr	フランス語
it	イタリア語
ja	日本語
ko	韓国語
ru	ロシア語
zh	中国語

カントリーコード

US	アメリカ
FR	フランス
IT	イタリア
JP	日本
KR	韓国
RU	ロシア
CN	中国

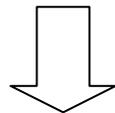
文字コード

ja_JP.EUC
ja_JP.UTF-8
ja_JP.SJIS
ja_JP.ISO_2022_JP
ja_JP (=ja_JP.EUC)



地域化、国際化と多言語化

- 地域化 localization (L10n)
 - 例: 英語ソフトを日本語化する
- 国際化 internationalization (i18n)
 - ソフトウェアから言語異存部分を分離する



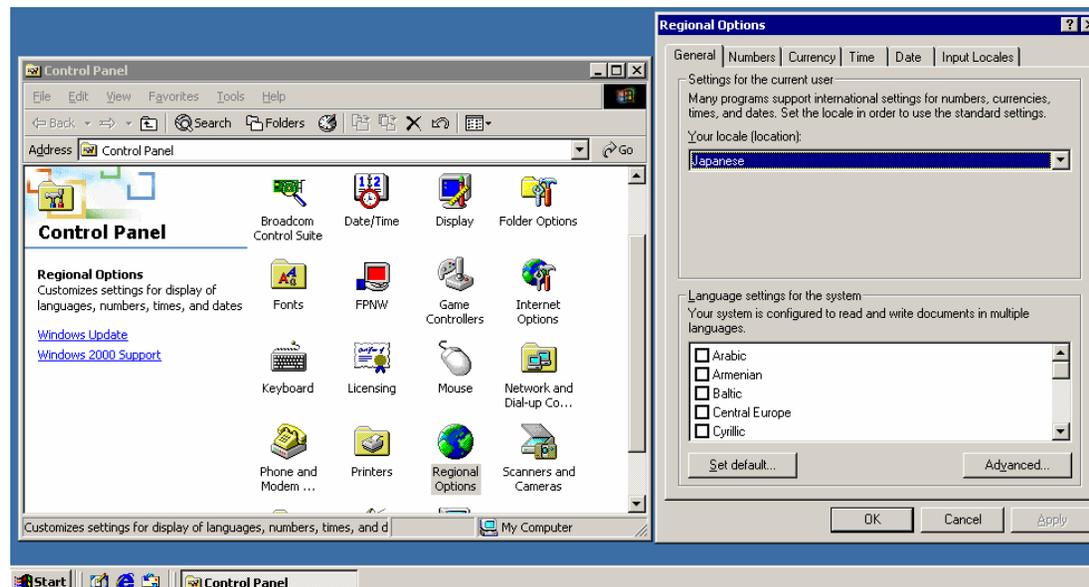
国際化に基づく地域化

- 多言語化 multilingualization (m17n)
 - 複数の言語を同時に扱う

Windows OSのロケール(LCID)

OSのロケール

- 言語バージョンにより決定
- システムのロケール
 - コントロールパネルの地域オプション
 - 複数言語をインストールすれば多言語化が可能



文字集合(1)

- ASCII
 - 7ビット(128種)の文字集合
 - アルファベット・数字・記号・制御文字
 - 文字用:94+制御用:34(改行、スペースなど)
- ISO 8859
 - 8ビット拡張ASCII(+128種)
 - 8819-1~8859-16の16のバリエーション
 - 0059-1(Latin-1)が頻用(西ヨーロッパ言語)

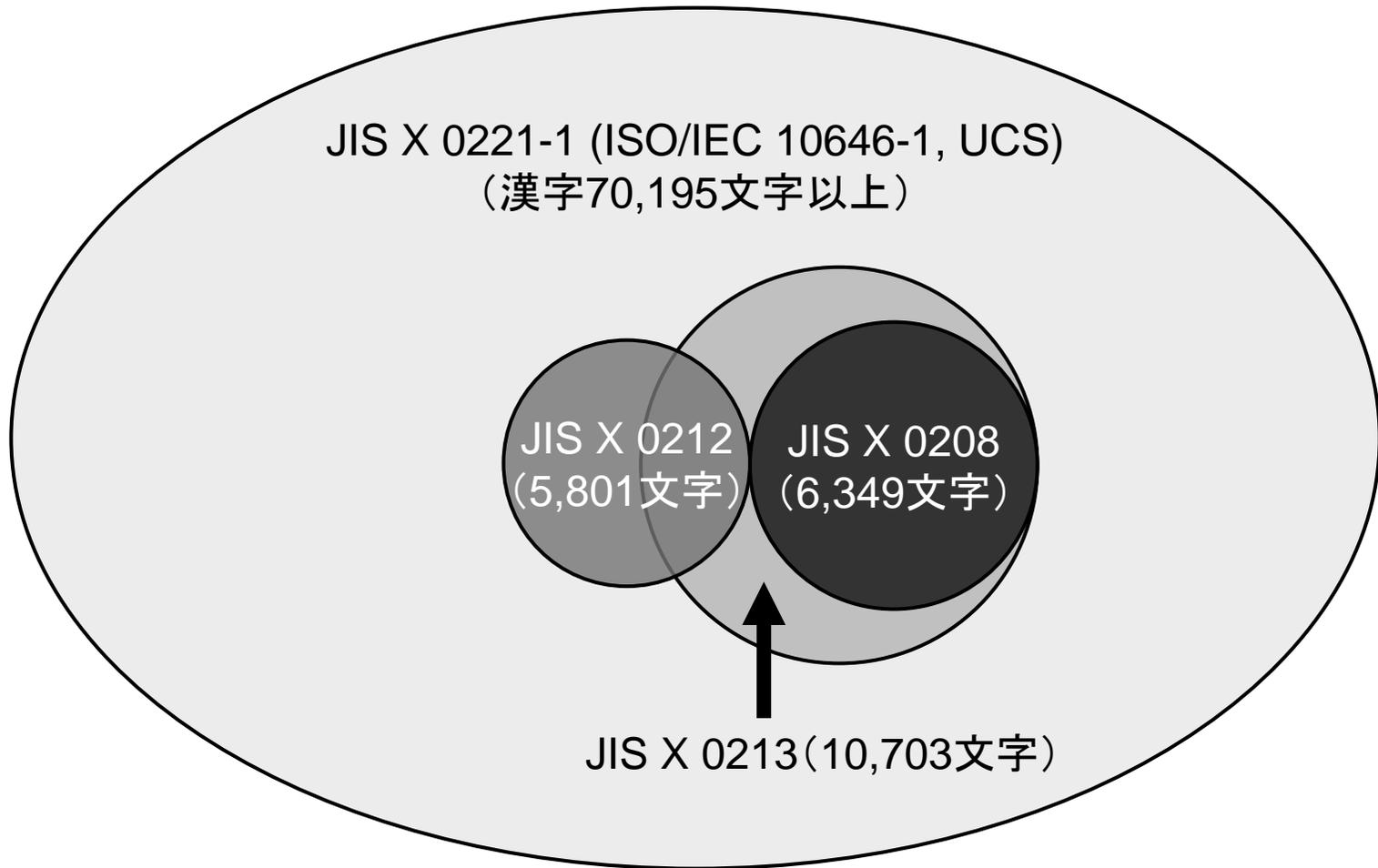
文字集合(2)

- JIS X 0201:1997
 - 1969年に策定されたISO 8859の日本版
 - 8ビット拡張分に半角カナ
 - ASCII非互換あり(バックスラッシュとオーバーライン)
- JIS X 0208:1997
 - 1978年に策定された全角ひら、カタ、漢字を扱うための文字集合
 - 「94の区」と「94の点」の組み合わせ
- JIS X 0212:1990
 - 補助漢字を追加(6,067文字)
 - JIS X 0208とセットで用いる(単体利用不可)

文字集合(3)

- JIS X 0213:2004
 - 2000年に策定
 - いわゆるJIS第3、第4水準を追加
 - 区と点十面
 - 1面: JIS X 0208+2000文字を追加(第3水準)
 - 2面: 2400文字を追加(第4水準)
- JIS X 0221-1:2001
 - ISO/IEC10646(UNICODE)の翻訳

JIS文字集合の包含関係



文字エンコーディング

- ASCII文字集合との互換性を保つ
- 代表的な文字エンコーディング
 - ISO-2022-JP
 - (JIS)EUC-JP
 - Shift JIS
 - Codepage 932
 - SJISの拡張: NEC特殊文字、IBM拡張文字などを含む

UNICODE/ISO 10646

- 文字集合 Universal Multiple-Octet Coded Character Set
 - UCS-4 32bit (群+面+区+点 各8bit)
 - UCS-2 16bit (UCS4の00群00面=BMPのみ)
- 文字エンコーディング UCS Transformation Format
 - UTF-8
 - 可変長マルチバイト(ASCII互換)
 - Plan9 でUCSを利用するために Ken Thompson が考案
 - UTF-16
 - 16ビット固定(ASCII非互換)
 - UTF-7,UTF-32

<http://www.unicode.org/charts/>



UNICODEの文字集合UCS-4

- 00群00面 (Plane0)
 - 基本多言語面 Basic Multilingual Plane (BMP)
- 00群01面 (Plane1)
 - 補助多言語面 (BMPに入っていない補助的文字)
- 00群02面 (Plane2)
 - 補助漢字面 (BMPに入っていない漢字)

UTF-8のマルチバイト処理

UCS-4	UTF-8			
	1バイト目	2バイト目	3バイト目	4バイト目
0000 0000-0000 007F	0xxxxxxx			
0000 0080-0000 07FF	110xxxxx	10xxxxxx		
0000 0800-0000 FFFF	1110xxxx	10xxxxxx	10xxxxxx	
0001 0000-0010-FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

任意のバイトの最上位ビットが1であれば非ASCII

Surrogate Pair

- UTF-16でBMP以外を表現する方法
- UTF-16をペアで使う(32bit)
- BMPの予約領域 U+D800～DFFFを用いる
 - high surrogate U+D800～DBFF
 - low surrogate U+DC00～DFFF
- $(hs-0xD800) \times 2^{10} + (ls-0xDC00) + 0x10000$

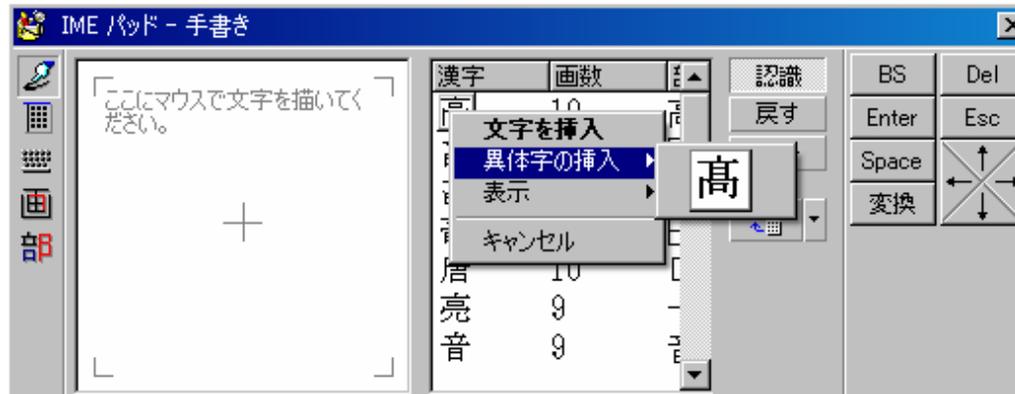
UTF-8もUTF-16もマルチバイトでUNICODEは16bit固定長ではないことに注意



吉野家 店舗検索

インターネットならお中元ギフトが
日本全国送料無料1,600点!
高島屋のお中元

8月15日(月)まで



JIS文字コードとOSの関係

OS名	Windows Me	Windows XP	MacOS	MacOS X
主な符号化方法	シフトJIS UCS-2(*)	UTF-16 (**)	シフトJIS(***)	UTF-16 (**)
基本フォントの文字集合	JIS X 0208 JIS X 0212	JIS X 0208 JIS X 0212	JIS X 0208	JIS X 0208 JIS X 0213 + α

*.....Windows Meでは文字の描画などで一部でUnicode(JIS X 0221)を採用している。UCS-2は全体のうち約65,000文字だけ扱える方式。

**.....UTF-16はJIS X 0221の符号化方法で約150万文字を取り扱い可能。これはすべてのJIS文字コードをふくむので、原理的にはJISに収録されたすべての文字が取り扱い可能。

***.....シフトJISはJISの漢字としてはJIS X 0208だけしか扱えない。

異体字と包摂

- 異体字

- 同等(起源が同じ)の文字だが違う字形をもつもの
- 略字・正字・俗字・古字・本字など
 - 手書きによるバリエーション: 草かんむりの3画と4画
 - 略字: 「國」と「国」、しんによりの1点と2点
 - 俗字: 間違った書き方が定着した
 - 形は異なっても由来が同じ字: 「坂」と「阪」、「弑」と「一」
 - 似ていても由来が異なる文字: 「土」と「土」
- 文字とグリフの間にある概念
- 地域、文化や歴史に依存するあいまいな概念

- 包摂

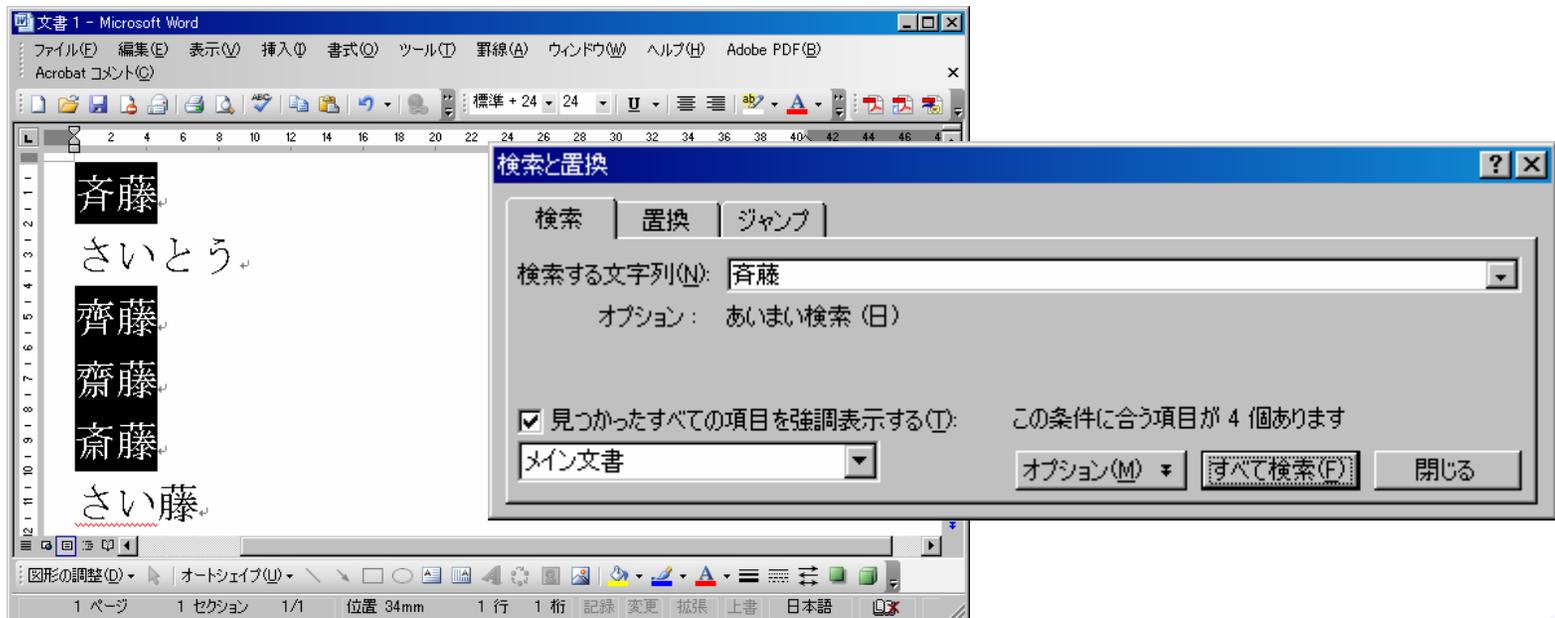
- 形の異なる文字を同じ文字と定義すること
- JISでは同じコードポイントを振るという意味で使っている

UNICODEでの異体字の扱い

- 原則として同じコードポイントを与える
(JISも基本的には同じ考え方をとる)
- CJK(中日韓統一コード)
 - 字形が似ているものは共通のコードポイント
- ラウンドトリップ変換の互換性
 - 日本で別字、中国で包摂の場合は分離
 - UNICODE3.0
 - Plane14に言語タグ、異体字タグ

異体字検索

- 別コードを振った異体字は文字コードレベルでは区別される
- 標準化された異体字辞書の必要性



20B00

CJK Unified Ideographs Extension B

20BFF

	20B0	20B1	20B2	20B3	20B4	20B5	20B6	20B7	20B8	20B9	20BA	20BB	20BC	20BD	20BE	20BF
0	𪛀 20B00	𪛁 20B10	𪛂 20B20	𪛃 20B30	𪛄 20B40	𪛅 20B50	𪛆 20B60	𪛇 20B70	𪛈 20B80	𪛉 20B90	𪛊 20BA0	𪛋 20BB0	𪛌 20BC0	𪛍 20BD0	𪛎 20BE0	𪛏 20BF0
1	𪛐 20B01	𪛑 20B11	𪛒 20B21	𪛓 20B31	𪛔 20B41	𪛕 20B51	𪛖 20B61	𪛗 20B71	𪛘 20B81	𪛙 20B91	𪛚 20BA1	𪛛 20BB1	𪛜 20BC1	𪛝 20BD1	𪛞 20BE1	𪛟 20BF1
2	𪛠 20B02	𪛡 20B12	𪛢 20B22	𪛣 20B32	𪛤 20B42	𪛥 20B52	𪛦 20B62	𪛧 20B72	𪛨 20B82	𪛩 20B92	𪛪 20BA2	𪛫 20BB2	𪛬 20BC2	𪛭 20BD2	𪛮 20BE2	𪛯 20BF2
3	𪛱 20B03	𪛲 20B13	𪛳 20B23	𪛴 20B33	𪛵 20B43	𪛶 20B53	𪛷 20B63	𪛸 20B73	𪛹 20B83	𪛺 20B93	𪛻 20BA3	𪛼 20BB3	𪛽 20BC3	𪛾 20BD3	𪛿 20BE3	𪜀 20BF3
4	𪜁 20B04	𪜂 20B14	𪜃 20B24	𪜄 20B34	𪜅 20B44	𪜆 20B54	𪜇 20B64	𪜈 20B74	𪜉 20B84	𪜊 20B94	𪜋 20BA4	𪜌 20BB4	𪜍 20BC4	𪜎 20BD4	𪜏 20BE4	𪜐 20BF4
5	𪜑 20B05	𪜒 20B15	𪜓 20B25	𪜔 20B35	𪜕 20B45	𪜖 20B55	𪜗 20B65	𪜘 20B75	𪜙 20B85	𪜚 20B95	𪜛 20BA5	𪜜 20BB5	𪜝 20BC5	𪜞 20BD5	𪜟 20BE5	𪜠 20BF5
6	𪜱 20B06	𪜲 20B16	𪜳 20B26	𪜴 20B36	𪜵 20B46	𪜶 20B56	𪜷 20B66	𪜸 20B76	𪜹 20B86	𪜺 20B96	𪜻 20BA6	𪜼 20BB6	𪜽 20BC6	𪜾 20BD6	𪜿 20BE6	𪜀 20BF6
7	𪜱 20B07	𪜲 20B17	𪜳 20B27	𪜴 20B37	𪜵 20B47	𪜶 20B57	𪜷 20B67	𪜸 20B77	𪜹 20B87	𪜺 20B97	𪜻 20BA7	𪜼 20BB7	𪜽 20BC7	𪜾 20BD7	𪜿 20BE7	𪜀 20BF7

<http://www.unicode.org/charts/>

	230	231	232	233	234	235	236	237	238	239	23A	23B	23C	23D	23E	23F	
0																	
1																	
2																	
3																	
4																	
5																	
6																	
7																	
8																	
9																	
A																	
B																	
C																	
D																	
E																	
F																	



参考

1. 日本語環境は4つのポイントで制覇せよ
UNIX USER vol.14, No.4, 2005.
2. 文字の海、ビットの舟
<http://internet.watch.impress.co.jp/www/column/ogata/index.htm>
3. 日本語フォントや文字コードについて
<http://www.zukeran.org/shin/jdoc/>
4. 文字コード問題を考える
<http://www.horagai.com/www/moji/moji000.htm>
5. 異体字の迷宮
<http://hp.vector.co.jp/authors/VA011700/moji/codeitai.htm>
6. はいぱーワークブック
<http://hwb.ecc.u-tokyo.ac.jp/current/>





コンピュータの日本語環境

—特に文字コードについて—

新潟インターネット研究会勉強会 Jun. 06, 2005 鈴木一郎

Copyright©2005 Niigata Internet SOCIETY. All Rights Reserved.

本日の内容

- ロケールに関する復習
- 文字集合とは
- 文字エンコードとは
- UNICODEの文字集合とエンコード
 - UCSとUTF
- 異体字と包摂



ポイント

- ✓ロケールとは何か？
- ✓なぜ国際化が重要なのか？
- ✓英語版windowsで一太郎は動くか？
- ✓UNICODEは漢字を何バイトで表現する？
- ✓タカシマヤの「タカ」、よしのやの「ヨシ」をワープロで入力するには？



ロケールとは？

- 言語や文化的なルールの集合(セット)
 - 使用言語
 - 通貨の書式
 - 日付の書式
 - 数値の書式
- コンピュータ環境の切り替えはロケールで行なう
 - 複数の国で同じ言語を利用している
 - 同じ国で複数の言語を利用する



ロケールに関連する環境変数

- LC_COLLATE 文字列のソート順
- LC_CTYPE 扱う文字の種類 (大文字小文字の区別など)
- LC_MESSAGE 表示されるメッセージ
- LC_MONETARY 通貨書式
- LC_NUMERIC 数値書式(区切り書式など)
- LC_TYME 時刻と日付の書式
- LC_ALL, LANG すべての環境変数



ロケール名

- <言語コード>_<カントリーコード>.<文字コード>
(例: LC_ALL=ja_JP.eucJP)

言語コード		カントリーコード		文字コード
en	英語	US	アメリカ	ja_JP.EUC
fr	フランス語	FR	フランス	ja_JP.UTF-8
it	イタリア語	IT	イタリア	ja_JP.SJIS
ja	日本語	JP	日本	ja_JP.ISO_2022_JP
ko	韓国語	KR	韓国	ja_JP (=ja_JP.EUC)
ru	ロシア語	RU	ロシア	
zh	中国語	CN	中国	



地域化、国際化と多言語化

- 地域化 localization (L10n)
 - 例: 英語ソフトを日本語化する
- 国際化 internationalization (i18n)
 - ソフトウェアから言語依存部分を分離する



国際化に基づく地域化

- 多言語化 multilingualization (m17n)
 - 複数の言語を同時に扱う

新潟インターネット研究会
Niigata Internet Society



Windows OSのロケール(LCID)

OSのロケール

- 言語バージョンにより決定
- システムのロケール
 - コントロールパネルの地域オプション
 - 複数言語をインストールすれば多言語化が可能



新潟インターネット研究会
Niigata Internet Society



文字集合(1)

- ASCII
 - 7ビット(128種)の文字集合
 - アルファベット・数字・記号・制御文字
 - 文字用: 94 + 制御用: 34 (改行、スペースなど)
- ISO 8859
 - 8ビット拡張ASCII(+128種)
 - 8819-1~8859-16の16のバリエーション
 - 0059-1(Latin-1)が頻用(西ヨーロッパ言語)

新潟インターネット研究会
Niigata Internet Society



文字集合(2)

- JIS X 0201:1997
 - 1969年に策定されたISO 8859の日本版
 - 8ビット拡張分に半角カナ
 - ASCII非互換あり(バックスラッシュとオーバーライン)
- JIS X 0208:1997
 - 1978年に策定された全角ひら、カタ、漢字を扱うための文字集合
 - 「94の区」と「94の点」の組み合わせ
- JIS X 0212:1990
 - 補助漢字を追加(6,067文字)
 - JIS X 0208とセットで用いる(単体利用不可)

新潟インターネット研究会
Niigata Internet Society



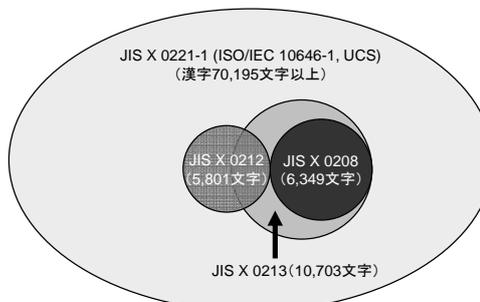
文字集合(3)

- JIS X 0213:2004
 - 2000年に策定
 - いわゆるJIS第3、第4水準を追加
 - 区と点十面
 - 1面: JIS X 0208+2000文字を追加(第3水準)
 - 2面: 2400文字を追加(第4水準)
- JIS X 0221-1:2001
 - ISO/IEC10646(UNICODE)の翻訳

新潟インターネット研究会
Niigata Internet Society



JIS文字集合の包含関係



新潟インターネット研究会
Niigata Internet Society

<http://internet.watch.impress.co.jp/www/column/ogata/sp18.htm> より改変



文字エンコーディング

- ASCII文字集合との互換性を保つ
- 代表的な文字エンコーディング
 - ISO-2022-JP
 - (JIS)EUC-JP
 - Shift JIS
 - Codepage 932
 - SJISの拡張:NEC特殊文字、IBM拡張文字などを含む

新潟インターネット研究会
Niigata Internet Society



UNICODE/ISO 10646

- 文字集合 Universal Multiple-Octet Coded Character Set
 - UCS-4 32bit (群+面+区+点 各8bit)
 - UCS-2 16bit (UCS4の00群00面=BMPのみ)
- 文字エンコーディング UCS Transformation Format
 - UTF-8
 - 可変長マルチバイト(ASCII互換)
 - Plan9 でUCSを利用するために Ken Thompson が考案
 - UTF-16
 - 16ビット固定(ASCII非互換)
 - UTF-7,UTF-32

<http://www.unicode.org/charts/>

新潟インターネット研究会
Niigata Internet Society



UNICODEの文字集合UCS-4

- 00群00面 (Plane0)
 - 基本多言語面 Basic Multilingual Plane (BMP)
- 00群01面 (Plane1)
 - 補助多言語面 (BMPに入っていない補助的の文字)
- 00群02面 (Plane2)
 - 補助漢字面 (BMPに入っていない漢字)

新潟インターネット研究会
Niigata Internet Society



UTF-8のマルチバイト処理

UCS-4	UTF-8			
	1バイト目	2バイト目	3バイト目	4バイト目
0000 0000-0000 007F	0xxxxxxx			
0000 0080-0000 07FF	110xxxxx	10xxxxxx		
0000 0800-0000 FFFF	1110xxxx	10xxxxxx	10xxxxxx	
0001 0000-0010-FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

任意のバイトの最上位ビットが1であれば非ASCII

新潟インターネット研究会
Niigata Internet Society



Surrogate Pair

- UTF-16でBMP以外を表現する方法
- UTF-16をペアで使う(32bit)
- BMPの予約領域 U+D800~DFFFを用いる
 - high surrogate U+D800~DBFF
 - low surrogate U+DC00~DFFF
- $(\text{hs}-0\text{x}D800) \times 2^{10} + (\text{ls}-0\text{x}DC00) + 0\text{x}10000$

UTF-8もUTF-16もマルチバイトでUNICODEは16bit固定長ではないことに注意

新潟インターネット研究会
Niigata Internet Society



吉野家 店舗検索

インターネットならお中元ギフトが
日本全国送料無料1,600点!
高島屋のお中元
8月15日(月)まで



JIS文字コードとOSの関係

OS名	Windows Me	Windows XP	MacOS	MacOS X
主な符号化方法	シフトJIS UCS-2(*)	UTF-16 (**)	シフトJIS(***)	UTF-16 (**)
基本フォントの文字集合	JIS X 0208 JIS X 0212	JIS X 0208 JIS X 0212	JIS X 0208	JIS X 0208 JIS X 0213 + α

.....Windows Meでは文字の描画など一部でUnicode(JIS X 0221)を採用している。UCS-2は全体のうち約65,000文字だけ扱える方式。
UTF-16はJIS X 0221の符号化方法で約150万文字を取り扱い可能。これはすべてのJIS文字コードをふくむので、原理的にはJISに収録されたすべての文字が取り扱い可能。
シフトJISはJISの漢字としてはJIS X 0208だけを扱えない。

新潟インターネット研究会
Niigata Internet Society

<http://internet.watch.impress.co.jp/www/column/ogata/sp13/hyou1b.htm>



異体字と包摂

- 異体字
 - 同等(起源が同じ)の文字だが違う字形をもつもの
 - 略字・正字・俗字・古字・本字など
 - 手書きによるバリエーション: 草かんむりの3画と4画
 - 略字: 「國」と「国」、しんじょうの1点と2点
 - 俗字: 間違った書き方が定着した
 - 形は異なっても由来が同じ字: 「坂」と「阪」、「弌」と「一」
 - 似ていても由来が異なる文字: 「土」と「土」
 - 文字とグリフの間にある概念
 - 地域、文化や歴史に依存するあいまいな概念
- 包摂
 - 形の異なる文字を同じ文字と定義すること
 - JISでは同じコードポイントを振るという意味で使っている



UNICODEでの異体字の扱い

- 原則として同じコードポイントを与える
(JISも基本的には同じ考え方をとる)
- CJK(中日韓統一コード)
 - 字形が似ているものは共通のコードポイント
- ラウンドトリップ変換の互換性
 - 日本で別字、中国で包摂の場合は分離
 - UNICODE3.0
 - Plane14に言語タグ、異体字タグ



異体字検索

- 別コードを振った異体字は文字コードレベルでは区別される
- 標準化された異体字辞書の必要性



20B00		CJK Unified Ideographs Extension B																												20BFF			
		20B00	20B01	20B02	20B03	20B04	20B05	20B06	20B07	20B08	20B09	20B0A	20B0B	20B0C	20B0D	20B0E	20B0F	20B10	20B11	20B12	20B13	20B14	20B15	20B16	20B17	20B18	20B19	20B1A	20B1B	20B1C	20B1D	20B1E	20B1F
0	𪛀	𪛁	𪛂	𪛃	𪛄	𪛅	𪛆	𪛇	𪛈	𪛉	𪛊	𪛋	𪛌	𪛍	𪛎	𪛏	𪛐	𪛑	𪛒	𪛓	𪛔	𪛕	𪛖	𪛗	𪛘	𪛙	𪛚	𪛛	𪛜	𪛝	𪛞	𪛟	𪛠
1	𪛡	𪛢	𪛣	𪛤	𪛥	𪛦	𪛧	𪛨	𪛩	𪛪	𪛫	𪛬	𪛭	𪛮	𪛯	𪛰	𪛱	𪛲	𪛳	𪛴	𪛵	𪛶	𪛷	𪛸	𪛹	𪛺	𪛻	𪛼	𪛽	𪛾	𪛿	𪜀	𪜁
2	𪜂	𪜃	𪜄	𪜅	𪜆	𪜇	𪜈	𪜉	𪜊	𪜋	𪜌	𪜍	𪜎	𪜏	𪜐	𪜑	𪜒	𪜓	𪜔	𪜕	𪜖	𪜗	𪜘	𪜙	𪜚	𪜛	𪜜	𪜝	𪜞	𪜟	𪜠	𪜡	
3	𪜢	𪜣	𪜤	𪜥	𪜦	𪜧	𪜨	𪜩	𪜪	𪜫	𪜬	𪜭	𪜮	𪜯	𪜰	𪜱	𪜲	𪜳	𪜴	𪜵	𪜶	𪜷	𪜸	𪜹	𪜺	𪜻	𪜼	𪜽	𪜾	𪜿	𪝀	𪝁	
4	𪝂	𪝃	𪝄	𪝅	𪝆	𪝇	𪝈	𪝉	𪝊	𪝋	𪝌	𪝍	𪝎	𪝏	𪝐	𪝑	𪝒	𪝓	𪝔	𪝕	𪝖	𪝗	𪝘	𪝙	𪝚	𪝛	𪝜	𪝝	𪝞	𪝟	𪝠	𪝡	
5	𪝢	𪝣	𪝤	𪝥	𪝦	𪝧	𪝨	𪝩	𪝪	𪝫	𪝬	𪝭	𪝮	𪝯	𪝰	𪝱	𪝲	𪝳	𪝴	𪝵	𪝶	𪝷	𪝸	𪝹	𪝺	𪝻	𪝼	𪝽	𪝾	𪝿	𪞀	𪞁	
6	𪞂	𪞃	𪞄	𪞅	𪞆	𪞇	𪞈	𪞉	𪞊	𪞋	𪞌	𪞍	𪞎	𪞏	𪞐	𪞑	𪞒	𪞓	𪞔	𪞕	𪞖	𪞗	𪞘	𪞙	𪞚	𪞛	𪞜	𪞝	𪞞	𪞟	𪞠	𪞡	
7	𪞢	𪞣	𪞤	𪞥	𪞦	𪞧	𪞨	𪞩	𪞪	𪞫	𪞬	𪞭	𪞮	𪞯	𪞰	𪞱	𪞲	𪞳	𪞴	𪞵	𪞶	𪞷	𪞸	𪞹	𪞺	𪞻	𪞼	𪞽	𪞾	𪞿	𪟀	𪟁	

<http://www.unicode.org/charts/>



Miscellaneous Technical

2300		Miscellaneous Technical																												23FF			
		2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	230A	230B	230C	230D	230E	230F	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	231A	231B	231C	231D	231E	231F
0	𪛀	𪛁	𪛂	𪛃	𪛄	𪛅	𪛆	𪛇	𪛈	𪛉	𪛊	𪛋	𪛌	𪛍	𪛎	𪛏	𪛐	𪛑	𪛒	𪛓	𪛔	𪛕	𪛖	𪛗	𪛘	𪛙	𪛚	𪛛	𪛜	𪛝	𪛞	𪛟	𪛠
1	𪛡	𪛢	𪛣	𪛤	𪛥	𪛦	𪛧	𪛨	𪛩	𪛪	𪛫	𪛬	𪛭	𪛮	𪛯	𪛰	𪛱	𪛲	𪛳	𪛴	𪛵	𪛶	𪛷	𪛸	𪛹	𪛺	𪛻	𪛼	𪛽	𪛾	𪛿	𪜀	𪜁
2	𪜂	𪜃	𪜄	𪜅	𪜆	𪜇	𪜈	𪜉	𪜊	𪜋	𪜌	𪜍	𪜎	𪜏	𪜐	𪜑	𪜒	𪜓	𪜔	𪜕	𪜖	𪜗	𪜘	𪜙	𪜚	𪜛	𪜜	𪜝	𪜞	𪜟	𪜠	𪜡	
3	𪜢	𪜣	𪜤	𪜥	𪜦	𪜧	𪜨	𪜩	𪜪	𪜫	𪜬	𪜭	𪜮	𪜯	𪜰	𪜱	𪜲	𪜳	𪜴	𪜵	𪜶	𪜷	𪜸	𪜹	𪜺	𪜻	𪜼	𪜽	𪜾	𪜿	𪝀	𪝁	
4	𪝂	𪝃	𪝄	𪝅	𪝆	𪝇	𪝈	𪝉	𪝊	𪝋	𪝌	𪝍	𪝎	𪝏	𪝐	𪝑	𪝒	𪝓	𪝔	𪝕	𪝖	𪝗	𪝘	𪝙	𪝚	𪝛	𪝜	𪝝	𪝞	𪝟	𪝠	𪝡	
5	𪝢	𪝣	𪝤	𪝥	𪝦	𪝧	𪝨	𪝩	𪝪	𪝫	𪝬	𪝭	𪝮	𪝯	𪝰	𪝱	𪝲	𪝳	𪝴	𪝵	𪝶	𪝷	𪝸	𪝹	𪝺	𪝻	𪝼	𪝽	𪝾	𪝿	𪞀	𪞁	
6	𪞂	𪞃	𪞄	𪞅	𪞆	𪞇	𪞈	𪞉	𪞊	𪞋	𪞌	𪞍	𪞎	𪞏	𪞐	𪞑	𪞒	𪞓	𪞔	𪞕	𪞖	𪞗	𪞘	𪞙	𪞚	𪞛	𪞜	𪞝	𪞞	𪞟	𪞠	𪞡	
7	𪞢	𪞣	𪞤	𪞥	𪞦	𪞧	𪞨	𪞩	𪞪	𪞫	𪞬	𪞭	𪞮	𪞯	𪞰	𪞱	𪞲	𪞳	𪞴	𪞵	𪞶	𪞷	𪞸	𪞹	𪞺	𪞻	𪞼	𪞽	𪞾	𪞿	𪟀	𪟁	



参考

1. 日本語環境は4つのポイントで制覇せよ
UNIX USER vol.14, No.4, 2005.
2. 文字の海、ビットの舟
<http://internet.watch.impress.co.jp/www/column/ogata/index.htm>
3. 日本語フォントや文字コードについて
<http://www.zukeran.org/shin/jdoc/>
4. 文字コード問題を考える
<http://www.horagai.com/www/moji/moji000.htm>
5. 異体字の迷宮
<http://hp.vector.co.jp/authors/VA011700/moji/codeitai.htm>
6. はいばーワークブック
<http://hwb.ecc.u-tokyo.ac.jp/current/>

