

IDNの「言語」問題と IDN-Admin Guideline

2003年6月20日

第1回 IDN ポリシー研究会

JPNIC IDN-TF

米谷嘉朗 <yone@jprs.co.jp>

IDNのおさらい

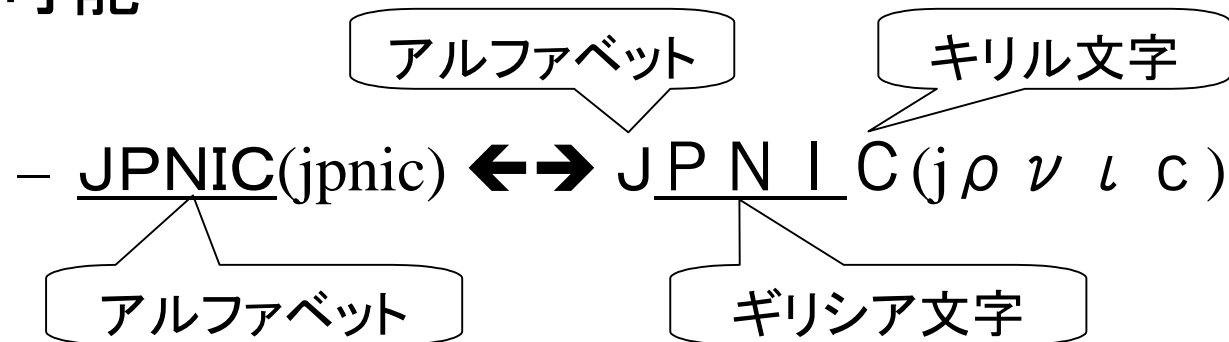
- Internationalized Domain Name
- RFC3490-3492で技術仕様を規定
 - 2003年3月に発行
- ドメイン名を表現するのに使用可能な文字を英数字以外にまで拡張
 - 従来はASCII文字の英数字とハイフンの37文字
 - IDNではUnicodeの約10万文字
 - あくまでも使用可能な「文字のレパートリー」を増やしたということ

文字、言語、用字

- 「文字」は「言語」を書き表すためのもの
 - 言語の一部だが、言語そのものではない
- 一定の規則にしたがって作られた・集められた文字の一群が「用字」
 - 平仮名、ギリシア文字、アラビア文字など
- 「用字」は複数の「言語」で使用され得る
 - 中国語、日本語、韓国語における漢字など
- IETF的定義はRFC3536の2章

IDNで起こり得る問題(典型例)

- 用字分類の異なる任意の文字の組み合わせが可能



- 同一用字分類の文字で、字形は異なるが言語的に同意・同一の文字(異体字)が存在
 - 機 = 机 (中国語) vs 機 ≠ 机 (日本語)

IDN Admin Guideline

- 漢字(CJK)の異体字を「登録時の処理」として統一的に扱う方式の提案
- すべてのゾーン管理者を対象
- 「言語」と「登録可能文字」(Valid Code Point)のセット(異体字テーブル; Variant table)を定義
 - 登録可能文字が異体字(Character Variants)を持つ場合は「優先」(Preferred Variants)文字を登録し、異体字は「予約」
- Registrantが複数の「言語」を指定した場合は、それぞれの「言語」の異体字テーブルを適用した結果を使う

適用例

| 日本語テーブル | | |
|---------|----|-----|
| VC | PV | CV |
| 氣 | 氣 | 氣、氣 |
| 氣 | 氣 | 氣、氣 |

| 中国語テーブル | | |
|---------|----|-----|
| VC | PV | CV |
| 氣 | 氣 | 氣、气 |
| 气 | 氣 | 气、氣 |

| 入力 | | 日本語 | 中国語 | 日本語+ 中国語 |
|----|----|-----|-----|-------------|
| 氣 | ZV | 氣、氣 | 氣 | 氣、氣 |
| | CV | — | 气 | 气 |
| 氣 | ZV | 氣 | × | × |
| | CV | 氣 | | |
| 气 | ZV | × | 氣、气 | × |
| | CV | | — | |

ZV: Zone Variants

CV: Character Variants

開発状況

- Informational RFC化を目標に作業中
- 最新のI-Dはdraft-jseng-idn-admin-03.txt
- 02以降、John Klensinをエディタに迎えて編集集中
 - JETメンバーMLで議論
 - 04くらいで収束させたい

各国のVariant Table開発状況

- 日本(JPNIC/JPRS)
 - draft-yoneya-idn-jachar-00
 - JISの平仮名、片仮名、漢字、準仮名・漢字をベースに文字を選択
 - 異体字(Character Variants)は「なし」
- 韓国(KRNIC)
 - I-Dなし
 - UnicodeのCJK統一領域とハングルシラブルをベースに文字を選択
 - 異体字は「なし」
- 中国・台湾・マカオ・香港(CDNC)
 - I-Dなし
 - UnicodeのCJK統一領域と拡張領域をベースに文字を選択
 - 異体字が多いのでPreferred Variantsを減らすアルゴリズムを開発中
- gTLD(VGRS)
 - IDN-Adminの遵守を表明
 - Variant Tableは上記ccTLD作成のものを使用