

日本語ドメイン名の状況

ドメイン名の最新動向 @ IW2003

2003年12月3日

米谷嘉朗 <yone@jprs.co.jp>

<http://日本レジストリサービス.jp/>

背景

国際化ドメイン名とは

- IETFで標準化されたプロトコルの名前
 - Internationalized Domain Name
または略してIDN
- 2003年3月に一連のRFCが発行
 - RFC3490(IDNA)
 - RFC3491(NAMEPREP)
 - RFC3492(Punycode)
- ドメイン名を表現するのに使用できる文字をUnicodeに拡張する規格
 - 従来はASCIIの英数字とハイフンのみ

日本語ドメイン名とは

- IDNの技術を使用して、日本語で使われる文字で表現されたドメイン名
 - ひらがな、かたかな、漢字など
- ドメインレジストリのサービス仕様
 - 使用可能な文字の定義(一覧)
 - 使用可能な文字数
 - など
 - JPドメイン名では「汎用 JP ドメイン名登録等に関する技術細則」で規定

IDNの例†

华人.公司.cn

華人.商業.tw

高島屋.会社.jp

삼성.회사.kr

三星.회사.kr

الاهرام.م

viagénie.qc.ca

ישראל.קום

ที่เชชนิค.พานิชย์.ไทย

現代.com

ヤフー.com

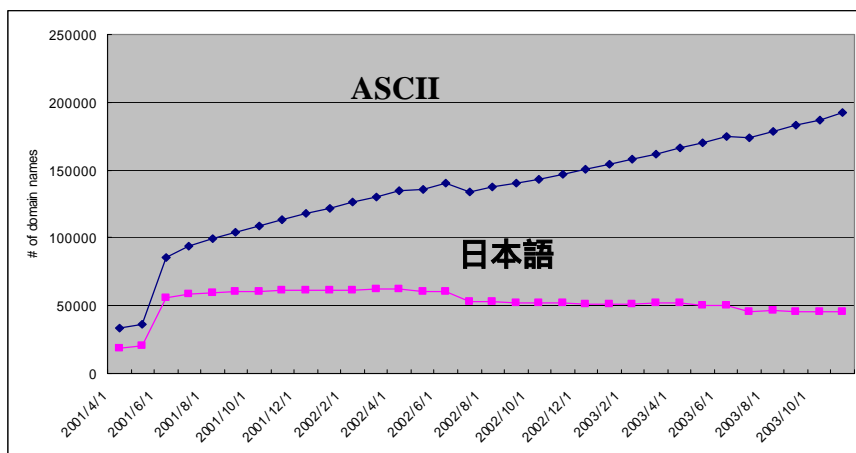
出典 <http://www.jdna.jp/event/jdn-tutorial/IDNSDK.pdf>

日本語ドメイン名の過去1年間の動向

重要な出来事

- 2月 Punycode Prefixの決定
- 3月 IETF IDN Standards RFC発行
- 6月 ICANN "Guidelines for the Implementation of Internationalized Domain Names" (Version 1.0)発行
- 7月 「汎用 JP ドメイン名登録等に関する技術細則」改訂
JPRS Punycode移行開始
Netscape7.1リリース
i-Nav4.0リリース
- 8月 「汎用JPドメイン名における予約ドメイン名」改訂
- 9月 JPRS Punycode移行終了
Opera7.1リリース

ドメイン名登録数の推移



技術仕様

IDNの実現方法(概要)

- 利用者インターフェースでの表現とネットワーク上での表現を分離
 - 表示は日本語(ローカルエンコーディング)、通信はASCII
 - 従来のドメイン名との下位互換性維持のため
 - RFC2825、RFC2826の遵守
- 日本語とASCIIの変換はIDNを扱うアプリケーションソフトウェアで実施
- IDNでは以下の文字列は等価(1対1対応)
 - 日本語ドメイン名 ↔ xn--eckwd4c7c777u7mwo4bc84j

表示

通信

RFC3490(IDNA)

Internationalizing Domain Names in Applications

- IDNの処理はアプリケーションプログラムで行うというアーキテクチャと、具体的な処理方式を規定
 - IDNを処理する際の文字セットはUnicode3.2
 - 利用者インターフェース層での入出力文字セットは特に規定しない
 - ネットワーク上で、プロトコル要素としてIDNを使用する場合の正規化方式と符号化方式を指定

RFC3491(NAMEPREP)

A Stringprep Profile for Internationalized Domain Names

- Unicodeの文字コードで表現された文字列(IDN)の正規化処理を規定
 - RFC3454(Stringprep, 国際化された文字列の前処理方式)をIDNに適用するための具体的な定義
 - 表示上同一の文字列(ドメイン名)の照合結果を一致させるため
 - 処理手順
 - map: 文字種(大文字・小文字)の統一
 - A→a
 - normalize: 合成記号の合成、互換文字の統一
 - ū→ü
 - ガ→ガ
 - Prohibit: 使用禁止文字のチェック
 - 空白文字など

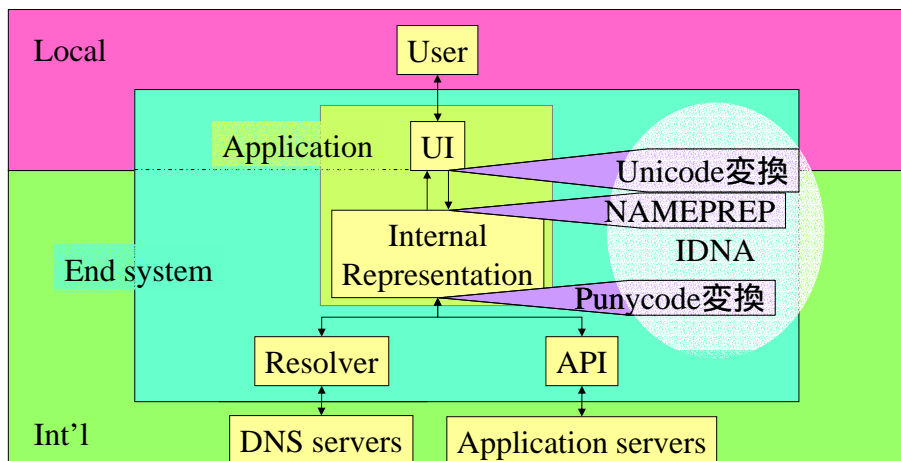
RFC3492(Punycode)

A Bootstring encoding of Unicode for Internationalized Domain Names in Applications

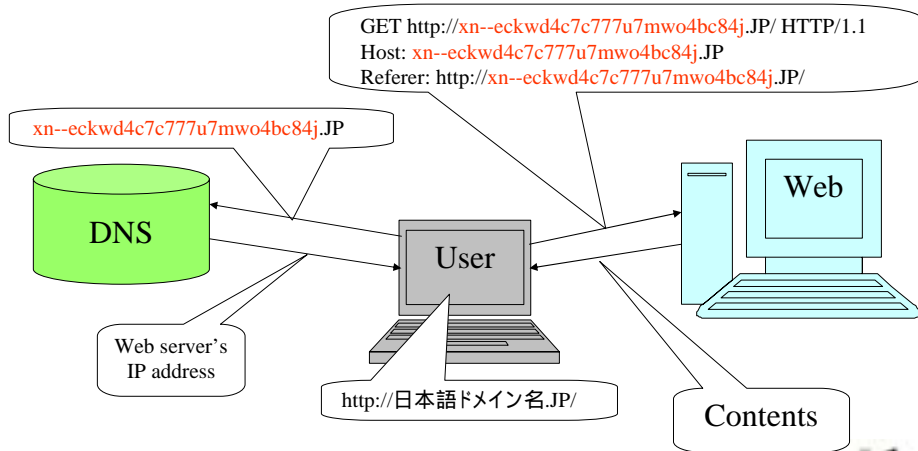
- Unicodeの符号化(encoding)アルゴリズムの一つ
 - ASCIIの英数字(a-z0-9)とハイフン(-)のみでUnicodeの文字列を表現
 - ASCII Compatible Encoding (ACE)
 - プロトコル要素でIDNを使用する場合でも下位互換性を維持
- 変換例
 - unicode文字列 \leftrightarrow xn--unicode-no0lo10eejx

Punycodeで変換されたことを示す識別子でRFC3490で規定

位置づけ



Webサイトのブラウズ



アプリケーションの対応状況

- Microsoft Internet Explorer
 - i-Nav4.0などプラグインを入れることで対応
- Netscape Navigator / Mozilla
 - Netscape 7.1 / Mozilla-1.4 以上で対応
- Opera

IDN対応アプリケーションとライブラリ

- Netscape7.1 (Mozilla-1.4)以降
 - <http://www.netscape.com/>
 - <http://www.mozilla.org/>
- Opera7.2以降
 - <http://www.opera.com/>
- Internet Explorer 5以降 + i-Nav4.0以降
 - <http://jprs.jp/i-Nav/>
- JPNIC idnkit
 - <http://www.nic.ad.jp/ja/idn/mdnkit/download/>
- GNU libidn
 - <http://www.gnu.org/software/libidn/>
- VeriSign SDK
 - http://www.verisign.com/nds/naming/idn/sdk_form.html

IDNConnect

- IDN Standards (RFC3490-92)のプロトコル相互運用性確認イベント
 - 2003年9月23-27日にオンラインで実施
- 主催者が用意したテストデータを参加者の実装に適用し、期待される結果と一致するか確認
 - 結果は参加者が主催者に報告
 - 参加者名や実装名は非公開
- 相互運用性は確認された
 - 最終レポートとテストデータは公開されている
<http://idnconnect.jdna.jp/>

課題

技術で解決できない問題

- Unicodeは多数の「似て非なる」文字を持つ
 - NI(ni) ↔ ()
 - ↔ □ ↔ □ ↔ □
- いくつかの言語は歴史的、文化的事情により「異体」文字を持つ
 - 電気通信 ↔ 電氣通信 ↔ 电气通信
- IDNの技術仕様ではそれら文字の「混用」を避けることはできない
 - 運用上の「制約」が必要

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-Admin

- IDNにおける「漢字」の扱い方について、日中韓台 (JP/CN/KR/TW)のNICおよび有識者で議論し作成した、IDN登録時の制約アルゴリズム
 - DNSのゾーン管理者を対象
 - IDNを「言語」と関連付ける
 - 「言語」は「登録可能文字」と「その異体字」の集合から定義される
 - IETFのInternet Draftとして発行されている
 - <http://www.ietf.org/internet-drafts/draft-jseng-idn-admin-05.txt>
 - Informational RFC化を目指して作業中
 - 日本語の異体字テーブル
 - <http://www.ietf.org/Internet-drafts/draft-yoneya-idn-jachar-00.txt>

ICANN Guideline

- ICANNがIDN登録を許可する条件
 - ICANNと契約関係を持つTLD(Top Level Domain)レジストリが対象
 - IDN-Adminの考え方を踏襲
 - 記号の使用を原則禁止
 - 2003年6月20日にVersion 1.0が公開
 - <http://www.icann.org/announcements/announcement-20jun03.htm>
 - <http://www.icann.org/general/idn-guidelines-20jun03.htm>

諸外国の状況

アジア

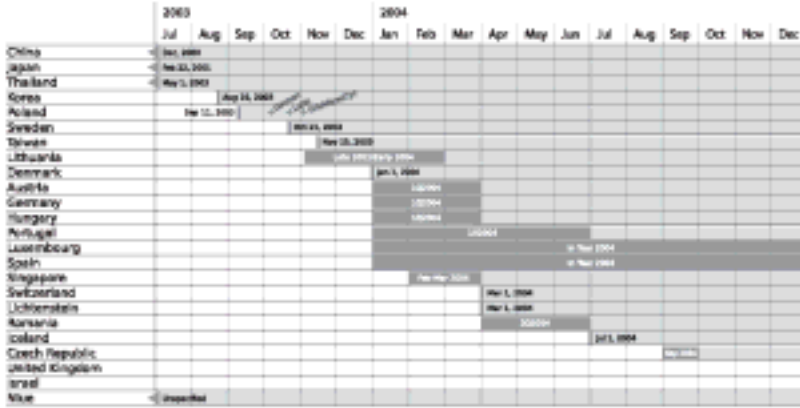
- CNNIC (.CN)
 - 2003年5月から中国語ドメイン名登録、名前解決開始
- KRNIC (.KR)
 - 2003年8月からハングルドメイン名登録、名前解決開始
- TWNIC (.TW)
 - 2003年11月から中国語ドメイン名登録、名前解決開始

ヨーロッパ

- NASK (.PL)
 - 2003年9月からポーランド語ドメイン名登録開始
 - その後ドイツ語、ラテン語、アラビア語も追加
- NIC-SE (.SE)
 - 2003年10月からスウェーデン語ドメイン名登録開始
- Affilias (.INFO)、DENIC (.DE)、SWITCH (.CH、.LI)、NIC.AT (.AT)もまもなく開始

その他各国の実施予定

IDN Deployment Timetable



出典: CENTRサーベイ結果(2003年10月末~11月上実施)

Q&A