

JPRS トピックス&コラム



■国際化ドメイン名を実現する3つの技術

国際化ドメイン名は `http://ドメイン名例.jp/` のように、従来英数字と一部記号のみが使用可能であったドメイン名を、それ以外の文字列も使用できるように機能拡張したものです。今回は、国際化ドメイン名を実現するための基本的な3つの技術について解説します。

■国際化ドメイン名を実現する3つの技術

国際化ドメイン名は、Punycode、Nameprep、IDNA という3つの技術の組み合わせにより実現されています。それら3つのしくみについて、順に説明していきましょう。

1. インターネット基盤への影響を最小限に

～Punycode～

インターネットにおいてプロトコルの国際化を行う場合、文字コードとして Unicode が使用され、文字列の符合化方式として RFC 3629 で規定されている UTF-8 が使用されることが主流となっています。UTF-8 は従来の ASCII と上位互換性があるため、従来のプロトコル仕様に与える影響を少なくすることができます。

国際化ドメイン名においても、文字コードに Unicode を採用することは早い段階で合意されていました。しかし、標準化の過程において、DNS はインターネットの基本部分を構成する基盤技術として重要な役割を担っており、符号化方式として 8 ビットコードを含む UTF-8 をそのままの形で採用した場合、既存の DNS インフラに悪影響をおよぼすことが懸念されました。

そのため、国際化ドメイン名では Unicode 文字列の符合化方式として UTF-8 ではなく、**Punycode** (ピュニコード) と呼ばれる方式が開発・採用されました。Punycode は RFC 3492 で規定されており、Unicode 文字列をドメイン名として従来から使用可能であったアルファベット・数字・ハイフンで表現するものです。

図 1 に Punycode を用いた符号化の例を示します。

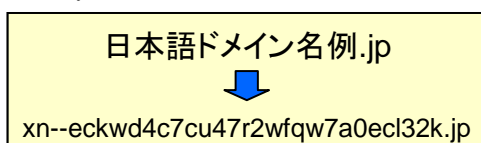


図 1: Punycode を用いた符号化

2. 名前の前処理 ～Nameprep～

Unicode では、「é」のようなアクセント記号のある文字が「文字」と「アクセント」の 2 つの文字コードで構成されている場合と、「アクセントつき文字」の 1 つの文字コードで構成されている場合があります。これらは全く同一の文字であるため、異なる文字コードであっても別のドメイン名と認識されるべきではありません。

また、全角文字と半角文字のような、ドメイン名として取り扱う場合に同一であると判断したいものもあります。

このような要求を実現するため国際化ドメイン名では入力された文字列に対し、一定の方法で事前処理を行うことにより、予期しない不都合の発生を防止する方式が採用されました。この変換は**名前の前処理** (Name Preparation) と呼ばれており、文字種の統一 (Mapping) 正規化 (Normalization)、禁止文字の排除 (Prohibited Output) という 3 つの処理により構成されています。またこれらの 3 つの処理を合わせて単に正規化と呼ぶ場合もあり、以後本稿ではこの表現に従います。

国際化ドメイン名では RFC 3491 で規定されている **Nameprep** (ネームプレップ) という方法で正規化が行われます。Nameprep による正規化のうち、主なものを図 2 に示します(注 1)。

- ・アルファベットは全て半角小文字に変換
例) ABC → abc
- ・全角数字は全て半角に変換
例) 123 → 123
- ・半角カタカナは全て全角に変換
例) ｶﾀｶﾅ → カタカナ
- ・濁点・半濁点は可能な限り合成
例) ｶﾝﾊﾞ ㇿ ﾆｰ → カンパニー
- ・全角句点(。)および全角ピリオド(.)は全て半角ピリオド(.)に変換
例) 日本語.jp → 日本語.jp

図 2: Nameprep により行われる主な正規化

(注 1) 半角ピリオドへの変換は IDNA(後述)により行われています。

3. 全ての処理はアプリケーションで ~IDNA~

国際化ドメイン名ではこれまでに説明した Nameprep による正規化、およびドメイン名データを送受信する際の Punycode による符合化・復号化は全て、Web ブラウザ等の各ユーザ側アプリケーションで行う必要があります。この方式は **IDNA** (IDN in Applications) と呼ばれ、RFC 3490 で規定されています。

ここまでで説明した Punycode、Nameprep、IDNA の関係を図 3 に示します。

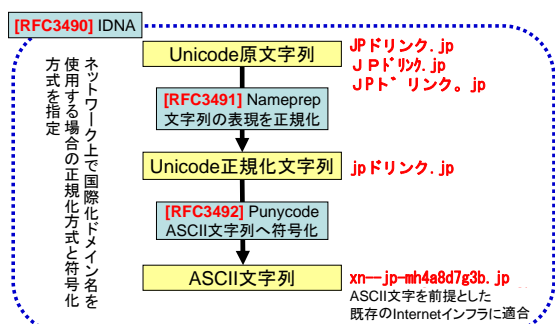


図 3: Punycode、Nameprep、IDNA の関係

■国際化ドメイン名と日本語ドメイン名

ここで「国際化ドメイン名」と「日本語ドメイン名」の関係について説明しておきましょう。

国際化ドメイン名 (Internationalized Domain Name) はインターネットプロトコルの正式名称として、標準的に使用されています。それに対し日本語ドメイン名は、日本語の文字列で表現された、国際化ドメイン名の一部ということになります。

また、JPRS が提供している「**日本語 JP ドメイン名**」では、国際化ドメイン名として利用可能な文字列のうち、ひらがな、カタカナ、漢字 (JIS コードの第一水準/第二水準)、およびそれらに準ずる記号 (々、・など) のみを取り扱えるように、JP ドメイン名の技術細則により登録可能な文字列を定めています。

登録可能文字	従来のASCII文字 + ひらがな・カタカナ・漢字・一部記号
文字数	全半角問わず15文字まで
表現の統一 (正規化)	以下の表現は区別せず取り扱う 英数字の全角/半角、大文字/小文字 片仮名の全角/半角
「。」句点「.」(全角)	ピリオド(半角)「.」に変換

図 4: 日本語 JP ドメイン名の主な仕様

■国際化ドメイン名への対応

国際化ドメイン名の標準化方式として IDNA が採用されたことにより、国際化 (日本語) ドメイン名を利用する場合、Web ブラウザやメールソフト等のアプリケーションにおいて国際化ドメイン名に対応する作業を**それぞれ個別に行う必要があります**。



図 5: 国際化ドメイン名に対応した主な Web ブラウザ

■サーバ側からみた国際化ドメイン名

一方、DNS サーバや Web サーバ等のサーバ側からみた場合、国際化ドメイン名に関する全ての処理はアプリケーション側で行われるため、サーバ側の設定ファイルにおける国際化ドメイン名の記述として Punycode を使用することにより、**従来のサーバプログラムを改変することなく国際化ドメイン名に関する設定を行うことができます** (図 6)。

```

zone "xn--eckwd4c7cu47r2wfwq7a0ec132k.jp" {
    type master;
    file "NIHONGO-jp.zone";
    allow-transfer { 192.168.2.2; };
}
  
```

図 6: BIND における設定例 (named.conf)

■より便利なインターネットをめざして

現在 ICANN において、トップレベルドメイン (TLD) の国際化の議論が進められています。ICANN からは、早ければ 2009 年には「.日本」のような、国際化された TLD が使用開始になる可能性が示されています。

これまで、さまざまなインターネットプロトコルが国際化されることにより、非英語圏のユーザがインターネットを使用するための障壁が取り除かれてきました。今後もさまざまな分野において、より便利なインターネットをめざした努力が続けられていくことでしょう。

なお、日本語 JP ドメイン名に関する最新情報につきましては、<http://日本語.jp/> をご参照ください。