



‘国際化’ドメイン名 について

田代 秀一

産業技術総合研究所
tashiro@m.aist.go.jp

国際化ドメイン名とは

インターネットのドメイン名は“ipsj.or.jp”のように、英文字（大文字小文字の区別なし）、数字、ハイフオン“-”とドット“.”のみからなる文字列によって表記すると規定されている¹⁾。これを、たとえば“情報処理学会.jp”のように、英文字以外の漢字、かな、ハングル、その他多様な文字セットを使えるように拡張しようという動きが最近活発となっている。

このように拡張されたドメイン名表記は一般に“国際化ドメイン名”^{☆1}、あるいは“多言語ドメイン名”^{☆2}と呼ばれている^{☆3}。本稿のタイトルはIETFのワーキンググループ名に合わせて「国際化ドメイン名」とした。

ドメイン名空間が拡大することは商標権所有者など、ドメイン名を他人に取られないように防御しなくてはならない立場の人にとっては必ずしも歓迎されない場合もある。しかし、雑誌や電車のつり広告などで簡単に覚えてもらい、後に打ち込んでもらいやすいようなドメイン名への需要は急速に高まっており、同音意義語やローマ字表記法のぶれなどに悩まされることのない、母国語でよく使われる文字（日本ではかな漢字）によるドメイン名表記への期待は高い^{☆4}。

このように、個々の利用者が求めているものは母国語でよく使われる文字の利用であり、その意味ではドメイン名のローカライゼーションなのであるが、同一の規格が多様なローカライゼーションに適用できることをもって、ドメイン名規格の「国際化」と呼んでいる。

‘国際化’ドメイン名規格化の要件

グローバル・アクセス性

国際化ドメイン名の導入に対する反対意見に、「我々は〇〇文字は読めないし、入力できない。そんな文字をドメイン名に使われては困る。グローバル・アクセス性の破壊である。」というものがある（この〇〇の部分には、主張者により、「漢字」、「ディバナガリ」、「タイ」などが入る）。しかし、人間にとってのコミュニケーションの分断と、インターネットの分断とは分けて考える必要がある。

たとえばメールやWWWなどのインターネットのコンテンツにはすでに多様な文字セットが用いられているが、すべてのアプリケーションがすべての文字を表示できるとは限らないし、また、表示できたとしても、それをすべての人が読めて理解できるとは限らない。しかし、その文字を表示/入力できるアプリケーションさえ用意すれば、全世界どこからでもこれらのコンテンツにアクセスできることは保証されている。コンテンツに含まれる特定の文字コードを通さない回線とルータといったものがないように配慮されているからである。その意味で、グローバル・アクセス性は維持されている。

ドメイン名の国際化についても同様な考え方がなされている。非英文字を使うと、そのドメイン名を読めない人や打ち込めない人がいたり、また表示できない

☆1 IETFの対応ワーキンググループの名称はInternationalized Domain Name-WGである

☆2 VeriSign社や(社)JPNICなどではMultilingual Domain Nameあるいは多言語ドメイン名という用語が使われている

☆3 本来「文字」と「言語」は独立のものである。これはたとえば、日本語をローマ字で書くことができることから分かる。この視点からは文字セットが拡大したことがすなわち多言語化であるという言い方は誤りといえる

☆4 ドメイン名の本来はネットワークの物理的位置付けから独立な一意名であり、そこにセマンティクスを持ち込むことは良くないという考え方も根強く存在する

アプリケーションがある可能性はあるが、そのこと自体はドメイン名の技術規格を考えるうえでは考慮されていない。「非英文字によるドメイン名」を使っているような場所にあるコンテンツは、どうせ「その文字」で書かれているに違いなく、もともとその文字を読めない人がアクセスすることは考慮されていないのだ、と割り切って考えられている。考慮されるべきは、適切なアプリケーションを用意すればインターネット上のどこからでもそれを利用可能であるという意味でのグローバル・アクセス性であり、そのようなアプリケーションを誰でも作ることができる規格のオープン性である。また、その文字セットを表示できないアプリケーションがその文字コードに出くわした場合にもそれがシステムダウンなどの障害を引き起こさないようにすべきであるという点である。

シングル・ルート性

ドメイン名は、ルート（根）を1つとする樹構造を持っている。このルートが唯一であることは非常に重要であり、それによりドメイン名の一意性が保証されている。ルートが複数あると、利用しているアプリケーションやその設定状態、接続しているプロバイダ、接続している地域などの条件の違いにより同一のドメイン名でもその指す対象が違ったものになってしまう可能性がある。

このシングル・ルート性は、現在、それを管理する唯一の組織であるICANN（Internet Cooperation for Assigned Names and Numbers）^{★5}によって維持管理されている。

国際化ドメイン名を導入する際にもこのシングル・ルート性を維持し、ルートの複数化による混乱を招かないよう注意する必要がある。

これまでの動き

ドメイン名に対する非英文字利用への動きは、1998年頃からシンガポール国立大学の多言語情報処理研究

グループを中心に活発化し始めた。同年夏にはアジア太平洋地域のインターネット関係者の会合であるAPNG^{★6}に研究グループが作られテストベッドの構築などが行われた。翌1999年10月にはシンガポール国立大学のグループが中心となり米国シリコンバレーに国際化ドメイン名技術の事業化を目指した民間会社^{★7}が設立されている。

1999年11月にワシントンDCで開催された第46回IETFにおいて国際化ドメイン名の規格化について議論するBoF^{★8}が初めて開催され、2000年3月の第47回IETFからは正式なワーキンググループ、IDN-WG（Internationalized Domain Name Working Group）、となり規格化への議論が開始された。

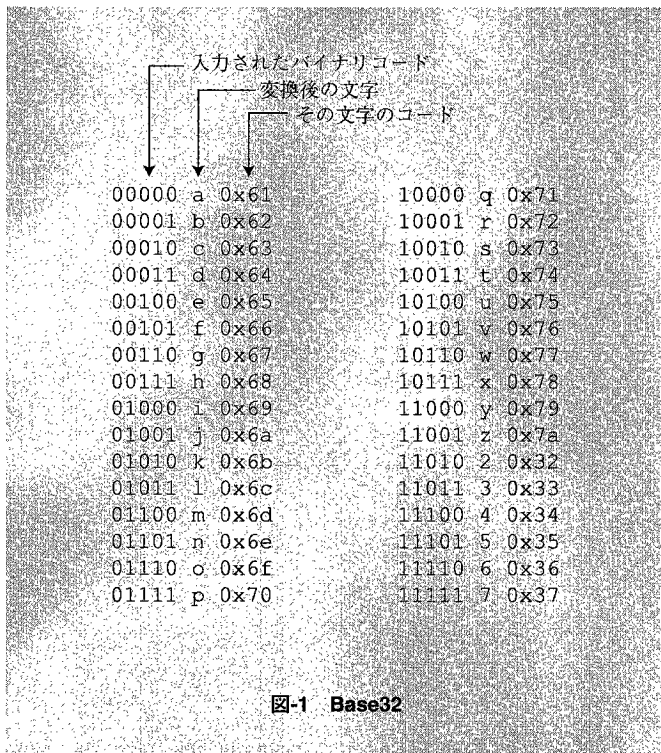
2000年11月にはVeriSign社^{★9}が、IETFの技術方針に従うとの方針の下、国際化ドメイン名の「テストベッド」の運用を開始した。これは一般利用者を対象としたものではなく、ドメイン名登録受業者とVeriSign社との間のドメイン名登録用プロトコル（国際化ドメイン名用に拡張されたRRP）をテストするのが目的とされているが、一般利用者からのドメイン名登録受付も同時に開始されている。日本の（株）JPRS^{★10}においても日本語ドメイン名の受付が2001年2月から開始された。ただし、これらはいずれも登録が可能だけで、DNSによるドメイン名→IPアドレス変換のサービスは行われていない。

1999年から2000年にかけては、いくつかの「怪しい」動きも見られた。

“○○商店.会社.xy”というドメイン名を考えてみる。“xy”というトップ・レベル・ドメイン名は実在しないが^{★11}、これが正式にICANNに認可されたものだと仮定する。したがって、“xy”の部分には問題ない。

サブドメイン名に2バイトコードが用いられているので現在のドメイン名表記規則¹⁾は満たさない。そのため、世の中にはこのコードを通さない部分、たとえばリゾルバ、ローカルネームサーバ、プロキシサーバなど、があり得るが、「運が良ければ」このサブドメイン名文字列は“xy”の管理するDNSまで到達し、処理される。また、最近のWebブラウザは、URLに漢字などの多バイト文字を入力可能で、それをUTF-8などにより送出するものが多い。送出文字コードやそのエンコーディング法に国際規格がまだできていないため、どのようなコードが送出されるかは必ずしも一定ではないが、よく普及したいくつかのブラウザが使用されると仮定すれば、送出されるコードをある程度絞ることはできる。

★5 <http://www.icann.org/> 参照
 ★6 <http://www.apng.org/> 参照
 ★7 <http://www.i-dns.net/> 参照
 ★8 Bird of Feather, 同好の士の会合。IETFにおいて正式WGができる前段階の暫定グループ
 ★9 <http://www.verisign-grs.com/> 参照
 ★10 <http://www.jprs.jp/> 参照
 ★11 <http://www.iana.org/domain-names.htm> 参照



そこで、すべての利用者が利用できるとは限らないが営業的に成立する程度の利用者数が期待できさえすれば良いと割り切って考える会社がいくつか現れ、それらによる「国際化ドメイン名」サービスが開始された。

ところで、“.com”や“.jp”に変わる、“.会社”や“.日本”のような、非英文字表記のトップ・レベル・ドメイン名への「憧れ」がある。これを実現するためにはドメイン名の非英文字表記の規格統一に加え、ICANNによる新トップ・レベル・ドメイン名の認可の2つが必要である。

技術的規格の決定を待たない動きについては上記のとおりであるが、それに加え、さらにトップ・レベル・ドメイン名の認可を待たずにこれを擬似的に実現しようとする動きも現れた。

これは、ZLD (Zero Level Domain)²⁾ と呼ばれる手法を用いたものである。

ZLDは、“.”で区切られた階層構造を持つドメイン名の文字列のうち、ある階層から上、上記の“〇〇商店.会社.xy”の例においてはたとえば“.xy”の部分、をアプリケーションにおいて利用者の目から隠してしまおうというアイデアである。

ZLDを実現するソフトウェアはドメイン名解決のためのライブラリ中に組み込まれるなどにより利用者のシステム内に常駐する。このソフトウェアの作用により、利用者は、“http://〇〇商会.会社/”とだけ打ち込むこと

により、“〇〇商会.会社.xy”へアクセスすることができるようになり、あたかも“会社”がトップ・レベル・ドメイン名であるかのように感じることができる。

しかし、この方式はインターネット分断への大きな危険をはらんでいる。ZLDを実現するソフトウェアが全世界統一であればまだ良いが現実にはそうならない。すると、利用者がたまたまのソフトウェアをインストールしたかにより¹²⁾、同じURL (と利用者が信じるもの) を打ち込んでも異なるアドレスへと誘導されてしまう可能性が出てくるからである。

このような動きに危機感を持ったIAB (Internet Architecture Board) は、インターネット分断防止への注意を喚起する声明を2000年5月にRFCとして発信した³⁾。しかし、残念なことに、その後もこのようなサービスが、時たま、あたかも正規なサービスであるように宣伝され、また報道されることが後を絶たない。しかし、正規にこれを実現するには、技術的問題をクリア (幸運な人だけ使えるというのでは不十分である) したうえで、当該トップ・レベル・ドメイン名がICANNによって認可されなくてはならないはずなのである。

規格化の動向

2000年3月、IETFにおいて国際化ドメイン名体系と既存のドメイン名体系とを両立させるための重要な提案がなされた。RACE (Row-based ASCII Compatible Encoding)⁴⁾ と呼ばれる方式である。

RACEは、2バイトコードをBase32 (図-1参照) によってASCIIコード列に変換し、変換結果を現在適用されているドメイン名表記規則に完全に合致した文字列とする。また、文字列中の2バイト文字の各上位ビットパターンが同一である場合、その重複した出力を省略するといった簡易な圧縮機能を持ち、出力文字数を削減する。

通常のASCIIドメイン名とRACEによりエンコードされた結果のASCII列によるドメイン名とを区別するために、RACEによるドメイン名には、識別用の特殊な文字列 (たとえば“bq-”) がプレフィクスとして付加される。

RACEが提案された後、次々とその改良版が提案された。その例に、LACE⁵⁾、BRACE⁶⁾、DUDE⁷⁾ などがある。

¹²⁾ 「ここをクリック!」などと書かれたボタンを安易にクリックすることにより、利用者が意識しないうちにインストールされてしまうような危険もある。

これらはいずれも圧縮方式に対して改良が加えられている。RACE, LACE, BRACEの変換結果を比べたものを図-2に示す^{★13}。いずれの方式も、ひらがな文字列のようにUNICODEのコード表上で近い位置にある文字が並ぶと圧縮率が上がるのが分かる。BRACEではASCII文字を生で出力するため、入力にASCII文字列が混じた場合の効率が良い。

このように既存のドメイン名体系 (ASCIIドメイン名) に互換性を持つ文字列によって非ASCII文字をエンコーディングする方式を総称してACE (ASCII Compatible Encoding) と呼ぶ。ACE方式を用いれば、DNSの基幹システムは既存のものをそのまま使うことができ、そこへは一切手を入れる必要がない。アプリケーション側の対応のみで国際化ドメイン名の利用が可能となる。

2000年12月の第49回IETFにおいて、国際化ドメイン名としてACEを採用することへのコンセンサスが固まった。なお、ACEへエンコーディングする前の文字コードにはUNICODE (ISO10646) を使用することが決定している^{★14}。続々提案されているACEのうちどの方式を規格として採用するかは未定であるが、その選択基準として、

- アルゴリズムが簡単であること
- 変換、逆変換が1対1対応となること
- よく使うパターンで圧縮率が良いこと

が示され、現在IDN-WGの下にACE選択チームが作られ、検討が進められている。

なお、国際化ドメイン名をACEではなく、8ビットコードをそのまま用いる規格によって実現すべきという主張も強かった。しかし、現状のDNSの多くで用いられているBIND (Berkeley Internet Name Domain) ^{★14}の最新版であるバージョン9は8ビットコードを扱えるものの、世界で使われているBINDの20%程度はいまだにバージョン4以下であるといわれ、これらは8ビットコードを扱えない。また、最新バージョンを使っているプロバイダなども多いといわれる。したがって、これら旧バージョンのBINDの置き換えを前提とする8ビットコード方式を採用することは現実的ではないと判断された。

また、ドメイン名表記において、その階層を書く順序は、“ipsj.or.jp”のように左に枝端のドメインを書き、

電子技術総合研究所	
RACE	bq--3c1pww2gmkaigu35z5kaq6aupj3geqa
LACE	bq--761pww2gmkaigu35z5kaq6aupj3geqa
BRACE	wqzfrp2sn2j3brvmur53w3aymqeb22-8q9
でんしぎじゅつそう	
RACE	bq--gbitzqv2oicowixkg
LACE	bq--beygpe2xizmikzc5iy
BRACE	b1eybpgmksipk7u1j1-8q9
電子ぎじゅつそうご	
RACE	bq--3c1pww2qgbhdawbqqquygimc5gbdava
LACE	bq--ag1pwak3kadatatsyqvsf2rsu
BRACE	j1gdhrpqn4gd43d6qr57a-8q9
ABCカンパニー	
RACE	bq--gd7ud72c75b2x46rzp6a
LACE	bq--amaecgsdauyqx46rzp6a
BRACE	31p-ABC-hmnerj2-8q9

図-2 RACE/LACE/BRACEの比較

右にその上位ドメインを書くという順序になっているが、この順序についても「国際化」すべし (ローカライズ可能とすべし) という議論もあった。つまり、日本もそうであるが、中国などでも郵便の住所は通常、“東京都千代田区外神田”のように、ルート→末端の順に書く。そういった習慣を持つ人にとっては、現在のドメイン名表記法の階層記述順序は親しみ難いという主張である。しかし、記述順序の「国際化」についてはIDN-WGとしては扱わない方向となっている。

登録制限文字と正規化

ドメイン名に利用できる文字セットを拡大する際に考慮すべき問題がいくつかある。

その1つに、異文化圏文字間の干渉がある。たとえば印刷された“IPSJ”という文字列を見たとき、我々は通常“アイ、ピー、エス、ジェー”だと考えるが、これはギリシャ文字やキリル文字は混在しないという暗黙の了解があるからである。この了解がなければ、上記文字列は“イオタ、エル、エス、ジェー”かもしれない。これは文字面を眺めただけでは判断できない。また、“高”(9AD9) (“高”(9AD8)の異体字で、JISには採録されていない)^{★15}のような「似て非なる文字」



★13 変換用ソフトウェアは、(社) JPNICの米谷嘉朗氏から提供いただいた。
 ★14 <http://www.isc.org/bind.html> 参照。
 ★15 () 内のコードはいずれもUNICODEの16進表現である。

による混乱の可能性も多い。

この混乱を回避するために、たとえば(社)JPNICでは、UNICODE文字セットのうち、JIS X0208の文字セットからギリシャ文字などを除外したサブセットに対応する文字だけを登録可能とする方針を決めている。しかし、このような対策は“jp”という、日本語だけを扱うと決めることが可能な特殊なドメインについて考えるからこそ可能なことであり、“com”のように世界中からドメイン名登録を受け付けることが期待されている“generic”ドメインについてはこういった対策は困難であろう。「どのような文字を制限すべきか」は言語依存な問題である。このため文字列にその言語に関する情報を追加したランゲージタグをつける案も検討されたが採用に至っていない。

他の問題に、同一文字の表記法が複数存在する場合がある。たとえば、半濁点付きの文字“べ”は、2バイトコード1つ(30DA)で表現することと、“へ”+“°”のように、2バイトコード2つ(30D8+309A)で表現することができる。このような例は欧州やアジアで使われる合成文字あるいは結合文字に数多く見られる。また、半角“A”(0041)と全角“A”(FF21)や、半角“へ”(FF8D)と全角“へ”(30D8)なども、表記法のふれといえる¹⁶。

このようなふれをそのままにしてドメイン名が登録されないよう、許される記述法を統一する正規化規則が決められつつある⁹⁾。また、この規則に基づく変換処理をアプリケーションソフト、リゾルバソフトなど、ドメイン名解決に関与するソフトウェア群のどのような位置で実施すべきかについての検討も始まった¹⁰⁾。

今後の課題

以上述べたように、国際化ドメイン名について数多くの商業的動きはあるものの、規格化に関しては既存のDNS基幹システムをそのまま使える方式とするということが決まったばかりであり、その具体的エンコード法については検討中という状況である。アプリケーションの対応については一部議論が始まったもの^{10), 11)}、細部はまったく手付かずといってよい。

たとえば、WWWブラウザで閲覧するためのHTMLドク

ュメントにおいて、リンク部分に

情報処理学会はここです。

のような記述をするとき、いったい、“地図.情報処理学会”の部分をもどのようなコードで書くべきなのだろうか？

HTMLドキュメントの本文はシフトJISなどのローカルコードで記述される場合が多いが、リンク部分のコードもその周辺と同じローカルコードで書くべきなのだろうか？あるいは、リンク部分だけRACEエンコードし、

情報処理学会はここです。

のように書くべきなのだろうか？ HTTPヘッダはどう書くべきか？

URLのうちドメイン名部分はよいが、“/”の右の部分はどうするのか？この部分について現状ではhttpサーバと、それが載っているOSの実装依存である。

メールのヘッダはどうするのか？正規化やACE化変換処理は誰がどこでいつやるのか？

メールアドレスにおける“@”の左側(ユーザ名部分)はどうするのか？氏名を漢字で書きたいという需要は多い。

問題は山積みである。

パンドラの箱は開けられつつある。

まもなく、こういったアプリケーション側の対応法や関連する規格について集中的な議論が開始されることになるだろう。

参考文献

- 1) Mockapetris, P.: Domain Names — Concepts and Facilities, RFC1034 (Nov. 1987).
- 2) Duerst, M.: Internationalization of Domain Names, Draft-duerst-dns-118n-02.txt (July 1988).
- 3) IAB: IAB Technical Comment on the Unique DNS Root, RFC2826 (May 2000).
- 4) Hoffman, P.: RACE: Row-based ASCII Compatible Encoding for IDN, Draft-ietf-idn-race-03.txt (Nov. 2000) (2000年3月の提案の改訂版).
- 5) Davis, M. and Hoffman, P.: LACE: Length-based ASCII Compatible Encoding for IDN, Draft-ietf-idn-lace-01.txt (Jan. 2001).
- 6) Costello, A.M.: Bi-mode Row-based ASCII-Compatible Encoding (BRACE), Draft-costello-idn-brace-00.txt (Sep. 2000).
- 7) Welter, M.: DUDE: Differential Unicode Domain Encoding, Draft-ietf-idn-dude-01.txt (Mar. 2001).
- 8) Wenzel, Z. and Seng, J.: Requirements of Internationalized Domain Names, Draft-ietf-idn-requirements-04.txt (Oct. 2000).
- 9) Hoffman, P. and Blanchet, M.: Preparation of Internationalized Host Names, Draft-ietf-idn-nameprep-03.txt (Feb. 2001).
- 10) Falstrom, P. and Hoffman, P.: Internationalizing Host Names In Applications (IDNA), Draft-ietf-idn-idna-01.txt (Mar. 2001).
- 11) Kang, M.: Internationalizing Domain Names in Mail User Agents, Draft-ietf-idn-mua-00.txt (Feb. 2001).

(平成13年4月2日受付)

¹⁶ カタカナの“へ”(30D8)とひらがなの“へ”(3078)というものがあるが、これは「似て異なる文字」問題の方である。

