

## ブラウザフォンにおける 多国語ドメイン利用に関する 1 考察

三浦 史光† 高橋 修 ‡

† ‡ 株式会社 NTT ドコモマルチメディア研究所  
〒239-8536 神奈川県横須賀市光の丘 3-5 NTTDoCoMo R&D センタ

E-mail: † fm@mml.yrp.nttdocomo.co.jp

‡ osamu@mml.yrp.nttdocomo.co.jp

あらまし ドメイン名は多くのプロトコルで利用されており、それらには膨大な実装がある。このドメイン名を多国語化する検討が IETF IDN(Internationalized Domain Name)ワーキンググループで行われている。この検討の中間結果を受け、多国語ドメインを URL に適用し、iMode を始めとするブラウザフォンで利用する場合の問題点を検討した。ドメイン名の多国語化が完了したとしても、それだけでは URL としては使いづらく、URL としての標準化活動が必要であることがわかった。また、HTTP や SSL での扱いにも標準化が必要であることもわかった。

キーワード ブラウザフォン, DNS, 国際化, URL

### The study on internationalized domain name and its adaptation for browser phone

Fumiai MIURA † Osamu TAKAHASHI ‡

† ‡ Multimedia Laboratories, NTTDoCoMo, Inc.  
3-5, Hikarinooka Yokosuka-shi Kanagawa-ken, 239-8536 Japan

E-mail: † fm@mml.yrp.nttdocomo.co.jp

‡ osamu@mml.yrp.nttdocomo.co.jp

**Abstract** domain name is utilized for many protocols, and there are many implementations. IETF IDN WG are working for internationalization of domain name. Using IETF's studies, We studied on IDN-URL from the viewpoint of browser phone, We studied service specifications. If IDN will accomplish there work, IDN-URL is not good enough. Internationalized URL requires father studies. HTTP and SSL requires father studies, too.

**Key words** browser phone, DNS, internationalization

## 1. 始めに

### 1.1 IDN とは

IDN とは Internationalized Domain Name の略であり、IETF IDN ワーキンググループで検討されている[1]。現在の DNS は、ASCII 文字の一部で記載されたドメイン名しか扱えない[2]。IDN ワーキンググループは既存の DNS と互換性を保ったまま、ドメイン名に ISO-10646 文字を使えるように拡張することを意図している[3]。IDN が利用できるようになると、「http://商品名.biz/」「http://会社名.jp/」といった URL を利用できる可能性が出てくる。その結果、英語に不慣れな人がインターネットを利用しやすくなったり、特定商品の情報にすばやくアクセスできるようになったりすると期待されている。

本研究は主としてブラウザフォンでこの IDN を URL に利用する場合における技術課題を検討する。

### 1.2 既存の DNS における制約

DNS が扱うドメイン名の制約の内、IDN 化に関係する制約には以下のものがある。[4]。

- 1) ドメイン名は'/'で区切られている。
- 2) '/'で区切られた各部分は英文字・数字・'-'で記載され、最大 63 文字である。
- 3) 英大文字と英小文字は区別されない。
- 4) ドメイン名は 255 文字までである。

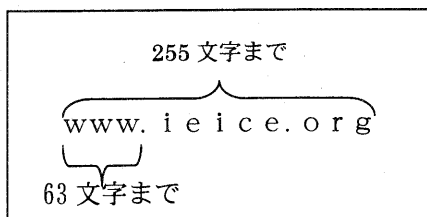


図 1: 既存ドメイン名の制約

### 1.3 既存の DNS が ISO10646 ドメイン名を扱えない理由

ISO10646 文字の表現方法には UTF-8 などの既存の方式があるが、いずれも 1.2 節の制約を満たさない。すなわち、IDN が利用できる新しい表現形式が必要である。そのような表現形式をとることで、既

存の DNS も(ちんぷんかんぷんな文字列として)IDN を扱うことができるようになる。その方式は多数提案されており、[5]に詳しい。どの方式を採用するにせよ、ISO10646 による表現と、[2]の規定する ASCII による表現の 2 つが存在することに変わらないので、2 通りの表現を適宜使い分ける必要がある(問題 1)。

なお、多バイト文字を扱う上で、63 文字制約は以下のようになる。

仮に 6B8B 変換とし、また IDN である識別子をつけないとして(多くの方式では必要である)、UCS2 を用いる場合は 23 文字が限界となる。また、UCS2 ではなく UTF-8 を使い、全て多バイト文字を用いると 15 文字しか扱えない(つまり実際にはもっと少ない文字になる)。商品名には 15 文字で充分かもしれないが、例えば映画やゲームのタイトルには窮屈ではないだろうか。

### 1.4 ブラウザフォンにおける機能分担

ブラウザフォンサービスにはいくつかあるが、現在は proxy server を用いることが多い。本稿も proxy server の存在を前提とする。

何らかの機能は端末ではなく proxy server に実装することが考えられる。しかしながら、proxy server にはあまり機能を搭載することはできない。その理由は SSL にある。

2001 年 1 月に発売された iMode 503 シリーズを筆頭に、現在のブラウザフォンは HTTP over SSL の機能を提供している。この機能において、CONNECT[6]等を用い、proxy server で平文に戻すことなく伝送することによって end to end security を実現している。昨今セキュリティはますます重要になっているのでこのサービスを degrade することは考えられない。従って、proxy server に機能を実装する方法は次の 2 通りしかない。

1. CONNECT 処理に関係する機能だけを実装する。
2. 1 に加え、端末から(proxy server でなくてもよいが)server に対し、変換等の処理を依頼し、返ってきた結果を用いて改めて HTTP(または HTTP over SSL)を処理する。

方法 1 は当然実装されなくてはならない。しかし、ここでエラーが起こることが考えられ、その処理は複雑である(問題 2)。

また、方法 2 は呼び出されるサーバーのスケーラ

ビリティ等の問題があるほか、遅延の大きい無線網を用いる場合には受け入れ難い。本節の結論として、大半の機能は端末に実装するものとする。

## 2. ドメイン名正規化の問題点

### 2.1. ドメイン名の入力方法

ドメイン名は様々な方法で入力される。HTML のアンカーのように比較的信頼性の高い方法もあるが、メールその他のテキストから抜き出す方法、何かを見ながら引き写す方法、「何とかの情報はこの URL だろう」と見当をつける方法など、誤解や入力ミスなどが入り込みやすい方法もある。実サービスを行うためにはこれらの入力方法をどれも扱えることが好ましいが、そのためには誤解や入力ミスなどへの対応が必要となる。

### 2.2. 得られたドメイン名の理解

変換はよく研究されているが、原理的に誤りを避けられない場合がある。しかも多くの文字コード自動変換は日本語を入力されたドメイン名はそのままでは使えない。得られたドメイン名は ISO-2022 や SJIS など、様々な文字コードで記載されている。文字コードの自動前提としている。その上、URL の場合は更に問題は複雑になる。

% HEX HEX encoding[7]されているかもしれない。例えば~に対応する 1 バイトは'%7E'に変換されているかもしれない。

```
http://www.ieice.org/~ken/  
http://www.ieice.org/%7Eken/  
http://www.ieice.org/%7eken/
```

図 2: 同一 URL の様々な表記

更に、入力経緯が異なる文字列を CGI などで合成している場合、部分によって文字コードが異なったり、一部だけエンコードされていたりするかもしれない。

これらの処理はかなり難しく、また誤りは避けられないだろう。つまり、既存の CGI を始めとするかなりの処理系は改良を必要とする。そればかりでなく何らかの運用規定も必要であると考えられる。

### 2.3. 文字コード変換

前節の処理を経てドメイン名を構成する文字がわ

かったとして更に文字コード変換が必要であるが、この際文字ごとに特別な処理が必要である。

- SJIS を例にとると、アルファベットや数字、記号、カタカナなどに複数の code point が割り当てられている。これらは ISO10646 への変換作業の一部として正規化が必要である。
- 例えば('株')の 3 文字を株(U+3231)の 1 文字に変換すべきか。常に変換すべきではないとしても、変換最終結果が長すぎる場合はこのようなことを試みるべきか。
- 対応する文字がない場合はどうすべきか。

ISO10646 への変換には大きな表を用意する必要があるのに、更にこのようなルールを実装するのは、利用できる資源に制約のあるブラウザフォンには重荷である。また、このルールが処理系ごとに異なると URL が universal ではなくなるため、ここにも標準化が必要であると考えられる。

## 3. エラー処理

### 3.1. エラーの起こる場所に依存する困難

変換はうまくいったとしても、他にもエラーはありうる。端末でエラーが起こった場合は入力された(変換前の)URL がわかっているため困難はないが、proxy server や HTTP server は変換後の URL しかわからない。従って proxy server や HTTP server が生成したエラーメッセージに入っている URL は次のどれかになる(1.4 節問題 2)。

1. 変換後の、人間には読めない文字列
2. 逆変換した文字列
3. (HTTP server の場合)設定された文字列

1.3 節で問題 1 として触れておいたが、ブラウザに与えられた入力は必ずしも変換前とは限らないので 1)を排除できない。また、2)や 3)は 2 章で述べた様々な変換に対応する URL を生成できない場合がある。エラーメッセージには入力した URL が表示されると期待されるが、それは難しい。ここにも新たな運用ルールが必要である。なお、この問題は log にも生じる。

## 4. IDN のそもそもの狙いは達成できているか?

### 4.1. IDN 化 URL はうまく扱えない時がある

以上の検討で、IDN を URL に用いる場合の課題を挙げた。ブラウザフォンには重荷であり、様々なノウハウが必要であり、にも関わらずうまく扱えない場合がある。さて、1.1 節で述べたそもそもの狙い

は達成されているであろうか。

#### 4.2. 見当をつけて手入力する時に進化を発揮するはずだが...

1節で述べたように、URLは様々な方法で得られる。bookmark や HTML などに含まれている場合は、URL が目に触れる機会が少ないので IDN 化 URL の利点は少ないと考えられる。雑誌などから引き写す時も、書いてある通りに入力するだけであるから比較的利点は小さい。利益が大きいのは「あの商品はきっとこんな URL だろう」とあてずっぽで入力する場合であろう。IDN 化 URL がうまく使えないことがあることを幾度か経験した利用者が続けて試すとは思えない。サービスとして提供開始する時点でかなり「当たる」ようにしなくてはならない。

ところが、サービス開始当初はそもそもそのような URL があまりない。卵と鶏の関係になっている。

#### 4.3. IDN 化すれば URL として有用なのか

URL にはドメイン名以外の要素もある。IDN はドメイン名の規定だから、ドメイン名でない部分は今までのままである。すると、以下のような URL はどのように扱うのだろうか。

`http://会社名.jp/製品名`

「会社名」は IDN に従い、「製品名」は他の規約に従うのだろうか。「製品名」を扱う規格はドメイン名も扱うのではないだろうか。それともこのような URL は許さないのだろうか。

#### 4.4. HTTP への対応

今までプロトコルの検討をしてこなかったが、HTTP で IDN 化 URL を利用する場合に残っている課題がある。HTTP/1.1 では Host ヘッダでホスト名を渡さなくてはならないが、ここにはどのようなドメイン名を渡すのだろうか。IDN 化されてないサーバーであれば ASCII 化したドメイン名を渡すのだろうが、そうでないなら変換前の多バイト文字を使ったドメイン名を渡す必要がある(逆変換で元に戻らないから)。それとも片方を試み、エラーになったらもう片方を試みるのだろうか。

#### 4.5. SSL への対応

SSL は X.509 形式の電子証明書を用いて認証を行っている[8]。2003年12月31日以降に発行される

X.509 電子証明書は、DN 等に UTF-8 を用いると規定している[9]。IDN が広く利用されるようになった場合、X.509 電子証明書にも多国語ドメイン名が記載されることが考えられるが、どのように記載されるのだろうか。現在までに X.509 や IETF PKIX ワーキンググループ側での議論はなされていないようである。このまま調整がつかないままだと電子証明書の検証時に多バイトドメイン名をうまく検証できない可能性が多分にある。当面、SSL を用いる場合は IDN でない URL を用いるのが無難だろう。すると、IDN 化 URL でアクセスし始めたのに、いつのまにか異なる(見かけでは全く関係のない)URL にすぐ変わっているのである。これは不審に思われまいだろうか。

#### 4.6. メールに使うとどうなるか

本研究はここまで URL の scheme[11]として、http と https しか扱ってこなかった。WWW と共に双壁をなすサービスにメールがある。メールアドレスを扱う URL には IDN は使えるのだろうか。2章や3章で述べたような問題はあがあるが、HTTP に用いる場合と同様の手法を適用し、SMTP では ASCII 化してから扱うことにすれば domain[10]部分に使うことはできなくはない。しかし人命に相当する local-part[10]に使うことはできない。4.3 節と同様の問題がある。

### 5. 結論

ドメイン名はインターネットにおいて根幹をなす規格である。多くの規格が参照しているし、それ以上に多くの実装が利用している。従って、ドメイン名を国際化するだけでなく、各プロトコルや実運用まで検討しないと使ってもらうのに充分とはいえない。

本研究は、IDN ワーキンググループが目標の1つとしている URL へ適用における問題点を具体的に指摘し、IDN 化 URL の有効性についても疑問を投げかけている。

今後はここで指摘した問題点を検討してゆくと共に、IDN 化 URL とは異なるアプローチをも検討したい。アプローチはどうあれ、(多国語化 IDN でなく)多国語化 URL の実用化をめざしたい。

## 文 献

- [1] "Internationalized Domain Name (idn) Charter," <http://www.ietf.org/html.charters/idn-charter.html>
- [2] P. Mockapetris, "DOMAIN NAMES - CONCEPTS AND FACILITIES," RFC1034, Nov.1987.
- [3] P. Faltström, P. Hoffman, A. M. Costello, "Internationalizing Domain Names in Applications (IDNA)," draft-ietf-idn-idna-07, Feb.24, 2002.
- [4] P. Mockapetris, "DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION," RFC1035, Nov.1987.
- [5] P. Hoffman, et al. "Report of the IDN ACE Design Team," draft-ietf-idn-ace-report-00 Jun.14, 2001(expired)
- [6] R. Khare, S. Lawrence, "Upgrading to TLS Within HTTP/1.1," RFC2817, May 2000.
- [7] R. Fielding, et al., "Hypertext Transfer Protocol - HTTP/1.1," RFC2616, Jun.1999.
- [8] A. O. Freier, et al., "The SSL Protocol Version 3.0," <http://www.netscape.com/eng/ssl3/draft302.txt>, Nov.18, 1996.
- [9] R. Housley, et al., "Internet X.509 Public Key Infrastructure Certificate and CRL Profile," RFC2459, Jan.1999.
- [10] P. Resnick, et al., "Internet Message Format," RFC2822, Apr.2001.
- [11] T. Berners-Lee, et al., "Uniform Resource Locators (URL)," RFC1738, Dec.1994.