

# インターネット 10分 講座

## ルートゾーンラベル生成ルール(LGR)とは ～ IDN TLDにおける日本語LGRの策定 ～



### はじめに

2020年12月11日、JPNIC Blogで「完成間近:日本語ルートゾーンラベル生成ルール(LGR)提案書ドラフトまでの道のり<sup>※1</sup>」と題して、日本語ルートゾーンLGRについてご紹介しました。このブログ記

事で紹介できたのは、タイトルの通り、ルートゾーンにおける日本語のLGRの策定に関してだけでしたが、本稿では国際化ドメイン名の起りから現在に至るまでの流れも含めて、紹介しようと思います。

※1 完成間近:日本語ルートゾーンラベル生成ルール(LGR)提案書ドラフトまでの道のり <https://blog.nic.ad.jp/2020/5523/>

### 2

### ドメイン名の多言語化の揺籃期

今でこそコンピュータ上では日本語やその他の言語を含む多言語を、普通に自然な形で取り扱うことができ、それに疑問を抱くこともないと思います。しかし、日本でインターネットが急速に広がった1990年代半ばの時点では、文字化けやバグに悩まされることが多くありました。この25年間をかけて、コンピュータにおける多言語対応は成熟したと言ってよいと思います。国際化ドメイン名(IDN)は、ドメイン名における多言語対応ということになります。ASCII文字しか利用できなかったドメイン名に、多言語対応を施していく活動は、いろいろなフェーズを経て、最終局面に入ったところだと言えると思います。

1990年代半ばの時点で、ドメイン名ラベルにはASCII文字しか使えませんでした。ASCII文字以外をドメイン名で使えるようにする、ドメイン名の多言語化は、1990年代終盤に始まります。日本レジストリサービス(JPRS)を分社化する前のJPNICでは、1999年5月にiDNS-TFを立ち上げ、多言語化・国際化ドメイン名への対応を開始します。最初の課題は、ASCIIによる文字列としてプロトコル上規定されているドメイン名ラベルに対する、多言語のエンコード方式でした。結果として採用された方式は、文字コードとしてUnicode<sup>※2</sup>を採用し、「ASCII以外の文字を含むラベルの文字列から、まず大文字や小文字、全角・半角などといった表記の揺れをNAMEPREPという方式により正規化した後、さらにPunycodeという方式で変換することで、「xn-」から始まるASCII文字列を得る」

というもので、この後のドメインネームシステム(DNS)での処理は、ASCII文字列のドメイン名とまったく同様となります。一方で、このような技術的な方式をユーザーが使えるようになるためには、PCのブラウザなどの端末機器で、ユーザーが入力した多言語表記をPunycodeに、Punycodeをユーザーが読める多言語表記に相互変換できる必要があるため、ブラウザベンダーによる方式対応が必要です。

ユーザが入力した文字列

日本語ドメイン名EXAMPLE.jp

↓ NAMEPREPによる正規化

日本語ドメイン名EXAMPLE.jp

↓ Punycodeによる変換

xn--example-6q4fyliikhk162btq3b2zd4y2o.jp

図1 多言語表記とPunycodeの変換例

ここでは紙幅を費やすことは控えますが、国際化ドメイン名の技術標準化と実用化に関しては、技術的な方式の設計と実装に関しても、

日本以外の各国のインターネット関係者、ビジネスプレイヤーなどとの議論、協議、交渉、協調、標準化活動、さらに、ベンダーに対する実装交渉、利用者にも門戸を開いた運用実験といった、当時の関係者

による多岐にわたる不断的努力の賜物です。日本語ドメイン名協会(JDNA)監修、宇井隆晴氏著の「日本語ドメイン名」<sup>※3</sup>という書籍には、当時の関係者の営みが克明に詳細にまとめられています。

※2 JPNIC - インターネット用語1分解説 Unicodeとは <https://www.nic.ad.jp/ja/basics/terms/unicode.html>

※3 日本語ドメイン名協会(JDNA)監修、宇井隆晴著の「日本語ドメイン名」インプレスR&D 2006年 ISBN4844322605

### 3

## IDNの登録規則 - .JPの場合

ドメイン名ですから、利用するためにはレジストリに登録され、ネームサーバに対する問い合わせで解決が可能である必要があります。JPドメイン名の登録管理業務を引き継いだJPRSが日本語ラベルの登録を開始したのは、2003年7月10日。IDNに関する標準化が完了したのを見届けてからのことでした。JPドメイン名では、当時新設されて間もなかった汎用JPドメイン名に対して、日本語ラベルの登録が可能となりました。それにあたっては、ドメイン名ラベルに対するルールが必要です。例えばASCIIの場合、英字AからZ、数字0から9、およびハイフン、63文字以下、先頭と末尾にハイフンは来ない、といったルールがあります。このようなルールが、ラベル生成ルール(Label Generation Rule,LGR)です。汎用JPドメイン名におけるLGRは、「汎用JPドメイン名登録等に関する

規則」が参照する技術細則<sup>※4</sup>の中に、他の技術的条件とともに明らかにされています。

汎用JPドメイン名の日本語LGRに関して着目すべきは、異体字<sup>※5</sup>の設定がないことです。異体字とは、文字としては異なる見た目を持つものの意味や発音が同じ文字のことです。日本語では漢字に対して旧字体が存在する場合があります、例えば「国沢」「国澤」「國沢」「國澤」はすべて「くにさわ(くにざわ)」と読み、持つ意味合いも同じです。しかしながら、これらが姓名に対して使われる場合、このうちのどれか一つにだけ自身のアイデンティティを感じるというケースが多いと思います。異体字を設定しないという方針は、このような日本語の実情に合致します。

※4 JPRS - 汎用JPドメイン名登録等に関する技術細則 <https://jprs.jp/doc/rule/saisoku-1-wideusejp.html>

※5 JPNIC - インターネット用語1分解説 異体字とは <https://www.nic.ad.jp/ja/basics/terms/variant.html>

### 11

## IDNの登録規則 - ICANNでは

ここまで、JPドメイン名におけるIDNのLGRを見てきました。日本を示すccTLDである.JPでは、日本の公用語が日本語だけであることから、IDNとしては日本語だけを規定すればよさそうです。しかし、公用語が複数ある国のccTLDでは、公用語の数だけIDNを導入したいでしょう。gTLDであれば、需要が見込めるさまざまな言語を導入したいかもしれません。さらに、トップレベルにIDNを適用する場合には、すべての言語がTLD空間を共有することになり、考慮しなければならないことが一気に増える結果となります。

JPドメイン名で日本語による登録が可能になった後、ICANNでは2000年と2003年の小規模なgTLDの追加に引き続き、2005年11月に新gTLDプログラムの導入に関するポリシー策定プロセスが始動され、これが2012年の新gTLDプログラム実施につながります。この間、7年近くがポリシー検討と実施準備に費やされました。その間に、ICANNの国コードドメイン名支持組織(ccNSO)でもIDNによるccTLDの導入検討が進み、暫定的なポリシーでIDN ccTLDの登録を行う、IDNccTLDファストトラックと呼ばれる制度が、2009年11月に始まりま

す。このファストトラックには、早速アラビア語圏のアラブ首長国連邦、エジプト、サウジアラビア並びにロシア連邦が申請の上、承認されますが、その次に承認された中国、香港、台湾<sup>※6</sup>が、その後のルートゾーンLGR検討のきっかけを作ります。

2010年6月に承認されたこれら三つのIDN ccTLDはいずれも漢字によるccTLDラベルを採用していますが、中国は「中国」と「中國」、台湾は「台湾」と「台灣」というように、異体字の関係にある繁体字と簡体字をいずれも承認することになりました<sup>※7</sup>。これらはSynchronized IDN ccTLDs<sup>※8</sup>と呼ばれ、異体字TLDのすべてを同じIPアドレスに解決させることとしました。この措置は大きな人口を擁する中国語圏に対するIDN TLDの需要に対してある程度妥当なものと考えられますが、国名にとどまらないさまざまな文字列を扱うことになるgTLDにおけるIDNの取り扱いに対して、一般的なアプローチを示すものとしては不十分と、ICANN理事会は考えました。その結果、理事会内に常設のIDN異体字作業部会(Board IDN Variant Working Group)を設置し、事務局エキスパートとともに、gTLDも含めたTLD空間における異体字管理に取り組むこととしました<sup>※9</sup>。また、これらの検討が完了するまでの間、

異体字TLDの登録は行わないことを決めました。

その結果として2013年3月にまとめられたのが、「Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels (IDNAドメイン名ラベルに関するルートゾーンののためのラベル生成ルールの策定と維持の手順)」<sup>※10</sup>と呼ばれる文書です。あらましとしては、以下のような要素から成り立ちます。

- 1) 各言語の利用者コミュニティごとに、各言語の生成パネル (Generation Panel, GP) を組成して、LGR案検討策定を行う
- 2) GPIはその言語のドメイン名ラベルとして利用するすべての文字を列挙したMaximal String Repertoire (MSR) の定義から作業に着手し、MSRからの文字の組み合わせでラベルを作るルールとして、LGR案を策定する

※6 JPNIC -「ICANN理事会がファスト・トラック プロセスによる、IDN ccTLDの導入を承認」 <https://www.nic.ad.jp/ja/topics/2009/20091106-01.html>

※7 JPNIC -「ICANNが中国語圏のIDN ccTLDを承認」 <https://www.nic.ad.jp/ja/icann/topics/2010/20100628-01.html>

※8 ICANN - Adopted Board Resolutions | Nairobi, 12 Mar 2010, 第13決議  
<https://www.icann.org/resources/board-material/resolutions-2010-03-12-en#13>

※9 ICANN - Conclusion of Working Group on Equivalent Strings Support, ICANN理事会決議  
<https://features.icann.org/2010-12-10-conclusion-working-group-equivalent-strings-support>

※10 ICANN - Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels  
<https://www.icann.org/en/system/files/files/draft-lgr-procedure-20mar13-en.pdf>

※11 ICANN - Root Zone Label Generation Rules <https://www.icann.org/resources/pages/root-zone-lgr-2015-06-21-en>

- 3) パブリックコメントを経て検討されたLGR案は、ルートゾーン全体のLGRの策定に責任を持つ統合パネル (Integration Panel, IP) に提出され、承認を経てルートゾーンLGRに組み入れられる

執筆時点でのルートゾーンLGRは第4版 (RZ-LGR-4) で、18の文字種を含んでいます。

ICANNはルートゾーンLGRに関して、ルートゾーンLGRページ<sup>※11</sup>を開設して、各種定義文書、作業状況、作業済みLGRなどの各種情報を集積、公開しています。

## 5

### 日本語GPによる日本語LGRの検討

このような形で、ルートゾーンにおけるLGRの検討要領が定まったのを受けて、日本語に関するルートゾーンLGRの検討も始まります。日本では、その起こりからグローバルなレベルでIDNに関する議論を先導してきたJPRSの呼びかけによって、レジストリ/レジストラ関係者、技術、言語、ポリシー策定等に関する専門家、知的財産権専門家など各コミュニティを代表するメンバーからなる日本語生成パネル (JGP) が2014年に発足しました。JGPIは、その活動に関する情報提供のためにWebサイトを開設しています<sup>※12</sup>。JPNICは、事務局機能をJPRSと分担するとともに、著者の前村昌紀がパネルメンバーとして参画しています。

日本語ルートLGRを検討する上での議論の焦点は、三つほどありました。中国語、韓国語との共用を踏まえた異体字の定義、異体字ラベルの削減、視覚的類似ラベルの抑制です。

1点目は異体字の扱いに関して、ccTLDであるJPドメイン名に導入されている、日本語JPドメイン名の第2、第3レベルの文字列規則には異体字の定義はありません。例えば「広」と「廣」などは別の文字として扱われています。一方、同じく漢字を利用する中国語については、ccTLDである.cnや.twでは「广」や「廣」、「広」などは異体字として定義され、同じ文字として扱われます。今回のルートゾーンLGRでは、異体字を特に必要としない日本語に対して、中国語、韓国語

ではルートゾーンLGRでも異体字を定義する方向にありました。TLDには複数の言語が乗り入れるため、中国語や韓国語で異体字である文字は、日本語でも異体字としなければ混乱の元になると考えられます。そこで、日本語LGRでは、中国語LGR、韓国語LGRで異体字とされるものは、異体字として受け入れることにしました。この中には「機」と「机」、「葉」や「叶」のように、日本語では別の意味を持つにもかかわらず、中国語では繁体・簡体の関係にある異体字といったものも含まれます。

2点目は、異体字ラベルの削減です。異体字を設けると、異体字を持つ文字複数からなるラベルには、組み合わせの数だけ異体字ラベルが発生するということとなりますが、利用者の混乱の元になります。これに対して提案書ドラフトは、申請されたラベル自体と、これを常用漢字に置き換えた異体字ラベルのみを使用可能として、それ以外の使用を禁止することで、異体字ラベルの削減を図りました。

3点目の視覚的類似ラベルの抑制は、もともとJGPが取り組んでいた「同じ意味で異なる形」の異体字に加えて、統合パネルからの要請に基づいて、「異なる意味だが視覚的に極めて似通った」文字も異体字として定義する方針を受け入れ、Unicode consortium<sup>※13</sup>のConfusable Characters List (錯視が起こりやすい文字の



リスト)<sup>※14</sup>から、認知実験を経て10組の文字を異体字として定義しました。

ひらがな	カタカナ	漢字	記号文字	漢字
へ	ヘ		ー	一
べ	ベ		、	、
ぺ	ペ			
	ニ	二		
	ハ	八		
	カ	力		
	ト	卜		
	ロ	口		
	タ	夕		
	エ	工		

図2 視覚的類似性から異体字として定義された文字

この3点のうち、特に1点目の3言語での共用にかかる調整と、3点目の認知実験の実施に時間を要したのが、日本語ルートゾーンLGRの検討に6年間もの時間がかかっている要因です。

日本語ルートゾーンLGR提案書のドラフトは、2020年10月15日に統合パネルに提示される<sup>※15</sup>とともに、JGP Webサイトで公開されました<sup>※16</sup>。

統合パネルから新たな検討項目が提示されており、今後は提案書の詰め作業が進んでいきます。意見も募集されていますので、ご関心のある方は提案書ドラフトをご覧ください、お気づきの点があれば日本語生成パネル事務局までぜひお知らせください。

※12 日本語生成パネル <https://j-gp.jp/>

※13 Unicode Consortium <https://unicode.org/>

※14 Unicode Consortium - Confusable Characters List

<https://unicode.org/Public/security/13.0.0/confusables.txt>

※15 日本語生成パネル - 日本語LGR提案書ドラフト(v0.15)をICANN関係者に提示

<https://j-gp.jp/topics/20201015-01>

※16 日本語生成パネル - 日本語LGR提案書(v0.15)

<https://j-gp.jp/%E6%97%A5%E6%9C%AC%E8%AA%9ELGR%E6%8F%90%E6%A1%88%E6%9B%B8>

日本語LGR提案書(v0.15)の概要

<https://j-gp.jp/%E6%97%A5%E6%9C%AC%E8%AA%9ELGR%E6%8F%90%E6%A1%88%E6%9B%B8%28v0.15%29%E3%81%AE%E6%A6%82%E8%A6%81>

## 6

## おわりに

ここまで、完成間近である日本語ルートゾーンLGRの策定状況を紹介するにあたり、JPNIC Blogの記事では含めなかった、国際化ドメイン名の標準化段階から、ICANNにおけるルートゾーンへの適用も含めた流れを説明しました。現在のICANNにおけるIDN関連の状況としては、ルートゾーンLGRの策定は、日本語を含め、まだLGRの策定完了を待つものもあるものの、ルートゾーンLGRのTLDへの適用要領に関しては、コミュニティでの検討を経て、2019年3月、ICANN神戸会議における理事会で承認されました<sup>※17</sup>。現在はこの適用要領を、gTLD、ccTLDのポリシーとして組み入れるための議論が、GNSO、ccNSOで進んでいます。これが完了すると、ようやく2010年以来取り組んできたルートゾーンにおける異体字の管理というテーマに対する取り組みがひと段落することになります。

ICANNにおけるIDN活動<sup>※18</sup>は、各言語のLGR策定も終盤に差し掛かり、徐々に軸足をユニバーサルアクセプタンス(Universal Acceptance, UA)<sup>※19</sup>に移しつつあります。UAは、IDNによるドメイン名(目新しく馴染みのないASCIIのgTLDも含む)が、広く一般に対して、ドメイン名(あるいは電子メールアドレス)として認知され、適切に処理されるように促していく活動です。これはICANN事務局だけでなく、コミュニティメンバーによるUniversal Acceptance Steering Group(UASG)<sup>※20</sup>によって推進活動が展開されていますが、ダイバーシティを推進する施策として、インターネット関係者のみならず、広く高い関心を集めているのが印象的です。

(JPNIC インターネット推進部 前村昌紀)

### [ ICANNにおけるIDN関連タイムライン ]

- 2000 — 2000年gTLD追加
- 2003 — 2003年gTLD追加
- 2005 — 新gTLDプログラムポリシー承認
- 2009 — IDN ccTLDファストトラック開始
- 2010 — 中国、香港、台湾へのIDNccTLD承認(Synchronized TLD)  
ICANN理事会内「IDN異体字作業部会」設立
- 2012 — 新gTLDプログラム施行
- 2013 — IDNTLDラベル生成手順書策定
- 2014 — 日本語生成パネル設立
- 2019 — ルートゾーンLGRのTLDへの適用要領文書が理事会承認

図3 ルートゾーンへのIDN組み入れの流れ

※17 ICANN - Approved Board Resolutions | Regular Meeting of the ICANN Board, 14 Mar 2019, 決議2.a

<https://www.icann.org/resources/board-material/resolutions-2019-03-14-en#2.a>

※18 ICANN - Internationalized Domain Names

<https://www.icann.org/resources/pages/idn-2012-02-25-en>

※19 ICANN - Universal Acceptance

<https://www.icann.org/ua/>

※20 Universal Acceptance Steering Group

<https://uasg.tech/>