

IETFが策定する国際化技術とIoT技術国際化の課題

根本 貴弘¹

¹青山学院大学

インターネットがさまざまな地域で利用されるようになり、国際化ドメイン名や国際化メールアドレスをはじめASCII文字集合の範囲外の文字を含む識別子の利用が求められるようになった。IETFではこのような需要に対していくつかの国際化技術を策定してきた。現在、標準化が進められているIoTサービスの情報資源の参照時に使用する識別子には、国際化文字列の利用が許容されていることから、利便性向上および安全性向上の一手法として国際化技術の利用が有効である。本稿では、IETFが行っている標準化活動のうち筆者が参加してきた国際化技術とその活用が期待されるWGで得た知見をもとに、日本からの貢献も求められているIETFが策定する国際化技術とその活用が期待されるIoT技術の標準化動向を紹介する。

1. はじめに

IoTサービスの普及に伴い、我々の身の回りにあるさまざまなモノが情報資源として活用されるようになった。IoTサービスの情報資源の中には閉じた系の中での利用を想定したものが一方、今後、それら情報資源をインターネット上の情報資源として広く活用されることが期待されており、インターネットとの親和性も求められるようになりつつある。また、IoTサービスの情報資源名は身の回りのモノを参照する際に利便性の観点から母語の利用が許容されるべきであり、特定の地域や文化圏と親和性が高くかつインターネット上で広く参照されることが求められる。

インターネット技術に係る技術標準を策定しているInternet Engineering Task Force (IETF)[1]では、インターネットがさまざまな地域で利用されるようになり、国際化ドメイン名[2][3][4][5][6]や国際化メールアドレス[7][8][9][10][11][12][13][14]をはじめASCII文字集合の範囲外の文字を含む識別子を利用するための国際化技術を策定してきた。一方で、IETFではIoT技術の標準化にも取り組んでおり、現在、標準化が進められているIoT技術のサービスディスカバリーに使用する識別子には、国際化文字列の利用が許容されている[15][16]。このような国際化文字列を含む識別子を利用するIoT技術の利便性および安全性の向上に有効な一手法としてIETFが策定してきた国際化技術[17][18][19][20]がある。

本稿では、一般社団法人情報通信技術委員会の2017年度「デジュール及びフォーラム標準に関する標準化活動の強化に資する調査等」にて筆者が取り組んだ「IETFが策定する国際化技術とそれらを活用するIoT技術の動向調査」をもとに、インターネット技術の主要なプロトコルの策

定を行っている標準化団体であるIETFが策定する国際化技術とその活用が期待されるIoT技術における国際化技術の動向および課題を論ずる。

2. IETFが策定する国際化技術

2.1 IETFにおける国際化

IETFではインターネットの利用地域拡大に伴い、電子メールやHTMLの本文で利用する文字列をはじめ、情報資源を参照するための識別子や認証に用いるIDやパスワードとして利用する文字列にASCII文字集合の範囲外の文字を含むことが求められ、いくつかの国際化技術に関する提案がされてきた。

たとえば、1990年代初頭には、それまで個々に実装・提案されてきた電子メールにおける国際化手法が、RFC 2045, RFC 2046, RFC 2047, RFC 2048（現RFC 4288およびRFC 4289）、RFC 2049によって統一された電子メールの拡張方式であるMultipurpose Internet Mail Extensions (MIME)として標準化され、電子メールやHTMLの本文でASCII文字集合以外の文字を扱うことが可能となり、今日では本文に国際化文字列を含んだ電子メールやWebページを日常的に利用することが可能となった。

IETFでは、RFC 2277にて主にアプリケーションプロトコルやセキュリティプロトコルで国際化文字列を使用する場合は、文字集合としてISO/IEC 10646 (Unicode)を使用することとし、その符号化方式はUTF-8とする必要があることが提案され、必要に応じてRFCに"Internationalization considerations"というセクションを用意すべきであるとしている。また、RFC 3629 (旧RFC 2279)にてASCII文字集合と互換性を持つUTF-8の扱いについて定義がされている。そして、RFC 6365 (旧RFC 3536)にてIETFにおける国際化はプロトコルで非ASCII文字集合を扱えるようにすることでであると定義が行われ、UTF-8を許容するプロトコルが策定されるようになっている。

なお、IETFでは、国際化文字列を使用するプロトコルごとに国際化手法を標準化することや標準化した仕様の維持、管理による負担を軽減するために、各プロトコルで共通して利用可能な国際化手法をフレームワークとして標準化している。このフレームワークには、2002年に策定されたStringprepと2015年に策定されたPRECIS Frameworkと呼ばれるものがある。

2.2 Stringprep概要および課題

IETFの国際化技術のうち最も多くのプロトコルが採用している国際化技術にRFC 3454として標準化されたStringprepがある。UTF-8を識別子やパスワード等のプロトコル要素として使用する場合、利便性や安全性の観点から文字列の照合の精度を向上させるための文字列変換処理やプロトコル要素として不適切な文字が含まれていないかを確認するための仕組みが必要となる。そのため、国際化文字列を含む識別子やパスワードを扱うための汎用的な国際化技術としてStringprepが標準化され、さまざまなプロトコルで国際化文字列の利用を可能とした。Stringprep概観図を図1に示す。Stringprepでは、プロトコル要素に国際化文字列を含む場合に、その文字列の比較および確認をどのように行うかを定めている。文字列の比較では、表1に示す大文字や小文字、全角文字や半角文字、合成済み文字や結合文字列といった文字の区別の有無によって実施する文字列変換処理を定めている。

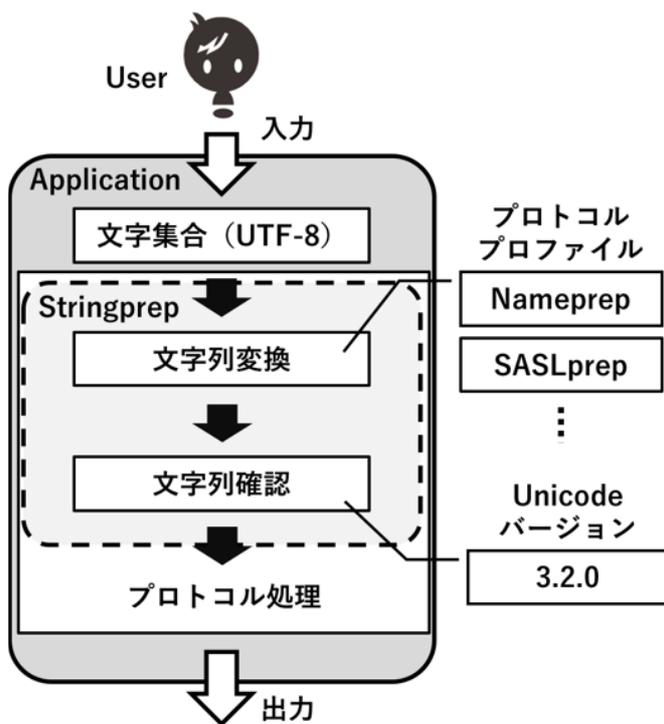


図1 Stringprep概観図

表1 国際化文字列の変換処理例

変換処理	例	
文字種 (大文字・小文字)	A (U+0041)	a (U+0061)
文字幅 (全角・半角)	ア (U+FF71)	ア (U+30A2)
合成済文字・ 結合文字列	カ [◌] (U+30AB U+3099)	ガ (U+30AC)

Stringprepでは、Unicode 3.2に基づく文字列変換表および禁止文字確認表が用意されており、プロトコルによって必要な変換および確認処理を選択し、Stringprepプロファイルを作成する。文字列変換では以下の3通りの表が用意されている。

- 削除する文字
- 正規化 (NFKC) および文字種統一による変換
- 正規化を行わず文字種統一のみによる変換

また、文字列の確認として、空白文字や制御文字、私用文字、サロゲートコードポイント等のプロトコル要素として不適切な文字が文字列中に含まれていないかを確認する禁止文字の確認や左から右に表記する文字と右から左に表記する文字が混在する文字列に対する双方向性確認が用意されている。

Stringprepを使用する代表的なプロトコルには、2003年にRFC 3490, RFC 3491, RFC 3492 として標準化された Internationalizing Domain Names in Applications (IDNA2003) がある。IDNA2003では、ドメイン名を構成するラベルと呼ばれ

るピリオド(.)で区切られた文字列に対して、IDNA2003のStringprepプロファイルであるRFC 3491として標準化されたNameprepを使用する。また、IDNA2003のほかにもStringprepを使用するプロトコルにはSASLやLDAP、XMPP等がある。

しかし、Stringprepでは以下に示す複数の課題があることが明らかとなっている。

- 正規化処理によって正規等価性が失われる文字がある
- 禁止文字確認表にないプロトコル要素として不適切な文字が文字列確認の結果利用可能となる
- 文字列の双方向性の確認によって特定の文化圏で使用する文字列に制約が生じる
- 特定の文字集合のバージョン依存にしている
- プロトコルごとにプロファイルを定義するため文字列処理が重複するプロファイルが存在する

そのため、Stringprepを使用する国際化ドメイン名は2010年にIDNA2003からRFC 5890, RFC 5891, RFC 5892, RFC 5893, RFC 5895として標準化されたIDNA2008へと更新されることとなった。IDNA2008では、Stringprepの課題の一つであったUnicodeのバージョン依存問題を解決すべく、StringprepおよびNameprepを使用せず、Unicode Consortiumが定義するUnicode Character Database (UCD) を使用して、国際化識別子の照合を行うための文字列変換および確認方法を定義しているという特徴がある。

また、Stringprepを利用するSASLやXMPP等の他のプロトコルについては、precis working group(WG)でStringprepに代わる国際化の枠組みおよびその適用方法について標準化が行われた。

なお、Stringprepは用いないが国際化ドメイン名と関連のある国際化技術として電子メールのメールアドレスの国際化がある。国際化ドメイン名同様に、国際化文字列を扱いたいという要望が中国語圏やアラビア語圏を中心としてあがり、EAI (Email Address Internationalization) と呼ばれるメールアドレスの国際化技術がRFC 6530, RFC 6531, RFC 6532, RFC 6533, RFC 6855, RFC 6856, RFC 6857, RFC 6858として標準化された。IDNA2008とEAIの違いは、IDNA2008は下位互換性を確保するために、国際化文字列をPunycodeと呼ばれるASCII互換エンコーディングを使用しASCII文字列に変換する仕組みを持つが、EAIでは、国際化文字列をASCII文字列へ変換することにより、@の左側のローカルパートの文字列が、意図的にASCII文字を利用してるのか、ASCII互換エンコーディングを使用した結果であるのかについて明示的な区別が不可能となるため、国際化文字列をASCII文字列へ変換しないこととしている。そのため、国際化技術に対応していないメールサービスとは原則互換性を確保しないものとしているIDNA2008との違いがある。

2.3 PRECIS Framework概要

PRECIS Frameworkは、precis WGにて標準化が行われたStringprepに代わる国際化文字列を含む識別子やパスワードをプロトコルで扱うための国際化技術である。

precis WGでは、Stringprepに代わる国際化技術としてPRECIS FrameworkをRFC 8264として標準化しており、Stringprepを利用するSASLやXMPP、LDAP等のプロトコルでPRECIS Frameworkを利用するためのプロファイルも標準化している。プロファイルには、認証で使用するユーザ名やパスワードで使用することを想定したRFC 8265: Preparation, Enforcement, and Comparison of Internationalized Strings Representing Usernames

and Passwordsやチャットルーム名等での使用を想定したRFC 8266 : Preparation, Enforcement, and Comparison of Internationalized Strings Representing Nicknamesがある。また、新たにプロファイルを設計するためのガイドラインとしてRFC 7790 : Mapping Characters for Classes of the Preparation, Enforcement, and Comparison of Internationalized Strings (PRECIS)もあり、RFC 7790ではプロトコルに依存した文字列処理や言語や文脈によって必要な文字列処理が異なる文字について説明が記載されている。

PRECIS FrameworkのStringprepからの主な変更点は以下の通りである。

- 文字列変換処理の改善
- 文字列の確認方法を禁止文字列の確認から使用可能文字列の確認へ変更
- 文字列の双方向性確認方法の改善
- 特定のUnicodeのバージョンに依存しない
- 2つのサブクラスをサポート

PRECIS Frameworkでは、表2に示す順を定義しており、プロトコルで必要となる処理を選択して、その変換処理をPRECIS Frameworkプロファイルとして標準化する必要がある。Width mappingでは、文字幅に関する文字列変換をUCDのUnicodeData.txt中のDecomposition_TypeおよびDecomposition_Mappingに従い実施する。この変換は、主にNormalization（正規化）にて正規化（NFC）を選択した場合に必要となる。Stringprepで用意されている正規化（NFKC）では、互換等価な文字へ変換を行うが、正規化（NFC）では、正規等価な文字への変換を行うため、日本語の半角カタカナや全角カタカナ等は異なる文字として区別される。Additioanl mappingでは、区切り文字の変換を行うDelimiter mappingや制御文字の削除や全角空白文字等を半角空白文字（U+0020）に変換するSpecial mapping、トルコ語やドイツ語、ギリシャ語の一部の言語や文脈によって字形が異なる文字を変換するLocal case mappingが用意されている。Case mappingでは、大文字を小文字に変換する文字種統一処理を行う。Normalization（正規化）では、UAX#15にて定義される正規化方式を選択する。

表2 PRECIS Frameworkの文字列変換処理手順

#	変換処理	変換前	変換後
1	Width mapping	ア (U+FF71)	ア (U+30A2)
2	Additional mapping	Delimiter mapping 。	。 (U+002E)
		Special mapping 制御文字 non-ASCII SPACE	削除 ASCII SPACE
		Local case mapping İ	i (U+0069)
3	Case mapping	A (U+0041)	a (U+0061)
4	Normalization	カ [◌] (U+30AB U+3099)	ガ (U+30AC)

文字列の確認処理では、IDNA2008同様にUCDで定義される各文字コードのプロパティ情報に従い、プロトコルで使用可能（PVALID）、使用禁止（Disallowed）、未割り当て（Unassigned）、文字列の構成により使用可能ないし禁止（Contextual Rule Required）等

の文字に分類を行う。変換処理後の文字列はPVALIDな文字コードのみで構成されているか確認を行う。なお、文字の分類方法については、文字コードのプロパティ情報から各文字コードを(A)-(R)までのカテゴリごとに分類し、図2に示す判定方法を適用する。これにより、Stringprepで問題となっていた特定のUnicodeのバージョンに依存している点や禁止文字の確認による本来禁止されるべき文字が禁止文字表に登録されずに使用できてしまう問題を解決している。また、プロトコルによって使用可能な文字列が異なることから、PRECIS Frameworkでは IDENTIFIERCLASS と FREEFORMCLASS の2つのサブクラスを用意している。IDENTIFIERCLASSでは、識別子等で利用する文字列を想定して用意されたサブクラスである。FREEFORMCLASSでは、パスワード等で利用する文字列を想定したサブクラスである。なお、図2中の、ID_DIS or FREE_PVALはIDENTIFIERCLASSでは使用禁止、FREEFORMCLASSでは使用可能となるカテゴリを示している。

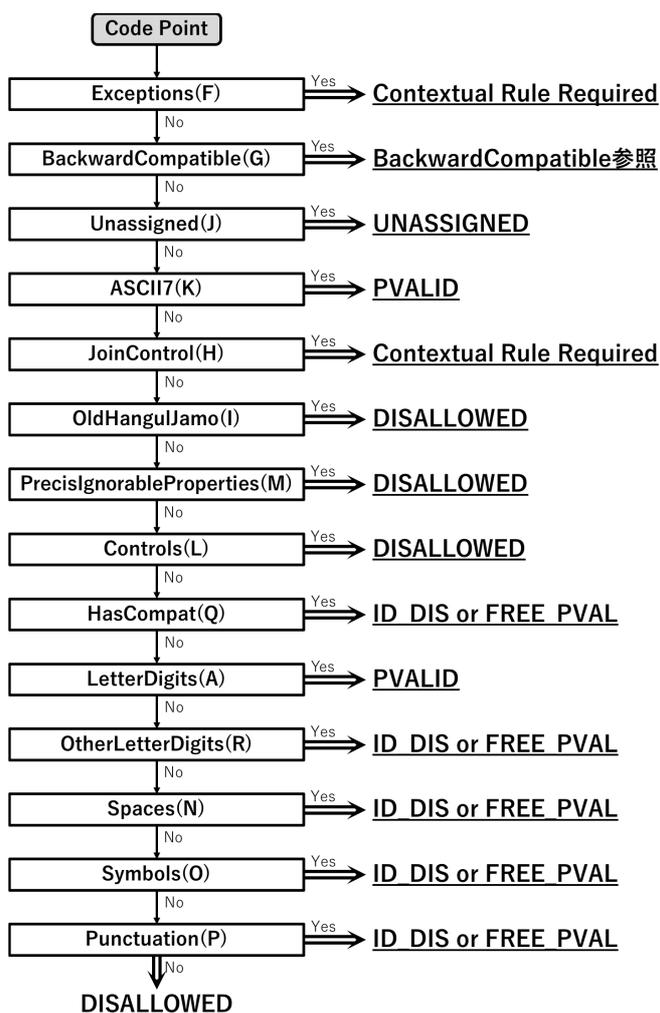


図2 PRECIS Frameworkのカテゴリによる判定方法

また、文字列確認のうちの双方向性の確認方法は、IDNA2008で採用されているRFC 5893 : Right-to-Left Scripts for Internationalized Domain Names for Applications (IDNA) を採用することで、Stringprepでは使用禁止とされていた、モルディブのディベヒ語を表記する際に使用されるターナ文字等が使用可能となっている。

2.4 IETFにおける国際化技術の課題

IETFでは、PRECIS Frameworkの標準化でStringprepによる問題を解決を可能とした。しかしUnicode 7.0以降、IETFで策定したPRECIS FrameworkやIDNA2008等の国際化技術等と整合しないプロパティ情報を持つ文字が収録されたことによる新たな問題が生じている。Unicode 7.0で新たに収録されたARABIC LETTER BEH WITH HAMZA ABOVE (U+08A1)という文字のプロパティ情報が「08A1;ARABIC LETTER BEH WITH HAMZA ABOVE;Lo;0;AL;;;;;N;;;;;」となっており、**図3**に示す通り同一の字形を表すARABIC LETTER BEH (U+0628) およびARABIC HAMZA ABOVE (U+0654)を用いた結合文字列と等価の関係になっておらず、利用者の混乱を生む可能性があることが明らかとなっている。

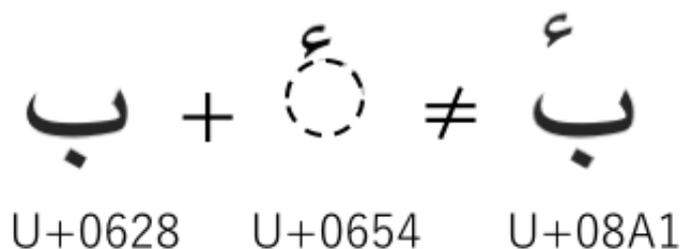


図3 等価とならない文字

IETFのいくつかの Protokol では、識別子として用いる文字列の比較一致の機会を増やすために、文字列変換処理の一環として文字列に対して正規化処理を行うことがある。たとえば、**図4**のLATIN SMALL LETTER A DIAERESIS (U+00E4)であれば、Unicodeのプロパティ情報は「00E4;LATIN SMALL LETTER A WITH DIAERESIS;Li;0;L;0061 0308;;;;;N;LATIN SMALL LETTER A DIAERESIS;;00C4;;00C4」と記述されており、U+00E4とU+0061 U+0308の対応関係が明らかとされおり、同じ字形であるが異なる文字コードとして入力された場合でも、等価の関係にある文字は同一の文字として扱うことが可能となるのだが、このARABIC LETTER BEH WITH HAMZA ABOVE (U+08A1)はそのような等価の関係を示す情報を持たないため、識別子として使用する場合、利用者の文字入力環境によっては意図しない情報資源を参照してしまう問題が生じる。

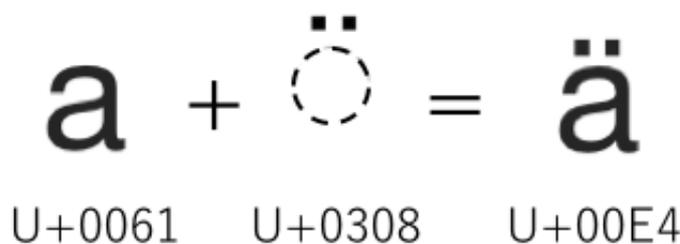


図4 等価となる文字

なお、IETFではUnicode Consortiumと整合した技術提案を継続するための対応について議論がされており、今後はUnicode ConsortiumやW3C等の国際化技術に関連する他の標準化団体との連携も視野に入れ対応方針を検討していくこととしている。

3. IETFが策定するIoT技術のサービスディスカバリ

3.1 IoTサービスディスカバリ要件

IoTサービスには、物流における配送追跡を行うシステムやビルの空調管理システムのように機器からネットワーク、アプリケーションまでをサービス事業者が独自に構築する垂直統合型のサービスや異なるサービス事業者が開発する機器やアプリケーションを相互連携させた水平統合型のサービスがある。水平統合型のIoTサービスは、従来の垂直統合型のサービスと比較してサービス開発コストを低くすることが可能であることから各標準化団体で関連する技術の標準化が進められている。

IETFでは、水平統合型のIoTサービスの中でも、以下の4点において、機器やネットワークに制限があるIoTサービスに焦点を当て、それに係る通信技術の標準化に取り組んでいる。

- IoT機器同士の通信 (図5)
- IoT機器とクラウド間の通信 (図6)
- IoT機器とゲートウェイ間の通信 (図7)
- バックエンドでのデータ共有の通信 (図8)

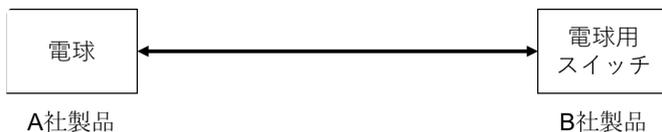


図5 IoT機器同士の通信

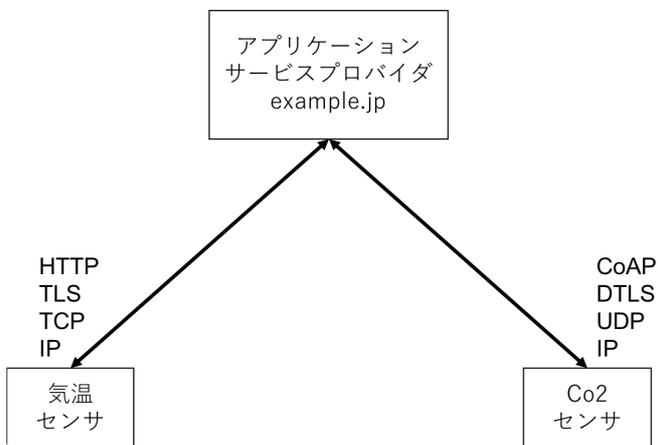


図6 IoT機器とクラウド間の通信

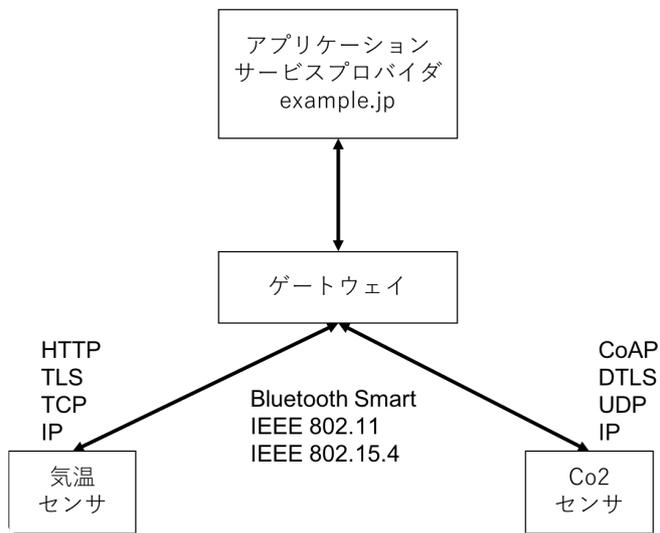


図7 IoT機器とゲートウェイ間の通信

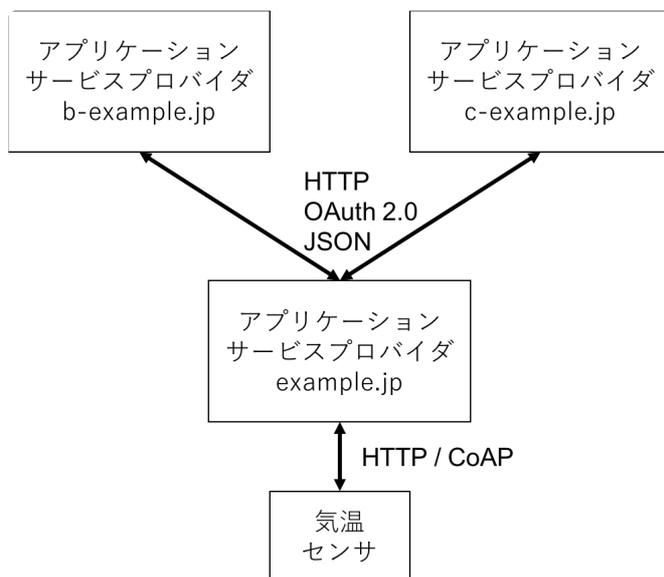


図8 バックエンドでのデータ共有の通信

この制限があるIoTサービスとは、IETFではConstrained RESTful Environments (CoER)と呼ばれており、バッテリーの駆動時間やCPUの処理能力、メモリ量、無線通信速度等が制限された環境下を想定している。

たとえば、必要な情報を効率的に交換を行うためのRESTfulアクセス可能なアプリケーション技術の標準化がcore WGにて行われている。RFC 7252 : The Constrained Application Protocol (CoAP)を主要プロトコルとして、その他の拡張プロトコルとして、エンドポイント・リソースの発見方法、IoT向けの暗号化通信等について標準化を行っている。また、IPv6適合レイヤーと制限付き無線リンクに適したヘッダ圧縮の定義や制限付きノード・ネットワークのルーティング・プロトコルの定義等に取り組むWGもある。

特にIETFが行うIoTサービスディスカバリには、図6, 7, 8のドメイン名を持つグローバルな領域とそれ以外のローカルな領域によって求められる要件が異なってくる。グローバルな領域においては既存のDNSによる名前解決が求められており、一方で、ローカルな領域においては利用者がホスト名として任意に命名した名前を利用することが想定されている。

3.2 IETFにおけるWG (dnssd)

DNS-Based Service Discovery (DNS-SD) は、Apple社の Bonjour 等を礎に RFC 6762 として標準化された mDNS を使用し、IP アドレスやホスト名を知らなくても同一ネットワークセグメント内のサービスを発見する手法である。DNS-SD の拡張を行う WG である dnssd WG では、これを複数のネットワークセグメントに拡張するための手法が提案されている。

mDNS では、「.local」 Top Level Domain (TLD) とホスト名の組合せを用いて名前解決を行い、ホスト名に該当するコンピュータが名前解決クエリを受信すると自身の持つ IP アドレスを返す仕組みとなっており、ホスト名には UTF-8 の使用が許容されている。

また、DNS-SD では、DNS による名前解決でサービスを特定可能とするための構造化されたサービスインスタンス名を標準化している。DNS-SD での定義されているサービスインスタンス名は SIP 等で使用される SRV RR と同様の形式で「<Instance>.<Service>.<Domain>」と記述する。たとえば、aoyamagakuin.jp の http サービスは「_http._tcp.aoyamagakuin.jp」となる。このサービスに対応するサービスインスタンス名を PTR レコードに記述し、サービスへのアクセスには SRV RR を使用することとなっている。たとえば、「_http._tcp.aoyamagakuin.jp RTP OfficialWebPage. _http._tcp.aoyamagakuin.jp.」とすると、「 OfficialWebPage. _http._tcp.aoyamagakuin.jp SRV 0 100 80 www.aoyamagakuin.jp.」が返される。

この mDNS および DNS-SD を組み合わせることにより、同一ネットワークセグメント内において、サービスの発見を可能としている。たとえば、プリンタで印刷をしたい場合、プリントサービスを _ipp とした場合、「_ipp._tcp.local」として候補となるプリントサービスを発見することが可能となる。

この仕様を複数ネットワークセグメントで対応させるための主要プロトコルとして draft-ietf-dnssd-hybrid-06 : Discovery Proxy for Multicast DNS-Based Service Discovery が提案されている。この I-D では、ネットワークセグメントごとにドメイン名を用意し、DNS および mDNS プロキシとしての機能をディスカバリプロキシとしてルータに追加するものである。たとえば、ネットワークセグメント A からネットワークセグメント B のサービスを使用する場合、「_ipp._tcp.netB.aoyamagakuin.jp」というクエリをディスカバリプロキシが受信すると、そのクエリをネットワークセグメント B に対して「_ipp._tcp.local」クエリを送信し、その応答クエリをネットワークセグメント A に返すことになる。

3.3 IETFにおけるWG (core)

core WG では、機器やネットワークに制限がある環境下で、必要な情報を効率的に交換を行うための RESTful アクセス可能なアプリケーション技術の標準化が core WG にて行われている。RFC 7252 : The Constrained Application Protocol (CoAP) を主要プロトコルとして、その他の拡張プロトコルとして、エンドポイント・リソースの発見方法、IoT 向けの暗号化通信等について標準化を行っており、その中の提案技術の一つに国際化技術を必要とする draft-ietf-core-rd-dns-sd-01 : CoRE Resource Directory: DNS-SD mapping という、既存の DNS インフラストラクチャを使用してサブドメイン内の CoAP サーバ等のサービスを検索するための PTR および SRV, TXT RR の記述方法を定義する提案がある。この提案では、draft-ietf-core-resource-directory-12 : CoRE Resource Directory と呼ばれる、スリープサーバや帯域制限のある環境下における機器のリソース情報を広域に展開するために、Resource

Directory (RD) というエンティティが制限された環境下のリソース情報を集約し、リソース検索可能とするために定義されたリソース情報の記述方法をDNS-SDにマッピングする方法が提案されている。

DNS-SD mappingでは、リソースインスタンスの「ins」属性をDNS-SDサービス名の<Instance>部分に、「rt」属性を<Service>部分に、「d」属性を<Domain>部分にマッピングすることとしている。

リソースインスタンスの「ins」属性を<Instance>部分にマッピングする際に、RFC 5198として標準化されたNet-Unicodeに従い正規化（NFC）を行うと明記がされている。

4. 国際化技術の観点から見るIoTサービスディスカバリの課題

4.1 dnssd WGにおける課題

dnssd WGのサービスディスカバリである、draft-ietf-dnssd-hybrid-06 : Discovery Proxy for Multicast DNS-Based Service Discoveryでは、UTF-8を許容し、機器のホスト名を識別子として利用するため、IDNA2008やPRECIS Framework等を適用する必要がある。しかし、そのような国際化技術についてはこのI-Dやそれが参照するRFC 6762やRFC 6763でも、触れられていないため、適切な文字列変換や確認を処理を経ていない利用者が任意に決めた文字列が識別子として利用される場合、このアプリケーション利用者は意図しないサービスを参照してしまうという利便性および安全性の観点から課題がある。この課題の影響を受ける言語圏としては、見かけ上ほぼ差異のない合成済み文字や結合文字列を扱う日本語やヨーロッパの言語圏等が想定される。なお、この合成済み文字や結合文字列に関する他の問題はWebブラウザ等の汎用アプリケーションから直接サービスインスタンス名指定してサービスを使用しようとするとサービスに到達できないという課題もある。DNS-SD用のアプリケーションを通じてサービスを指定していく場合は、サービスへの到達は可能であるが、直接サービスインスタンス名を入力しようとした場合は、利用者の文字入力環境によっては、合成済み文字もしくは結合文字列のどちらかでしか入力できない場合がある。またこの問題は全角文字や半角文字においても同様であるほか、制御文字等の識別子としての利用が不適切な文字についても、0x00-0x1Fおよび0x7FのASCII制御文字については禁止としているが、それ以外の制御文字については禁止されていないためU+202EのようなRight-to-Left Override (RLO) 制御文字を含むことが可能となってしまう悪意あるインスタンス名を容易に作成できてしまう。

一方で、DNS-SDを実装するアプリケーション側に文字列処理機能を持たせた場合、本来は正しい情報資源であってもアプリケーション上で表示される、サービスインスタンス名の<Instance>部分が利用者の意図した文字列と異なって表示される可能性があるため、利用者の混乱を招く。また、DNS-SDを実装するアプリケーション側で、IDNA従い文字列処理を行うようにした場合、その文字列処理がIDNA2008に従った上でU-ラベルを保持する必要があるほか、IDNA2008の日本語ドメイン名を処理できるWebブラウザでは、IDNA2003との互換性を考慮し、UTS #46 Unicode IDNA Compatibility Processingで定められている句点(.)をドット(.)に変換する処理を実装可能としているため、ホスト名に句点が含まれている場合は名前解決が失敗することが生じるという実装依存の課題もある。

そのため、このようなホスト名を識別子として利用する場合は、本来ホスト名の設定時点で利用可能な文字の確認や文字列変換処理を行うことが上記の問題の解決に効果的な一つの方法ではあるため、そのような統一された命名規則を標準化する必要がある。

4.2 core WGにおける課題

core WGでは、提案技術の一つに国際化技術を必要とするdraft-ietf-core-rd-dns-sd-01: CoRE Resource Directory: DNS-SD mappingという、既存のDNSインフラストラクチャを使用してサブドメイン内のCoAPサーバ等のサービスを検索するためのPTRおよびSRV, TXT RRの記述方法を定義する提案がある。この提案では、draft-ietf-core-resource-directory-12: CoRE Resource Directoryと呼ばれる、スリープサーバや帯域制限のある環境下における機器のリソース情報を広域に展開するために、Resource Directory (RD) というエンティティが制限された環境下のリソース情報を集約し、リソース検索可能とするために定義されたリソース情報の記述方法をDNS-SDにマッピングする方法が提案されている。

DNS-SD mappingでは、リソースインスタンスの「ins」属性をDNS-SDサービス名の<Instance>部分に、「rt」属性を<Service>部分に、「d」属性を<Domain>部分にマッピングすることとしている。

リソースインスタンスの「ins」属性を<Instance>部分にマッピングする際に、RFC 5198として標準化されたNet-Unicodeに従い正規化(NFC)を行うと明記がされているため、4.1節における課題にて指摘した、合成済み文字や結合文字列に関する問題は起こらない。しかし、正規化(NFC)では、日本語の全角カタカナや半角カタカナの変換はされないため、日本語対応に課題がある。この課題に対しては、正規化(NFC)に加えて、別途、UCDのUnicodeData.txt中のDecomposition_TypeおよびDecomposition_Mappingに従い、文字幅に関する変換処理を実施することで解決可能である。また、Net-Unicodeでは、0x00-0x1Fおよび0x7FのASCII制御文字については禁止としているが、それ以外の制御文字については禁止されていないため、dnssd WGでの指摘同様に入力された文字列が利用可能な文字であるか確認する処理が必要である。

なお、「rt」属性を<Service>部分にマッピングすることから、「rt」属性では、アンダースコア(_)およびピリオド(.)の仕様が禁止されており、<Service>部分にマッピングする際には、「rt」属性の文字列の前にはアンダースコア(_)を付加し、後ろには、「.udp」を付加すると提案されているため、非ASCII文字でアンダースコア(_)およびピリオド(.)と対応関係のある文字の変換や使用禁止も必要となる。

4.3 IoTサービスディスカバリの識別子におけるその他課題

4.1節および4.2節では、主に識別子を構成する文字コード列の比較に際した課題を指摘した。一方で、すでに運用されている国際化ドメイン名の事例等から文字コード列以外の国際化文字列を含む識別子に関する課題が指摘されている[21]。

たとえば、利用者にとって視覚的に同じ文字列に見える同形異義語問題がある。これは、利用者が意図している識別子の一部を視覚的に似た文字に置き換えることで、利用者が意図していない情報資源を参照してしまうという問題である。この問題は小文字の「l」と大文字の「i」がArial フォント等で同じ字形に見えるフォントを用いた際に生じる国際化ドメイン名が誕生する前から存在していた問題である。しかし、国際化ドメイン名が誕生したことにより、ラテン文字

とキリル文字のように字形が似た文字を混在させたドメイン名が仕様上作成可能となっており、同形異義語問題による影響が増してしまうため、gTLDレジストリは、このようなリスクを制限するための登録ポリシーの策定を行い現在は運用によりこの問題への対処が行われている。

一方で、本稿で述べた標準化技術の中には、利用者が任意でつけたホスト名を識別子として利用しようとしているものもあるため、前述のような登録ポリシーが及ばないことが想定される。そのため、利用者が誤った情報資源を参照しないようにする一手法として、IoTサービスのサービスディスカバリ用アプリケーションに、異なる文字セットが混在するホスト名に対してセキュリティ警告を表示し、利用者に接続するか否かの選択肢を与える機能を実装する方法がある。

また、ホスト名を識別子として登録する際にPRECIS Frameworkにより文字列を変換および確認した上で登録するという手法も、利用者が意図しない情報資源に接続しないようにする手段として有効である。たとえば、PRECIS Frameworkの文字列確認処理では、制御文字を禁止文字として処理するため、左から右に表記する識別子中にRLO (Right-to-Left Override) 制御文字という右から左へ表示させる制御文字を用いて文字コード上は異なるが見かけ上同じ識別子を構成することを防ぐことが可能となる。また、国際化ドメイン名で課題となった、中国の簡体字や台湾の繁体字のように同一の文字として見なしたい異体字や、等価な意味を持つ異なる文字列としてアメリカ英語と"theater"とイギリス英語の"theatre"等同義語についても、PRECIS FrameworkのAdditional mappingで同義語や異体字等の対応関係を定義したファイルを作成し、それらを使用するプロファイルを設計することも必要と考えられる。

5. 標準化活動を通じて得た知見

現在IETFでは、PRECIS FrameworkがRFC 8264として標準化され、それを利用するプロトコルがRFC 8265およびRFC 8266として標準化されたことから、IETFにおける国際化技術に関する主要プロトコルの策定は完了し、今後はStringprepを使用していたプロトコルのPRECIS Frameworkへの移行作業や新たに策定されるプロトコルでのPRECIS Framework使用に関する検討作業が行われる段階となっている。新たに策定されるプロトコルでのPRECIS Framework使用に関する検討については、プロトコルデザイナー向けのガイドラインがRFC 7790として発行されている。一方、Stringprepを使用していたプロトコルのPRECIS Frameworkへの移行については指針となるものが提案されていないことから、国際化技術移行における互換性を整理したガイドライン等の提案が、国際化技術の普及の面から求められている。特に、UTF-8を許容するRFCないしI-Dのいくつかは、RFC 5198にて説明されているNet-Unicodeという文字列処理方法を引用しており、Net-Unicodeでは、正規等価性を保持する文字列処理方法として正規化（NFC）を推奨しているが、特に日本語においては正規化（NFC）を利用する際には、4.2節の core WGにおける課題で述べた通り、Width mappingと組み合わせると利用しないと利用者の意図しない情報資源を参照してしまうという利便性および安全性の課題がある。そのため、Net-Unicodeではなく、PRECIS Frameworkのような文字列変換処理を使用する提案も必要である。

この問題は、筆者がPRECIS Framework策定時にも提案してきたことであるが、この全角・半角問題の影響を受けるのは日本や一部の言語圏の国のみであり、世界的に見るとWidth mappingがないことによる影響を受ける国や地域はほぼない。そのため、実際にその文字を利用する言語圏の人が標準化会議に出席し、提案を行っていくことが我々にとっても使いやすい標準技術を策定していく上で重要となる。また、Width mappingについてはIDNA2008の前例が

あったことから、PRECIS Frameworkへの採用については参加者間での合意形成が容易であったが、筆者がもう一つ提案していたAdditional mappingのLocal case mappingについては、インターネットの相互運用性の観点から議論が難航した。Local case mappingは、ドイツ語のエスツェットやギリシャ語のファイナルシグマ、トルコ語のドット無しi、ドット付きI等の他の言語圏の変換規則とは異なった文字変換規則を持つ文字のための文字種変換処理である。これらの文字に対して一般的な文字種変換としてCase mappingした場合、その言語圏の利用者が意図した文字変換が行われないという問題が生じるため、IETFでもこの変換処理は重要であるとの認識はあったものの、Local case mappingを採用するとその言語圏以外の利用者にとって意図しない文字列変換が行われるという問題が生じるとのことで、一時はインターネットの相互運用性を重要とし、Local case mappingを除外する話もあった。ただし、PRECIS Frameworkの処理順序を見直し、Local case mappingを実施する順番を精査した上で、Local case mappingについては、オプションで選択可能ということで合意を得た。なお、この合意を得る際に、precis WGの参加者にLocal case mappingの対象となるすべての文字について精通している人がいなかったため対象文字の利用事例等を説明できるように準備を行う必要があり、筆者も標準化提案を通じて初めて知ることとなった言語について説明する際に大学の留学生等やUnicode Consortiumのオープンフォーラム等には大変助けられた。そのため、他国に知人がいない人がこれから国際化に係る活動を行う際に、他国の文字についても触れようとした場合、留学生の多い大学と共同研究等の体制を組めたり、言語の専門家が集まるUnicode Consortiumのオープンフォーラム等を利用したりすると研究や標準化活動が円滑に進むことが分かった。

6. おわりに

本稿では、IETFが行っている標準化活動のうち筆者が参加してきた国際化技術とその活用が期待されるWGで得た知見をもとに、IETFが策定する国際化技術とその活用が期待されるIoT技術国際化の課題について論じた。

本稿で述べたIoT技術には国際化文字列を使用するいくつかの提案技術には課題があることが分かっている。特に、IETFで策定するIoT技術の中でも主要プロトコルとなるCoREとそれをDNSインフラストラクチャを利用して情報資源発見を行うDNS-SDについては、情報資源名に利用者が任意につけた国際化文字列が含まれることから、日本語圏や他の英数字以外の文字を使用する言語圏の利用者にとっても便利かつ安全に利用可能なIoT技術を標準化するために、国際化の観点から改善提案を行う必要もある。

なお、2018年3月2日にprecis WGを含む国際化技術およびDNSに関するWGのメーリングリストにUnicode 7.0以降の問題の解決について検討を行う専門家チームを設置することをICANNが計画しており、今後、国際化技術に関する議論が活発化する見通しである。また、IETFのメーリングリストではこれら国際化に関する問題を解決するために、英語を母語とする人以外の参加も必要としていることが話題に上がっており、この問題解決には日本からの貢献も求められる分野となっている。

将来、日本をはじめとしたさまざまな国において安全かつ便利に利用できるIoTサービスの普及には、英語圏以外の言語圏の人々による標準化活動への参加が求められている。

国際化技術の標準化は日本からの貢献も求められている

謝辞 本稿は、一般社団法人情報通信技術委員会「IoT/BD/AI時代に向けたデジュール及びフォーラム標準に関する標準化動向調査」による支援を受けたものです。深く感謝いたします。

参考文献

- 1) IETF: Internet Engineering Task Force, <https://www.ietf.org/> (2018年5月7日現在).
- 2) Klensin, J.: Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework, RFC 5890 (2010).
- 3) Klensin, J.: Internationalized Domain Names in Applications (IDNA): Protocol, RFC 5891 (2010).
- 4) Faltstrom, P.: The Unicode Code Points and Internationalized Domain Names for Applications (IDNA), RFC 5892 (2010).
- 5) Alvestrand, H. and Karp, C.: Right-to-Left Scripts for Internationalized Domain Names for Applications (2010), RFC 5893 (2010).
- 6) Resnick, P. and Hoffman, P.: Mapping Characters for Internationalized Domain Names in Applications (IDNA) 2008, RFC 5895 (2010).
- 7) Klensin, J. and Ko, Y.: Overview and Framework for Internationalized Email, RFC 6530 (2012).
- 8) Yao, J. and Mao, W.: SMTP Extension for Internationalized Email, RFC 6531 (2012).
- 9) Yang, A., Steele, S. and Freed, N.: Internationalized Email Headers, RFC 6532 (2012).
- 10) Hansen, T. Ed., Newman, C. and Melnikov, A.: Internationalized Delivery Status and Disposition Notifications, RFC 6533 (2012).
- 11) Resnick, P., Newman, C. and Shen, S. (Ed.): IMAP Support for UTF-8, RFC 6855 (2013).
- 12) Gellens, R., Newman, C., Yao, J. and Fujiwara, K.: Post Office Protocol Version 3 (POP3) Support for UTF-8, RFC 6856 (2013).
- 13) Fujiwara, K.: Post-Delivery Message Downgrading for Internationalized Email Messages, RFC 6857 (2013).
- 14) Gulbrandsen, A.: Simplified POP and IMAP Downgrading for Internationalized Email, RFC 6858 (2013).
- 15) Cheshire, S. and Krochmal, M.: DNS-Based Service Discovery, RFC6763 (2013).
- 16) Lynn, K., van der Stok, P., Koster, M. and Amsuess, C.: CoRE Resource Directory: DNS-SD mapping, draft-ietf-core-rd-dns-sd-01 (2018).
- 17) Saint-Andre, P. and Blanchet, M.: PRECIS Framework: Preparation, Enforcement, and Comparison of Internationalized Strings in Application Protocols, RFC 8264 (2017).
- 18) Saint-Andre, P. and Melnikov, A.: Preparation, Enforcement, and Comparison of Internationalized Strings Representing Usernames and Passwords, RFC8265 (2017).
- 19) Saint-Andre, P.: Preparation, Enforcement, and Comparison of Internationalized Strings Representing Nicknames, RFC 8266 (2017).
- 20) Yoneya, Y. and Nemoto, T.: Mapping Characters for Classes of the Preparation, Enforcement, and Comparison of Internationalized Strings (PRECIS), RFC 7790 (2016).
- 21) 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/lgr-procedure-20mar13-en.pdf> (2013年3月20日現在) .

根本 貴弘 (正会員) nemoto@aim.aoyama.ac.jp

博士（メディアデザイン学）。慶應義塾大学メディアデザイン研究科後期博士課程修了。2014年 青山学院大学附置情報メディアセンター助手として着任し現在に至る。計算機ネットワークにおける国際化された識別子および認証技術とその活用に関する研究に従事。2011年よりIETFにおける標準化活動にも参加。RFC 7790共著者。

投稿受付：2018年5月7日

採録決定：2018年9月5日

編集担当：佐藤三久（理化学研究所）