

議会情報の利用促進 とWebの多言語表示

—アジア太平洋議員フォーラム「APOINT 2001」について

山内康英・小町祐史・松下周二
長村 玄・大久保彰徳・上村圭介



▶ 1 はじめに

アジア・太平洋議員フォーラム (Asia Pacific Parliamentary Forum: APPF) は、国際社会の相互理解の推進を目的として、アジア・太平洋地域の議員が1993年に設立した国際組織である。アジア・太平洋議員フォーラムは、情報技術とりわけコンピュータネットワークが、21世紀の産業社会に与える影響の大きさを考え、1998年からインターネットを利用した立法府間の情報交流プロジェクトに着手した。

国際大学グローバル・コミュニケーション・センター (GLOCOM) が主催する「APFont研究会」は、アジア・太平洋議員フォーラム日本事務局¹⁾と協力して、1997年以降、議会情報の利用促進とWebの多言語表示について調査研究を行なっている。この活動は、国際的な政治組織がインターネット利用の推進とその技術的な問題を取り上げた、という点でユニークであり、国際政治における情報技術の利用と言語の問題について一つのケーススタディになる、と考えられる。本稿では、このアジア・太平洋議員フォーラムの活動とAPFont研究会の研究内容について紹介したい。²⁾

▶ 2 アジア・太平洋議員フォーラムと多言語環境

(1) APOINT2001

アジア・太平洋議員フォーラムの参加国は、2000年11月現在、26カ国³⁾になっており、毎年1月に持ち回りで年次会合を開いて共通課題について政策協議などを行なっている。一般に国家間の外交は、主として行政府（内閣や大統領府）が担当し、代表民主制によ

脚注

1. アジア・太平洋議員フォーラム日本事務局は現在、中曽根事務所が運営している。本研究会の活動にご支援を頂いているアジア太平洋議員フォーラムのメンバー議員各位、とりわけ中曽根康弘会長、衆議院、参議院、外務省の担当者各位、中曽根事務所の中田茂秘書に御礼申し上げたい。

2. 本稿の元になった「99年度APFont研究会報告書」の原文を、APPF日本事務局のWebに掲載している。
<http://www.glocomnet.or.jp/appf/>

って国民が選出する立法府（国会や議会）のメンバーは、通常は条約の批准や、予算の審議などを通じて外交に参画している。他方で、実際の外交過程で、立法府メンバー（衆・参両院，上・下両院等の議員）の個人的なネットワークが重要な役割を果たすことがあり、また、選挙区の人々と直接、接する機会の多い議会人の「国際感覚」は、二国間の外交関係などにとっても重要である。このような議員間の相互交流を目的とする組織としては、1889年創設のInter-Parliamentary Union (<http://www.ipu.org/>) などが有名である。

既述のように、アジア・太平洋議員フォーラムは、情報技術とりわけコンピュータネットワークが、21世紀の産業社会に与える影響の大きさを考え、1998年からインターネットを利用して立法府間の情報交流を促進する計画に着手している。1998年1月のAPPF総会（韓国）は、特に決議を採択し、この計画を「アジア・太平洋オープン・インフォメーション・ネットワーク（Asia Pacific Open Information Network: APOINT 2001）」と名付けた。1998年9月の執行委員会（ペルー）では、「APPF技術ワーキング・グループ」の会合を開いて、「APOINT 2001プロジェクト」の進捗状況について討議した。この中で、各国が分担して行動計画を策定し、2001年にレビューを行なうことが決まった。具体的に取り上げるようになった課題は以下のようなものである。

- ・ 議員間の電子メールによる情報交換のためのアドレス・ディレクトリー作成
- ・ APPFスタッフ間の事務作業の電子化
- ・ 議会Webを持たないメンバー国への支援／各国APPF事務局のWebの開設
- ・ APPFホームページの設置と運営
- ・ Web表示の多言語化
- ・ 各国議会情報の部分的な英語翻訳および多言語検索機能の導入

ペルー議会は、APPFメンバー国のうち議会がWebサイトを持たない国を支援する計画に取り組むことを決め、1998年以降、フィジー、ベトナムなどの議会と協力して、個別に議会Webサイトを立ち上げた。この結果、ラオス、マーシャル諸島などを除くAPPFメンバー国の議会は、何らかの形でWebサイトを持つようになっている。

APPF日本事務局は、APOINT 2001プロジェクトの一環として、ネットワークの多言語環境の整備を担当することを決め、国際大学グローバル・コミュニケーション・センターが主催するAPFont研究会は、「APPF技術ワーキング・グループ」と協力して、ネットワークの多言語環境について調査研究を行なった。⁽⁴⁾

(2) 1998年度の活動

アジア・太平洋各国の議会が、インターネットを通じて提供する情報の内容は急速に高度化している。たとえば日本の衆議院では、本会議と主要な委員会について、Real Videoを使ったインターネットのライブ放送を行なうようになった。⁽⁵⁾

このような議会の情報を相互に参照する場合には、技術標準やソフトウェアの共通資

脚注

3. タイ、インドネシア、マレーシア、フィリピン、シンガポール、ベトナム、ラオス、カンボジア、日本、韓国、中国、ロシア、米国、カナダ、メキシコ、チリ、ペルー、オーストラリア、ニュージーランド、パプアニューギニア、モンゴル、コロンビア、マーシャル諸島、ミクロネシア、フィジー、ブルネイ（オブザーバー）

4. APFont研究会には、このほかに林紘一郎（慶応義塾大学教授）、

檜山正幸（檜山オフィス代表）、宮澤信一郎（秀明大学教授）、小山裕司（GLOCOM講師）、海田茂（ネクストソリューションズ主任研究員）などのメンバーが参加している。

5. 記録のために付記すれば、11月20日から21日早朝にかけて森政権の不信任決議案をめぐる衆議院が紛糾したが、衆議院TVは21日4時過ぎまで本会議の様子をインターネットで実況中継していた。

源を前提としているが、なかでもそれぞれの「文字」が問題となる。コンピュータのシステムは、各国言語にローカライズしており、相手国の議会の異なる文字環境をパーソナル・コンピュータの上で再現することは必ずしも容易ではない。

このためにAPFont研究会では、1998年度の活動として、メンバー各国の言語仕様のコンピュータ上で、各国の言語を相互に表示する多言語表示ツールについて、1999年1月にリマで開かれたAPPFの総会でデモを行なうための準備に着手した。この多言語表示用ツールは、Windows 95のブラウザ（具体的にはNetscape Ver.3とExplorer Ver. 2）をベースにし、その上で、他言語表示への切り替えを行なうもので、以下の内容からなっていた。

- ・各国の工業標準に基づいたフォント各1種類
- ・OSの文字コードと表示言語文字コードとの変換テーブル
- ・ブラウザと組み合わせて使用する表示切り替えツール

以下では、このブラウザ上での言語表示の切り替えツールの開発などをめぐる1998年の調査研究について紹介し、多言語環境を考える一助としたい。

(3) アジア言語のフォント：文字の迷宮？

よく知られているように、東アジアの主要な言語——中国語、日本語、朝鮮語——は、2バイトコードを使用している。ASCIIに代表される1バイトコードは、ギリシャ・ラテン文字に由来するアルファベットの処理には充分であっても、東アジアの言語が日常用いている数千の文字を扱うには、全く無力である。

実際に膨大な字数を持つ漢字（中国では Han-Zi、朝鮮語では Hanja）の言語処理は、一筋縄では行かない。また、日本語処理では、ひらがなおよびカタカナと、6000字以上の漢字を取り扱わなくてはならない。同様に、中華人民共和国とシンガポールが使用するスタンダードの「国標」文字セット GB 2312-80 には、文字数が7445個存在する。（その内、3755字がレベル1の漢字、3008字がレベル2の漢字、残り682字が非漢字である。）また、主に台湾と香港で使用する事実上の標準は、Big5であるが、これは、5401字のレベル1漢字、7652字のレベル2漢字、470字の非漢字を含んでいる。また、韓国の標準文字セットは、KSC 5601-1992と呼ばれており、その内訳は、2350字がレベル1のハンゲル、4880字のレベル2漢字、986字の非ハンゲル・非漢字になっている。

近年、マイクロソフト社などが後押しして普及に動いているユニコード（ISO-10646）は、国際標準の主流として無視できない。ユニコードは、混沌を極める日本語、中国語、韓国語の漢字の世界を統一するひとつの試みとして、たしかに注目に値する。非漢字ユーザーが、コードの全文字数を節約するために、日本語・中国語・朝鮮語の漢字をまとめて扱うことが可能だと考えるのは、当然のなりゆきである。他方でユニコードは、いわゆる異体字などにとまなう画や点の位置を無視しているために、繊細な漢字ユーザーの文字感覚を逆なでしている感がある。このためにユニコードは、東アジアのユーザーにはどうしても賛同を得られないでいる。

(4) 日本語文字セットのコーディング

日本語で使用される文字セットは、JIS X 0208-1990であるが、このJISレベル1には2965字の漢字、レベル2には3390字の漢字が定義されている。この他、524字の非漢字、および1バイトの文字（94字のASCII文字、63字の半角カナ、34のコントロール記号）が含まれている。このような膨大な文字数に加えて、日本語処理を混乱させているのが、多数のコーディング体系である。現在では、もっとも代表的な体系は一つではなく、

EUC-JP, JIS, シフトJISという3種類が並立して拮抗している。

この中でも、ISO-2022-JPとも呼ばれるJISコーディングは、日本語のインターネット電子メールでの標準という立場にある。EUC-JPは、ほとんどのUNIXシステムで使用されている。しかし最も広範囲で使用されているのはシフトJISで、ウインテルであれアップルであれ、ほとんどのパソコンではこのコーディングが主力になっている。実際、日本語のWebサイトの大多数にはシフトJISが用いられており、JISやEUCコーディングのページを設けているサイトは比較的少数にとどまっている。Netscape Navigator や Internet Explorerなどの主要なブラウザは、この日本語の複雑なコーディング環境にある程度、自動的に対処する機能を備えるようになった。しかしながらブラウザが国境を越えて、異なる文字セットを持つWebを読みに行った場合、OSがそのコード体系に対応していない場合には、多くの場合、何らかの手動による対応が必要である。

▶ 3 多言語Web環境のユーザ要求

以上のようなWebの多言語環境についてのユーザ要求を満たすために、コンピュータは次のような機能を備えなければならない。

(1) 多言語文字の表示機能

Webのフロントエンド・アプリケーションであるWebブラウザが、多言語文字を表示する機能を備えていること。

(2) 多言語文字の表示でブラウザが使用するフォント・リソース

ブラウザは、文書中の文字形をフォント・リソースの中から探し出し、画面上に再構成する。したがって一般の利用者が、多言語環境でブラウザを使用するためには、必要なフォント・リソースを無料で配布しなければならない。アルファベット（いわゆる英数字やキリル文字）は文字数が少ないために、インターネット上で入手可能な多様なフォント・リソースを持つ。他方で、中国語、日本語、韓国語の文字などのフォント・リソースの入手は、依然として有料である。その理由は、中国語、日本語、韓国語の文字は多数のフォントを持つため、労働集約的なフォント作成が高コストになるためであろう。

アジア太平洋議員フォーラムの情報交流に際しては、メンバー国全ての公的文書または国家文書をカバーできるように、少なくとも各一組のフォント・リソースが必要になる。われわれは将来的には、そのようなフォント・リソースが、各国の議会から各国の言語情報に添付する形で無料で配布されるべきだと考えた。このような「国際フォント環境」によって、インターネットの利用者は、ネットワークを通じた「フォント降下 (font precipitate)」状態、つまり必要に応じて欲するところのフォントがネットワーク経由で随時入手できる状態に近づくであろう。

(3) 外字や他言語の文字の表示機能

多言語表示で忘れてはならないのが、工業規格の文字表外の文字や販売会社が独自で作成した文字、いわゆる「外字」である。中国、日本、韓国で使用されている2バイトの文字は、継続的に新しい文字や記号が追加されるために、いわば「開かれた集合」を成している。言い換えれば、特定のスピーチコミュニティで使用される全ての文字を含む、有限の文字集合は定義不可能であり、インターネット全体としてこの問題を解決す

るための仕組みが必要になる。

(4) 多言語文字の入力機能

コンピュータ上では、ある言語の表示と入力とは別個の機能である。ベンダーはOSをローカライズした段階で、キーボード入力に特定の配列を割り付けている。したがって画面上の他言語表示に対して、検索や質問などを入力する場合には、画面からアルファベット入力を行なうなど、フィードバックチャネルの確保のためには別個の機能が必要になる。

(5) 整形済み文書を発信するためのシステム：「ポータブル形式」

議会文書の多くは、プリントアウト・イメージに相当する整形済み文書になっている。このような文書については、既述のようにフォントをOSに依存するのではなく、Adobe社のPDFファイルのように、個々の文章ファイル自体にフォントの一部を内蔵させる方法（フォント埋め込み型のポータブルドキュメント・ファイル形式）で多言語環境を実現することができる。このシステムが普及すれば、ブラウザ上で多言語機能が提供されない環境でも、オリジナル言語で発信した文書をプリントアウト・イメージで読むことができる。なお、この「フォント・ポータブル方式」（後述する「サブシステム②」）と、既述の「フォント降下方式」（以下では「サブシステム①」とする）は、ネットワークで多言語環境を実現するための二つの実際の方式であり、両者は、各々の得失に併せて使い分けるのが便宜である。⁶⁾

(6) 議会情報を交換するための共通形式

議会情報発信のための新しい形式としてXMLは有効である。また、XMLの能力を最大限利用するために、文書作成ツールや文書型定義（DTD）が必要である。XMLはHTMLに代替する技術であり、インターネット上での情報の配布および改訂に使用されている標準技術である。

(7) 多言語文書の配布モデル

多言語文書の配布においても、以下のような電子化文書の三形態を考慮に入れる必要がある。

- (1) 改訂可能な論理的文書（以下、形態(1)文書）
- (2) 整形情報を伴う改訂可能な文書（形態(2)文書）
- (3) 整形済み文書（形態(3)文書）

特に、文書が分散的に作成される場合、形態(1)文書と形態(2)文書の交換が不可欠になる。HTMLまたはSGML/XMLは、形態(1)文書の交換を十分にサポートできる。⁷⁾

定義済みの整形指定を含むHTML処理システムに対して、HTMLは形態(2)文書の交換をある程度サポートできる。より複雑な整形指定が要求される場合は、XSLまたはDSSSLの使用が望ましい。⁸⁾

形態(3)の文書は、実装されているPDFまたはPDLで記述可能である。PDF文書を使用する場合、圧縮データで配布することが可能である。⁹⁾

脚注

6. グローバルフォントサーバの構想については以下を参照。石川直太、檜山正幸、小町祐史「SGML文書処理系におけるフォント資源情報交換方式」
http://www.y-adagio.com/public/confs/iieej_ac/ac25_23/ac25_23.htm
7. HTML：ハイパーテキストマーク付け言語、SGML：標準一般

化マーク付け言語、XML：拡張可能なマーク付け言語
8. XSL：拡張可能なスタイル言語、DSSSL：文書スタイル意味および指定言語
9. PDF：ポータブル文書ファイル、PDL：ページ記述言語

▶ 4 システム実装案とデモンストレーション

(1) サブシステム

本システムは、上記(2)整形情報を伴う改訂可能な文書（下図「サブシステム①」）、および、(3)整形済み文書（「サブシステム②」）の二つのサブシステムによって構成される。

「サブシステム①」で実現できるHTML文書の表現能力は、使用する文字体系によってはかなり低い。また、とくに漢字を用いる国においては、標準文字セット以外の文字を使用せざるを得ないケースも多い。この場合「サブシステム①」には限界がある。

「サブシステム②」は、上記問題を解決する手段として採用するもので、基本的には、フォント埋め込み型のポータブルドキュメント・ファイルシステムである。このファイルシステムは、組版（フォーマット）された文書情報を、その組版処理に用いたアプリケーションとは無関係にフォーマットし直し、受け手側では、組版処理に用いたアプリケーションに依存せずに、そのファイルを開くことができ、かつ送り手側が組版した状態に復元できる。また、文書自体にフォント実体を埋め込むことによって、文字の中に標準以外の「外字」が入っていても、あるいは特殊な書体を用いる場合にも、組版スタイルを確実に再現できる。このサブシステムでは、発信国側の文書データを所定のポータブルドキュメント・ファイルフォーマットに変換しておくこと、および受信国側では、このファイルを開くことのできるツール（読むことができる、という意味で「ポータブルドキュメントリーダー」と総称する）を用意する必要がある。

(2) ソフトウェアの仕様

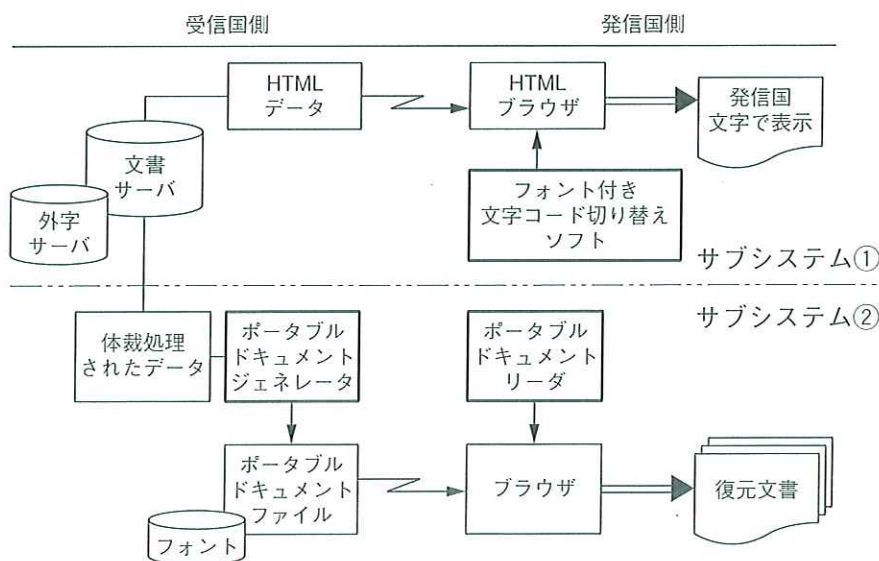
a. 動作環境

各国語仕様のMicrosoft Windows 95または Windows 3.1が動作するPC

b. 対応アプリケーション

- Netscape Navigator 3.0, Microsoft Internet Explorer 3.0

「APPF多言語表示システムのブロック図」



c. 基本操作

本ソフトをインストールし、これを立ち上げると登録された言語のスイッチ（ポップアップメニュー）が画面上に現れる。

(3) リマ総会でのデモンストレーション

リマ総会で行ったデモでは、スペイン語Windowsマシンの上で、各国の議会のWebを表示し、文字化けが生じていることを示した上で、中国語、タイ語、ロシア語、韓国語、日本語に切り替えて「Webの多言語表示」の意味を示した。なお、この表示切り替えツールは、台湾に本社を置く「ダイナフォント」社が開発し、「Asia Surf」として商業化していたもので、APFont研究会では「ダイナフォント」社と協議の上、「APOINT2001計画」で必要な機能を切り出した上、別途、開発等を委託して、フリーソフトとして利用する準備を行っていた。

▶ 5 1999年度以降の進展について

99年1月のリマ総会で行ったデモに基づいて、多言語化ツールの開発を進める前に、APFont研究会では商用アプリケーションを中心に類似のソフトウェアを比較検討した。この中で、5月にリリースされたMicrosoft Internet Explorer Ver.5 および、これと連動して働くMicrosoft Font Centerが、多言語Web環境のユーザ要求をほぼ満たしていることを発見した。このため、「APPF技術ワーキング・グループ」に諮って、各国OS上でこの機能を確認した結果、APFont研究会で独自開発する必要はなくなったと判断した。現在、APFont研究会および研究会のメンバーは、以下の課題についてAPPF日本事務局などと協力しながら調査研究を続けている。

(1) 非商業化言語のフォント開発

アジア太平洋諸国のマイナー言語については、Microsoft Font Centerでも十分には対応していない。このような言語の中から、カンボジアを取り上げて、クメール語のフォントおよびタイプメソッドについて調査研究を行なっている。¹⁰⁾

(2) AAFontの利用促進

東京外国語大学アジア・アフリカ言語文化研究所（AA研）では、本研究会メンバーの松下教授などが中心になって主としてメインフレーム向けに、研究対象となる各国言語の文字情報をデジタル化して「AAFont」を開発してきた。AA研が開発中のシステムでは、現時点で以下の文字が利用できる。Arabic (Arabic, Hausa-Ajami, Persian, Urdu, Pashtu, Uighur and Malayan subsets), Traditional Mongol, Thai, Khmer, Devanagari, Bengali, Tibetan, Laotian, Burma-Mhong, Tamil, Kannada, Malayalam, Telugu。現在、APFont研究会では、この「AAFont」（ビットマップ・データ）をアウトライン・フォントに変換し、インターネットを通じて利用可能にする作業を進めている。

脚注

10. この活動は、長村委員が参加するアジア太平洋研究会などが行なっている。

(3) Linuxでの利用

Windows以外のOSとしてLinuxを選び、AAFontの利用を例として、Linuxの多言語化のための標準化活動を行なっている。¹¹⁾

(4) 国際グリフ登録システムの開発

「サブシステム①」との関連で述べたように、AAFontなどの言語や記号（グリフ（glyph））を、ネットワーク上で広く利用するためには、グリフの画像情報をオープンソースとして登録し、識別用の一意的なコードを付与した上で、データベースとしてオンラインで利用可能にする仕組みが必要になる。ISO/IEC 10036に基づく国際グリフ登録作業は、従来、米国のAFII（Association for Font Information Interchange）が行なっていたが、現時点ではその活動が休止しているために、APFont研究会／GLOCOMでは、ISO/IEC JTC1/SC34に諮って、ISO/IEC 10036の内容を一部修正した上で、国際グリフ登録システムを運営するためのデータベース用アプリケーションの開発などを行なっている。

- (山内 康英 国際大学グローバル・コミュニケーション・センター教授)
 (小町 祐史 松下電送システム株式会社 ドキュメント技術研究所副理事)
 (松下 周二 東京外国語大学アジア・アフリカ言語文化研究所教授)
 (長村 玄 株式会社ドキュメント・エンジニアリング研究所代表取締役)
 (大久保彰徳 株式会社リコー研究開発本部RD推進センター担当課長)
 (上村 圭介 国際大学グローバル・コミュニケーション・センター助手)

11. この活動は、本研究会の小山委員がGlocomで主催するLinux研究会などで行なっている。