

文字情報の画像化による携帯電話向け多言語情報配信システム

—在留外国人向け災害情報提供に関する提案—

Cell Phone-Based Multilingual Information System with Image Conversion from UTF-8 Text
- Proposal Designed for a Method of Providing Disaster Information to Foreigners -平松 初珠†
Hatsumi HIRAMATSU堀 一成‡
Kazunari HORI石島 悌†
Dai ISHIJIMA高階 美行‡
Yoshiyuki TAKASHINA

1. はじめに

地方自治体をはじめとする行政機関において、さまざまな情報を外国人向けに発信することが求められるようになってきた。「内なる国際化」という言葉に象徴されるとおり、国内の在留外国人ならびに海外からの観光客は増加している[1, 2]。このような状況に対応するために、外国人向けの情報発信の重要性は増している。実際に街中でも日本語と英語(ローマ字)にとどまらず、中国語や朝鮮語(ハングル)を含めた4ヶ国語表記の案内板などを多数見かけるようになってきた。

また、行政機関のウェブサイトにおいても、日本語だけではなく、中国語やハングル表記など、多言語でページを開いている組織が増えてきている。これらのページは、パソコンなどの多言語対応の進んだ情報端末であれば閲覧することが可能である。しかし、多くの人にとって最も身近な情報端末である携帯電話では閲覧することができない。

国内における携帯インターネットの普及状況は、2004年度末において、契約数が7,515万契約となっている[3]。すなわち、携帯電話は最も普及している情報端末の地位を確立したといつてよい。このような携帯電話の普及を受け、さまざまな場面で携帯電話を用いた情報配信サービスが提供されている。たとえば、大阪府警では、犯罪発生情報や防犯情報を携帯電話に発信する「安まちメール」という情報配信サービスを2006年1月より開始した[4]。

しかし、これらの携帯電話向けサービスは、日本語での情報配信を前提としており、外国人向けのサービスはほとんど提供されていない。現在国内で使われている携帯電話では、中国語をはじめとする諸外国の言語を表示できないことが、このことの大きな原因となっている。

携帯電話を使った外国人向けのサービスを実現するため、一部の自治体では、ローマ字を使ったピンイン表記による中国語の情報配信などを実施している[5]。とはいえ、この方法では、発信する情報を、パソコンなど多言語対応の端末向けの外国語表記と、携帯電話向けのローマ字表記の両方で用意しなければならない。また、受信者の読みやすさという点でも若干問題が残っていると言わざるをえない。結果として、言語による情報の格差が生じている。

さて、大阪外国語大学と大阪府立産業技術総合研究所では、「多言語同時処理によるアジア系言語の自然言語翻訳

に関する基礎研究」という共同研究を実施し、多言語データベースの構築ならびに蓄積されたデータを活用したアプリケーションの開発に取り組んできた[6]。この共同研究の目標は、「言葉の壁」を乗り越えたユニバーサル・サービスを実現する基礎を築くことにある。そして、その成果の一つとして、世界中のさまざまな国や地域の言語を、携帯電話で表示するシステムを開発した。このシステムは、各言語で書かれた文字(テキスト)情報を、ウェブサーバで画像に変換することによって、携帯電話に表示するものである。このシステムの大きな特徴は、携帯電話の機種や携帯電話会社(キャリア)に依存せず、インターネットにアクセスすることができ、画像を表示できればどの機種でも利用できることである。

本稿では、この携帯電話向け多言語情報配信システムについて説明し、その応用例として、在留外国人向けの災害情報提供システムに関する提案を紹介する。

2. 多言語情報提供システムの仕組み

本システムは、携帯電話で動作するブラウザのリクエストに応じて、UTF-8で記述されたテキスト情報を画像情報に変換することによって、携帯電話での多言語表示を実現するものである。

最近発売された携帯電話の中には、Javaなどのプログラミング言語を用いてアプリケーションを作成できるものがある。これらの携帯電話であれば、携帯電話やアプリケーション自身に中国語やハングルなどの字形情報(文字フォント)を組み込んでおき、携帯電話だけで多言語表示を実現することができるだろう。同種のシステムとして、本来は日本語を表示できない(日本語文字フォントを持たない)海外で使われている携帯電話で、日本語のメールを表示できるシステムもある[7]。

しかし、そのようなシステムは、携帯電話の機種およびキャリアに大きく依存したものとなり、アプリケーションが動作しない機種をサービスの対象外とするか、あるいは多くの機種に対応するために複数のプログラムを用意しなければならない。また、既存の携帯電話には、そもそもそのようなアプリケーションを動作させることができないものも多数存在する。

そこで、携帯電話の機種やキャリアに依存せず、現在出回っている多くの機種に対応するため、我々は、携帯電話は画像を表示するだけのシンクライアントとして利用し、ウェブサーバ上でテキスト情報を画像に変換する手法を用いてシステムを開発した。本システムの概略を図1に示す。

†大阪府立産業技術総合研究所,
Technology Research Institute of Osaka Prefecture

‡大阪外国語大学, Osaka University of Foreign Studies

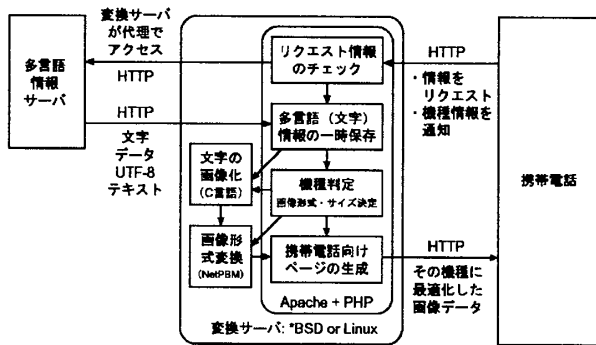


図1 本システムの概略

本システムは、情報端末である図1右側の携帯電話、リクエストに応じて多言語情報を画像化する変換サーバ（同図中央）、多言語情報を保持している多言語情報サーバ（同図左）の3つの要素から構成されている。

携帯電話のブラウザは、多言語情報サーバに格納された多言語情報が掲載されているテキストのURLを、変換サーバへ通知する。この際に、携帯電話の機種名や画面の解像度などが同時に通知される。

変換サーバは、携帯電話のブラウザから送られてきたリクエストを元に、多言語情報サーバに代理アクセスし、UTF-8 エンコーディングされた文字情報を取得し、作業ディレクトリに格納する。次に、送られてきた文字情報を、携帯電話の画面サイズに合わせたビットマップ画像に変換する。その後、このビットマップ画像を携帯電話に応じた形式（GIFあるいはPNG形式）に変換して携帯電話に送信する。

最後に、携帯電話は、送信されてきた画像を表示する。このようにして、機種やキャリアに依存することなく、インターネットにアクセスして画像を表示することのできるほとんどすべての携帯電話で、多言語情報の表示が可能となる。

変換サーバでは、携帯電話からのリクエストを処理するために、ウェブサーバであるApacheが動作している。また、変換サーバのプログラム記述には、PHPを用いた。なお、文字情報から画像データを生成する部分は、C言語を用いて記述している。文字情報から画像データを生成する際に用いた文字フォントは、各国語の字形情報を格納したGNU Unifontである。携帯電話に応じた画像を生成する部分は、NetPBMを用いている。サーバは、これらのソフトウェアが動作するものなら何でもよい。本システムでは、FreeBSDおよびLinuxをサーバとして用い、動作検証を行った。

本システムは、OSやウェブサーバ、そして文字フォントなどに、インターネットで無償公開されているフリーウェアやオープンソースソフトウェアを使っている。本システムも、それらのソフトウェアと同様に、仕様などのドキュメントを含め、そのすべてを無償で公開する予定である。

3. システムの動作検証と実証実験

本システムで、中国語をはじめとする諸外国の言語を携帯電話で正しく表示できることを確認するため、動作検証と実証実験を行った。以下にその内容を紹介する。

3.1 旅行ダイアログを使った動作検証

大阪外国語大学では、日本語を含む専攻25言語すべてについて、基本単語5000語と、旅行ダイアログ1000文のデータベース化を行っている。本システムの動作を検証するため、そのダイアログの中から、「日本大使館に行きたいのですが」という文の中国語訳およびタイ語訳を表示した結果を図2に示す。

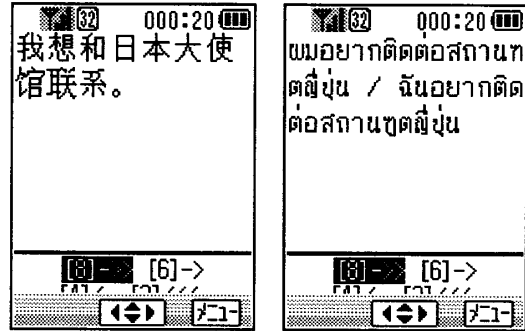


図2 携帯電話で中国語（左）とタイ語（右）を表示（日本語では「日本大使館に行きたいのですが」）

変換サーバは、携帯電話のリクエストに応じて、多言語情報サーバに格納されている旅行ダイアログを取得し、適切に変換した画像を携帯電話に送信していることが、この結果から確認できた。

3.2 大阪府産技研における実証実験

大阪府立産業技術総合研究所では、所内の見学などを随時受け入れているが、近年は海外から、特に中国や韓国などの東アジア諸国からの見学者が増えている。そこで、図3に示すようなQRコードを実験施設の表札に貼り付け、中国語とハングルによる実験施設の簡単な説明文を、携帯電話で提供するサービスを実施している。

QRコード読み取り機能付きの携帯電話を使えば、表札に貼付されたコードを取り込むことによって、その実験施設の説明文を簡単に見ることができる。また、QRコード読み取り機能のない携帯電話でも、研究所内の携帯電話向けウェブサイトから図4（左）のメニューを表示し、そこから実験施設を選択することによって、説明文を参照することができる。



図3 QRコード付きの実験施設の表札の例

図3の無響室についての説明文をハングルで表示させた例を図4（右）に示す。



図4 多言語表示ポータル(左)とハングルで無響室の説明を表示(右)(日本語では「あらゆる方向からの反射音がないように設計された特殊な実験室です. …」)

4. XML テキストコーパスを活用した自動翻訳による災害情報提供システムについての提案

次に、本システムの機種およびキャリア非依存という特徴を活かした応用についての提案を紹介する。

本システムは、携帯電話を使った情報提供における、言語による情報格差の解消を目指したシステムである。このシステムのアプリケーションとして最適なものに、在留外国人向けの災害時情報提供がある。はじめに述べたとおり、携帯電話は情報端末としての普及度が非常に高い。また、電源も内蔵しているため、パソコンなどと比較して、災害によって停電になった時でも使える可能性が高い。

一方、災害時に携帯電話を使った情報提供システムが提案されてはいるが、これは日本語での情報提供に限定されている。阪神淡路大震災など、過去の大型災害発生時、特に災害発生初期段階においては、言語による情報の格差が問題となった。本システムを活用すれば、このような情報格差を解消することが可能となる。

特定の携帯電話やキャリアであれば情報提供を受けることはできるが、他の機種やキャリアでは利用できないという状態は、このような災害情報提供システムにおいては、社会的公平性の観点から避けなければならない。本システムにより、その問題を克服することができる。

地震や台風などの災害発生時に配信すべき情報のほとんどは定型文であり、対訳文の雛形を用意しておき、状況に応じた情報を提供するシステムが提案され、実用化されつつある[8, 9]。

しかし、これらのシステムで生成された情報は、多言語対応の進んだ OS を搭載したパソコンなどでは閲覧が可能であるが、多言語化されていない現在の携帯電話では利用できない。対訳文の自動生成システムと、本システムを組み合わせることによって、携帯電話向けの情報提供が可能となる。

たとえば、地震が発生した場合に、まず最初に伝えるべき情報として、地震を観測した地点と震源地などの情報を提供し、今後の情報へ注意を払うことをうながすような文章を考えることができる。文章の例として「近畿地方で大きな揺れを観測しました。震源地は熊野灘沖と推測されます。今後の地震情報に注意してください。」を想定する。

このような文例を図5のような XML 形式で蓄積しておけば、図6のような中国語訳を機械的に生成することが可能となる。この場合は、「placename district_no」タグで示

される観測地名と「placename hypocenter_no」タグで示される震源地名だけを置き換えればよい。このようにして生成された XML 表記の翻訳文から UTF-8 プレインテキストを生成し、本システムから呼び出せば図7のように地震発生時の速報文の表示が可能となる。

```

<?xml version="1.0" ?>
<data>
<su>
<adp>
<p>placename district_no="5">近畿地方</p>placename
</p>placename
<ad>て</ad>
</adp>
<adp>
<np> <ajp> <aj>大きな</aj> </ajp> <n>揺れ</n> </np>
<ad>を</ad>
</adp>
<adp>
<v>観測しました</v>
</su>
<su>
<adp> <np> <n>震源地</n> </np> <ad>は</ad> </adp>
<adp>
<p>placename hypocenter_no="50">熊野灘沖</p>placename
</p>placename
<ad>と</ad>
</adp>
<adp>
<v>推測されます</v>
</su>
<su>
<adp>
<np>
<adp> <np> <n>今後</n> </np> <ad>の</ad> </adp>
<n>地震情報</n>
</np>
<ad>に</ad>
</adp>
<adp>
<v>注意してください</v>
</su>
</data>
    
```

図5 XML テキストコーパスの例

```

<?xml version="1.0" ?>
<data>
<su>
<p>placename
<p>placename district_no="5">近畿地区</p>placename
</p>placename
<v>
<v>観測</v>
<np> <n>強烈震感</n> </np>
</v>
</su>
<su>
<np> <n>震源</n> </np>
<v>
<v>推測</v>
<p>placename
<p>placename hypocenter_no="50">熊野灘</p>placename
</p>placename
</v>
</su>
<su>
<v>請</v>
<np> <n>各位</n> </np>
<vp>
<v>继续</v>
<vp>
<v>关注</v>
<np>
<n>地震消息</n>
</np>
</vp>
</vp>
</su>
</data>
    
```

図6 自動生成された中国語 XML テキスト

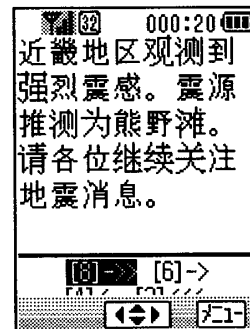


図7 地震発生時に提供する中国語の注意喚起情報

大阪外国語大学では、専攻 25 言語の基本単語、旅行ダイアログのテキストならびに音声を含むマルチメディア情報のデータベース化に引き続き、GDA 方式[10]を多言語向けに拡張した XML タグにより、テキストにメタ情報を付加する作業を行っている。今後、この作業によって作成された XML テキストコーパスを活用することによって、より広範な行政情報の配信サービスを提供することが期待できる。

5. システムに関する考察と今後の課題

本システムは、携帯電話をシン・クライアントとして利用し、多言語表示に関する処理をサーバで行っている。このことは、機種やキャリアに依存しないという大きなメリットを生んでいる。このシステムは災害時の情報提供だけでなく、行政情報の提供や観光情報の提供などにも応用することができる。

さらには、日本語を表示できない海外の携帯電話を持っている在外邦人への情報提供システムへも応用することが可能である。

一方、機種とキャリアに依存しないことを実現するために、本システムではサーバ側でテキスト情報を画像情報に変換している。このことは、携帯電話に取り込むべき情報量が増えることを意味している。

たとえば、「こんにちは」を意味する中国語の「ニーハオ」は、簡体字表記では 2 文字になる。これが UTF-8 テキストで表現された場合は、1 文字 3 バイトなので、本来送信すべき情報は 6 バイトに過ぎない。しかし、本システムでは、1 文字を 16 ドット×16 ドットのビットマップ画像に変換しているため、圧縮を行わなかった場合は 2 文字で 64 バイトとなる。

実際には、GIF 形式や PNG 形式で圧縮されてデータが配信されるため、データ量が 10 倍以上に膨れ上がることはないが、携帯電話が読み込むべきパケットは確実に増加する。とはいえ、現在使われている様々な携帯電話を活用するためには、サーバで画像化するのがやはり簡単な方法であるといえよう。

また、実際の行政サービスとして提供するにあたっての課題として、実災害発生時を想定したシステムのテストが必要となる。たとえば、同時に多数の携帯電話からのアクセスが集中した場合に、応答時間がどの程度になるかを検証しておく必要がある。その結果に応じて、たとえば、個々の携帯電話からのリクエストを受けてから、応答を返すまでの時間を数秒以下にするためには、クライアントの台数に応じてどの程度のサーバの能力を見積もっておくかを明らかにしなければならない。これについては、講演においてその評価結果を報告する予定である。

さらに、災害発生時に運用するシステムならば、サーバそのものが災害発生時にも稼働している必要がある。大阪府には、このような災害時にも対応できる府立インターネットデータセンターがあり、こういった組織と連携して、実証実験を行う必要があると考えている。

そして、今回紹介したシステムでは、UTF-8 で書かれたプレインテキストを携帯電話で表示する機能のみを実装したが、より実用的なシステムとするため、ハイパーリンクなどのリッチコンテンツへの対応も予定している。

6. おわりに

本稿では、サーバで文字情報を画像化することによって、携帯電話で中国語をはじめとする、世界中のさまざまな国や地域の言語を表示することのできるシステムについて報告した。また、その応用例として、災害発生時の在留外国人向け情報提供システムの提案を行った。

本システムは、情報端末として最も普及している機器の一つである携帯電話を多言語表示可能にするシステムであり、言語による情報格差の解消を実現することができる。

また、その応用範囲は、災害情報の提供にとどまらず、行政情報や観光案内など、さまざまなものへの適応が可能である。

本システムの非常に大きな特徴は、機種やキャリアに依存していないことである。インターネットに接続可能で、かつ、画像を表示可能な携帯電話で幅広く利用することができる。国や自治体が目指しているユニバーサル・サービスの実現に大いに役立つだろう。本システムはソースプログラムやドキュメントなど、すべてをインターネット上で公開する予定である。在留外国人や海外からの観光客をターゲットとしたシステムとして幅広く活用してもらえることを期待している。

謝辞

本研究は文部科学省科学研究費補助金（基盤研究（B）課題番号 14310220）および大阪外国語大学特別研究費の補助を受けています。また、本システムの作成にあたって利用したフリーウェアやオープンソースソフトウェアの作成者ならびにコミュニティの方々、そして、本稿を執筆するにあたって有益な助言をいただいた大阪外国語大学の山崎直樹助教授に感謝します。

参考文献

1. 法務省入国管理局、「平成 16 年末現在における外国人登録者統計について（概要）」、(2005/06)、<http://www.moj.go.jp/PRESS/050617-1/050617-1.html>
2. 国際観光振興機構（JNTO）、「訪日外客数・出国日本人数」、(2006/04)、<http://www.jnto.go.jp/info/statistics/pdfs/060410stat.pdf>
3. 総務省、「情報通信白書」、携帯インターネットの普及、(2005/06)、pp. 85-86.
4. 大阪府警察本部安全なまちづくり推進室、「安まちメール」、(2006/01)、<http://www.police.pref.osaka.jp/15topics/seian/images/anmachimail.pdf>
5. 東大阪市国際情報プラザ、「WIP メール」、http://www.city.higashiosaka.osaka.jp/~bunkoku/wip-j/index50_7.html
6. 堀 一成, 青野 繁治, 藤家 洋昭, 石島 悌, 脇田 由実, 高階 美行, 「『多言語同時処理』研究の射程と言語間バリアフリー」、情報処理学会第 65 回全国大会, 講演論文集 第 5 分冊 (2003/03), pp. 347-350.
7. オムロンソフトウェア、「和メール」、(2004/07)。
8. 宮城県国際交流課、「災害時外国人サポートウェブ構築事業」。
9. 多言語防災情報研究開発コンソーシアム、「多言語防災情報翻訳システム」、<http://tagengo.seis.nagoya-u.ac.jp/>
10. 橋田 浩一、「GDA. DTD (Global Document Annotation)」、<http://i-content.org/gda/>