

東京都区市町村ウェブサイトの多言語化対応状況調査

2009年7月7日
国際化JP株式会社

1. はじめに

国際化JPでは、ウェブサイトの国際化状況調査の一環として、東京都区市町村（以下、自治体と称す）の公式サイトが多言語対応について調査を行った。自治体を調査対象とした理由は、住民の多国籍化、多民族化が進む中、ウェブによる多言語情報サービスが不可欠となっているからである。

調査の結果、多言語対応の傾向と問題点が明らかになった。本報告書ではその結果を分析し、問題点を明確にし、改善のための提言を行う。

■ 東京の区市町村

東京都には23区、多摩地区、島しょ部の3地域に合計62の区市町村がある。

東京の市区町村

地域	区	市	町	村	合計
23区	23	0	0	0	23
多摩地区	0	27	2	1	30
島しょ部	0	0	2	7	9
合計	23	27	4	8	62

■ 調査日時

2009年6月1日～2009年6月30日

■ 調査項目

多言語対応に関する以下の項目について調査を行った。

調査項目

No.	調査項目	内容
1	多言語対応	多言語対応をしているか
2	対応言語	どの言語をサポートしているか
3	対応方式	どのように多言語対応しているか
4	言語切り替え	言語の切り替えはどのように行うのか
5	多言語検索	日本語以外のコンテンツの検索は可能か

■ 調査方法

クライアント PC から自治体のウェブサイトにブラウザでアクセスして各調査項目を確認した。

- クライアント PC
 - DELL DIMENSION 9200C, 2GB RAM
- Windows Vista Business 32bit
 - ブラウザ
 - Mozilla Firefox 3.0.X

■ 参考情報

自治体の人口と外国人登録者数は以下の情報を参考にした。

東京都の人口（推計）

- 住民基本台帳による世帯と人口 平成21年1月
毎月1日現在の住民基本台帳上の人口・世帯数、外国人登録者数
<http://www.toukei.metro.tokyo.jp/juukim/2009/jm091a0000.xls>
- 区市町村別、国籍別外国人登録人口 平成21年1月
<http://www.toukei.metro.tokyo.jp/gaikoku/2009/ga09ea0300.xls>

自治体サイトのページ数（参考値）は Google 検索サービスのドメイン内検索オプションを利用してカウントした。

- Google 検索の基本：便利な検索オプション
<http://www.google.co.jp/support/websearch/bin/answer.py?answer=35890>

例えば千代田区の場合、`site:city.chiyoda.tokyo.jp` をキーワードとして検索を実行し、ヒット数（ドメインに `city.chiyoda.tokyo.jp` を含むページの数）を取得した。その際、ドメイン名から `www` は取り除いた。

なお、この値は Google が独自の基準で検索対象と判断したページの数であり、かつ概算である。あくまでも参考値と捕えていただきたい。

2. 調査結果

以下、各調査項目に関する調査結果、問題点、提言を示す。調査結果のデータは以下の URL を参照のこと。

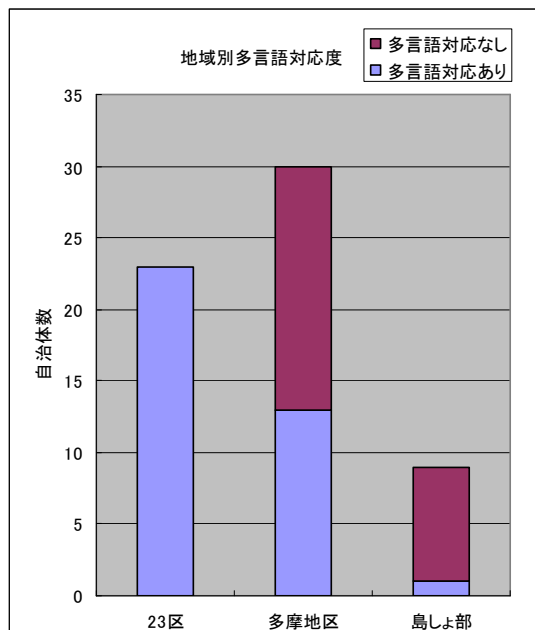
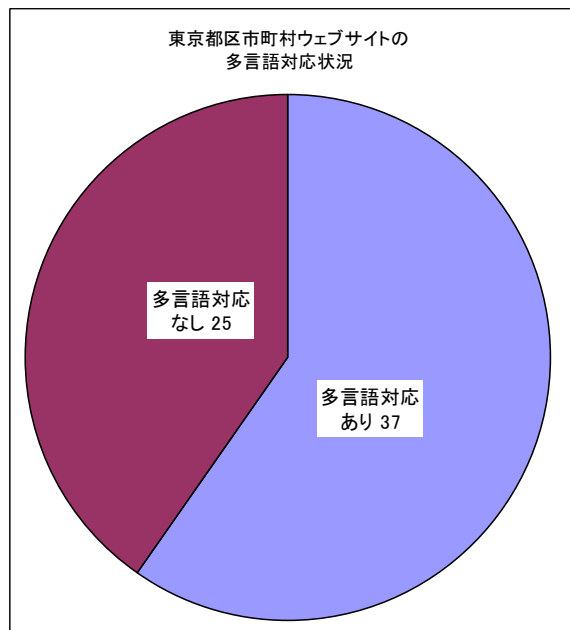
<http://www.kokusaika.jp/images/kjp/info/web-i18n-tokyo-lg-data.pdf>

なお本報告において、多言語コンテンツとは、多言語対応サイト上の日本語以外のコンテンツを指すものとする。同様に、多言語ページは日本語以外のページを指す。

2.1 多言語対応

(1) 調査結果

62 ある自治体のうち、多言語に対応しているのは全体の約 6 割にあたる 37 自治体であったが、地域別に見ると、23 区が 100%、多摩地区は 43.3%、島しょ部では 11.1% と、地域によって大きな違いがあることが明らかになった。



地域別多言語対応状況

地域	自治体数(A)	多言語対応自治体数(B)	割合(A/B)
23 区	23	23	100.0%
多摩地区	30	13	43.3%
島しょ部	9	1	11.1%
全体	62	37	59.7%

(2) 問題点と提言

外国人登録人口(注)が 4 千人を超えるにもかかわらず多言語対応していない自治体が見受けられた。このようなニーズの高い自治体では早急な対応が必要である。

注：

日本語の読み書きが不自由な日本国籍所持者もいればその逆もいるわけであるが、多言語対応の重要性を示す指標のひとつであると考えます。

2.2 対応言語

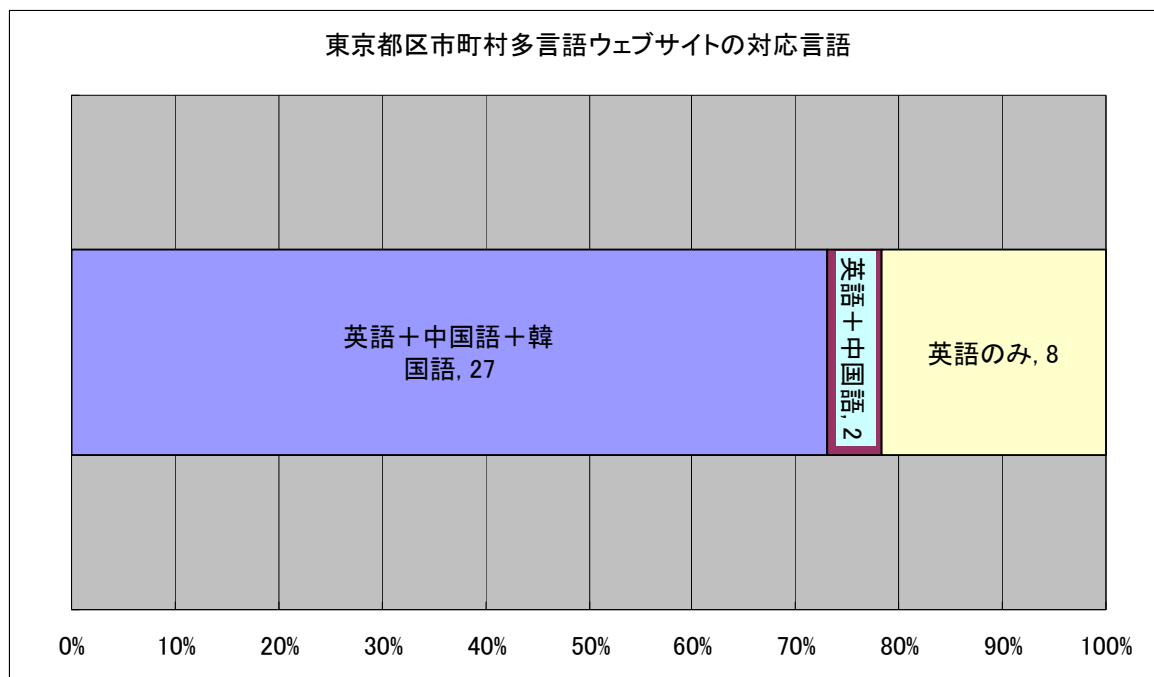
(1) 調査結果

対応言語は、最も多いのが英語ですべての多言語対応自治体に対応、次いで中国語（簡体字、以下同様）、僅差で韓国語の順であった。それ以外の言語はどの自治体でも対応していなかった（一部の観光情報サイトを除く）。

対応言語

言語	対応自治体数	多言語対応自治体中の対応率	全自治体中の対応率
英語	37	100.0%	59.7%
中国語	29	78.4%	46.8%
韓国語	27	73.0%	43.5%

言語の組み合わせで見ると英中韓 3 言語対応が 27 自治体、英中 2 言語対応が 2 自治体、英語のみ対応が 8 自治体であった。英韓、中韓、中のみ、韓のみの自治体はなかった。



対応言語の組み合わせ

対応言語	自治体数
英中韓	27
英中	2
英	8
合計	37

(2) 問題点と提言

英語にしか対応していない自治体の中には、数千人の中国人、韓国・朝鮮人登録者を有する自治体が含まれている。これらの自治体では中国語および韓国語に対する追加対応が急がれる。

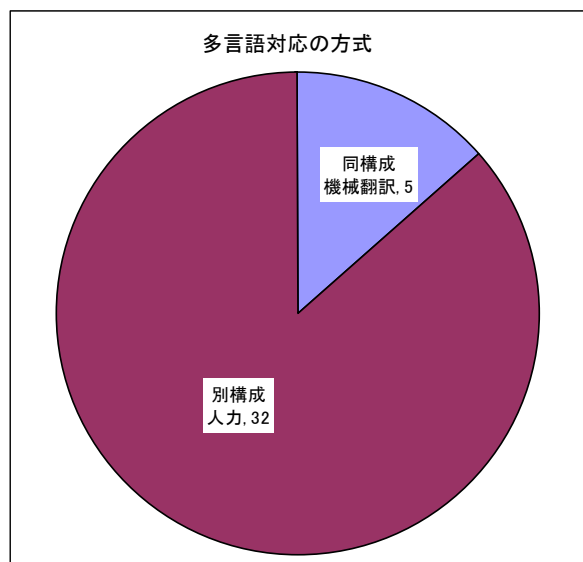
また英語、中国語、韓国語以外を母国語とする住民は都全体ではかなりの数に及ぶため、今後対応が必要となる。その中には、都の統計上中国人扱いのため数字に表れない台湾人（繁体字中国語）や、ブラジルをはじめとするポルトガル語・スペイン語圏住民などが含まれる。

2.3 対応方式

多言語コンテンツの構成と生成方式について調査した。

(1) 調査結果

多言語コンテンツの構成を見ると、日本語と同構成（同階層）とするサイトが 5 つ、別構成とするサイトが 32 であった。一方、生成方式は、同構成サイトがすべて機械翻訳であるのに対し、別構成サイトはすべて人力による生成・翻訳であった。すなわち多言語サイトは、機械翻訳による同構成サイトと人力による別構成サイトの 2 種類に分類できることが明らかになった。



(2) 問題点と提言

機械翻訳に関して以下の問題が確認された。

■ 翻訳精度が低い

ウェブ上の機械翻訳サービスの利用が広まってきたが、翻訳精度は人力翻訳に及ばないのが現状である。実際に機械翻訳方式の自治体サイトにアクセスしてみると、翻訳の精度は高くなく、中には理解不能なものも見受けられた（注）。それは機械翻訳サイト側も認識しており、多言語コンテンツ参照時に「翻訳は 100% 正確とは限らない」旨の断り書きが表示され、了解するとページが表示される仕組みとなっていた。なお、辞書の活用などにより、翻訳精度を高める努力がなされているサイトも見受けられた。

注：

これは翻訳精度の問題ばかりでなく、翻訳ソースとなる日本語の言い回しも関係する。たとえば表現が曖昧であったり文法的に正しくない場合、翻訳精度が下がるようである。機械翻訳の精度を上げるには翻訳ソース側を翻訳されやすい表現にする必要がある。

■ 表示に時間がかかる

アクセス時に随時翻訳しているらしく、翻訳ページのレスポンスが悪かった。ただし、キャッシュを有効にしていると思われる一部のサイトでは 2 回目以降はレスポンスが良くなった。

■ 画像ファイルが日本語用のまま（一部のサイト）

画像ファイル内に画像として取り込まれた日本語テキストは、一般に機械翻訳不可能である。このような画像ファイルを使用していると、多言語コンテンツの中で、その部分だけ日本語表示となってしまう。対策として、文字部分を画像と分離したり、言語ごとの画像ファイルを前もって用意しておき、CGI などのサーバープログラム側で切り替えて表示するといった方法が考えられるが、一部のサイトではすでに実行されていることが確認された。

(3) 機械翻訳に関する考察

問題点が見つかった機械翻訳ではあるがメリットも多い。

■ 翻訳コストが低い

機械翻訳では辞書メンテナンスや画像などのカスタム対応が必要であるが、人力翻訳に比べて一般に翻訳コストは低い。そのメリットはページ数が多いほど大きくなるが、自治体サイトのコンテンツは数千～数万ページ（ファイル）に及ぶため、全ページを対応するとなると、コスト的には機械翻訳の方が圧倒的に有利である。自治体ページの場合コストは住民にも跳ね返ってくる。

■ 翻訳のタイムラグがない

人力翻訳ではどうしても発生する翻訳タイムラグが、機械翻訳には存在しない。

機械翻訳システムは本格導入が始まったところでまだ完成度が高くなく、機能向上の余地が十分にあると思われる。今後の改善に期待したい。

(4) 同構成方式と別構成方式の比較

本来、利用者にとって重要なのは必要な情報にアクセス可能かどうかであって、多言語コンテンツの構成が同構成か別構成かは本質的な問題ではない。しかし、自治体サイトにおいては日本語コンテンツがオリジナルである以上、同構成の多言語コンテンツ提供がいろいろな意味で望ましいのは確かである。

しかし同構成の場合、全ページを翻訳するとなると、翻訳精度の高い人力翻訳はコスト的に現実的ではなく、機械翻訳が採用されているのが現状である。

それでも重要なページに関しては高い翻訳精度が求められる。したがって、重要なページは人力で翻訳し、それ以外は機械翻訳といった、いわばハイブリッド翻訳方式が、現状取りうる最善の方式と考えられる。

2.4 言語切り替え

(1) 調査結果

多言語サイトでは言語の切り替えは言語リンク（言語ボタン）で行う。日本語から他の言語への切り替えについて見ると、言語リンクを日本語の各ページに設置しているのが 9 自治体、主なペ

ージにのみ設置しているのが1自治体で、残りはトップページのみでの設置であった。

日本語ページの言語リンク設置状況

設置状況	自治体数
各ページに設置	9
主なページのみ設置	1
トップページのみ設置	27
合計	37

言語切り替え時の遷移先は、すべての自治体で選択言語のトップページとなっていた。

日本語以外のページでは状況がまちまちであったが、「トップ」や「ホーム」へのリンクで日本語トップページに遷移した。

(2) 問題点と提言

言語切り替えはサイト内のどのページからも行えるべきである。しかしそのような自治体は9自治体と少数派であり、大多数のサイトではトップのみ、または一部のページからしか切り替えができない（日本語ページの場合）。これは非常に不便であり、早急な改善が望まれる。たとえばページ共通のヘッダ部に言語リンクを組み込めば、容易に実現可能となるはずである。

言語切り替え時の遷移先は、別構成方式の場合トップページへの遷移でよいが、同構成方式の場合対応する選択言語にページに遷移するべきである。機械翻訳システムによる実装でも容易に可能なはずなので、早期改善が望まれる。

さらに、言語選択が記憶されないという問題も明らかになった。たとえばトップページで英語サイトに切り替え、ブラウザを閉じて、再びブラウザを開いてトップページにアクセスした場合、英語サイトが表示されるのが望ましいが、すべての自治体サイトにおいて日本語ページが表示された。また英語ブラウザでトップページにアクセスした場合も、英語ページが表示されるのが適切であるが、すべての自治体サイトで表示されたのはやはり日本語ページであった。

2.5 多言語検索

今日のインターネット環境では、情報アクセスは検索サービスを介して行うのが当たり前となっている。これはまた、検索でヒットしない情報はアクセスの機会が限られてしまうことを意味する。このことは自治体サイトの多言語コンテンツにもあてはまる。すなわち日本語以外のコンテンツも日本語と同様、検索でヒットしなければならないのである。この点をふまえて調査を行った。

(1) 調査結果

自治体サイトを検索できる検索サービスには、ウェブ検索とサイト内検索の2種類がある。

● ウェブ検索

Google や Yahoo! など、広くインターネット全体を対象とした検索サービス。自治体サイトはインターネットに開かれたサービスなのでウェブ検索の対象となる。

- サイト内検索
いわゆる検索窓のこと。当該サイトのコンテンツのみを対象とした検索サービス。ウェブ検索サービスを利用したものとサイト固有の検索機能によるものがある。

それぞれについての調査結果を以下に示す。

■ ウェブ検索

機械翻訳によって生成されたコンテンツの場合、トップページなど一部のコンテンツしか検索できなかった。一方、人力で生成された多言語コンテンツは、問題なく検索できた。

■ サイト内検索

最初に検索窓の有無を確認したところ、37ある多言語対応サイトのうち、日本語ページに検索窓があるサイトが35あったが、多言語ページにも検索窓があるのは15サイトにとどまった。

検索窓の設置ページは、日本語ページでは各ページに設置されているのが28サイト、トップページだけに設置されているのが7サイトであった。多言語ページでは各ページが12サイト、英語ページのみが2サイト、トップページのみが1サイトであった。

検索窓設置状況

検索窓設置ページ	日本語サイト	多言語サイト
各ページ	28	12
英語ページのみ	N/A	2
トップページのみ	7	1
合計	35	15

次に、日本語ページの検索窓で多言語コンテンツが検索できるか調査したところ、35サイトの約半分にあたる18サイトで、全てまたは一部のコンテンツが検索できないことがわかった。

さらに多言語ページでその言語のコンテンツが検索できるか調査したところ、15サイト中6サイトで、全てまたは一部のコンテンツが検索できないことがわかった。

(2) 問題点と提言

■ ウェブ検索

機械翻訳コンテンツが検索できない原因のひとつに、URLにパラメータが付与されている等の理由で、ウェブ検索サービスの対象外となっていることがあげられる。この場合、パラメータを付与しない形式に改めれば検索可能となると思われる。

■ サイト内検索

検索窓がない、あるいは一部のページにしかないサイトが少なからず見受けられたが、自治体サイトのコンテンツ数を考えると、検索窓は全言語の全ページにあってしかるべきである。

多言語コンテンツに対してサイト内検索が可能になるには、以下の条件がすべてそろわなければならない。

- A. 当該コンテンツが検索の対象となっている。
- B. 当該コンテンツの文字コードを検索サービスがサポートしている。
- C. 検索窓が当該文字コードをサポートしている。
- D. 検索結果画面が当該文字コードをサポートしている。

しかし、結果からは、多くのサイトがこの条件をクリアしていないことが推察される。文字コードや国際化・多言語化の知識があればクリアできる内容なので、早急な改善を望みたい。

興味深いのが、機械翻訳サイトの多言語ページにおける検索で、たとえば英語ページ上での検索では、英語の検索キーワードを和訳し、日本語ページに対して検索を実行し、結果を英訳して表示しているようである。そして検索結果を選択した時に表示されるページは、当然機械翻訳システムで英訳したものとなっている。翻訳による検索精度が上がれば、今後もっと広く使われるようになる技術であると思われる。

3. おわりに

以上、多くの問題点が見つかったが、全体を見渡すと、いかにコストをかけずに機能を向上するか、といった点が改善のキーポイントになると思われる。それにはコンテンツ・マネジメント・システム (CMS)、翻訳システム、検索システムといった各要素システムの機能向上が必須である。同時に、これらのシステムの機能・性能を正しく、最大限に引き出す実装および運用が不可欠である。

具体的には、コンテンツを CMS で効率的に管理しながら、機械翻訳と人力翻訳を適宜使い分け、検索エンジンを活用したコンテンツアクセスを、全言語、全ページで高速に実現する、というのが一つの到達点となるであろう。それは、IT 技術者、翻訳者、ウェブ制作者、運用管理者、発注者など、すべての関係者の努力と協力があって初めて実現するものである。

そのような多言語対応が当たり前となり、言語を問わず、住民への情報サービスが十分に行われるようになる日が来ることを願って止まない。

[End]