

多言語同時処理とデータベースに関する基礎研究

キーワード：多言語同時処理，自然言語翻訳，データベース，携帯情報端末

はじめに

情報通信関連分野において、最近よく使われるキーワードとして、「情報バリアフリー」と「ユビキタス」があります。前者は、情報利用におけるさまざまな障壁(バリア)を取り除いて、多くの人々が情報通信技術の恩恵を受けられるようにすること、そして、後者は、いつでもどこでも情報通信技術を利用できる環境を整備することを意味しています。

この「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」は、大阪外国語大学を中心として、大阪府立産業技術総合研究所、松下電器産業先端技術研究所などのメンバーが産学官共同で取り組んでいる研究です。そして、インターネット利用における障壁の一つである「言語の壁」を取り除くことと、携帯情報端末(PDA)に代表されるモバイル機器を用いてユビキタス環境を実現することを目標としています。なお、本研究は、平成14~16年度 文部科学省 科学研究費 基盤研究(B) 課題番号 14310220の補助を受けています。



図1 言葉の壁を乗り越えて世界中の人々とコラボレーションを

研究の概要

コンピュータによる多言語間の自動翻訳や自動通訳を実現するためには、その基盤となるデータベースの整備が欠かせません。現在、日本語や西欧諸国の言語については、データベースの整備が比較的進んでいますが、アジア系言語ではそうではありません。

アジア系言語に関するデータベースの整備が遅れている原因としては、これらの言語をコンピュータ上で扱うことがこれまで難しかったことがあげられます。しかし、インターネットの急速な普及とともに、アジア系言語を手軽に、しかも同時に複数の言語を扱える環境が徐々に整備されつつあります。

そこで、これらの環境を利用して、中国語、タイ語、ベトナム語などをはじめとするアジア系言語のデータベースを整備し、多言語同時処理やマルチメディアコンテンツを誰もが手軽に扱えるシステムを構築することが、本研究の大きな目標です。

大阪府立産業技術総合研究所では、この共同研究の分担課題として、多言語同時処理が行えるデータベースのマルチメディアコンテンツの作成と管理手法の開発を行い、その応用システムとして、携帯情報端末からのデータアクセス手法の開発に取り組んでいます。

研究の成果

これまでに、データベース、ウェブサーバ、ウェブブラウザを組み合わせた応用システムとして、複数のアジア系言語を同時に扱うことのできるデータベースアプリケーションを作成しました(図2)。このシステムでは、各言語の文字情報だけではなく、音声情報といったマルチメディア情報も扱うことができるようになっていきます。

ウェブを利用した単純なサービスは、さまざまな場所での導入がかなり進み、現在はその次の応用が模索されている段階です。その一つとして多言語の利用があげられます。政府機関や自治体に限らず、一般企業においても、海外、とくにアジアへの進出をにらんで、多言語を扱えるシステムが徐々に導入されるようになりつつあります。本システムの開発によって得られた技術は、今後、多言語応用システムに関する技術支援に多いに役立つと考えられます。

本システムでは、サーバ側のシステムは、オペレーティングシステム、データベース、ウェブサーバまで、すべてインターネットで無償公開されているオープンソースソフトウェアを用いて開発を行っています。またウェブブラウザはパソコンに標準装備されているか、インターネットで無償配布されているものを利用します。

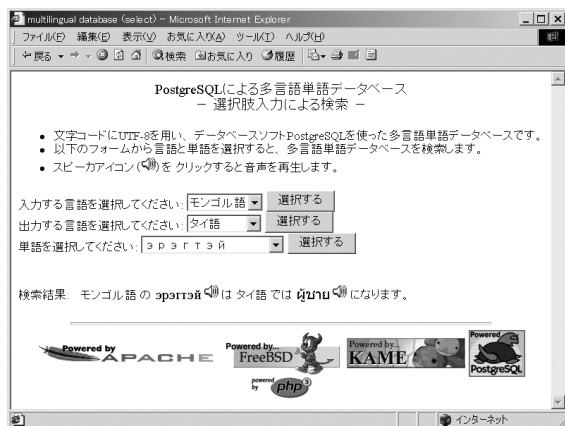


図2 アジア系言語を検索できるウェブアプリケーション

また、携帯情報端末を応用したシステムをあわせて開発しました。このシステムでは、日本語・中国語・タイ語の相互間で、基本となる約5,000の単語と、海外旅行での利用を想定した約1,000の会話集をそれぞれ検索することが可能となっています。先に紹介したウェブ応用システムを利用するには、ネットワーク環境とノートパソコンが必須であり、これでは、いつでもどこでも使えるシステムとはいえません。しかし、この携帯情報端末を用いたシステムは、すべてのデータが携帯情報端末に登録されているので、ネットワーク環境が利用でき

ない場所でも、多言語データベースを検索することができます。このシステムを図3に示します。

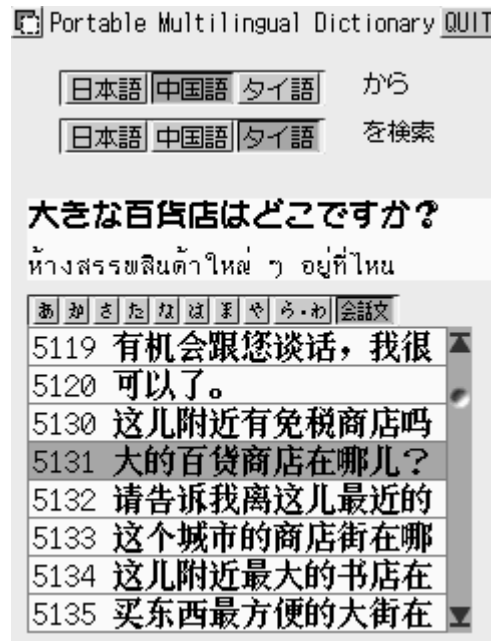


図3 携帯情報端末用多言語電子辞書

今後の展開

大阪外国語大学では、専攻25言語について、文字データやネイティブスピーカの発話による音声データの電子化を進めています。大阪府立産業技術総合研究所では、そのデータの整備に合わせて、より実用的な応用アプリケーションの充実を目指しています。そして、最終的にはコンピュータを使ったアジア系言語を対象とした多言語音声自動翻訳の実験システムの構築を図っています。

参考文献

- 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.
- 堀 一成, 石島 悌; 「PostgreSQLを用いた多言語文字・音声データベースの構築とアプリケーションの開発」, 情報処理学会 第63回全国大会 講演論文集第2分冊, (2001/09), pp.193—194.