

1R-2

## 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

堀 一成           大阪外国語大学  
石島 悌           大阪府立産業技術総合研究所

### 1. はじめに

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、その基盤となる多言語間の言語データベースの整備が不可欠である。前回の第 6 2 回大会において我々は、PostgreSQL が多言語データを手軽に扱える UTF-8 コードデータの処理をサポートしていることに着目し、簡易な多言語単語データベースを作成したこと、さらに検索 Web アプリケーションを作成したことを発表した。今回、PostgreSQL が数値や文字列といった基本データ型以外に、音声情報や画像情報を格納できるラジオオブジェクトをサポートすることに着目し、文字データに加えて音声データ、文字画像データをデータベースに蓄積し、Web アプリケーションを拡張した。また、利用形態の拡大をはかるために、iモード携帯電話から利用できるアプリケーションを作成した。

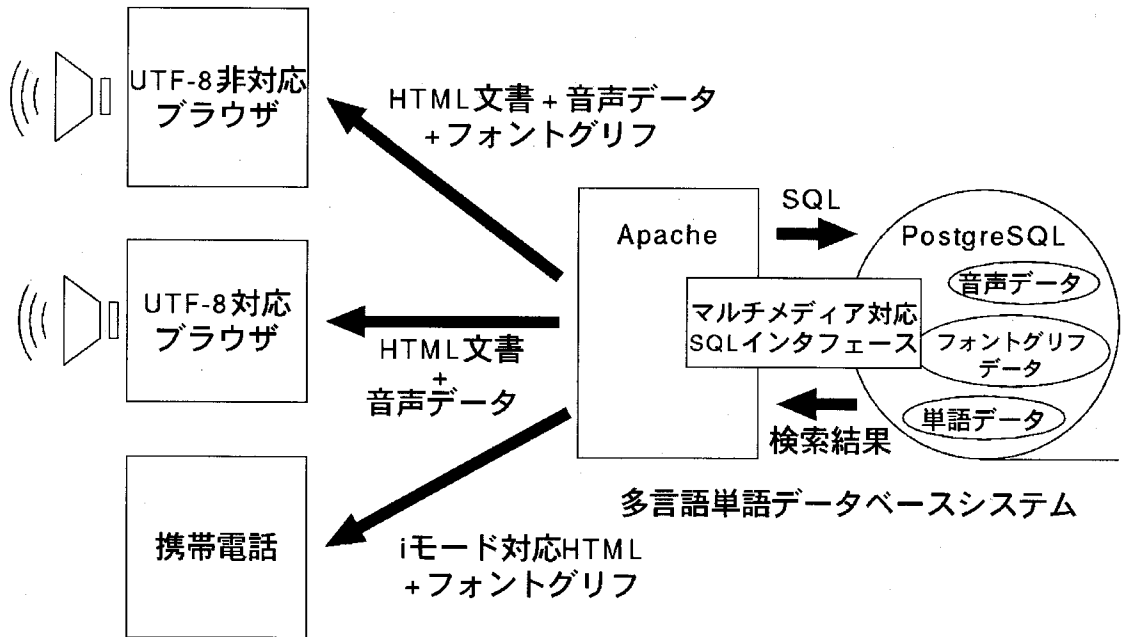


図 1 システムの概念図

Multilingual Word and Voice Database and Application Construction Using PostgreSQL

Hori Kazunari

Osaka University of Foreign Studies

Ishijima Dai

Technology Research Institute of Osaka Prefecture

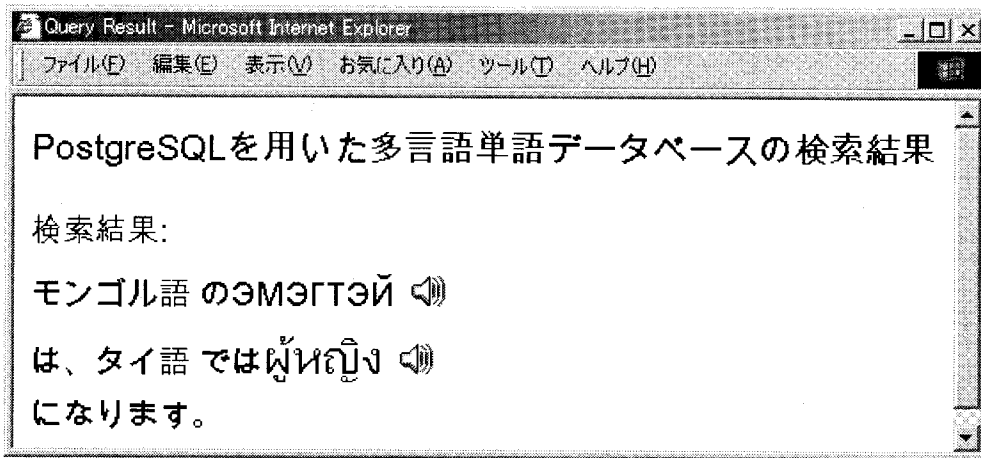


図 2 単語検索の結果  
(検索された結果の単語の右に表示されるスピーカのアイコンをクリックすると対応する音声再生される。)

## 2. 音声データの蓄積とアプリケーション開発

前回構築した多言語単語データベースに、大阪外国語大学 多言語処理室において録音された、各言語のネイティブスピーカーによる各単語の発話音声データを、ラージオブジェクトを使って追加格納した。検索アプリケーションにおいては、単語検索と同時に、音声データを一時ファイルに取り出し、ブラウザから転送の要求があれば転送する。上図はその検索結果表示画面の一例である。

## 3. iモード携帯電話用アプリケーションの開発

携帯電話などのモバイル情報機器との連携をはかることによって、このデータベースの応用範囲をさらに拡大することができる。我々は、iモード対応 HTML を出力するアプリケーションもあわせて作成し、パソコンだけでなく携帯電話からも利用できるようにした。右にその検索出力画面の一例を示す。現時点では iモード対応 HTML は UTF-8 文字コードが利用できないので、多言語データは文字フォントを転送することによって表示させている。

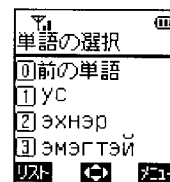


図 3 携帯電話上の入力画面



図 4 携帯電話上の検索結果

## 謝辞

本研究は平成 12 年、13 年度 科学研究費 基盤研究 (B) 課題番号 12410125 に基づく研究である。

## 参考文献

- [1] 石井達夫: PostgreSQL 完全攻略ガイド 改訂第 3 版, 技術評論社 (2001)
- [2] 錦見美貴子、高橋直人、戸村哲、半田剣一、桑理聖二、向川信一、吉田智子: マルチリンガル環境の実現, プレンティスホール, (1996)