プログラミング教材としての音声認識ウェブシステム

Exercises of Web programming with the Speech-input Enabled Web-based system

西村 竜一
Ryuichi NISIMURA
和歌山大学 システム工学部
Faculty of Systems Engineering, Wakayama University
Email: nisimura@sys.wakayama-u.ac.jp

あらまし：本発表では、音声認識に対応したウェブシステムを題材としたプログラミング教育の課題を紹介する。2007 年から和歌山大学システム工学部デザイン情報学科で実施している演習授業では、3 名が構成するチームが 3 週間で音声認識ウェブシステムを開発する。本課題は、ウェブプログラミングの基礎の習得と同時に、音声認識ユーザインタフェースの設計を学ぶことができる内容として構成されている。また、2014 年度からは教材を HTML5 を用いた実装に更新することで、最新のウェブ技術を体験できる内容となった。

キーワード：ウェブプログラミング, 音声認識, ユーザインタフェース設計, 演習課題

1 はじめに

著者の「w3voice システムを用いた音声ウェブアプリケーションの開発」は、課題解決型のプログラミング演習課題である。音声認識プログラムと連動したしたウェブシステムの開発を課題としている。本課題を学部 3 年生対象の演習授業の一部として採用しており、3 名が一つのチームを構成し、システムの仕様作成から実装、評価と報告書の作成を行う。2007 年度から 2013 年度までに 40 チームが本課題を履修し、全チームがシステム開発を遂行している。

受講生が開発したシステムの例を図 1 に示す。本システムは、プロ野球の試合結果を表示するウェブサイトと連動しており、例えば「中日の 8 月の試合結果を見せてください。」等の発話によるクエリに対して、結果のウェブページを表示するものである。この他にも、Google Map や占い、ウェブ通販サイトなど、さまざまなサイトと連動したシステムを作成した事例がある。

2 音声ウェブシステム w3voice

著者が 2007 年に公開した w3voice システム [1, 2] は、音声認識等を内部処理とする音声入力ウェブシステムを作成するための簡易なフレームワークである。クラウドサービスに拡張可能なサーバ・クライアント型のアーキテクチャを採用している。ウェブブラウザ側のプログラムは Java アプレットとして実装されており音声信号の録音と、録音された信号をウェブサーバに送信する機能を担当する。一方、ウェブサーバ側は一般的な CGI プログラムである。受信した音声信号に対して認識や加工処理をし、ユーザに最終的に提示する HTML ドキュメントを生成する。音響信号は、HTTP の POST メソッドによって、無圧縮のビットストリームのままサーバに送られる。

w3voice システムの公開時は、音声認識システムは今ほど一般的なものでなく、多くのクラウドサービスは登場していなかった。今では、音声認識システムを手軽に利用できる。Google Chrome ブラウザは、同社のクラウドサービスと連動することで、音声認識機能を標準搭載している。このため、w3voice システムが実現してきたウェブベースの音声認識システムの開発は難しくなくなっている。

(1) 音声録音時画面

(2) 結果表示画面

図 1: 完成したシステムの例（この例では「日本野球機構オフィシャルサイト」を一部引用して使用）

しかしながら、Google 社の音声認識サービスは未公開な部分が多く、システムの規模が巨大である。ウェブ API として公開されている部分は限定的であり、ブラックボックスとしての利用が求められている。w3voice システムを用いた開発では、認識処理はウェブサーバ上の CGI プログラムに過ぎないため、一般的なウェブプログラミングの知識を利用することができる。また、音声認識プログラムに Julius[3] を利用することで、オープンソースソフトウェアでシステムすべてを実装することができる。仕様がすべて公開されているため、初歩的なものから、発展課題となる高度システムまで幅広くカバーすることができる。このため、本システムを用いることで、ウェブプログラミングの基礎の習得とともに音声認識システムの内部構造を扱う課題を提供することができる。

3 課題構成

本課題は、大きく a)〜c) までのステップで構成される。全内容を 3 週間で遂行することを課題としている。

a) システムの仕様作成　提供するサービスと、連動して表示するウェブサイトを決定する。システムが応答できるクエリを最低 10 個決め、仕様書に明記すること

を課している。10 クエリは、f) の評価で用いるとともに、受講者が構想するシステムの複雑性を知る材料とする。場合によっては、仕様の再検討を要求する。

クエリは、自然文章を基本とするが、孤立単語（キーワード）の入力も許可する。文章や複数の単語の組み合わせ（AND 検索）が受理できる複雑なシステムを完成させた場合は高い評価を与える。また、仕様書の段階で音声認識結果から抽出するキーワードに対応する URL のリストを作成することを課している。例えば、クエリが「和歌山大学」の場合は、`http://www.wakayama-u.ac.jp/`が対応する URL となる。URL の文字列内にパラメータを埋め込むことが可能である。パラメータを URL 内に直接埋め埋め込めない場合は、cookie 等を用いた外部ウェブサイトとの連動が可能である。

b) 音声認識用文法の作成　上記の仕様に基づき、音声認識用の単語辞書、文法を作成する。Julius は大語彙連続音声認識が可能であるが、あらゆるドメインの発話の受理は困難である。一般に、受理するドメインに適応した言語モデルのカスタマイズが必要である。Julius では記述文法を利用できるため、受講者に想定される単語の選定と記述文法の作成を指導している。自然文章では、同じ内容を意図する場合でも、さまざまな言い回しが考えられる。単語の省略形なども存在し、それらを柔軟に受理する文法を記述するためには試行錯誤が必要である。

なお、言語モデルの他に音響モデルを用意する必要があるが、音響モデルのカスタマイズは複雑であるため、本課題では、フリー配布のモデルを使用している。

c) Perl 言語によるコアプログラムの実装　Julius は、音声信号をテキスト文字列に変換するプログラムである。後処理として、Julius が出力したログから不要部分を削除し、認識結果の文字列からキーとなる単語を抽出、対応する URL を出力するプログラムを実装する。簡単なテキスト処理を理解する必要があり、現状では、実装言語には Perl を使用している。Perl は CGI との相性が良いこと、テキスト処理の参考書が豊富に存在することが Perl を採用している理由ではあるが、PHP や Ruby, Python 等の多言語を使用することも可能である。本学科では、このときに多くの学生がプログラミングでの正規表現をはじめて使用することになる。

なお、この段階までは、プログラムは Linux のシェル上で動作する。

d) サンプルプログラムの解読　w3oice システムの開発キットに含まれるサンプルを解読し、音声認識ウェブシステムの仕組みを理解するためのステップである。サンプルには、w3voice システムの Java アプレットと HTML ドキュメント、Julius の認識結果を出力する CGI プログラム（Perl）が含まれる。取得したファイルをサーバ上の `public_html` 等のディレクトリにコピーし、プログラム中のパスを適切に設定する必要がある。CGI プログラムの基本的な設置方法を確認するためにも必要なステップとなっている。

e) システムの統合　c) で作成したプログラム（認識結果からの情報抽出、応答の出力）を w3voice システムに組み込むことでウェブアプリケーション化する過程である。ウェブページとして提示する HTML の出力部分を作成する。授業では、CSS 等を用いたデザインも採点対象としている。

f) 報告書の作文　完成したシステムを a) で定めた仕様に基づき評価し、報告書を作成する。期待通りに動作しなかった場合は、その理由を考察することを課している。これまでの事例では、b) で作成した文法が不十分であり、被験者の発話文章に含まれる言い回しを受理できないことが誤動作の原因として挙げられることが多い。言動の多様性を理解することがユーザインタフェースの設計を学ぶ上では重要であり、本課題はその理解に貢献していることを確認することができた。

4　HTML5 の導入による実装方法の更新

これまでの w3voice システムでは、さまざまな PC 環境でプラグインなしで動作することを期待し、コンポーネントを Java アプレットで実装していた。Java は、今でも Android の開発などで多用されているが、PC のブラウザ上ではセキュリティの理由により動作が制限されることが多い。この結果、これまでに開発してきた音声認識ウェブシステムが動作しない（動作が許可されない）ことが多くなってきている。

現在では、次世代 HTML 規格の集合である HTML5 の標準化が進み、音声信号のライブ (マイク) 入力を可能とする API（WebRTC や WebAudio API）がブラウザに搭載されている。この状況を踏まえ、w3voice システムを新たに HTML5 によって再実装し、2014 年から利用を開始した。新 w3voice システムは、PC 及び Android 端末の Google Chrome ブラウザで安定に動作する。不具合が残るが Firefox でも動作し、標準化が進むことでさらに動作環境が広がることが期待される。また、HTML5 化に伴い Javascript を利用した動的な HTML ドキュメントの生成を課題に追加した。ウェブサーバとクライアント間のプロトコル及びウェブサーバ上の CGI プログラムは、従来の w3voice システムと互換があるため、ノウハウなどのこれまでの教育資源を引き続き利用することができる。

5　今後の予定

従来の w3voic システムと同様に、HTML5 版の音声入力ウェブアプリケーションの開発キットをフリーソフトウェアとして 9 月に公開する予定である [4]。また、教材としての利用価値を高めるために、本課題でも利用しているチュートリアルの一般公開を予定している。フィードバックをいただければ幸いである。

参考文献

(1) R. Nisimura, et al., “Development of Speech Input Method for Interactive VoiceWeb Systems”, *Lecture Notes in Computer Science*, vol.5611, pp.710–719, Springer (2009)

(2) 西村竜一, 三宅純平, 河原英紀, 入野俊夫, “音声入力機能を有する対話型 Web アプリケーションの公開試験”, 情報科学技術フォーラム *(FIT2007)* 講演論文集, pp.319–322 (2007)

(3) A. Lee, et al., ”Julius – an open source real-time large vocabulary recognition engine”, *Proc. EUROSPEECH*, pp. 1691–1694 (2001)

(4) 田藤千弘, 西村竜一, “HTML5 による音声入力ウェブアプリケーションの開発キット”, 日本音響学会 *2014* 年秋季研究発表会講演論文集 (2014) (発表予定)