

令和3年度 卒業研究 中間レポート

中間報告

音声認識 API を用いた Google ブックスの音声操作

近畿大学工学部 情報学科

1810990082 大橋雄喜

1. 研究の背景と目的

1.1. 背景

今日では、人間の声をコンピュータに認識させることで、音声でスマホやパソコンを操作できる音声認識機能 [1] を使用することが増えてきた。Apple が提供している「Siri」や Microsoft が提供している「Cortana」などが例に挙げられる。現在、音声認識を利用したサービスは、企業、病院、自治体でも議事録の自動作成支援や学校向けの語学学習支援などの実用システムが活用されている。

研究を行うにあたっての前段階として、Web アプリの作成に必要な知識を深めるために、JavaScript についての学習を行った。JavaScript を学習するための書籍や Web サイトは多く存在しており、主に書籍を利用して学習を進め、プログラムを改変して理解を深めた。その際に、代表的な Web API として Web Speech API と Google Books API に触れた。

私は今年度入院しており、スマホやパソコンの操作ができないほど体調が優れないことがあった。その際、音声でデバイスを操作し、音声で読書ができればいいと感じた。また、身体障害者や目の不自由な方には、デバイスの操作が難しいことや紙の本がめくれないうなどの壁がある。そこで音声で操作することによって身体障害者や目の不自由な方にも気軽に読書をする機会を設けられるのではないかと考えた。

このような背景から、Web Speech API を使用し、Google ブックスを音声で操作したいと考えた。

1.2. 目的

以上の背景を基に、便利なものをより便利にすること、身体障害者や目の不自由な方に読書を楽しむ機会を設けることを研究の目的とし、アプリの制作を行う。

2. 研究内容

2.1. 概要

Web Speech API を使用して Google ブックスを音声で操作するアプリケーションを作成する。アプリを立ち上げてから一切のマウス、キーボード操作なしで読書ができればいいと考えている。

2.1.1. 機能

現在実装を考えている機能

- 音声認識で Google Books 内の本の検索

音声認識中にある程度の間ができたタイミングで音声認識を中止し、検索を開始する。上位検索結果の本の題名、著者を音声合成で読み上げる。

- 検索した本の内容の読み上げ

音声合成を行うことで実装可能だと考える。Google ブックスにプレビューがない場合は読み上げを行わず、表示しないことができるなら表示しない。

- 指定された文言での Google Books の操作

Switch 文を用いて、文章や単語を登録し、音声認識で読み取った文章や単語と一致した場合に指定の処理を行うプログラムで実装可能だと考える。「次のページ」や「戻る」などの文章を登録する予定である。

- カメラで読書している人の表情を見て、寝ていると判断した場合は再生をストップする

この機能については実装できるかわからないが、TensorFlow を利用することで実装できるのではないかと考えている。

2.2. 使用技術について

Microsoft の提供している無料のソースコードエディタである Visual Studio Code を使用し、以下の言語等をコーディングした。また、動作環境は Google Chrome を利用する。

2.2.1. HTML [2]

HTML とは HyperText Markup Language の略であり、Web のもっとも基本的な構成要素で、ハイパーテキストを記述するためのマークアップ言語の 1 つである。HTML は Web ページの基本レイアウトに従って Web ページのコンテンツを記述し定義するものである。HTML 要素をタグで囲むと、単語や画像をどこかにハイパーリンクさせたり、単語を斜体にしたり、フォントを大きくしたり小さくしたりすることができる。

「ハイパーテキスト」は Web サイト内でも Web サイト間でも、Web ページから別のページに接続するリンクを示す。

HTML では Web ブラウザのテキスト、画像、その他のコンテンツを記述するために「マークアップ」を使用する。

2.2.2. CSS [3]

CSS とは Cascading Style Sheets の略であり、Web ページのスタイルを設定するスタイルシート言語で、HTML や XML で記述された文書の体裁や見栄えを表現するために用いられる。CSS は、要素が画面上でどのように表現されるのかを定義する。

Web サイトは HTML と CSS から構成される。

2.2.3. JavaScript [4][5]

JavaScript は、クロスプラットフォームで動作するオブジェクト指向のスクリプト言語で、Web ページをインタラクティブにするために使用される。JavaScript を使用すると、静止したデータである HTML や CSS を、その場でリアルタイムに書き換えて、一部のコンテンツを入れ替えたり、画像のスライドショーのような動きを付けたりできる。

また、JavaScript 言語のコアをベースに多種多様なツールが作成され、最小限の労力で膨大な機能が利用できる。例えば、Web カメラからの動画ストリームの収集や操作、3D グラフィックやオーディオサンプルの生成などの機能を提供するブラウザのアプリケーションプログラミングインタフェース (API) などが利用できる。

JavaScript には、Array、Date、Math といったオブジェクトからなる基本的なライブラリや演算子、制御構造、文といったプログラミング言語の要素からなる主要な機能が含まれている。追加のオブジェクトを補うことで、マウスクリックやフォームへの入力などのユーザーイベントへの応答といったクライアントサイドの機能、アプリケーションのデータベースとデータのやり取りやサーバー上のファイルの操作といったサーバーサイドの機能を拡張できる。

2.2.4. Google Books API

Google が提供する API の一つである。世界中の書籍情報を収め、検索することによって詳細情報が取得できるサービス「Google ブックス」に登録されている書籍データを API 経由で取得できるサービスである。

2.2.5. Web Speech API [6]

Web Speech API は、音声データを用いて音声認識と音声合成を Web アプリに組み込むことができる JavaScript API である。

音声認識と音声合成の概要と主な機能を下記に示す。

- 音声認識 (SpeechRecognition API)

音声認識では、デバイスからの音声認識サービスを経由した音声入力から音声の文脈を認識し、テキストデータをブラウザに表示するなど適切に応答する機能を提供する。

- 音声合成 (SpeechSynthesis API)

音声合成では、アプリ内のテキストを音声に合成し、デバイスのスピーカーや音声出力接続から再生する機能を提供する。合成されるテキストや発話に使用される音声などを表現するための多くのインターフェースが存在する。

2.2.6. jQuery [7][8]

jQuery は、Web ブラウザ用の JavaScript コードをより容易に記述できるようにするために設計された JavaScript ライブラリである。様々なブラウザで動作する使いやすい API によって、HTML ドキュメントのトラバーサルや操作、イベントハンドリング、アニメーション、Ajaxなどをよりシンプルにすることができる。

3. 研究計画

3.1. 研究スケジュール

研究スケジュールは下記の表 1 に記す。また、詳しい説明は 3.2. で後述する。

表 1 研究スケジュール表

月	研究内容
2021 年 04 月	(参考書) 確かな力が身につく JavaScript 「超」 入門【第二版】 ・ HTML, CSS, JavaScript について
05, 06	・ HTML, CSS, JavaScript の学習継続 ・ 研究テーマの模索
07	・ 研究テーマの決定 ・ 卒業研究中間報告レポート作成
08	・ 卒業研究中間報告 ・ アプリの実装
09	・ アプリの実装
10, 11	・ ユーザテスト ・ アプリ改善
12	・ 卒業研究完成
2022 年 01 月	・ 卒業論文作成
02	・ 卒業研究発表

3.2. スケジュールの説明

- 7月までの研究内容

主な研究活動は、テーマの選定と HTML・CSS・JavaScript の学習である。これらは、入門書の「確かな力が身につく JavaScript 「超」 入門」を用いて学習し、サンプルの内容を改変し、理解を深めた。その中で、最も力を入れたのが「HIT AND BLOW」というゲームの作成である。これについては 4.1. で後述する。

- 8, 9 月の研究内容

中間報告後の研究活動は、音声認識の知識について参考書をもとに学習を進め、アプリに実装する機能の選定を行う。選定した機能をアプリに追加する方法を考え、すぐに実装できる場合は実装し、できない場合はプログラミング言語の追加や音声認識の理解を深めるなど別の方法を検討する。

- 10, 11, 12 月の研究内容

実際に機能を追加したアプリのユーザテストとそのユーザテストを基にしたアプリの改善を繰り返す。ユーザテストの詳細は決まっていないが、当研究室のメンバーへのアンケートを検討している。

- 1, 2 月の研究内容

12 月までと同様にユーザテストとアプリの改善を繰り返し、アプリを完成させ、卒業論文の作成に取り掛かる。研究活動よりも卒業論文作成に時間を費やしたいので、アプリの完成を 1 月中には終わらせておきたい。

3.3. 役割分担

当研究室では、各々が研究テーマを決め研究課題を進めるため、明確な役割分担は存在しない。しかし、得た知識や技術、情報を共有することで研究室内でのチームワークを強めていきたい。

4. 研究の進捗

4.1. 現状報告

中間報告までの研究活動状況は下記の通りである。

- スケジュールの進行状況

6 月までは、主に JavaScript の学習と研究テーマの模索を進めており、スケジュール表通りに進行している。しかし、研究テーマの選定がなかなかできず、7 月からの進行が遅れている。アプリの制作に手が付けられておらず、実装できるかわからない機能を検討している場合がある。

- JavaScript の学習について

2月から参考書やWebサイト等でJavaScriptの学習を進めていた。特に力を入れたのは「HIT AND BLOW」というゲームの作成である。「HIT AND BLOW」とは、指定された3桁の数字を予想し、数字と位置が一致すれば「ヒット(HIT)」、数字のみ一致すれば「ブロー(BLOW)」として計算され、それぞれの個数がわかる。本来は対戦相手がいて、先にそれぞれが指定した3桁の数字を当てたほうが勝ちだが、今回は対戦相手はなしでランダムに指定された3桁の数字を当てる脳トレゲームになっている。

4.2. 今後の予定と課題

今後の予定は3.1のスケジュール表の通りに、アプリの実装からユーザテストまでアプリの制作に取り掛かっていく予定である。課題として挙げられるのは下記の通りである。

- 実装を考えている機能の明確性

実装しようと考えている機能が技術的に可能なものかどうか明確になっておらず、参考書などで情報を集め、知識を深めていく予定である。

- スケジュール管理

研究テーマの決定が遅れていたことにより、アプリの制作に取り掛かれていないため、中間報告後は急ピッチで進める必要がある。また、アプリの制作に時間がかかると思われるため、実装する機能の選定など方向性を素早く決める必要がある。

5. 参考文献・参考サイト

[1] 音声認識 | Wikipedia

<https://ja.wikipedia.org/wiki/%E9%9F%B3%E5%A3%B0%E8%AA%8D%E8%AD%98#%E6%A6%82%E8%AA%AC>

[2] HTML | MDN – Mozilla (2021/07/14 閲覧)

<https://developer.mozilla.org/ja/docs/Web/HTML>

[3] CSS | MDN – Mozilla (2021/07/14 閲覧)

<https://developer.mozilla.org/ja/docs/Web/CSS>

[4] 入門編 - JavaScript | MDN – Mozilla (2021/07/14 閲覧)

https://developer.mozilla.org/ja/docs/Web/JavaScript/About_JavaScript

[5] JavaScript の基本 | MDN – Mozilla (2021/07/14 閲覧)

https://developer.mozilla.org/ja/docs/Learn/Getting_started_with_the_web/JavaScript_basics

[6] Web Speech API | MDN – Mozilla (2021/08/01 閲覧)

https://developer.mozilla.org/ja/docs/Web/API/Web_Speech_API

[7] jQuery | Wikipedia (2021/07/16 閲覧)

<https://ja.wikipedia.org/wiki/JQuery>

[8] jQuery (2021/07/16 閲覧)

<https://jquery.com/>