



2008 年度下期未踏 IT 人材発掘・育成事業 採択案件評価書

1. 担当PM

畑 慎也 PM (サイボウズ・ラボ株式会社 代表取締役社長)

2. 採択者氏名

チーフクリエイター : 塚田 翔也(フリーランス)

コクリエイター : 小林 裕幸(名古屋大学工学部電気・電子情報工学科4年)

コクリエイター : 馬場 達也(名古屋大学工学部電気・電子情報工学科4年)

3. プロジェクト管理組織

テクノロジーシードインキュベーション株式会社

4. 委託金支払額

5,560,000 円

5. テーマ名

音声認識 Web アクセスツール「Puppy」の開発

6. 関連Webサイト

なし

7. テーマ概要

「Puppy」とは利用者が一言話すだけで、その要求にあった Web アクセスを正確に実現することが出来る次世代音声認識 Web アクセスツールです。例えば、利用者が「地

図 コンビニ」と言うだけで、GPS の現在地付近のコンビニを検索し、結果を地図上に表示します。

iPhone、Android などの次世代モバイルの登場でモバイルは今後、より Web アクセスのためのデバイスとしてシェアを拡大していくものと考えられる。しかし、これらのデバイスはユーザインタフェースとして相変わらず利用者のタッチパネルへの操作や文字入力に頼っており、様々な改善が行われているものの既存のインタフェースからの脱却が出来ていません。

「Puppy」は「一言話すだけでいい」という究極に簡単なインタフェースでモバイル利用者が抱える画面操作や文字入力の煩わしさを解決します。

「Puppy」はクライアント・サーバ間を汎用的なインタフェースで構築することで様々なクライアントから利用できるプラットフォームを目指します。本提案では iPhone アプリケーションをクライアントとして開発しますが、将来的に Android や BlackBerry、docomo、au、softbank などの既存携帯電話 WindowsCE 搭載機、カーナビゲーションシステム、家電などをクライアントとすることができる拡張性も秘めています。

8. 採択理由

一言話すだけで要求にあった Web アクセスを正確に実現することを目的としたシステムである。この手のテーマはよく考案されるものであるが、対応端末を増やすための工夫がなされている点が実用性を重視していると感じさせる。プロトコルを規定し、端末側では音声要求をプロトコルに変換し、サーバーに送る。サーバー側で、音声要求に対応した Web アクセスを行う HTTP リクエストを作成して返すというものである。サーバー側でほとんどの処理を行うことは負荷分散の観点からは不利となるが、対応プラットフォームの拡大という点ではユーザーにメリットとなる。また、開発者側にも開発言語を絞れるという利点もある。これは携帯端末が主たるターゲットとなる場合は非常に有利に働く。iPhone、次に Android に対応したものから開発していくという辺り、開発者嗜好ではあるが、仕組み自体がクライアントの広がりを意識したものであるため、将来的な対応プラットフォームの広がりに期待できると思われる。ヒアリングを通じて実用的な便利さを感じ取れるものであった。以上の理由により採択としたい。

9. 開発目標

iPhone、Windows Mobile 上のアプリケーションから声と GPS を使って、現在地付近の地図検索が行えることを目標とし、開発を行った。

1. 開発環境

- iPhone アプリケーションの開発
 - ◇ OS: Mac OS X
 - ◇ IDE: XCode
 - ◇ SDK: iPhone SDK 3.0
 - ◇ 開発言語: Objective-C 2.0
- Windows Mobile アプリケーションの開発
 - ◇ OS: Windows XP
 - ◇ IDE: Visual Studio 2005 C#
 - ◇ SDK: Windows Mobile 6 SDK
 - ◇ 開発言語: Visual C#
 - ◇ 実行環境: .Net Compact Framework 2.0
- Puppy サーバーの開発
 - ◇ OS: Mac OS X
 - ◇ IDE: Vim
 - ◇ Web フレームワーク: Merb 1.0.8.1
 - ◇ 開発言語: Ruby 1.8.7
- 音声認識エンジンの開発
 - ◇ OS: Mac OS X
 - ◇ IDE: Vim
 - ◇ ベースとしたオープンソース音声認識エンジン: Julius 4.1
 - ◇ 開発言語: C

2. 機能要件

- iPhone アプリケーションの開発
- Windows Mobile アプリケーションの開発
 - 音声入力から音声ファイルを作成し、GPS 情報と音声ファイルをサーバーに送信する。サーバーから受信した検索結果を表示する。
- Puppy サーバーの開発
 - クライアントから音声ファイルのアップロードを受け付け、音声認識エンジンに依頼を行う。音声認識結果から、URL の生成や検索を行い、その結果をクライアントに送信する。
- 音声認識エンジンの開発
 - Web リクエストは複数同時に発行される可能性が高いので Julius をマルチスレッド化する。また、初期化処理が遅いので逐次起動ではなく常駐起動させておくためのカスタマイズも行う。

3. システム構成

10. 進 結 果、 目 標 の 機 能 要 件 を 満



たし、iPhone、Windows Mobile 上のアプリケーションから声と GPS を使って、現在地付近の地図検索が行えるシステム一式を開発した。さらに、以下の機能を追加した。

- iPhone アプリケーションでは、録音時に音量の変化を監視し、話し終わったことを検知する機能を開発した。
- リクルート Web サービスのホットペッパー API を利用し、飲食店検索をする機能を開発した。

11. 成果

音声認識処理をサーバー上で行っているのでクライアントアプリケーションを容易に開発できる。つまり多くの対応プラットフォームを開発することが可能で、広く人々に利用してもらうことができる。また、以下の機能要件を満たせば、どんな端末もクライアントとなることができる。

- マイク入力から音声ファイルへの録音が可能
- HTTP 通信が可能
- GPS (推奨)

音声認識と Web アクセスの組み合わせをサーバー上で行っているのでサービスを容易に開発し、追加することができる。例えば、1つのサービスを追加すれば、全てのクライアントアプリケーションで簡単に利用することができる。

音声認識や検索をサーバー上で行っているので CPU やメモリなどのリソースが問

題になることがない。もし問題になれば負荷分散や並列処理などで対応することが可能。逆にクライアント上で行っていると CPU やメモリなどが問題になることが多い。

サービスがクライアントに依存する API などを利用せずにサーバー上で行っているため、この端末では動作するが、この端末では動作しないということがない。

音声辞書ファイルをサーバー上で一括管理しているため、音声辞書ファイルを更新してもクライアントアプリケーションを再配布する必要がない。

12. プロジェクト評価

今回の未踏プロジェクトの期間ではクライアント端末としては iPhone と Windows Mobile 端末の 2 種類であった。クライアント側に実装する項目を極力少なくし、サーバー側に実装する項目を増やしたことは、クライアント端末依存の部分を少なくするというメリットがあるので良い着想だったと思う。つまり、対応端末を増やすフェーズにおいては有利となる。その代わりにサーバー側の負荷が増えることになり、負荷分散という点に対してはデメリットとなる。クライアント実装の難易度とのバランスが問われる。また、近年スマートフォンの CPU 速度もどんどんアップしているため、クライアントに依存できる要素が増している点もあげられる。

音声認識エンジンとして Julius を使用した。同時要求が起こるリクエストに関して出てきた課題に対して、マルチスレッド化などの対処が行えたことは評価してよい。

このようなサービスの場合、実用的なレスポンスが求められるが、当初実行速度は実用上問題がある速度であった。問題点の 1 つに音声ファイルの転送があげられたが、転送量を減らすためにクライアント側で特徴量を抽出する処理を行い、対処することができた。

このようにプロジェクトを進める過程において各種の課題が発生したが、1 つ 1 つ解決策を見つけて対処することができたことは評価に値する。

13. 今後の課題

この手の分野は非常に競争が激しい分野であるため、各種サービスに対応した音声認識サービスをすばやくリリースすることが求められる。また、競合サービスに対して音声認識精度などでリードすることも必要となる。

また、広範囲なキャリア、および端末への対応が求められる。