

A Kanji Information Server with a Camera-ready Mobile Phone

Access Interface for Japanese Children

「日本の子供のための、カメラ付携帯電話がアクセスできる

インタフェースを持つ漢字情報サーバ」

オンドラ・ベレック

VELEK ONDREJ

1. 要約(背景、目的)

長い間、人間はコンピュータが多くの仕事をこなすように願い、いまやそれは人間に向けて逆転された。文字認識もそのひとつである。すでに何年もの間に研究者は文字認識技術を向上させている。それは適度な効率性と90-99%の認識率を得ているが、実際に役立つアプリケーションはほとんど存在しない。郵便物宛先の認識は例外的なアプリケーションである。

このプロジェクトにおいて、我々は文字認識の新しいアプリケーションを創りだした。このアイデア自体が全く新しいというわけではない...人は既に、不明の文字を読む助けとなる新しいテクノロジーを長年夢見てきたのである。これは特に日本語や中国語の漢字認識にあてはまる。数千種類にも及ぶ漢字を全て読むことはネイティブの人々にとっても大問題であろう。子供達にとってはなおさらである。

我々のシステムは特殊で高価なハードウェアを必要としない。使用するものは携帯電話だけである...今や携帯電話は日本だけでなく開発途上国でもほとんどの人が所有する装備品である。誰でもすぐに、不明漢字の読み方や意味を知り、外国人には彼らの言語に翻訳もおこなうのである。使用方法は簡単で、不明の漢字、単語、文章の写真を撮るだけだ。どこでもどんな状況でも使えるのだ。

この10ヶ月の間に、我々は実用的なシステムを創り出した。まだ今後の作業(インタフェースの向上、辞書の追加等)は残るが、既にこのシステムは漢字を読む必要のある人全てに役立つものとなっている。これは、数千もの文字を持ちながら誰もその全てを記憶していない書記体系を持つ日本において、在住外国人、日本の子供やさらには成人のためのシステムである。

2. 概要

我々は、通常のカメラ付携帯電話でアクセス可能な、自由に入手できる漢字情報サーバを構築する。基本的なアイデアはFigure4で示す。

Figure4 : 不明漢字を遠隔で認識するプロセス

Remote Off-line recognition
Business application 遠隔でオフラインでの認識
ビジネスアプリケーション
Recognition server 認識サーバ
Internet インターネット

不明漢字が携帯電話のカメラで撮影され、認識エンジンが作動している漢字情報サーバに送信される。次にその漢字が認識され、携帯電話に返答が返信される(Figure4)。

サーバからの返答フォーマットはユーザの選択により、日本語での返答の代わりに漢字が異なる言語に翻訳されることも可能である(Figure5)。各携帯電話はスピーカーを搭載しているため、サーバからの返答は漢字の読みとともに音声ファイルも含めることができる。

Figure5: 不明漢字を遠隔で英語に翻訳するプロセス

今後、同じソリューションが外国語から日本語への翻訳に活用することが可能である(Figure6)。専門的なペン型辞書と同様の方法である。その違いは、翻訳のためにユーザが特殊な機器を購入する必要が無いということである。

大きな利点は、常に持ち運ぶ通常の携帯電話が利用でき、専門的な機器を購入する必要が無いということである。

Figure6 : 不明英語を遠隔で日本語に翻訳するプロセス

3. 開発成果の特徴

我々のソフトウェアは2つのモジュールから構成される。クライアントが携帯電話を作動させている間に認識エンジンと辞書を持つサーバが搭載された機器を作動させていることになる。

3-1 クライアント・モジュール

このモジュールは下記から構成されている

- ・ 写真撮影およびサーバへの送信用ユーザ・インタフェース
- ・ 予備的处理および分割化作業

この部分は現在、新型NTTドコモ携帯電話(505シリーズ以上の機種)で実行されている。開発環境はJavaに基づき、ドコモによる実行はDoJa3.0と呼ばれている。これら携帯電話機種の基本的能力と開発プラットフォームの特徴はFigure22~23で示されている。

Figure22 : NTTドコモ505シリーズ携帯電話の特徴

電話機的能力

- ・ プログラムサイズ 最大30kB
- ・ 記憶容量 最大200kB
- ・ HTTPまたはHTTPSプロトコル使用の通信
- ・ 286クラスのコンピュータレベルのプロセッサ速度
- ・ 浮動小数点演算無し

Figure 2 3 : NTTドコモ505シリーズ携帯電話の特徴

統合型カメラのモバイルソリューションの概括

携帯 NTTドコモDoJa3.0

- ・ + 比較的速いプロセッサ
- ・ ± メモリの限界
- ・ - 自社独自の規格
- ・ - ユーザの操作無しでカメラから画像の読み取り不可能
- ・ - より大きなメモリサイズの記憶装置にアクセス不可能 (メモリカード等)

DoJa3.0のプログラミング環境のもうひとつの限界は、アプレットがダウンロードされたサーバとしか通信できないことである。これを解決するため、プログラムのモバイル・ユーザへのダウンロードに対応する簡単なHTTPウェブサーバがサーバ・モジュールに導入された。近い将来、我々はクライアント・モジュールを他の大手プロバイダ2社の携帯電話にも対応させる予定である。全ての開発環境がJavaに基づいているため、この対応が大問題にはならない。KDDIのJava実行は、ヨーロッパGSMの携帯電話の規格に非常に類似しており、そのモジュールをヨーロッパの携帯電話に対応させることもまた可能と言える。次のFigure 2 4において、KDDIとVodafoneの携帯電話の特徴を比較することができる。

Figure 2 4 : KDDIとVodafone携帯電話の特徴

統合型カメラのモバイルソリューションの概括

携帯 KDDI au MIDP Java

- ・ + 規格が公開されている
- ・ - メモリの限界、速度等がメーカー依存

携帯 Vodafone MIDP Java

- ・ - 実行していない。操作にオペレーターが必要

3-2 サーバ・モジュール

このモジュールはデータの受け取り、漢字の認識、辞書から読み方と意味の検索、そして携帯電話機にデータの返信をする役割を持つ。この部分に関してユーザは入手不可能で、搭載された機器で作動している。そしてユーザアクセスはインターネットを通じHTTPプロトコルによってのみ提供される。

3-3 携帯電話での認識使用方法

クライアント部分はサーバからダウンロードされる：

<http://kohaku.anime.cz/docoreco.html>

クライアント・アプリケーションはNTTドコモ505シリーズおよびそれ以上の携帯電話機で利用できる。

(全機種で我々のアプリケーションをテストできなかったため、全ての機種でアプリケーションが正常に動作しない可能性があることをご承知いただきたい。テストをした機種にはF505iとSO505iを含む。成功例や失敗例の著者への報

告をお願いします。)

アプリケーションはドコモのF505iを使用して開発された。

クライアント・アプリケーションのインストール後、新しいiモードアプリケーションがiモードメニューで利用できる。アプリケーションの名前は、DoCoReCoである。

アクションボタンを押すと、カメラが起動する。

カメラのピントを漢字や単語、文章に合わせる。小さな文字の場合は、近距離撮影よりも搭載されたデジタルズームを使用したほうが良い。

漢字は写真の中央におさめる。適切な写真例：Figure 2 5

Figure 2 5 : 適切に撮影された写真
Taking picture 写真撮影
Take picture 撮影する

写真には、一様でない混乱した背景を左右に含むべきではない。(しかし上下には含むことができる。)不適切な写真例：Figure 2 6

Figure 2 6 : 不適切に撮影された写真
Taking picture 写真撮影
Take picture 撮影する
Left and Right.... 左右に混乱した背景を含んでいる。(上下はOK)

成功する画像入力プロセスのもうひとつの条件は、認識される漢字が画像

の中心に置かれることである。この条件にあてはまらない例がFigure 2 7で示されている。その結果、任意に選択された2行のひとつとなる。

Figure 2 7 : 不適切に撮影された写真
Taking picture 写真撮影
Take picture 撮影する
Text line isn't.... 行が画像の中心ではない

写真は、携帯電話が横向きの位置で撮影する。Figure 2 8 .

Figure 2 8 : 写真撮影のための横向きの位置

アクションボタンを再度押した後、写真は携帯電話で処理され分割化された結果が示される (Figure 2 9 ~ 3 0)。

Figure 2 9 : (不適切に) 分割化された写真

Figure 3 0 : (適切に) 分割化された写真

写真が不適切に分割化された場合 (Figure 2 9)、「撮影する」ボタンを押してもう一度撮影する。分割化が正しく行われると、データを確認してアクションボタンを再度押してサーバに送信するだけでよい。

約1~2秒後に漢字情報サーバからの返答を受け取ることになる。(Figure 3

1) この返答の内容は、

- (1) OCRの結果 - 認識された漢字
- (2) Exciteウェブサービスからの、英

語への翻訳

(3) 3つの漢字それぞれの読み方と意味

Figure 3 1 : Figure 2 8 ~ 2 9 の写真認識結果。(1) OCRの結果 - 認識された漢字(4) Exciteウェブサービスからの、英語への翻訳(5) 3つの漢字それぞれの読みと意味

4. 秘匿ノウハウの指定

我々の開発内容は、ビジネスアプリケーションとして成功する可能性があると考えている。多くの方のご推薦により、我々のいくつかのアイデアを特許申請することにより保護する予定である。今回東京農工大学共同研究開発センターの方とこの件について検討中であるため、このプロジェクトの詳細の何点かについては特許権保護申請の関係上公表できない。