

携帯端末用リアルタイム文字認識エンジンの開発
- 次世代の文字認識エンジン -
Development of real-time characters recognition engine
for mobile devices
- The next generation of characters recognition engine -

ペロフ アレクサンドル^{*1}・立田ルミ^{*2}
Alexander Belov, Lumi Tatsuta
Email: k02345@dokkyo.ac.jp

キーワード : 日本語学習、リアルタイム文字認識、筆順
Keywords : Japanese learning, Real-time OCR, Stroke order

日本語を学習している外国人にとって、リアルタイム文字認識エンジンは有用な勉強道具である。その学習者が未習熟の漢字の量が多すぎる場合、各漢字の読み方を漢字辞書でひとつひとつ調べながらテキストを解読するためには時間を要し、多大な苦勞を伴う。そしてそれが学習者にとって次の学習段階に進むことを阻む大きな原因となっている。本研究では、リアルタイムで漢字の読み方とその使用例を表示することで、日本語学習をより楽に行う環境の開発も目指している。又、2020年に東京オリンピックが開催され、多くの外国人観光客の日本訪問が見込まれる。しかし、レストランや食堂などのメニューや温泉の利用ルールなど、日本語のみの表示も多いため、さまざまなトラブルや無理解が生じる可能性がある。そのようなトラブルを防止するために、また訪日外国人の日本滞在をサポートするために、本研究では外国人を対象とした携帯端末カメラを利用するリアルタイム文字認識エンジンの開発を目的とする。

For foreigners who are learning Japanese, real-time character recognition engine is a useful study tool. If the learner does not studied the amount of too many Japanese characters, then it is take a considerable time and effort in order to translate the text, while examining one by one reading of each kanji in dictionary. It is a major cause for prevent the learner at proceeding to the next learning phase. In this research, by display in real-time examples of applying and reading of Japanese Kanji, we also aim to the development of more comfortable environment for Japanese learning.

In addition, Tokyo Olympic Games will be held in 2020. Many foreign tourists will be expected to visit to Japan. But many as menus of restaurants and cafeterias, and, for example, usage rules of hot springs, and other texts display in only the Japanese yet. There is a possibility that various troubles and lack of understanding occurs. In order to prevent such trouble, and in order to support the foreign visitors, our object of this research is a development of real-time character recognition engine for mobile devices with a camera.

*1: 獨協大学 経済学部
*2: 獨協大学 情報学研究所

1. はじめに

近年、世界のグローバリゼーションが急激に加速している。インターネットと携帯電話やフラットパソコンなどの普及により、地球のどの国でも何かの出来事が起きたら、その確認ができるテキストや写真、ビデオデータなどをほぼリアルタイムで閲覧する事ができる。このような状況の中、他言語・異文化理解が重要な役割を果たしている。幸い、大手企業が一般公開している「Google Translate™」、「Bing translator™」のような無料機械翻訳サービスの存在により、デジタル化されている情報を世界の多くの言語に訳すことができる。しかし、まだデジタル化されていない、ペーパー上などのテキストも数多く残っている。お店の看板、公園の案内図、病院のサービス説明書などを例として挙げられる。そこで、本研究は異文化の環境にいる外国人観光客と留学生などを身近にあるスマートフォンやタブレットを利用して、サポートするシステムの開発を目指している。例えば、携帯電話のカメラを使用し、ペーパー上のテキストを自動的に認識して翻訳させる方式もそのサポートの方法の一つである。以下ではその方式を備えている「リアルタイム文字認識エンジン Kanji Scan 3.0」の研究開発に関する説明をする。

2. リアルタイム文字認識エンジンとは

従来からでも外国語のテキストをスキャンして、文字認識を行い、機械翻訳をかける方法があった。しかし、現代社会では、作業を複数の段階に分けて行うことが遅れをとっているやりにかただと思われる。例えば、レストランを訪問する外国人がメニューのページ一つ一つをスキャン、文字認識、翻訳してから料理を選ぶ状態を考えられない。膨大な時間がかかるからである。他方、その3段階の作業が一瞬で終わることができる場合、その機能がレストランも例外ではなく、どこでも使用可能になる。その時間をとらずに、ほぼ即座に結果を出す作業が「リアルタイム」作業という。本研究で開発したリアルタイム文字認識エンジン「Kanji Scan 3.0」がスマートフォンのカメラを利用し、日本語のテキストをスキャンし、文字認識を行い、リアルタイムで翻訳結果を表示する外国人サポートシステムである。

3. リアルタイム文字認識エンジン「Kanji Scan 3.0」の開発

3.1 「Kanji Scan 3.0」の使い方

以下には「Kanji Scan 3.0」の使用手順を示す。

- (1) ユーザが「Kanji Scan 3.0」のアプリケーションをインストールして実行する。
- (2) スマートフォンの画面の上半分にリアルタイムでカメラからのビデオデータが表示される。
- (3) そのビデオデータのエリアの中心にある赤い目標枠を囲むようにして、読みたい単語の最初の文字に合わせる。
- (4) 自動的に漢字や単語の読み方が画面の半分が表示される。

- (5) 作業を止めずにリアルタイムで別の漢字に合わせて、そのデータを習得。

必要に応じて、赤いボタンをタッチして作業を一時停止することができる。目標枠が漢字を指さない限り、データ表示が更新しない仕組みになっている。

3.2 「Kanji Scan 3.0」のインターフェース

本アプリケーションのインターフェースは複数のボタンと二つの表示エリアで構成されている(図1参照)。

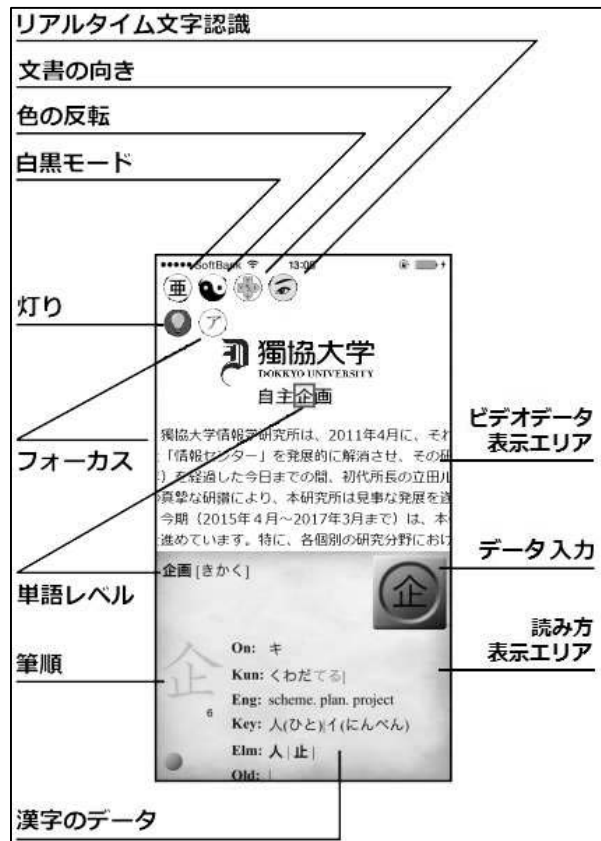


図1 「Kanji Scan 3.0」のインターフェース

スマートフォンの画面の上半分がビデオデータ表示エリアである。そのエリアではリアルタイムでカメラからの映像が白黒最高コントラスト化の状態に表示される。エリアの中心部には目標枠があり(図2参照)、その枠で取り囲む漢字から文字認識が開始する。必要に応じて、ビデオ入力を止める事も続けることもできる。

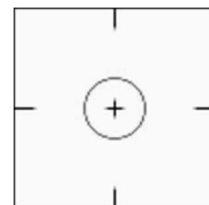


図2 「ビデオデータ表示エリア」の中心部での目標枠

画面の下半分が文字認識判定結果、及び翻訳結果と読み方表示エリアである。単語レベルの認識結果を示すと同時に、その単語の最初の文字の単字のレベルの結果も表示される仕組みである。

そのエリアでは認識された単語、その読み方、翻訳、及び漢字の筆順、音読み、訓読み、英語での意味などが表示される。筆順データがアニメーション化されているため、ユーザが認識された漢字の各線の書き方とその筆順の方向をリアルタイムで確認できる仕組みである。

各ボタンの役割が以下の通りである。

(1) 「白黒モード」ボタン

ユーザが手ぶれや他の認識エラーを目で確かめられるため、デフォルトではビデオデータが白黒最高コントラスト化されている。しかし、ユーザの希望により、普通のビデオデータ表示に切り替えることが可能になっている。この「白黒モード」ボタンがその切り替えモードボタンである。図3ではボタンのオンとオフ状態を示す。



図3 「白黒モード」ボタンのオンとオフ状態

(2) 「色の反転」ボタン

取得データの条件に応じて、テキストの背景を暗く、フォントが明るく印刷する。「色の反転」を使用するとビデオデータ表示エリアの全体の色が反転される。図4ではボタンのオンとオフ状態を示す。



図4 「色の反転」ボタンのオンとオフ状態

(3) 「文書の向き」ボタン

一般的には日本語の横書きの文書はヨーロッパ言語と同じ、左から右へ書かれている。但し、文学系のテキストでは縦書きが使われており、上から下への向きが多い。また、右から左への向きもあり、トラックの右側での広告や、看板がその例である。「Kanji Scan 3.0」でデフォルト設定は、自動的に左から右へのテキストと上から下へのテキストを同時に対象としている。しかし、ユーザの希望により強制的にテキストの向きの状態の設定を行うことができる。図5では「文書の向き」ボタンの各状態を示す。左からの順で、「横書き・縦書きモード」、「横書きモード」、「縦書きモード」、「右からの横書きモード」、「漢字単独モード」の状態の様子である。



図5 「文書の向き」ボタンの各状態

「漢字単独モード」は、漢字一文字だけの認識を行い、日本語学習者や研究者向きモードとなっている。単語レベルでの認識では形態素解析アルゴリズムも利用するので、一般的な辞書で登録されている単語を対象としている。しかし、時には古い形の珍しい漢字なども調べる必要がある。そのときは、形態素解析を一時的に止めることにより、漢字単独モードで普段使われていない字の認識を効率よく行うことができる。

(4) 「リアルタイム文字認識」ボタン

リアルタイム文字認識機能を一時的に止めるボタンである。その時、ビデオデータの入力がし続けられる。これが、ユーザがテキスト内で単語を検索するためのモードである。単語検索をより速く行うためには、ペーパー上の単語を指差すことが望ましい。その場合、ビデオデータ表示エリア上で単語の位置を見つけやすくなる。図6ではボタンのオンとオフ状態を示す。



図6 「リアルタイム文字認識」ボタンのオンとオフ状態

(5) 「灯り」ボタン

文字認識の正確度は、文字取込みに大きく影響を与える。効率の良い認識のためには、比較的明るい場所が必要だと思われる。しかし、事実上では、図書館や地下鉄、夜の街など、読み取りテキストが必ず明るい光を浴びているわけではない。幸い、現在のスマートフォンには光のフラッシュ機能を備えていることが多い。その機能のタッチモード（まぶしくない、中型光）を利用すると、テキストを輝かすことができる。図7ではボタンのオンとオフ状態を示す。



図7 「灯り」ボタンのオンとオフ状態

(6) 「フォーカス」ボタン

本や資料のテキストが小さいフォントで印刷されている場合に用いる。漢字を正確に認識するために一定の大きさが必要なので、スマートフォンとテキストの間の距離がとて短くなる。カメラのフォーカス距離より短くなると画像がぼやけるという問題がある。しかし機種にもよるが、現在のスマートフォンのカメラにはフォーカス自動合わせ機能がある。この機能を利用する場合、文字認識が正確に行える。図8ではボタンのオンとオフ状態を示す。



図8 「フォーカス」ボタンのオンとオフ状態

(7) 「ビデオ入力」ボタン

これは「ビデオ入力」のオン・オフを切り替えるボタンである。例えば、ユーザが現時点で表示されているデータを永く観察したい時、また、ページ全体を一回だけ撮り、複数の漢字と言葉を調べる場合、「ビデオ入力」のオン・オフにする必要がある。ビデオデータを止めた状態では、画面上のどの漢字でもタッチするとそのデータを確認できる。図9ではボタンのオンとオフ状態を示す。



図9 「ビデオ入力」ボタンのオンとオフ状態

図1で示すように、データが確認されている場合、ボタンの中央部に現時点で認識された漢字を表示する。

3.3 「Kanji Scan 3.0」の内部機能

主に「Kanji Scan 3.0」で設置されている機能や技術は、以下のように5つの部分に分けられる。

(1) ビデオデータ処理機能

現在、デスクトップスキャナと携帯端末のカメラで撮影する画像の質が大きく異なっている。携帯端末のカメラでは、手振れ、フォーカスの不具合、ノイズなど問題がある。文字認識の正確度を向上させるために、できるだけコントラストの高い画像を撮らなければならない。「Kanji Scan 3.0」では、最高のレベルのコントラストの白黒画像を使用している。ここでは、コントラストの画像をソフトウェアフィルターで作り出している。このアルゴリズムとしては、ビデオデータで取得した画像をグレースケールに変換してから、一番明るいピクセルと一番暗いピクセルを取得し、その平均をとる。最終的にはその平均より明るいピクセルを白色にして、平均より暗いピクセルを黒色に塗るという手法を使っている。このようなアルゴリズムによりできた画像は、手振れやノイズによるスキャンミスが消える。

(2) 文字認識エンジン

通常、OCRソフトでは各文字のデフォルトのマトリクスが使用されている⁽¹⁾。そのマトリクスの各ピクセルのデータを、テキスト上の漢字の領域と思われる画像の一部のピクセルデータと比較している⁽¹⁾。比較マトリクスの例を図10に示す。

認識率と計算量のバランスを考慮し、高速に文字認識を行うために、本アプリケーションでは8×8と32×32バイトの比較マトリクスを使用している。通常の白黒比較マトリクスと異なって、一バイトの「深さ」⁽¹⁾を備えているデータを使用した。

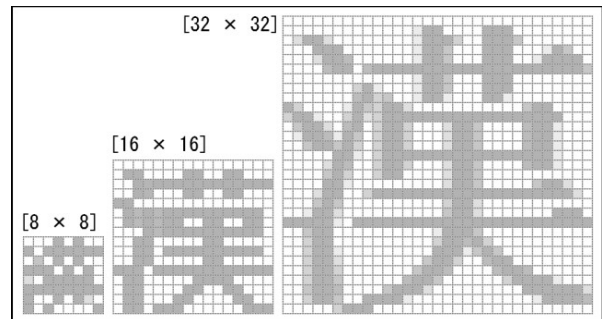


図10 「漢」の字の各サイズ比較マトリクス

(3) 形態素解析と単語レベル辞書

単語レベルでは認識率を高めるには、形態素解析を備える必要がある。万が一、文字認識に誤りが生じた場合、形態素解析を行うことで存在しない単語を見つける。そして、その間違った単語に一番近い正式単語を検索して、表示する仕組みを追加している。ここでは、形態素解析と単語レベル辞書として、国立国語研究所が開発した「UniDic」を使用した。

(4) 筆順データ

「Kanji Scan 3.0」は日本語学習支援システムとしても利用できる。日本語を効率よく学習するには、漢字の筆順を目で覚えさせるサポートツールの使用が有効である。しかし、現実には多くの筆順表示ツールが線の順番のみを示すものである。その場合、学習者が線の進む方向などを誤って覚えることもある。その問題の解決のためにアニメーションデータを導入し、線の順番と方向が示されるデータを作成した(表1参照)。筆順データのアニメーション作成ツールのメイン画面を図11に示す。



図11 筆順データのアニメーション作成ツールのメイン画面

表1 筆順データのスペック

字の数	7012 字
字の種類	カタカナ、ひらがな、ローマ字、ロシア文字、小学校で習う漢字、中学校で習う漢字、人名用漢字、常用漢字、JIS 第1水準漢字、JIS 第2水準漢字 など
キャンバス	200×200 ピクセル
データサイズ	~800 MB
元のフォント	株式会社リコーの「HG 教科書体」

筆順データを表示する場所をタッチすると本アプリケーションが茶色い各線を黒くならざるように少しずつ塗りつぶしていく。同時に線番号も表示する。

(5) 漢字辞典

本システムでは、7012 字の漢字を対象とした独自の漢字辞典を開発した。一般の漢字辞典と違って、音読み、訓読み、意味、部首と使用例だけではなく、複数の部首やエレメント、また 11 か国語に翻訳した意味などを備えているものである。複数部首を度入した理由は、以下のとおりである。筆者のような漢字圏以外の外国人にとっては、日本語の漢字の部首を見つけることが難しい。例えば、「宀」という字の部首は「心」または「宀」かが判断しにくい。ペーパー上の辞書ではどちらか一つを選んで使用しなければならないが、デジタルメディア上では両方を使用できる。また、複数の漢字から構成されている漢字の場合、その場で各エレメントを調べたい学習者が多い。例えば、「活」の字は「斗」と「舌」に分けられる。さらに、「舌」を「千」と「口」に分解ができる。各漢字のすべての部をすぐに調べられると、学習は楽に速く行うことが可能になる。「Kanji Scan 3.0」では漢字の各部をタッチすることで、その部を独自の字として表示する機能を備えている。他には字の古いスタイルと同類文字情報も調べられるようになっている。漢字辞典を作成するためのツールを独自に開発した(図 12)。

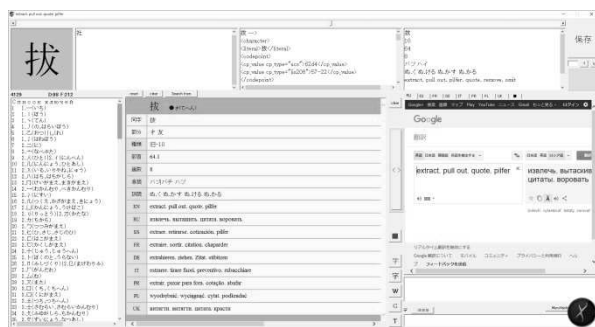


図 12 多言語に対応漢字辞典作成ツールのメイン画面

3.4 リアルタイム認識及び出力

3.3 で説明した機能で使用するデータの量は膨大であり、一般的なやりかたで処理を行うとかなりの時間を要する。そこで様々な工夫をこらしリアルタイム化ができるような工夫をした。様々な実験を行い、以下のように一番速い処理の方法を検出した。

まず、単語辞書、漢字辞典、筆順データを各漢字の UNICODE のコードを利用して細かく分けた。そのコードの Hi バイトでフォルダを作成し、Low バイトでサブフォルダを作った。それぞれの辞書のデータをそのサブフォルダに保存し、現時点で認識されている漢字のデータのみを読み上げて表示する方法を選択した。これらのアルゴリズムにより、ビデオデータのコントラスト化から、単語及び漢字データを表示する段階までの整理平均時間を 2 秒まで短縮することができた。事実上ではユーザがデータを読み上げる時間も必要なので、現時点ではデータの更新は 3 秒毎に行われる。すなわち、リアルタイムでデータを表示できるアプリケーションの開発としては成功している。

4. 各技術の研究、テスト実験と問題解決

リアルタイム化のため、膨大な処理作業をできるだけ速く行うことが必要である。本章では処理加速のための工夫と新しいアイデアについて述べる。

(1) ビデオデータ処理の新技術

一般的には画像を最高のレベルのコントラストにするために、すべてのピクセルの明るさを計り、最低レベルと最高レベルの平均をとる方法が使われている。しかし、画像が 1080×1080 ピクセルの大きいもの場合、その計算の仕方が非常に遅い。本研究では、様々なパターンを確かめながら、スマートフォンのカメラの画像での光マップの観察を行い、その特徴について調べた。図 13 ではその光マップの主なパターンを示す。

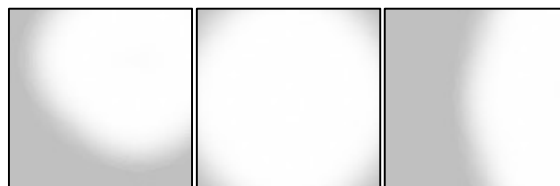


図 13 カメラの画像の光マップのパターン

光マップを考慮して、すべてのピクセルではなく、左上から右下の角までと右上から左下の角まで二つの線に並んでいるピクセルだけの光の明るさだけを計っても、従来の平均値の計算と結果がほとんど変わらないことが分かった。この方法により、この部分の処理スピードが 1000 倍ほど上がった(表 2)。

表 2 入力画像最高コントラスト化実験結果(単位: ミリ秒)

画像のサイズ	640 × 640	720 × 720	1080 × 1080
全てのピクセル	568	826	1244
4 角の 100×100 ピクセル	26	38	52
× 印の 2 線 ピクセル	5	8	11

(2) 文字認識エンジンの新技術

通常の文字認識エンジンでは、16×16 の比較マトリクスを備えている。しかし、携帯端末のカメラの画像の画質を対象に、それ以上の認識率が高いものを備えるべきである。文字認識エンジンで、32×32 バイトの比較マトリクスを使用すると認識率が高いが、7012 字を対象とすると最新の iPhone6 でも 20 秒もかかる処理となる。そこで、本研究では 2 種類の比較マトリクスを備えた。最初は 8×8 の認識率が低いマトリクスを使って 7012 字のリストから 20 字ほどを割り出す。その後、32×32 バイトの比較マトリクスで正確に 4 文字を判定する形をとった。最後は形態素解析の結果により適切な一文字を選択するアルゴリズムを決めた。このアルゴリズムにより、処理スピードが約 120 倍上がった(図 14)。

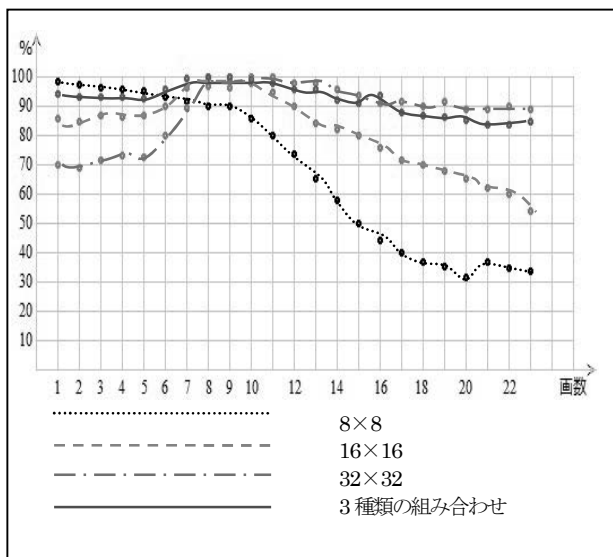


図 14 各サイズのマトリクスの備えた文字認識比較実験

(3) 単語辞書使用の新技術

形態素解析で使う辞書の UniDic のために、122MB のメモリを用意しなければいけない。単語の数が増えるほどメモリ容量も拡大化する。そして当然ながら、処理スピードも遅くなる。本研究では、辞書を文字毎に分けることにより、100KB 以下のメモリの使用に成功した(図 15)。

5. 同レベルの市販ソフトとの比較

ここでは、携帯端末用 OCR 技術を備えている市販ソフトとの比較実験を行った。その結果を表 3 でまとめた。世界各地では、同レベルソフトの開発が行われている。しかし、リアルタイムで日本語の単語辞書と漢字辞典を同時に備えたものは本アプリケーションのみである。短所としてはデータ更新時間が比較的に大きいことが挙げられる。このことより、処理スピードの向上が今後の課題として考えられる。

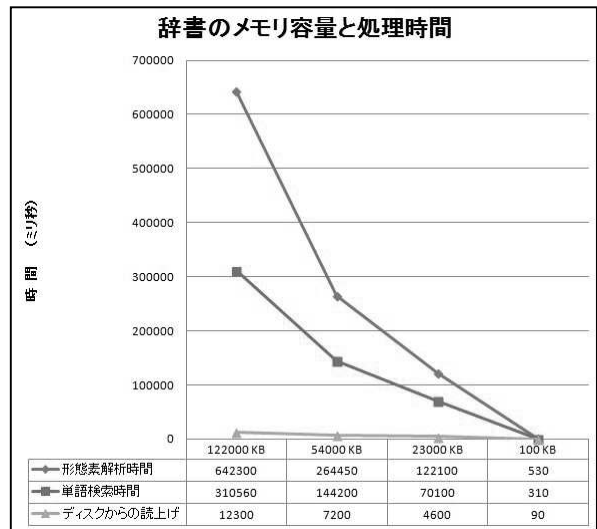


図 15 辞書用メモリの処理時間への影響の比較実験

表 3 市販のソフトとの比較

ソフト名	単語 OCR	漢字 OCR	リアルタイム	更新時間 (秒)
J-OCR	○	×	×	4
Pleco	○	×	○	2
OCR Scanner	○	×	×	4
Aedict	×	○	×	1
Japan Googles	○	×	○	2
Kanji Scan 3.0	○	○	○	3

6. リアルタイム文字認識エンジンのニーズ

人気のあるソーシャルネットワークの一つであるフェイスブック上で、201 人対象としたリアルタイム文字認識エンジンのニーズに関するアンケート調査を行った(図 16)。

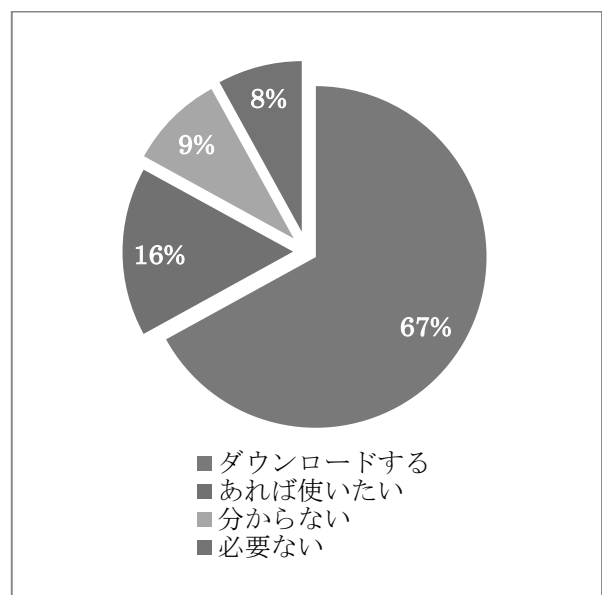


図 16 日本語対処リアルタイム文字認識エンジンのニーズに関するアンケート調査の結果

近年、日本のアニメ、漫画、コスプレ文化、J-POPなどが世界の若者の間に広く人気を得ている。また、2020年に開催される東京五輪中に、日本訪問を予定している外国人が増えている。そのため、日本語を勉強している学習者も増加している。これらのことから、本アプリケーションのようなソフトのニーズが拡大化するものと思われる。

7. おわりに

本稿では、リアルタイム文字認識エンジンの研究開発について述べた。様々な実験と調査を行い、エンジンの内部で使われる複数の機能の技術を発見した。また、そのエンジンのニーズに関する調査を行い、良好な結果を得た。このことより、全体的にリアルタイム文字認識エンジンの開発に成功したと思われる。将来的に、幅広い分野でこのエンジンの技術が使われることを期待したい。

謝辞

本研究の一部は、情報科学研究所研究助成によるものである。

参考文献

- (1) ベロフ・アレクサンドル, “ワンタッチエンジンの構造と解析アルゴリズム”, 獨協大学情報センター「情報科学研究」, 第22号, pp.1-12(2004)
- (2) 中野 康明, 花野井 歳弘, 丸山 稔, 宮尾 秀俊, 丸山 健一, “複数の文書理解システムを用いた文書理解の高度化(文字とドキュメントの認識・理解)”, 電子情報通信学会技術研究報告, PRMU, パターン認識・メディア理解 103(659), pp.55-60(2004.2)
- (3) 熊谷 勝彦, 鈴木 真一, 上野 浩司, “OCRの認識率アップ法とそのシステムの簡素化”, 全国大会講演論文集 第42回平成3年前期(2), pp.104-105(1991.2)
- (4) ベロフ アレクサンドル, 立田ルミ, “漢字 OCRシステムでの認識率の向上方法と考察”, 獨協大学情報学研究所「情報科学研究」, Vol.1. pp.60-64(2012.2)

(2015年9月29日受付)

(2015年12月2日採録)