

2008 年度 卒業研究

移動する人間の状況に基づく情報提供サービスの検討

担当教員 渡辺 恭人

千葉商科大学政策情報学部

学籍番号：0644193

倪 玮炜

目次

1. 背景と目的	4
1.1. 背景	4
1.2. 目的	4
2. 現状と問題点	5
2.1. 日本を訪問する外国人の現状	5
2.2. 交通機関における国際化の現状	7
2.2.1. 駅の状況	7
2.2.2. ホームページの状況	19
2.3. 行政機関における国際化の現状	22
2.4. 日本を訪問する外国人のニーズと国際化対応の問題点	24
3. 解決法	25
3.1. 解決へのアプローチ	25
3.1.1. 外国人にとっての障壁	25
3.1.2. 必要な機能のモデル	25
3.2. 解決法の提案	26
3.3. 期待されるサービスイメージ	27
3.4. 関連技術	29
4. 設計	31
4.1. 設計目標	31
4.2. 必要となる情報	31
4.3. 構成要素	31
4.4. 機能概要	32
4.5. データベース	34
4.5.1. 日本語漢字・ピンインテーブル	34
4.5.2. 漢字地名・ローマ字テーブル	34
5. 実装	35
5.1. 実装環境	35
5.1.1. クライアント環境	35
5.1.2. サーバ環境（学外： www.cyaneum.org ）	35
5.1.3. プログラミング言語	35
5.2. データの準備	35
5.2.1. ピンインと日本の漢字の対応データ作成	35
5.2.2. 日本の地名とローマ字のデータ	37
5.3. データベースの作成	38

5.3.1.	データファイルの転送.....	38
5.3.2.	データベースの設定	38
5.3.3.	「ピンイン⇄漢字」テーブルの作成.....	39
5.3.4.	「漢字地名⇄ローマ字」テーブルの作成	41
5.4.	各機能の実装.....	43
5.4.1.	ピンイン入力機能	43
5.4.2.	ピンインデータ受信機能.....	44
5.4.3.	ピンインから漢字を検索する機能	45
5.4.4.	漢字の選択機能とピンインの継続入力機能.....	48
5.4.5.	漢字から地名を検索する機能.....	52
5.4.6.	ピンイン入力機能から地名検索機能に地名文字列を送信する.....	54
5.4.7.	地名ローマ字検索機能に地名文字列受信機能を追加	55
6.	動作の検証と評価.....	58
6.1.	動作検証.....	58
6.1.1.	実装された機能の動作結果（PC）	58
6.1.2.	実装された機能の動作結果（携帯）	59
6.2.	一般的な利用者（中国人）による評価	62
7.	まとめ.....	64
	参考文献	65
	謝辞.....	66

1. 背景と目的

1.1. 背景

最近数年、中国と日本の関係が政治的・経済的に密接になるにつれて、人の交流も多くなる。留学、観光、ビジネスなどのさまざまな目的で、中国から日本に来る人は今後増加することになる。

中国から日本に来た外国人として、当然言葉の問題がある。特に、移動する場合の交通において、言葉の問題は非常に大きい。多くの外国人にとって日本語の地名は難しい。

日本を訪れる外国人のうち中国人は、日本語が同じ漢字を用いているという点で、比較的有利である。なぜなら、日本語の中の漢字の 90%以上の意味がわかるからである。しかし、日本語の漢字の読み方は中国の漢字の読み方と全く異なる。したがって、漢字の意味がわかっても読み方はわからないのである。地名や駅名では特に読み方が重要である。読み方がわかれば、他人に聞いたり、コンピュータや情報端末に入力することができる。

日本は情報通信の技術が高度に発達、普及している。携帯電話の機能はとても豊富で強力である。GPS から取得できる現在位置を入力し、ユーザの目的地を入力することにより、最適なルート案内を取得することと現地までのリアルタイムなナビゲーションができる。しかし、このような高度なサービスは、日本語を母国語としている人や日本語に慣れている人にとっては使いやすいが、それ以外の外国人が使うことは困難である。目的地の日本語の読み方がわからないため、入力できないからである。日本を訪れる外国人にとって、このような問題点が多く存在している。

1.2. 目的

本研究では、日本を訪れる外国人として特に中国人を対象とし、公共交通機関等で移動する場合の問題点を明らかにし、その解決法を検討し、インターネット上のサービスとして提供することを目指す。

具体的には、中国語のローマ字表記であるピンインによって、日本の漢字を表示し、単語の読み方情報を提供する機能を、設計・実装する。

本サービスを実現することにより、日本にいる中国人に対して、移動を支援し、移動における利便性をより高めることができる。

2. 現状と問題点

2.1. 日本を訪問する外国人の現状

法務省入国管理局の統計によると、「2006 年（平成 18 年）の外国人入国者数は、日本政府のビジット・ジャパン・キャンペーンや 2005 年 3 月からの韓国人及び台湾居住者に対する査証免除措置、及び中国に対する査証発給条件の緩和の効果等があいまって、2005 年比 8.8%増の 810 万 7963 人で過去最高となった。」という（図 2-1）。

2006 年末現在の外国人登録者数は、中・長期的に生活を送る者が増加し、2005 年比 3.6%増の 208 万 4919 人、総人口に占める割合も 1.63%で過去最高を更新した。

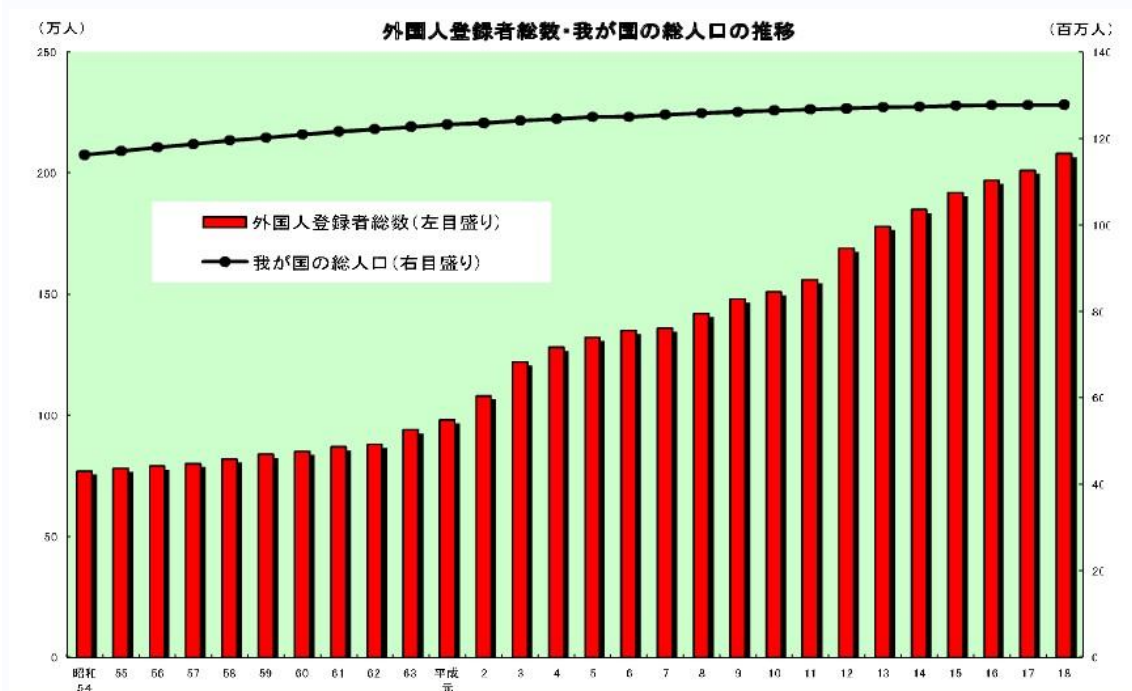


図 2-1 : 2006 年までの外国人登録者数

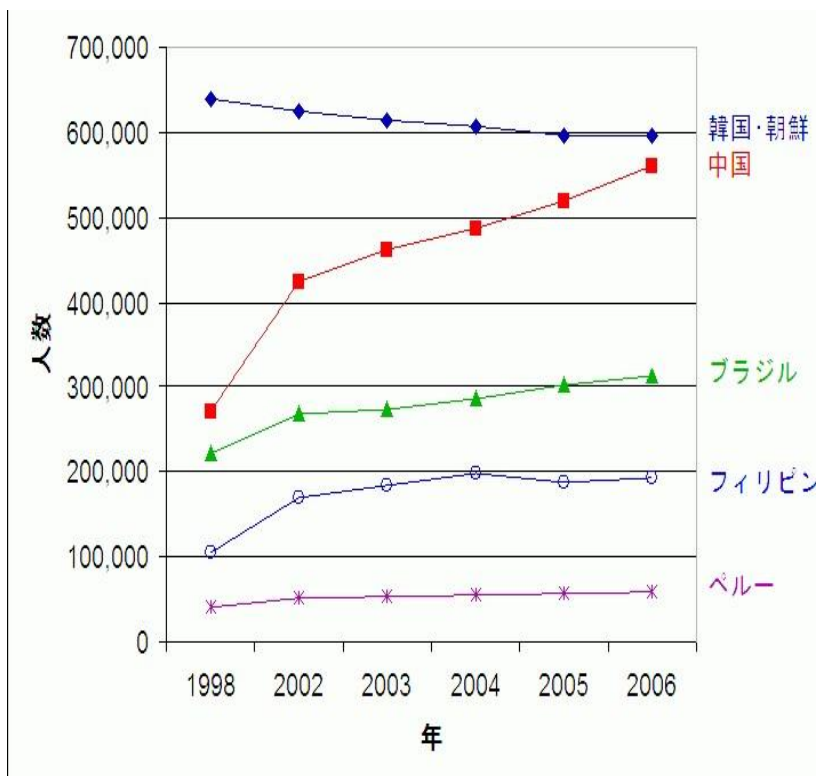


図 2-2：日本における外国人の人数の推移

2006年末現在の日本における
国籍別外国人登録者数

国籍	人数	構成比
韓国・朝鮮	598,219	28.7%
中国	560,741	26.9%
ブラジル	312,979	15.0%
フィリピン	193,488	9.3%
ペルー	58,721	2.8%
米国	51,321	2.5%
その他	309,450	14.8%
合計	2,084,919	100%

図 2-3：2006 年末現在の日本における国籍別外国人登録者数とその構成比

図 2-2 および図 2-3 の国籍別の外国人登録者数から、この傾向が継続すると、近い将来には中国人の比率最も高くなることが予想される。

2.2. 交通機関における国際化の現状

ここでは、日本の交通機関における、国際化対応の現状について、調査し、分析する。

2.2.1. 駅の状況

今の日本の交通機関に関して、まだ十分な国際化ができていない。交通機関では「文字に頼らない表示をしている」のが理想的な国際化であるとされている。文字に関してはアルファベットなら外国人も読めるとは限らず、例えば、中国大陸出身の人の中には、漢字は読めてもアルファベットは読めないという人が結構いる。

駅名・近隣施設などの案内看板を多国語表示にすることや、トイレ、車椅子、云々を表示する国際的統一表示アイコン等の「Universal Design」も含めるべきである。「多国語言語表示」と言っても、英語、中国語、韓国語だけで、実際は殆どの用が足りるわけではない。しかし、これ以上別の言語(仏、独、西、蘭、葡等)も書くと、看板だけが大きく見難くなるだけである。

本節では、交通機関における国際化の現状を、駅を対象として、以下のようないくつかの項目について調査した。

(1) 案内

駅構内においては、行く先や利用する路線によって、向かうべきホームが異なることがあり、そのための文字情報による「案内」が最も必要である。案内がどのように国際化されているかを調査した。

(2) 放送

駅構内では、固定的な情報である案内よりもリアルタイム性が求められる情報が多い。その場合は、文字よりも音声のほうが即時性があり、伝わりやすい。この音声による放送において国際化がどのように行われているかを調査した。具体的には、他言語による放送が行われているかである。

(3) 駅員（スタッフ）の対応能力

日本人であろうと外国人であろうと、案内でも放送でもわからなければ、駅員等のスタッフに直接尋ねるしかなくなる。この場合に、外国人と情報の伝達ができるかどうかを調査した。具体的には、日本語以外の言語を使えるスタッフがいるかどうかである。

① JRの場合

(a) 東京駅

案内	英語、中国語、韓国語
放送	日本語、英語
駅員の対応能力	中国出身の駅員がおり、英語と中国語の対応ができる



図 2-4：東京駅内の看板

東京駅内の案内（図 2-4）は、日本語、英語、中国語、韓国語、四種類の言葉を使うので、外国人でも分かりやすい。

(b) 神田駅

案内	英 語、中国語、韓国語
放送	日本語だけ
駅員の対応能力	日本語と英語（ほとんど日本語）



図 2-5：神田駅の南口

出口付近の町の案内（住所表示）（図 2-5）は、数も多く、日本語だけが記述されている。外国人には理解できない。

(c) 市川駅

案内	英 語、中国語、韓国語
放送	日本語だけ
駅員の対応能力	日本語と英語（ほとんど日本語）



図 2-6：市川駅内の総武線の看板

ホームへ向かう案内表示である（図 2-6）。各駅停車の部分が「Local Line」と英語化されている以外は日本語表記である。Local Line という単語で日本語の各駅停車の意味が伝わるかどうかにも気になるところである。



図 2-7：市川駅の1番線のホーム

図 2-7 は、市川駅の1番ホームにある案内板である。日本語と英語による表

記となっている。

JR 市川駅は、他の路線が接続する乗換駅ではないが、各駅停車と快速の 2 通りの電車が利用できる駅である。この乗り分け方は、日本人でも最初はわかりにくいらしい。市川市の外国人人口が増えていることを考えると、日本語、英語以外の言語への国際化対応が必要である。

(d) 秋葉原駅

案内	英 語、中国語、韓国語
放送	日本語、英語
駅員の対応能力	英語ができる



図 2-8 : 秋葉原駅内の看板

図 2-8 は秋葉原駅内の案内板である。英語に加えて、中国語、韓国語での表記が見られる。

(e) 上野駅

案内	英 語、中国語、韓国語
放送	日本語、英語
駅員の対応能力	英語ができる



図 2-9 : 上野駅不忍口

図 2-9 は、上野駅の不忍口の入り口である。英語表記のみである。



図 2-10 : 上野駅公園口

図 2-10 は、上野公園口の案内板である。英語に加えて、中国語、韓国語での表記が見られる。駅名は英語表記のみである。



図 2-11：上野駅不忍口内の案内板

図 2-11 は、駅構内にある案内板である。改札の方向と、のりば、トイレ、旅行サービス、切符売場などの案内が、日本語、英語、中国語、韓国語などの文字表記に加えて、図で表示されている。ひとつの案内板に盛り込みすぎのようにも感じるが、まずまずわかりやすいといえる。



図 2-12：ホームの階段付近に設置された案内板

図 2-12 は、ホームの階段付近に設置された案内板である。文字以外に図形での案内がなされている。階段付近にあるので、この案内板から上に行くのか下に行くのかは問題なく、階段があることが重要である。

この案内板では、日本語、英語、韓国、中国語の 4 ヶ国語で案内が行われており十分伝わると考えられる。不忍口に行きたい場合はわからない。

このように同じ上野駅内でも、案内の国際化対応に違いが見られた。外国人にとって必要な情報を併記しないと理解されないし、案内板ごとの表記に一貫性がなければ、かなりのストレスになる。統一されたフォーマットで作成するなどの工夫が必要である。



図 2-13：駅の外に設置されたきっぷの自動販売機

図 2-13 は、駅の外に設置されたきっぷの自動販売機である。ご覧の通り、日本語の表示だけしかない。外国人にとって、きっぷを買うことが非常に不便である。

② 京成電鉄の場合

(a) 京成上野駅

案内	英 語、中国語、韓国語
放送	日本語
祖度駅員の対応能力	英語ができる

京成電鉄の上野駅は、出発点であり、地下鉄や JR と連絡するターミナル駅である。観光地浅草に近いこともあり多くの外国人観光客が利用する。また、成田空港と往復するスカイライナーの発着駅でもあるので、外国人の利用者は多くなる。



図 2-14 : 京成上野駅の入口付近の案内板

図 2-14 は、京成上野駅の入口付近の案内板である。画面上部に日本語、下部に英語のみの表記である。



図 2-15 : 京成上野駅の入口の案内板

図 2-15 は、京成上野駅の入口の案内板である。日本語表記のみであり、肝心の英語表記はない。成田空港に行くスカイライナーのみが英語表記されている。



図 2-16 : 改札を通過した後の列車の出発案内

図 2-16 は、改札を通過した後の列車の出発案内である。電光掲示板による表示で、メッセージは、日本語と英語が定期的に切り替わる形式である。その他の言語での表示はない。



図 2-17：ホーム内に設置された案内板

図 2-17 は、ホーム内に設置された案内板である。日本語と英語のみによる表記である。



図 2-18：スカイライナー専用の切符売場への案内板

図 2-18 は、一般の切符売場と成田空港に行くスカイライナー専用の切符売場への案内板である。スカイライナーを利用する外国人を考慮して、この案内板は、日本語、英語、中国語、韓国語で表記されている。



図 2-19：出口に設置された案内板

図 2-19 は、出口に設置された案内板である。京成上野駅では、スカイライナー切符売場案内板とこの案内板が多言語表記であった。

京成電鉄のターミナル駅であるにも関わらず、案内板は日本語のみの古いものが多く、また日本語以外の言語への対応は JR よりさらに一貫性がなく、中途半端である。すべての案内板を見直すべきである。

駅員の対応では、英語を話することができる駅員がおり、最低限英語を話す外国人への対応はできるようである。

(b) 京成国府台駅

案内	英 語
放送	日本語
駅員の対応能力	日本語だけ

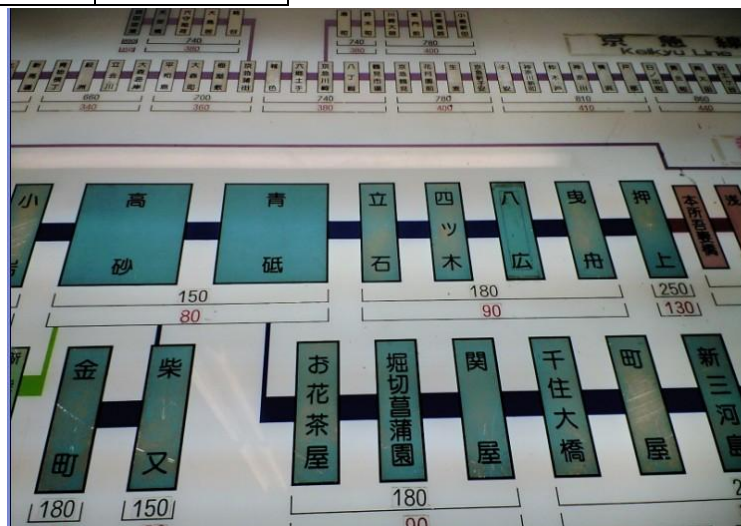


図 2-20：京成線の駅名および運賃が表記された案内板

図 2-20 は、駅名および運賃が表記された案内板である。日本語のみであるが、かなり駅数が多いこともあって、一つの図に他の言語を表記することは難しそうで

ある。別の方法が必要である。



図 2-21：国府台駅のホームへの案内板

図 2-21 は、国府台駅のホームへの案内板である。日本語と英語で表記されている。

京成国府台駅は、京成電鉄の京成本線にある各駅停車の電車が利用できる駅で、一般的な駅の一つである。外国人の観光客が利用することはまずないと考えられるが、市川市自体に中国人をはじめとする外国人が増えていることを考えると、国際化が今後必要になると考えられる。

駅における国際化対応のまとめ

駅における国際化の現状を表 2-1 に整理した。日本の交通機関に対する調査を通して、乗換駅やターミナル駅などの大きな駅が良い外国語のサービスを提供していることがわかった。しかし、中小の駅は日本語以外の外国語サービスをよく提供してない。たとえば、千葉商科大学に近い市川駅の場合、中国語ができる駅員（スタッフ）はいない。英語もそんなにうまくないようである。実際は市川で生活する中国人の人数は多いため、彼らが日本語を習得するのを待つのではなく、国際化の達成を急ぐべきである。

表 2-1：J R 駅と私鉄京成線の駅における国際化の現状

駅名	J R 東京駅	J R 神田駅	J R 市川駅	J R 秋葉原駅	J R 上野駅	京成 上野駅	京成 国府台駅
案内	英 語、中 国語、韓 国語	英 語、中 国語と韓 国語（少 ない）	英 語、中 国語、韓 国語	英 語、中 国語、韓 国語	英 語、中 国語、韓 国語	英 語、中 国語、韓 国語	英 語
放送	日 本 語、 英語	日 本 語	日 本 語	日 本 語、 英語	日 本 語、 英語	日 本 語	日 本 語
駅員	英語、中 国語がで きる	ほとんど 日本語	ほとんど 日本語	英語がで きる	英語がで きる	英語がで きる	日本語だ け

2.2.2. ホームページの状況

ここでは鉄道会社のホームページにおける国際化対応について調査分析した。

(1) J R 東日本のホームページ

J R 東日本の国際化されたホームページを図 2-22 に示す。



図 2-22：J R 東日本の中国語のホームページ

このホームページは、表紙の画面上部にある言語を選択することで表示される。英語、中国語（2 種類）、韓国語に対応しており、それぞれ選択すると各国語のページに切り替わる。ブラウザ情報から自動的に切り替わるかどうかは確認していない。日本語のコンテンツはすべて翻訳して提供されているようである。ただし、路線や乗換、運行時刻等に関する情報を見つけることができなかった。自宅等でパソコンからアクセスして利用する場合でも、自分が行きたいところにどのように行くべきかの情報は、このホームページからは得られない。もちろん、インターネットで検索すれば、乗換案内のようなサイトを発見し利用することはできるかもしれないが、そこで言葉の問題にぶつかることが予想できる。したがって、単純に翻訳したからといって、日本に来たばかりの外国人にはあまり意味がないと言える。

同社の広報担当者に国際化対応の考え方について問い合わせた結果、表 2-2 のような返答を頂いた。

表 2-2：国際化対応についての J R 東日本の応答

外国からのお客さまへは、電話による通訳案内サービス「J R イーストインフォライン」（電話 050-2016-1603 時間 10:00～18:00 -除く年末年始-）を行っております。

従来は英語のみでしたが、2001 年 10 月よりハングル・中国語によるサービスも開始いたしました。

現在、駅案内サインへのハングル・中国語の併記について整備を行っているほか、東京駅や新宿駅などに「インフォメーションセンター」を設置し、英語による対面でのご案内も行っております。

また、J R 東日本ホームページにおいては、外国からのお客さまへのご案内のページを以下のとおり設けておりますのでご覧下さい。今後も、外国からのお客さまへの案内を充実させていく予定です。

- 英 語 <http://www.jreast.co.jp/e/info/index.html>
- ハングル <http://www.jreast.co.jp/kr/index.html>
- 中国語 <http://www.jreast.co.jp/sc/index.html>

このように、電話での翻訳サービスや、対面での案内の強化などの即時性の高いサービスを拡充するなどの姿勢があり、評価できる。たとえば、この通訳案内サービスであるが、このサービスを利用してもらうためには、さまざまな形で周知をしていく必要があるが、その方法についての説明はなかった。成田エクスプレスに乗車した際に、これらのサービスに関する情報を車内で探してみたが、見つけることはできなかった。知ってもらわなければ利用されることはない。

(2) 京成電鉄のホームページ

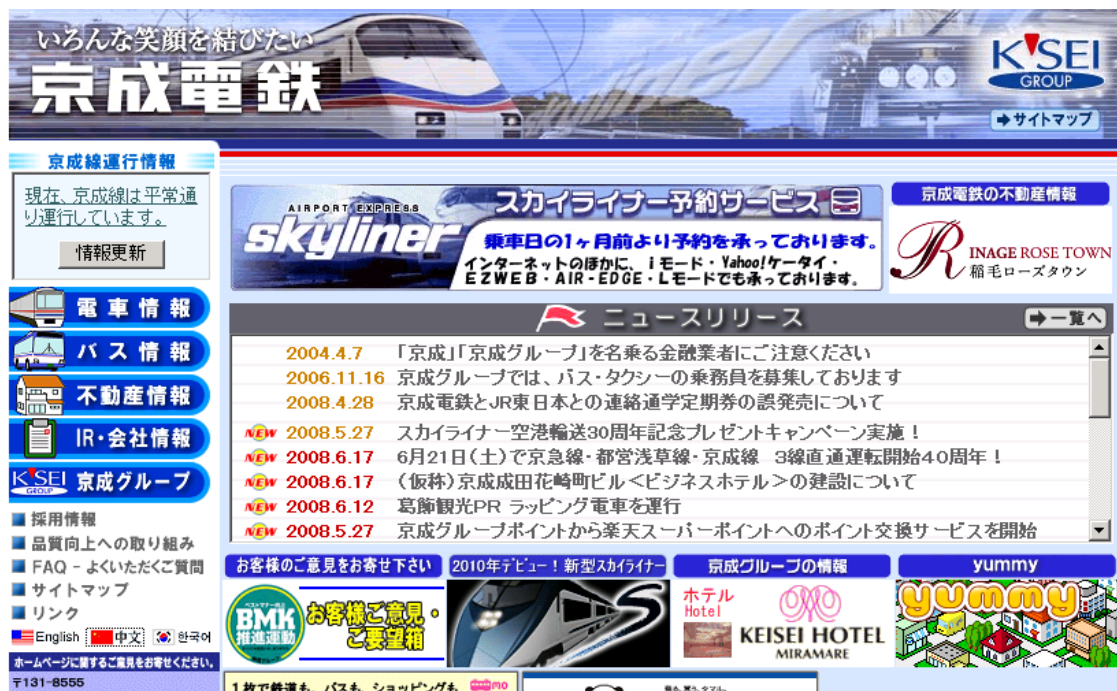


図 2-23：京成電鉄のホームページ



図 2-24：京成電鉄の中国語のホームページ

京成電鉄のホームページを図 2-23 と図 2-24 に示す。図 2-23 が京成電鉄のホームページの表紙である。画面左下に英語と中国語と韓国語のページへのリンクがある。ここをクリックすると、スカイライナーの情報ページが表示される。外国人であれば、スカイライ

ナーを利用する可能性は高いと言えるので、このような情報提供の仕方はニーズがあると考えられるが、通常は、日本語のホームページと同じ内容を、英語、韓国語、中国語に翻訳したものを用意するものであるし、利用者によっては、必要な情報にたどり着けない可能性がある。

ホームページにおける国際化対応のまとめ

JRにしても京成電鉄にしても、日本語のページを単に各国語に対応したことで終わっているように見える。このような取組は無駄ではないが、日本に来る前にどれくらいの外国人がこのページを見てから来るかどうかを考えるとかなり疑問である。実際には、成田空港に到着してすぐ、または上野駅についてすぐに、どう行動すべきかの情報が必要となるのであり、インターネットをパソコンで利用して情報にアクセスすることはあまり考えられない。利用者の利用形態や状況を考慮して、利便性を高める必要があると考えられる。

2.3. 行政機関における国際化の現状

行政機関のホームページの例として、市川市を取り上げ、その国際化状況を調査した。市川市は、一地方自治体であるが、住民にとってはもっとも身近な行政機関である。また、外国人にとっても「外国人登録」などの手続きや、生活上のさまざまな情報を得るために必要な役所である。特に、中国人をはじめとする外国人が多く住む市川市が、インターネットのホームページというメディアにおいて、どのように国際化して、情報提供しているかを確認する。



図 2-25 : 市川市市役所のホームページ

市川市欢迎您

欢迎大家来市川市！你们在日本生活得怎么样？离开住惯了的祖国在这里生活，由于语言和习惯的不同会产生种种不安和烦恼吧。为了多少能帮助大家消除一些不安和烦恼，每天过得更愉快，更舒适方便，市川市发行了这本外语版便民手册。手册中介绍了与大家关系密切的市役所的业务指南、公共设施、市内的情况等。希望大家把它作为熟悉市川市的信息来源随时加以利用。

ようこそ市川市へ

皆さん、市川市へようこそ！日本の生活はいかがですか。住みなれた母国を離れての生活では、言葉や習慣の違いからさまざまな不安や悩みが生じることと思います。そこで市川市では、皆さんの不安や悩みを少しでも解消し、毎日をより楽しく、快適に過ごせるように、この市民便利帳外国語版を発行いたしました。皆さんの日常生活に関わりの深い市役所業務の案内や、公共施設、市内の様子などをご紹介します。市川市を知る身近な情報源としてご活用いただければ幸いです。

図 2-26：市川市市役所の中国語のホームページ

図 2-25 では、市川市のホームページの表紙を示している。画面上部には、中国語、韓国語、英語、スペイン語、ドイツ語、インドネシア語へのリンクがある。クリックすると、図 2-26 や図 2-27 のようなページが表示される。

中文

- ▶ [市川市概要\(PDF\)](#)
- ▶ [市川生活指南\(PDF\)](#)
- ▶ [税金\(PDF\)](#)
- ▶ [垃圾\(PDF\)](#)
- ▶ [我们热爱市川市](#)

한국어

- ▶ [이치카와시\(市川市\)의 개요\(PDF\)](#)
- ▶ [생활핸드북\(PDF\)](#)
- ▶ [세금\(PDF\)](#)
- ▶ [쓰레기\(PDF\)](#)
- ▶ [우리는 이치카와시를 사랑합니다](#)

Español

- ▶ [Resumen de la Ciudad de Ichikawa\(PDF\)](#)
- ▶ [Manual de Preguntas y Respuestas\(PDF\)](#)
- ▶ [IMPUESTOS](#)
- ▶ [BASURA](#)

図 2-27：市川市のホームページの国際化対応

市川市のホームページにおける国際化対応では、日本語でのコンテンツとは別に各国語を利用する人々のために別のコンテンツを用意している。外国人が日本に来て、市川市で生活するために必要な情報を PDF ファイルにまとめている。しかし、本来ならば、日本語と同じコンテンツを各国語に訳して提供することが一貫性を保つという点では必要なことである。誰が見ても同じ情報にアクセスできることが大切である。外国人がある程度市川市に慣れた場合は日本人の市民と同様の情報を必要とすることになるが、まだ日本語がよくわからない場合は、必要な情報を得られないことになってしまうという問題がある。

2.4. 日本を訪問する外国人のニーズと国際化対応の問題点

日本のことを詳しく知らない外国人にとって重要なのは、ストレスなく目的地に移動できることである。目的地を決めても、どのように行動すればよいかわからないことがほとんどであるからである。日本語がわかる人たちは、事前に yahoo や Google で検索すれば、情報を得ることができるが、日本語ができない人の場合、とても難しい事である。

日本の交通機関において、例えば、私のような中国人であれば、日本の漢字の 90%以上の意味がわかる。したがって、苦手ではないと考えられる。しかし、日本語の漢字の読み方は中国の漢字の読み方と全く異なる。漢字を使う中国人ならまだしも、アルファベットしか使わない欧米人にとっては日本語のみの案内板は意味不明な文字の羅列に過ぎない。また、日本語を勉強しようとしている韓国人では、ひらがなを先に覚えてしまう人が多く、読みやすいひらがなは大きな情報となる。最低限でも、案内版における国際化はすべての駅で行うべきであり、各外国人のニーズに合わせた対応も必要となると考えられる。

3. 解決法

この節では、2 章で述べた国際化対応の現状と外国人にとっての障壁について触れ、外国人にとって必要な情報とその仕組みについて検討し、解決方法を探る。

3.1. 解決へのアプローチ

3.1.1. 外国人にとっての障壁

日本語を学習している外国人にとって漢字の日本語読みは、最大の壁である。日本語での日常会話に問題はなくとも、「日本語で書かれたサイトを読むことが出来ない」「送られてきたメールを読むことができない」といったことが多くある。同じ漢字文化圏の人でも、「文章の意味はわかるけど、日本語での発音がわからない」ということもよく耳にする。

分析した結果によって、外国人にとって最適な移動情報とは、まずはもちろん目的地の名称である、その後は目的地に行くための必要な乗り換え方法である。

このような言葉と文字の障害を解決するため、自分の国の言語と文字を使い、できるだけ短い時間で必要な情報を手に入れることを目指す。このようなシステムによって、日本語ができない人でも日本語の入力を行うことができるようになる。

3.1.2. 必要な機能のモデル

以上のような仕組みを携帯電話で利用することを目指す。図 3-1 にこの仕組みのイメージを示す。この仕組みの特徴は、ユーザ（外国人）が熟知している文字の入力法を利用して、日本語に変換することである。日本語ができないユーザでも日本の文字の入力が可能になる。

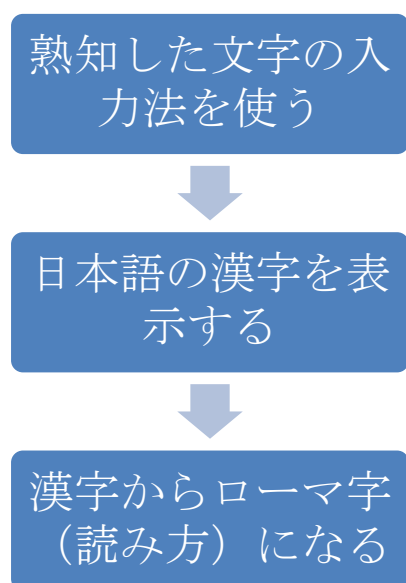


図 3-1：外国人が熟知している文字の入力法を利用して、日本語を読む仕組みのモデル

3.2. 解決法の提案

この解決のアプローチから、日本にやってくる多くの外国人のうち、特に多い中国人の移動を支援する情報サービスを検討する。筆者自身が中国人であり、中国人は日本の漢字の90%以上の意味がわかるため、このアプローチを採用することで他の言語より容易に支援ができると考えたからである。しかし、日本語の漢字の読み方は中国の漢字の読み方と全く異なる。だからまずこの仕組みでは中国のピンインの入力法を実現する必要がある。ピンインの入力法を使うことにおいて、自動的に日本語の漢字を変更することができる。しかも中国人のユーザがこの漢字を選定した後で、この漢字の日本語の読み方を示すことができる。同時に図 3-2 の①と②を解決することができるようになる。

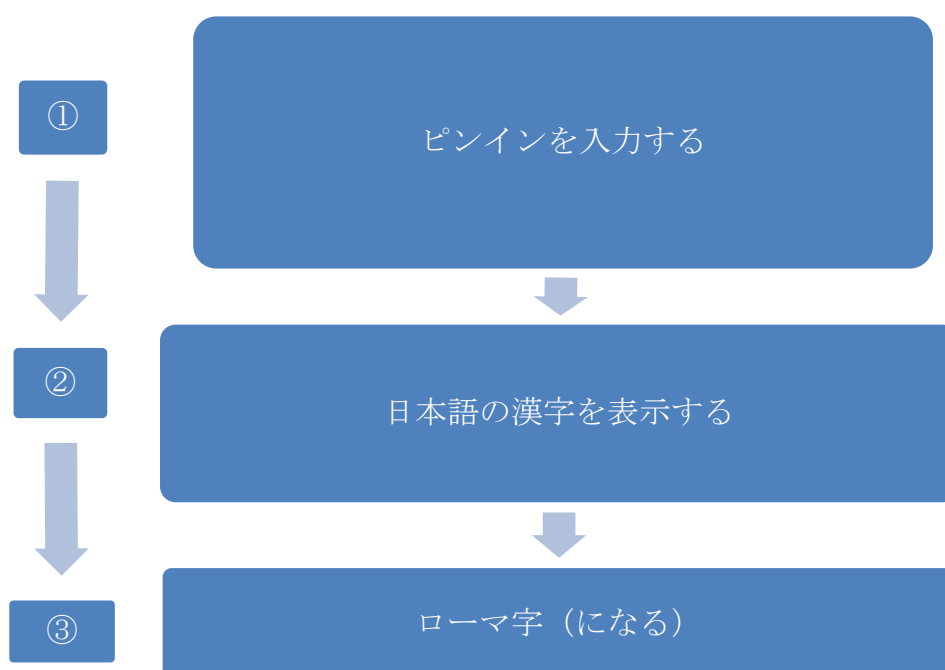


図 3-2 : 中国人に向けピンインから読める日本語になる仕組みのモデル

3.3. 期待されるサービスイメージ

以上の議論を基に、期待されるサービスイメージを検討する。このサービスは、インターネットに接続可能な携帯電話での使用を想定している。

図 3-3 にサービスイメージを実際の利用で示す。ここでは「千葉」と入力すると仮定する。①～⑧まで順を追って説明する。

- ① まず、探したい漢字のピンインを入力して、o k ボタンを押す
- ② 表示した漢字の中、探したい漢字を選択する
- ③ 終了ボタン、あるいは再入力ボタンを選択する
- ④ 再入力ボタンを選択して、次の漢字のピンインを入力する
- ⑤ 表示した漢字の中、探したい漢字を選択する
- ⑥ 終了ボタンを選択する
- ⑦ 漢字と読み方がローマ字で表示された。終了ボタン、あるいは **For help** ボタンを選択する
- ⑧ **For help** ボタンを選択すると、日本語の短文が表示されます。これを日本語が分かる人に見せることで教えてもらえます。

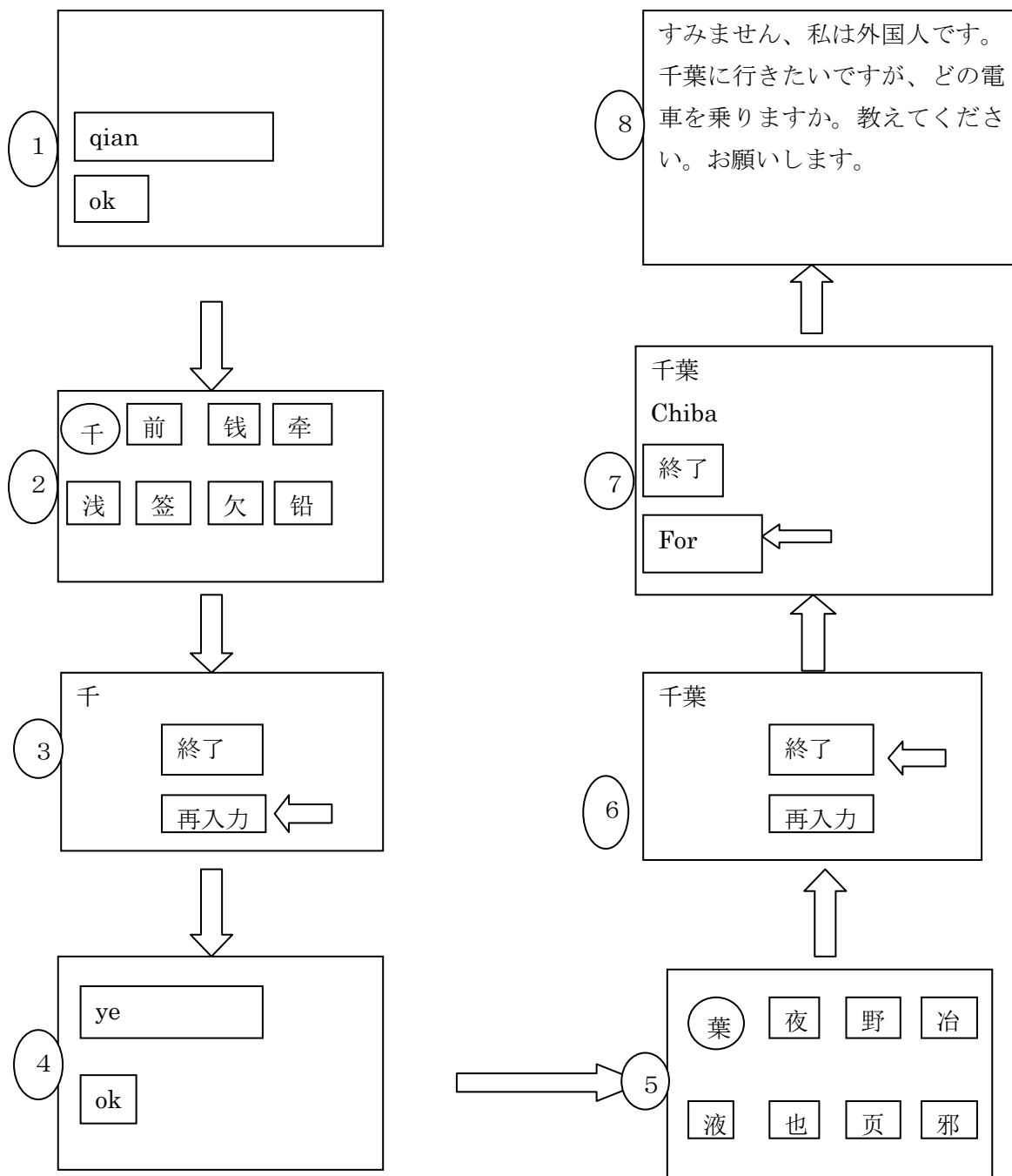


図 3-3 : サービスイメージの検討

3.4. 関連技術

ここでは、本研究と別の方法で似たような機能を持っている技術について述べる。

私の携帯電話は SOFTBANK の 920SH である。この携帯電話にはユーザ電子辞書の機能がある。ユーザは自分の用途によって、携帯電話の漢字の入力方式をユーザが好きな入力方法に変えることができる。だから中国人のユーザがピンインを使って、日本語の漢字に変えることができる。

インターネットの OJ88.com (http://www.oj88.com/forum/forum_posts.asp?TID=638)にその辞書が公開されている。以下にその使用法について述べる。

第一步:

进入短信输入, 输入法转为“漢”

```
options---input/conversion---reset learning
```

```
options---input/conversion---clear log
```

以上操作清除之前字库

网上下载字库文档, 解压缩把所有的*. suj文件复制到TF卡上

目录是: private/sharp/mobile/userdic

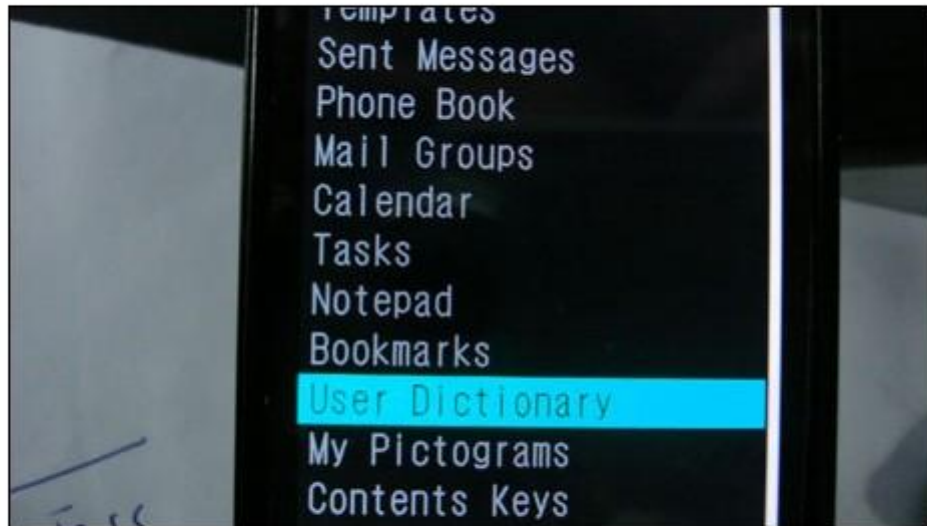
第二步:

然后打开手机移到setting (设定) ——Connectivity (外部接续) ——Memory Card (メモリカード)



図 3-4 : ユーザ辞書をメモリカードから本体メモリにデータをコピーする説明

接着选择 User Dictionary (用户字典)



第三步：然后他会出现你刚刚下载传到手机的字典文件 一个一个把A-Z都恢复进去。

恢复完会有错误提示，不用管他，实际还是恢复进去了。

図 3-5 : ユーザ電子辞書機能の選択画面

上記のホームページに記述された使用方法の日本語訳を以下に示す。

まず、入力法は漢字を選択して、1. options---input/conversion---reset learning
2. options---input/conversion---clear log を順番にやりなさい。
インターネットからダウンロードしたデータをメモ리카ードにコピーしてください。場所は private/sharp/mobile/userdic
ユーザ辞書の機能を選択して、メモ리카ードのデータ（a から z まで）をコピーしてください。

図 3-4 と図 3-5 はこのユーザの電子辞書の紹介と使用方法に関する中国語のホームページである。今使用しているユーザの電子辞書は SHARP 製の携帯電話の機能である。SHARP の携帯電話以外の携帯電話はすべてこの機能がなくて、この機能を実現することができない。ピンインを使って日本語の漢字に変わるという機能は、PC でもほとんどない。したがって、全ての携帯電話や PC で使うためには Web アプリケーションとして構築する必要がある。

4. 設計

本章では、3 章のモデルを踏まえて、必要なシステムの設計について述べる。

4.1. 設計目標

本システムでは、以下のような項目を実現することを目標として設計を行う。

- ピンインから日本語の漢字に変換できること
- 日本語の漢字から日本語の読み方が表示できること
- 以上の項目をインターネット接続された P C および携帯電話で利用できること

4.2. 必要となる情報

- ピンインデータ（ピンインと日本の漢字の対応表）

例えば、千葉の「千」であれば「qian」である。このデータをデータベースに蓄積することで、Web アプリ上で取り扱うことができ、ピンインから日本の漢字に変換できる。

- 日本の地名データ（漢字とローマ字の対応表）

日本の地名（漢字）をローマ字に変換できることである。例えば、「千葉」であれば「chiba」である。

4.3. 構成要素

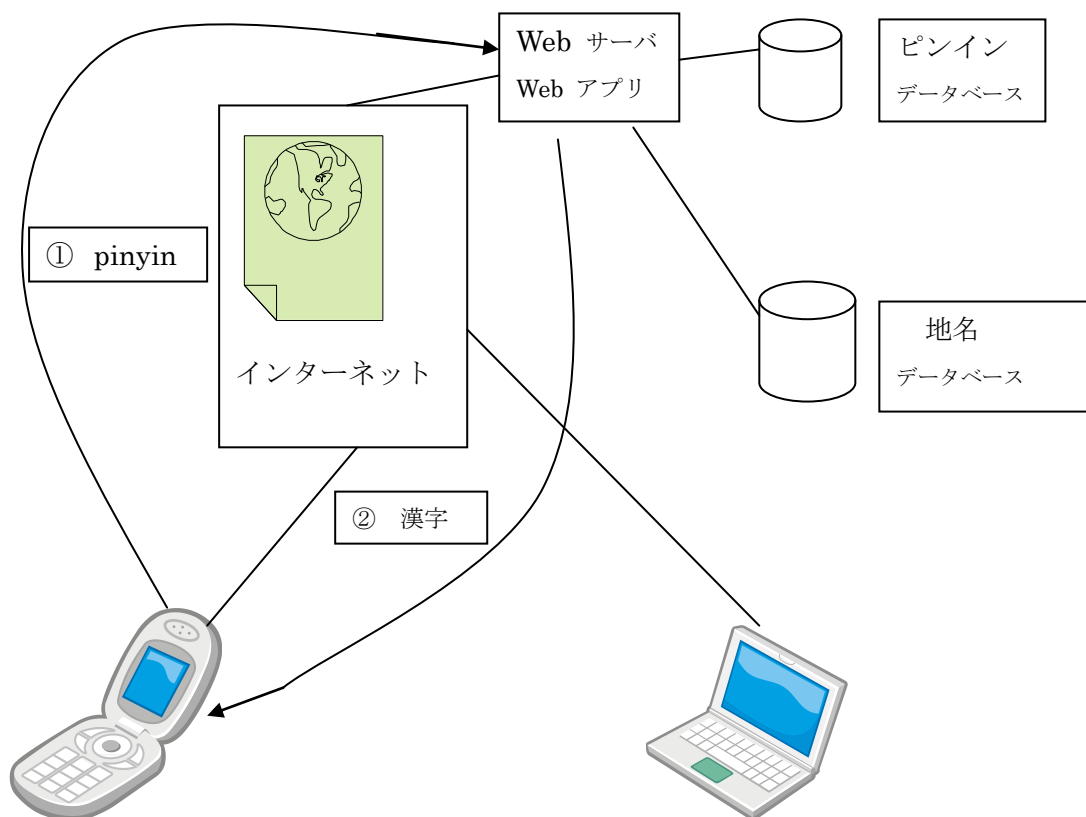


図 4-1：システムの構成要素と情報の流れ

システムの構成要素と情報の流れを図 4-1 に示す。設計するシステムは、インターネットを介して、クライアントとなる携帯電話、PC、サービスを提供する Web サーバ、ピンインのデータベースと地名のデータベースから構成される。

利用者はクライアントとなる携帯電話又は PC から、Web サーバ上の Web アプリにアクセスし、ピンインで入力を行う。このピンインデータは Web アプリに渡されて、日本の漢字に交換される。必要な文字数の漢字を入力したところでこれらの文字列を再び Web サーバに送信し、今度は日本の地名データベースを検索して、ローマ字に変換され、クライアントで表示される。

4.4. 機能概要

各機能の概要について述べ、機能間の連携、やりとりされるデータについて、図 4-2 に示す。

① ピンイン入力機能

クライアントにおいてユーザのピンイン入力を受け付ける機能

② ピンインデータ送信機能

ユーザが入力したピンインデータをサーバ側に送信する機能

③ ピンインデータ受信機能

サーバにおいてクライアントから送信されたピンインデータを受信する機能

④ ピンインデータベース検索機能

サーバにおいて、受信したピンインデータを鍵として、ピンインデータベースに問い合わせ、結果を受け取る機能。

⑤ 検索結果送信機能

検索した結果（日本語の漢字）をクライアントに送信する機能。

⑥ 検索結果選択機能

受信した検索結果（日本語の漢字）をクライアントにおいて選択し、サーバに送信する機能。

⑦ 地名（ローマ字）データベース検索機能

サーバにおいて、クライアントで選択した漢字を受信し、地名データベースに問い合わせ、結果（ローマ字）を受信する機能。

⑧ 地名データベース検索結果送信機能

地名データベースを検索した結果をクライアントに送信する機能

⑨ 地名データベース検索結果受信機能

クライアントにおいて、地名データベースの検索結果を受信し、表示する機能

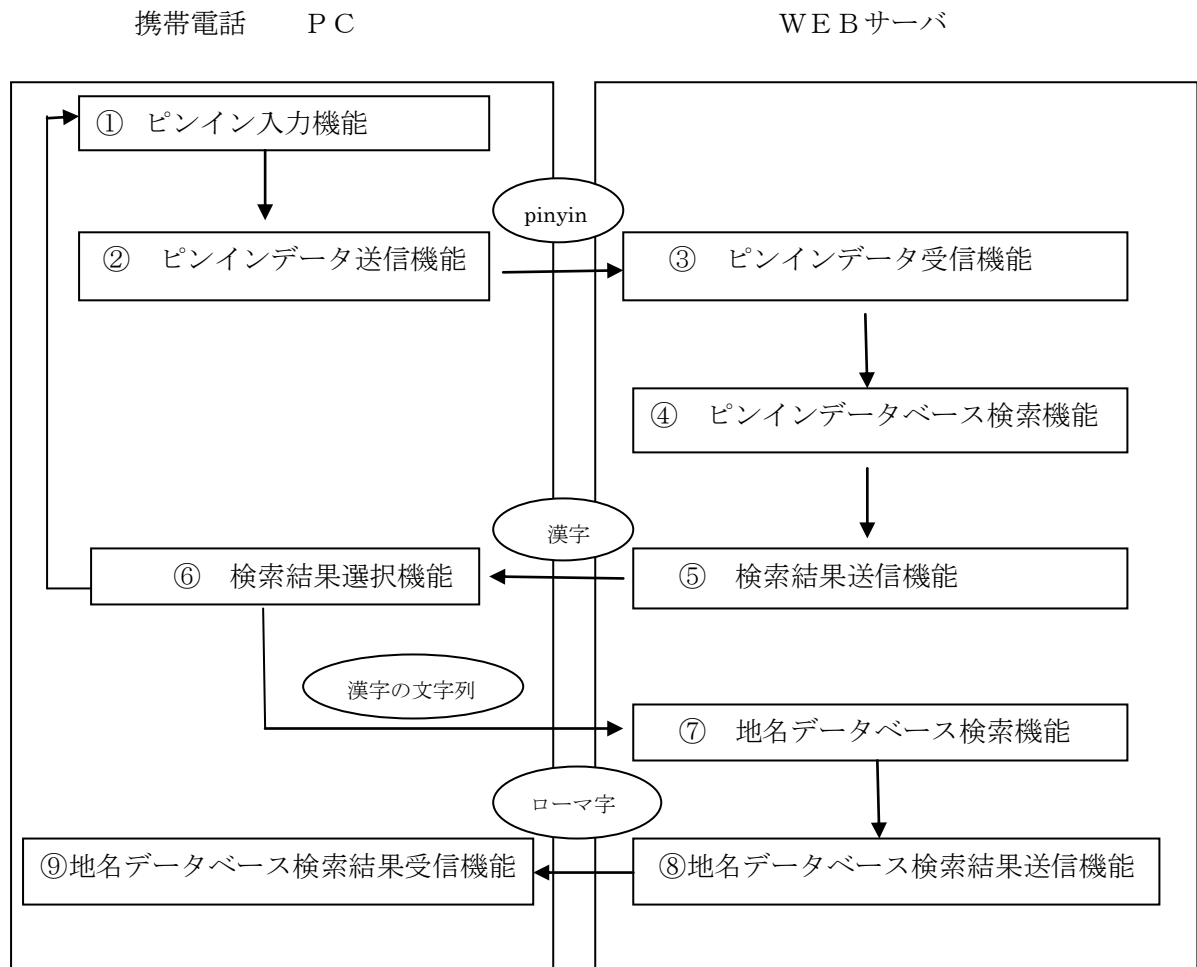


図 4-2 : 各機能の配置とやりとりされるデータ

4.5. データベース

本システムに必要な情報である、日本語漢字とピンインデータの対応表と、漢字地名とローマ字の対応表の設計について述べる。

4.5.1. 日本語漢字・ピンインテーブル

このテーブルでは、日本語の漢字（1文字）とピンインを格納する。各レコードは唯一であり、同じものは存在しないようにするため、重複回避のためにインデックスを設定する。

文字は Unicode の UTF-8 を使用するため、データ型は `varchar` を使用する。ピンインは最大 5 文字である。テーブルの設計を表 4-1 に示す。

表 4-1：日本語漢字・ピンインテーブルの設計

フィールド名	データ型	内容
renban	int	重複回避のためのインデックス
jkanji	varchar(3)	日本語の漢字
pinyin	varchar(5)	ピンイン

4.5.2. 漢字地名・ローマ字テーブル

日本語漢字・ピンインテーブルと同様にインデックスを設定する。文字列に対しては、`varchar` を使用し、それぞれ作成されるデータの最大文字数を考慮する。テーブルの設計を表 4-2 に示す。

表 4-2：漢字地名・ローマ字テーブルの設計

フィールド名	データ型	内容
renban	int	重複回避のためのインデックス
chimei_kanji	varchar(15)	漢字の地名
romaji	varchar(30)	ローマ字

5. 実装

この章では、4 章での設計に基づき、実装を行う。

5.1. 実装環境

5.1.1. クライアント環境

オペレーティングシステム : Microsoft Windows XP Home Edition Version 2002 Service Pack 2

CPU : Intel® Celeron® M 410 @1.46GHz

メモリ : 1GB

5.1.2. サーバ環境 (学外 : www.cyaneum.org)

オペレーティングシステム : Ubuntu 8.04 Server

CPU : Pentium M 1.6GHz

メモリ : 1GB

Web サーバ : Apache 2.2.8

データベース : MySQL 5.0.51

5.1.3. プログラミング言語

PHP 5.2.4

5.2. データの準備

5.2.1. ピンインと日本の漢字の対応データ作成

(1) ピンインと中国の漢字対応データ

本システムでは、まずユーザから入力されたピンインから日本の漢字に変換する必要がある。そこで、まずピンインと中国の漢字データが必要である。そのため、フリーソフトである多国語対応のターミナルである `cxterm`(<http://cxterm.sourceforge.net/>)に付属の対応ファイルがあったので利用した。ピンインと中国の漢字の対応表は、`PY.tit` というファイルである (表 5-1)。

表 5-1 : ピンインと中国の漢字の対応表 (`PY.tit` ファイル)

a	啊阿呵吡腌钢嘎
ai	埃挨哎唉哀捩鎧皑癌呆藹矮暖霰艾碍爰隘隘媛媛暖破
an	鞍氨安厂广谖庵桉鹌俺掩揞铵按暗岸胺案犴黯
ang	肮昂盎
ao	凹熬敖翱噐嗷廐遨熬磬磬磬磬磬袄拗媯傲奥懊澳坳舛鹜磬
ba	芭捌扒叭吧芭八疤巴岂耙拔跋芑菝魃靶把钋耙坝霸罢爸灞
ba	鲑

この論文で検討しているシステムでは、漢字とピンインが個別に対応付けられたデータが必要なため、このままではデータベースに登録して使うことができない。そこで、このデータを変換して、表 5-2 のようなデータを作成した。

表 5-2 : データベースに登録するために変換されたデータ

啊 , a
阿 , a
呵 , a
吡 , a
腌 , a
钢 , a
嘎 , a
埃 , ai
挨 , ai
哎 , ai
唉 , ai
哀 , ai
捱 , ai
鎗 , ai
皑 , ai
癌 , ai
呆 , ai

(2) 中国の漢字と日本の漢字の対応表

次に、中国の漢字のうち、日本では別の漢字を使用するものを変換する。フリーソフトの PinConv+ (<http://www.ctrans.org/gobi/1120984246>) に付属のデータを利用した。中国の漢字と日本の漢字の対応データは jis2gb.dat というファイルである。表 5-3 の左が日本の漢字、右が中国の漢字である。

表 5-3 : 中国の漢字と日本の漢字の対応データ (jis2gb.dat ファイルの一部)

齊	齐
齋	斋
質	质
遊	游
愛	爱

漢字とピンインの対応と、中国の漢字と日本の漢字の対応のデータから、ピンインと日本の漢字のデータを Excel で変換して作成した。作成されたデータから、日本語に存在しない漢字を取り除いたものが表 5-4 のようになる。

表 5-4：作成された日本の漢字とピンインの対応データ

阿 , a
呵 , a
嘎 , a
埃 , ai
挨 , ai
哀 , ai
皑 , ai
癌 , ai
呆 , ai
藹 , ai
矮 , ai

5.2.2. 日本の地名とローマ字のデータ

本システムでは、中国人が日本語の漢字で書かれた地名を読めるようにするため、ローマ字対応データが必要である。

ここでは、MS-IME95 用にフリーで配布されている地名辞書「全国地名テキスト MS-IME95 用」(http://ikomyaka.hp.infoseek.co.jp/soft/m95_zen.htm) を利用した。

これは、ひらがなと漢字で構成されている（表 5-5）。

表 5-5：地名辞書（抜粋）

ほっかいどう	北海道	県	県
あおもりけん	青森県	県	県
いわてけん	岩手県	県	県
みやぎけん	宮城県	県	県
あきたけん	秋田県	県	県
やまがたけん	山形県	県	県
ふくしまけん	福島県	県	県
いばらきけん	茨城県	県	県
とちぎけん	栃木県	県	県
ぐんまけん	群馬県	県	県

これらの、ひらがなをローマ字に変換する必要がある。

KIroumazi(<http://www2s.biglobe.ne.jp/~iryo/2vba/s2vba15.html>)という Excel 上でローマ字をひらがなに変換するマクロを使用して変換した。変換した結果は表 5-6 のようになった。

表 5-6：ひらがなをローマ字に変換した結果

北海道	hotukaidou
青森県	aomoriken
岩手県	iwateken
宮城県	miyagiken
秋田県	akitaken
山形県	yamagataken
福島県	fukusimaken
茨城県	ibarakiken
栃木県	totigiken
群馬県	gunmaken
埼玉県	saitamaken
千葉県	tibaken
東京都	toukyouto

このローマ字表記は、実際の地名のローマ字表記と異なる場合があるが、今回の場合、中国人が読むことを前提にしているので、特に問題はないが、利便性のために、ti は chi に、si は shi に変換する。

5.3. データベースの作成

5.2 で作成したデータベース用のファイルから、データベースを構築する。

5.3.1. データファイルの転送

WinSCP を使って、サーバ (www.cyaneum.org (学外)) に 5.2 で作成したデータファイルを転送する

日本語漢字・ピンインデータ : jp_pin_final.txt

漢字地名・ローマ字データ : chimei_romaji.txt

5.3.2. データベースの設定

サーバの MySQL モニタを使用し、データベースの設定を行う。まず、データベースを作成する。データベースは UTF-8 を使用するよう設定する。作成後、作成されたかどうか

かを確認する（表 5-7）。

表 5-7：データベースの作成と確認

```
mysql> create database db_a644193 character set utf8;
Query OK, 1 row affected (0.00 sec)

mysql> show databases;
+-----+
| Database          |
+-----+
| information_schema |
| db_a644193         |
| mysql              |
+-----+
3 rows in set (0.01 sec)
```

次に、データベースを使用可能とするための設定を行った（表 5-8）。

表 5-8：データベース利用の設定

```
mysql> use db_a644193;
Database changed
```

5.3.3. 「ピンイン⇄漢字」テーブルの作成

設計されたテーブルに基づき、データベース上にテーブル jkanji_pinyin の作成を行った（表 5-9）。

表 5-9：ピンイン・漢字テーブルの作成

```
mysql>
mysql> create table jkanji_pinyin (jkanji varchar(3), pinyin varchar(6));
Query OK, 0 rows affected (0.00 sec)
```

その後、テーブルが作成できたかどうか確認した（表 5-10）。

表 5-10：ピンイン・漢字テーブル作成の確認

```
mysql> show tables;
+-----+
| Tables_in_db_a644193 |
+-----+
```

```
| jkanji_pinyin          |
+-----+
1 row in set (0.00 sec)
```

また、フィールドが作成できたかどうかも確認した（表 5-11）。

表 5-11：ピンイン・漢字テーブルのフィールド作成の確認

```
mysql> show fields from jkanji_pinyin;
+-----+-----+-----+-----+-----+-----+
| Field | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| jkanji | varchar(3)    | YES  |     | NULL     |       |
| pinyin | varchar(6)    | YES  |     | NULL     |       |
+-----+-----+-----+-----+-----+-----+
2 rows in set (0.00 sec)
```

CSV ファイルである jp_pin_final.txt からデータを MySQL に取り込んだ（表 5-12）。

表 5-12：ファイルから jkanji_pinyin のフィールドにデータを取り込む

```
mysql> load data local infile "jp_pin_final.txt" into table jkanji_pinyin fields terminate
d by ',';
Query OK, 5459 rows affected (0.02 sec)
Records: 5459  Deleted: 0  Skipped: 0  Warnings: 0
```

次に、連番フィールドの設定を行った。まず、renban フィールドを追加した（表 5-13）。

表 5-13：jkanji_pinyin への連番フィールドの追加

```
mysql> alter table jkanji_pinyin add renban int;
Query OK, 5459 rows affected (0.02 sec)
Records: 5459  Duplicates: 0  Warnings: 0
```

次に、renban フィールドにインデックスを追加した（表 5-14）。

表 5-14：jkanji_pinyin への連番フィールドへインデックス追加

```
mysql> alter table jkanji_pinyin add index ind1(renban);
Query OK, 5459 rows affected (0.04 sec)
```



```
Records: 5459  Duplicates: 0  Warnings: 0
```

最後に `renban` フィールドの属性を変更して、追加でデータが登録された場合に自動的に連番が付けられるようにした（表 5-15）。

表 5-15 : `jkanji_pinyin` の連番フィールドへ自動更新の設定

```
mysql> alter table jkanji_pinyin modify renban int auto_increment;
Query OK, 5459 rows affected (0.07 sec)
Records: 5459  Duplicates: 0  Warnings: 0
```

5.3.4. 「漢字地名⇔ローマ字」テーブルの作成

漢字地名とローマ字のテーブルを `chimei_romaji` として、データベース上に作成した（表 5-16）。

表 5-16 : `chimei_romaji` テーブルの作成

```
mysql> create table chimei_romaji (chimei_kanji varchar(15), romaji varchar(30));
Query OK, 0 rows affected (0.00 sec)
```

また、テーブルが作成できたかどうか確認した（表 5-17）。

表 5-17 : `chimei_romaji` テーブルの作成確認

```
mysql> show tables;
+-----+
| Tables_in_db_a644193 |
+-----+
| chimei_romaji        |
| jkanji_pinyin        |
+-----+
2 rows in set (0.00 sec)
```

最後に、フィールドが作成できたかどうか確認した（表 5-18）。

表 5-18 : `chimei_romaji` テーブルのフィールド作成確認

```
mysql> show fields from chimei_romaji;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Extra | Privileges |
```

Field	Type	Null	Key	Default	Extra
chimei_kanji	varchar(15)	YES		NULL	
romaji	varchar(30)	YES		NULL	

2 rows in set (0.00 sec)

CSV ファイルである、chimei_romaji.txt からデータをテーブルに取り込んだ（表 5-19）。

表 5-19 : chimei_romaji テーブルへファイルからデータの取り込み

```
mysql> load data local infile "chimei_romaji.txt" into table chimei_romaji fields
terminate
d by ',';
Query OK, 70885 rows affected, 58 warnings (0.17 sec)
Records: 70885 Deleted: 0 Skipped: 0 Warnings: 58
```

次に、連番フィールドを設定した。まず、renban フィールドの追加した（表 5-20）。

表 5-20 : chimei_romaji テーブルへの renban フィールドの追加

```
mysql> alter table chimei_romaji add renban int;
Query OK, 70885 rows affected (0.12 sec)
Records: 70885 Duplicates: 0 Warnings: 0
```

次に、renban フィールドにインデックスを追加した（表 5-21）。

表 5-21 : chimei_romaji テーブルの renban フィールドにインデックスを設定

```
mysql> alter table chimei_romaji add index ind1(renban);
Query OK, 70885 rows affected (0.27 sec)
Records: 70885 Duplicates: 0 Warnings: 0
```

最後に、renban フィールドの属性を変更して、追加データの登録があった場合に、自動的に連番が付けられるようにした（表 5-22）。

表 5-22 : chimei_romaji テーブルの renban フィールドに自動更新の設定

```
mysql> alter table chimei_romaji modify renban int auto_increment;
Query OK, 70885 rows affected (0.48 sec)
Records: 70885 Duplicates: 0 Warnings: 0
```

5.4. 各機能の実装

5.4.1. ピンイン入力機能

ピンイン入力機能と言うのは、ユーザがピンインを入力し、サーバにデータを送信することである。これを `send.htm` (表 5-23) に作成した。HTML のフォームを用いて入力を受け付け、後述するピンイン受信機能である `rec.php` に対し、入力データを POST で送信する。

表 5-23 : ピンイン入力機能 (`send.htm`)

```
<html>
<body>
  請輸入 Pinyin。

  <br>
  <form method="post" action="rec.php">
    <input type="text" name="pinyin" value=""><br />
    <br>
    <input type="submit" name="start" value="OK">
  </form>
</body>
</html>
```



図 5-1 : ピンイン入力機能の実行画面

この画面はピンイン入力機能の実行後の画面である。図 5-1 の通り、ピンインを入力して、ok ボタンを使えば、ピンインの入力することができる。そして、ok ボタンを押すなら、ピンインデータの送信することができる。

5.4.2. ピンインデータ受信機能

この機能では、ピンインデータ入力機能から送信されたピンインデータを受信する。この機能を rec.php に実装した（表 5-24）。

表 5-24：ピンインデータ受信機能（rec.php）

```
<html>
<head>
</head>

<body>

Pinyin <?php print($_POST['pinyin']); ?>已被接收。
</body>
</html>
```



図 5-2：ピンインデータ受信機能の実行画面

ピンインを受信することが表示された（図 5-2）。

5.4.3. ピンインから漢字を検索する機能

ピンインから漢字を検索する機能では、データベース中の jkanji_pinyin テーブルにおいて、Pinyin フィールドをピンインデータ受信機能で受信したピンインの文字列(\$pinyin)で検索し、その結果として、jkanji フィールドの漢字を得る。

例えば、ni というピンインの漢字を検索する場合の、SQL 文は、select jkanji from jkanji_pinyin where pinyin='ni'; のようになる。表 5-25 に MySQL で直接入力し、実行した結果を示す。

表 5-25 : MySQL において、ピンイン ni で漢字を検索した結果

mysql> select jkanji from jkanji_pinyin where pinyin='ni';	
+-----+	
jkanji	
+-----+	
霓	
倪	
泥	
尼	
猊	
悌	
擬	
祢	
匿	
膩	
逆	
溺	
尿	
昵	
慝	
睨	
+-----+	
16 rows in set (0.00 sec)	

この SQL 文を PHP から MySQL にアクセスして、実行するプログラム rec2.php を表 5-26 に示す。

表 5-26：ピンインから漢字を検索する機能（rec2.php）

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
</head>
<body>
<?php
mb_language("uni");
mb_internal_encoding("utf-8");
mb_http_input("auto");
mb_http_output("utf-8");
$pinyin = $_POST['pinyin'];
echo "Pinyin: " . $pinyin . "<br />";
if ($pinyin<>""){
    $db = mysql_connect('localhost','root','20cuc08!');
    mysql_query("SET NAMES utf8",$db);
    mysql_select_db('db_a644193');
    $sql = "select jkanji from jkanji_pinyin where pinyin='$pinyin'";
    $result = mysql_query($sql);
    $rows = mysql_num_rows($result);
    if ($rows == 0){
        echo "<p>無法找到漢字</p>";
    } else {
        while ($row = mysql_fetch_array($result)){
            echo $row["jkanji"] . " ";
        }
    }
}
?>
</body>
</html>
```

このプログラムはピンインから漢字を検索する機能を実行する。文字コードとして UTF-8 を使う。この文字は単体の中国語でも対応する。

入力したピンインで `jkanji_pinyin` から漢字を検索している。`jkanji_pinyin` の中がない漢字の場合、“無法找到漢字”（漢字が見つかりません）を表示される。

また、ユーザが入力したピンインを送信する send2.htm においては、表 5-27 の太字部分のような文字化け対策を施した。

表 5-27 : send.htm に文字化け対策をした (send2.htm)

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
</head>
<body>
  請輸入 Pinyin。

  <br>
  <form method="post" action="rec2.php">
    <input type="text" name="pinyin" value=""><br />
    <br>
    <input type="submit" name="start" value="OK">
  </form>
</body>
</html>
```

これらを、ブラウザから実行して確認した。図 5-3、図 5-4 に実行結果を示す。

<http://www.cyaneum.org/~a644193/send2.htm>



図 5-3 : send2.htm において、ピンインを入力する画面



図 5-4 : req2.php で、検索された漢字を表示した画面

5.4.4. 漢字の選択機能とピンインの継続入力機能

前節までで、ピンインで検索した漢字が表示された。

次に、検索結果から必要な漢字をユーザに選択させる「検索結果選択機能」を実装する。

この機能では、地名を構成する複数の漢字を結合させるため、選択した漢字は、地名漢字文字列として保持する。この地名漢字文字列をピンイン入力機能に渡す。

一方、ピンイン入力機能では、検索結果選択機能から地名漢字文字列を受信し、表示する。ユーザからピンイン入力を受け付け、再び「ピンイン漢字検索機能」に送信する。

(1) ピンイン入力・送信機能への機能追加 (send2.htm⇒pinyin.php)

ここでは、GET で地名漢字文字列が送られているかどうかを確認し、地名漢字文字列を表示する。次に、ユーザからのピンイン入力を待ち、ピンインデータと地名漢字文字列を kanji_lookup.php に POST で送信する。

表 5-28 に、ピンイン入力側のプログラム (pinyin.php) を示す。

表 5-28 : ピンイン入力側のプログラム (pinyin.php)

```
<?php
    if ($_GET["chimei_kanji"]<>""){
        $chimei_kanji = $_GET["chimei_kanji"];
    }
?>

<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
```



```

</head>
<body>
已輸入の漢字: <?php echo $chimei_kanji ?>
<br>
請輸入 Pinyin。
<br>

<form method="post" action="kanji_lookup.php">
    <input type="text" name="pinyin" value=""><br />
    <br>
    <?php
    echo "<input type='hidden' name='chimei_kanji' value='";
    echo $chimei_kanji;
    echo "';";
    ?>
    <input type="submit" name="start" value="OK">
</form>
<br>
<?php
    echo "<a href='chimei_lookup.php?chimei_kanji=";
    echo $chimei_kanji;
    echo "';>輸入完了</a>";
?>
</body>
</html>

```

(2) ピンイン漢字検索機能に検索結果選択機能を追加 (rec2.php⇒kanji_lookup.php)

kanji_lookup.php では、pinyin.php から POST で送信される地名漢字文字列と、ピンインデータを受信する。受信したピンインデータでピンインデータベースを検索し、漢字を表示する。さらに、検索された漢字から、使う文字を選択するために、文字に GET で送信するリンクを設定する。この際、選択した文字を地名漢字文字列に加える。

表 5-29 に、kanji_lookup.php を示す。

表 5-29 : rec2.php に検索結果選択機能が追加された (kanji_lookup.php)

```

<?php
    if ($_GET["chimei_kanji"]<>""){

```

```

        $chimei_kanji = $_GET["chimei_kanji"];
    }
    ?>
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
</head>
<body>
    已輸入の漢字: <?php echo $chimei_kanji ?>
<br>
    請輸入 Pinyin。
<br>
    if ($pinyin<>""){
        $db = mysql_connect('localhost','root','20cuc08!');
        mysql_query("SET NAMES utf8",$db);
        mysql_select_db('db_riho_m');
        $sql = "select jkanji from jkanji_pinyin where pinyin='$pinyin'";
        $result = mysql_query($sql);
        $rows = mysql_num_rows($result);
        if ($rows == 0){
            echo "<p>無法找到漢字</p>";
        } else {
            while ($row = mysql_fetch_array($result)){
                $add_chimei = $row["jkanji"];
                echo "<a href=¥"pinyin.php?chimei_kanji=";
                echo $chimei_kanji . $add_chimei;
                echo "¥">";
                echo $row["jkanji"];
                echo "</a>  ";
            }
        }
    }
}

```

ブラウザから実行した結果を、図 5-5 に示す。



図 5-5 : pinyin.php の実行画面

1 文字目のピンインを入力した



図 5-6 : kanji_lookup.php の実行画面

検索された漢字の候補が表示された (図 5-6)。漢字 1 文字ずつには、それぞれリンクが貼ってあり、選択した文字を、これまで入力された文字と結合して、再び pinyin.php に GET で送信する。



図 5-7 : kanji_lookup.php から既に入力した「秋葉」を受信した pinyin.php の実行画面

kanji_lookup.php で漢字をクリックして、再び pinyin.php でピンインを入力することができる (図 5-7)。

5.4.5. 漢字から地名を検索する機能

この機能では、データベース中の chimei_romaji テーブルにおいて、Chimei_kanji フィールドを、地名データベース検索機能で受信した漢字の地名文字列(\$chimei_kanji)で検索する。その結果として、romaji フィールドのローマ字を得る。

例えば、MySQL での問い合わせ文は、select romaji from chimei_romaji where chimei_kanji='秋葉原'; のようになる。

検索結果は、複数の場合もあり、すべて表示する。

表 5-30 に、プログラム chimei_lookup.php を示す。

表 5-30 : 漢字文字列から地名を検索する (chimei_lookup.php)

```
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
</head>
<body>
<?php
mb_language("uni");
mb_internal_encoding("utf-8");
mb_http_input("auto");
mb_http_output("utf-8");
```

```

$chimei_kanji = "秋葉原";
echo $chimei_kanji . "<br />";
if ($chimei_kanji<>"){
    $db = mysql_connect('localhost','root','20cuc08!');
    mysql_query("SET NAMES utf8",$db);
    mysql_select_db('db_a644193');
    $sql = "select romaji from chimei_romaji where chimei_kanji='$chimei_kanji'";
    $result = mysql_query($sql);
    $rows = mysql_num_rows($result);
    if ($rows == 0){
        echo "<p>無効的地名</p>";
    } else {
        while ($row = mysql_fetch_array($result)){
            echo $row["romaji"] . " <br />";
        }
    }
}
?>
</body>
</html>

```

ブラウザから chimei_lookup.php を実行した結果を図 5-8 に示す。



図 5-8 : chimei_lookup.php の実行結果

5.4.6. ピンイン入力機能から地名検索機能に地名文字列を送信する

ピンイン入力機能（pinyin.php）で、必要な文字を入力したかどうかを確認し、入力完了なら、地名漢字文字列を GET で、chimei_lookup.php に送信する。

プログラムを表 5-31 に示す。

表 5-31：地名漢字文字列を chimei_lookup.php に送信する（pinyin.php）

```
<?php
    if ($_GET["chimei_kanji"]<>""){
        $chimei_kanji = $_GET["chimei_kanji"];
    }
?>

<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
</head>
<body>
已輸入的漢字: <?php echo $chimei_kanji ?>
<br>
請輸入 Pinyin。
<br>

<form method="post" action="kanji_lookup.php">
    <input type="text" name="pinyin" value=""><br />
    <br>
    <?php
        echo "<input type='hidden' name='chimei_kanji' value='";
        echo $chimei_kanji;
        echo "'>";
    ?>
    <input type="submit" name="start" value="OK">
</form>
<br>
<?php
    echo "<a href='chimei_lookup.php?chimei_kanji=";
    echo $chimei_kanji;
```

```
        echo "㊦">輸入完了</a>;  
    ?>  
</body>  
</html>
```

5.4.7. 地名ローマ字検索機能に地名文字列受信機能を追加

地名ローマ字検索機能(chimei_lookup.php)に、pinyin.php から送信された地名文字列を受信する機能を追加する。

地名文字列 chimei_kanji は GET で送信されるため、受信する部分は、次のようになる。

```
$chimei_kanji=$_GET['chimei_kanji'];
```

機能を追加した chimei_lookup.php を表 5-32 に示す。

表 5-32 : 地名文字列受信機能を追加した chimei_lookup.php

```
<html>  
<head>  
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">  
</head>  
<body>  
  
<?php  
mb_language("uni");  
mb_internal_encoding("utf-8");  
mb_http_input("auto");  
mb_http_output("utf-8");  
  
$chimei_kanji=$_GET['chimei_kanji'];  
  
echo  $chimei_kanji . "<br />";  
  
if ($chimei_kanji<>""){  
    $db = mysql_connect('localhost','root','20cuc08!');  
    mysql_query("SET NAMES utf8",$db);  
    mysql_select_db('db_riho_m');  
    $sql = "select romaji from chimei_romaji where chimei_kanji='$chimei_kanji'";  
    $result = mysql_query($sql);  
    $rows = mysql_num_rows($result);
```

```

        if ($rows == 0){
            echo "<p>無効的地名</p>";
        } else {
            while ($row = mysql_fetch_array($result)){
                echo $row["romaji"] . " <br />";
            }
        }
    }
?>
</body>
</html>

```



図 5-9 : 「秋葉原」という文字列を `kanji_lookup.php` から受信した `pinyin.php` の実行画面

図 5-9 のように地名の入力が完了した場合は、「輸入完了」をクリックする。すると、「秋葉原」という文字列が、`chimei_lookup.php` に送信される。



図 5-10 : pinyin.php から「秋葉原」を受信して、ローマ字表記の検索結果を表示した chimei_lookup.php

図 5-10 に示すように、chimei_lookup.php では、pinyin.php から地名文字列「秋葉原」を GET で受信し、地名ローマ字データベースを検索して、ローマ字表記を表示する。

6.1. 動作検証

6.1.1. 実装された機能の動作結果 (PC)

存在しない地名を入力した場合には、その地名が存在しない旨のメッセージが表示されることも確認できた。

The image shows a laptop screen with the Mozilla Firefox browser open. The address bar contains the URL `http://www.cyanum.org/a644193/kansu_lookup.php`. The browser's main display area is filled with a dense grid of small, blue, illegible characters, likely representing a large dataset or a list of items. The Windows taskbar is visible at the bottom of the screen, showing the Start button and several open applications, including Firefox and a file explorer.

図 6-1 : PC での動作画面

6.1.2. 実装された機能の動作結果（携帯）

本節では、日本の携帯電話会社各社（NTT DoCoMo、Softbank、au by kddi）の携帯電話を使用して、動作状況の確認を行った。

(1) NTT DoCoMo の場合（SH905i を使用）

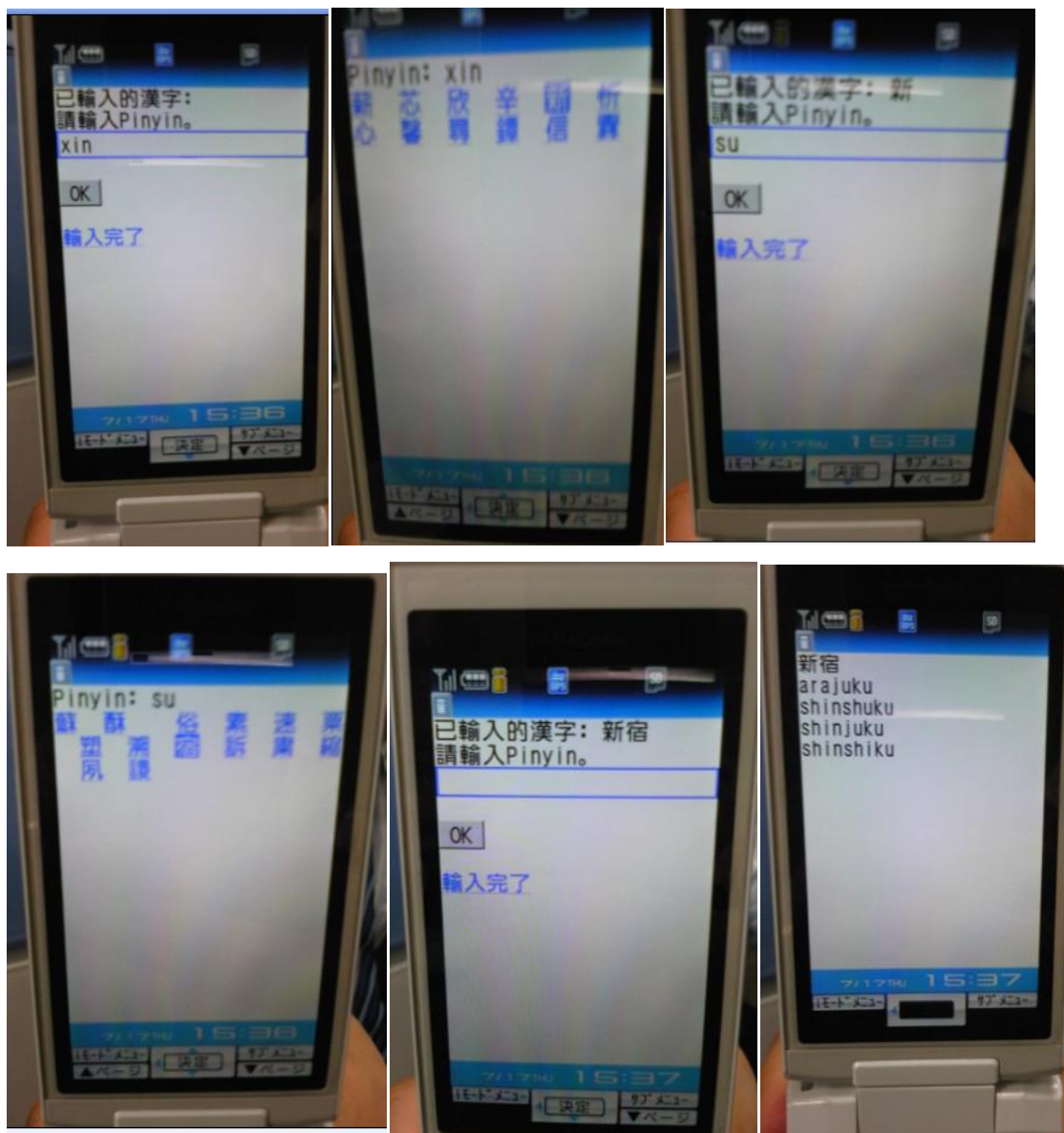


図 6-2 : NTT DoCoMo SH905i での動作画面

NTT DoCoMo の SH905i は、2007 年 11 月末に発売された比較的新しいタイプの携帯電話である。NTT DoCoMo は FOMA に移行してから、ブラウザの文字コードで UTF-8 に対応しているが、フォントとしては日本語しか表示できない。したがって、中国語の簡体字

が表示できないので、メッセージは、繁体字に近い日本の漢字を使うようにプログラムでは設定されている。

それ以外に、基本的に実装された機能は、問題なく動作した（図 6-2）。

(2) Softbank の場合（920SH を使用）

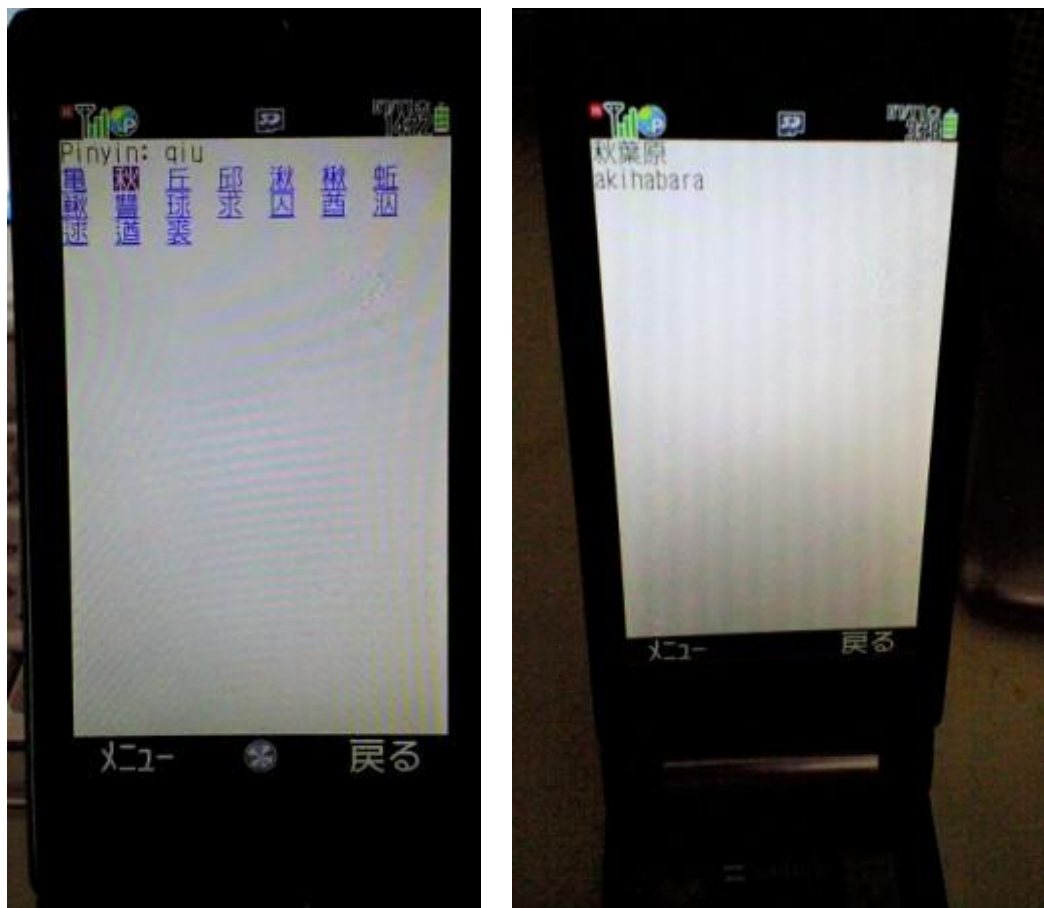


図 6-3 : Softbank 920SH での動作画面

Softbank の 920SH は 2007 年の 11 月に発売された携帯電話である。こちらの携帯電話も、NTT DoCoMo の場合と同様で、UTF-8 に対応しているが、簡体字は表示できない。しかし、日本の漢字で繁体字を代用できるので、特に問題はない。基本的に実装された機能は問題なく動作した（図 6-3）。

(3) au by kddi の場合 (W53CA)

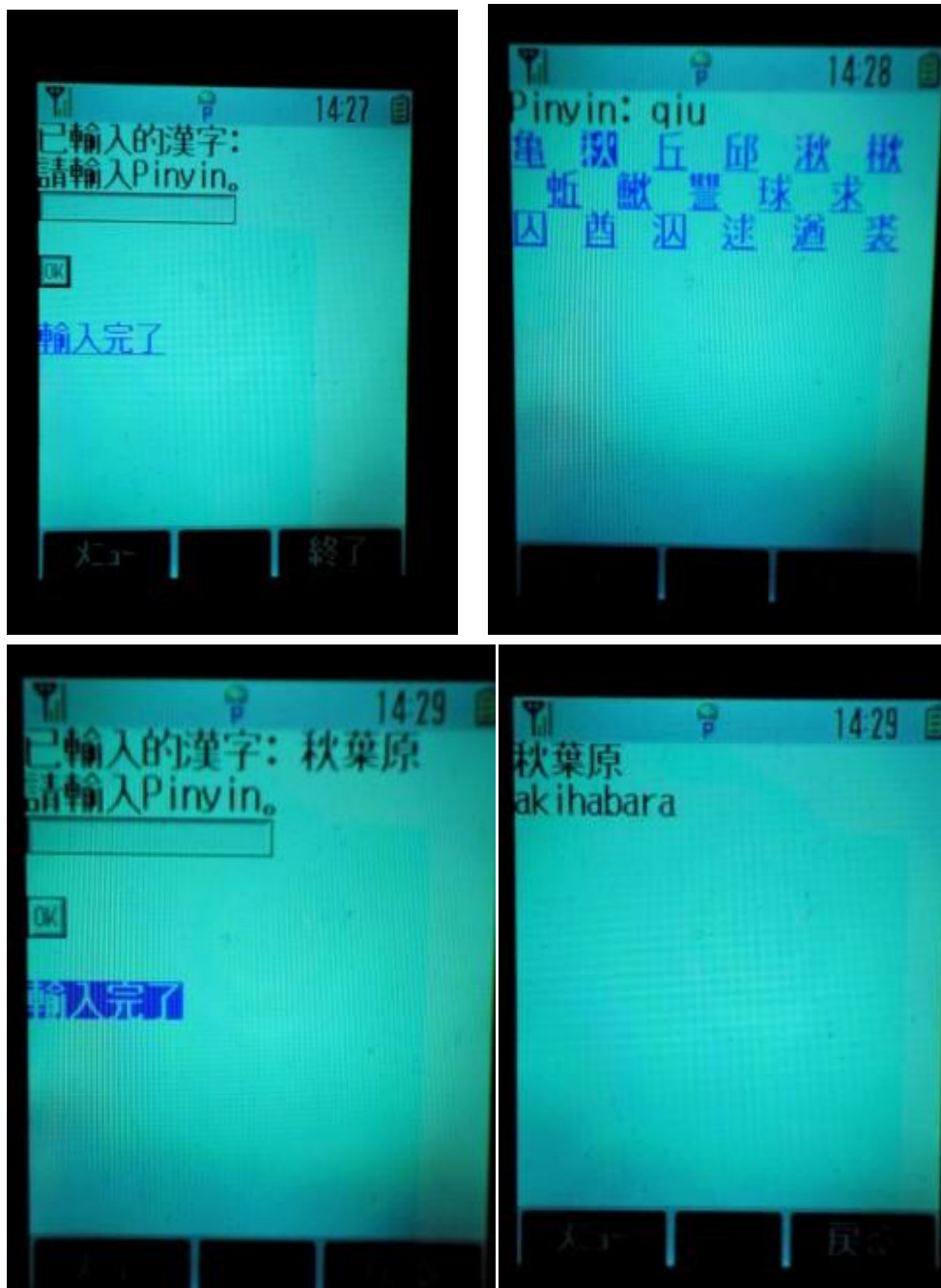


図 6-4 : au by kddi W53CA での動作画面

W53CAは、2007年8月に発売された携帯電話である。こちらも、NTT DoCoMo, Softbank の場合と同様で、UTF-8 に対応しているが、簡体字は表示できない。しかし、日本の漢字で繁体字を代用できるので、特に問題はない。基本的に実装された機能は問題なく動作した (図 6-4)。

6.2. 一般的な利用者（中国人）による評価

本節では、中国人を対象に本システムを利用してもらい、その評価について述べてもらった。利用者となった中国人は2名で、本学の学生である。

● 利用の様子



図 6-5 : 利用の様子

● 利用者の感想①

0644194 チョウ ヒンイ

日本への中国からの観光客にとっても便利なソフトだと思う。駅に、ネームが迷うときにただ入力すると、すぐ日本語のローマ字を出でくる。分かりやすいだから、たいへんありがたい。

● 利用者の感想②

0644189 チン キ

日本にいるが、日本語が分からない中国人にとって、このプログラムがとても便利だと思います。日本語の読み方が分からない時、ピンインで調べられて、日本の地名がすぐ分かる。

しかし、私は二つの提案があります。一つは、今は一つひとつの字から調べるしかできません。単語でも調べることができるならば、もっと便利になれると思います。もう一つは「入力した漢字:」の字が見落としやすいから、もっと目立つにしたほうがいいと思います。

政策情報学部は学生の問題発見、問題解決の能力を養成する学部です。倪さんの研究

が自分の身近なところの問題を発見し、そしてこの問題を解決しました。この点がとてもすごいと思います。

- 利用者評価のまとめ

利用者は、中国人ではあるが日本語をすでに学習しており、が理解できるほうである。しかし、そのような中国人であっても、日本の漢字を読むには、辞書が必要な場合がかなりある。辞書を持たない場合でも、インターネットで日本の漢字の読み方がピンインから検索することでわかることは価値がある。

システムとしては、ピンインを複数文字分まとめて入力したほうがいいという意見もあり、使いやすくするための工夫、改善が必要であることがわかった。

7. まとめ

本研究では、特に移動時における言葉と文字の障害を解決するため、中国の言語と文字を使い、日本語ができない中国人でも日本語の入力を行うことができるようになるということを目的とした。

そのため、交通機関の駅やホームページ等で国際化対応の現状を調査し、日本語のわからない外国人が目的地に到着するためには、現状の国際化対応がニーズにあってないことを明らかにした。

また、外国人、特に漢字の大半の意味がわかる中国人を対象として、ピンインから日本の漢字、漢字の地名からローマ字表記の地名を得る、日本の地名を読めるようにするためのモデルの構築、モデルに基づいたシステムの設計、実装を行った。

これまで作ってきたファイルは

① chimei_lookup.php ② kanji_lookup.php ③ pinyin.php, rec2.php ④ send.htm, send2.htm

である。このうち、最終的には、①②および③が最終的なプログラムのファイルである。

その結果、ピンインを入力すること、漢字を表示すること、漢字を選択すること、地名（ローマ字）データベース検索することが実現できた。このように、設計した機能は、Webアプリケーションとして、全部実装することができた。

この Web アプリケーションを使うことで、日本語がほとんどわからない中国人が、日本語の地名をローマ字で読むことができるようになった。

問題点としては、日本語をローマ字表記した場合に、中国語にない発音が存在することである。例えば、千葉の **chi** は、発音しにくく、少し異なる発音をする。しかし、日本人がその発音を聞いた場合に、似ているのでほぼ理解できるのではないかと考えられる。そのほかには、日本の地名では、同じ漢字で複数の読み方の地名があるが、本システムではすべての候補が表示される。日本語がわからない中国人には、どれに自分が行くべき地名なのかが判別できない。したがって、将来的には絞り込んで特定するなどの改善を検討する必要がある。

参考文献

- 法務省入国管理局、出入国者数および外国人登録者数、2007 年、
<http://www.immi-moj.go.jp/toukei/index.html>
- 内閣訓令第 1 号、「ローマ字のつづり方について」1954 年
- 文化庁、「ローマ字のつづり方（第 1 表・第 2 表）」、1954 年、文化庁ホームページ
<http://www.bunka.go.jp/>
- JR 東日本ホームページ <http://www.jreast.co.jp/>
- 京成電鉄ホームページ <http://www.keisei.co.jp/>
- 市川市ホームページ <http://www.city.ichikawa.lg.jp/index.html>
- 憩舞華、全国地名テキスト MS-IME95 用、1997 年、
http://ikomyaka.hp.infoseek.co.jp/soft/m95_zen.htm
- タケウチ、中国語変換ツール「PinConv+」、2005 年、
<http://www.ctrans.org/gobi/1120984246>
- ピンイン・日本漢字携帯電話用ユーザ辞書、
OJ88.com (http://www.oj88.com/forum/forum_posts.asp?TID=638)
- Excel 用「ローマ字・ひらがな変換マクロ KIroumazi」
<http://www2s.biglobe.ne.jp/~iryu/2vba/s2vba15.html>
- 多言語対応ターミナルソフト「cxtterm」
<http://cxtterm.sourceforge.net/>
- 西沢直木、「PHP による Web アプリケーションスーパーサンプル 第 2 版」、2006 年、
ソフトバンククリエイティブ

謝辞

本研究を行うに当たり、ご指導を頂いた渡辺 恭人准教授に感謝いたします。論文の構成・プログラミング・執筆に至るまで、親身にご指導いただいたお陰で、私の実力以上のすばらしい卒業論文となりました。感謝の念でいっぱいです。私自身の大きな自身になったと思います。

渡辺ゼミでは、モバイルコンピューティングとは何かの授業から始まり、最新のモバイル事情、今後のモバイル環境の変化について、おしえていただきました。その授業の中で、興味を持つテーマを聞く事ができ、その議題から本研究が始まりました。

研究を始めて、約 1 年が経ちました。本研究が卒業論文として提出出来たことは大変嬉しく思えます。ここまでの過程には多くの問題がありましたが、その都度、渡辺准教授にアドバイスをいただき、また手助けしていただいたおかげで、ここまで研究を進めることができました。

最後に、私に卒論に関わって下さった方全員にもう一度感謝を述べさせて頂き謝辞とさせていただきます。ありがとうございました。