JAIST Repository
https://dspace.jaist.ac.jp/

Title	JAPANESE - ENGLISH - SPANISH DICTIONARY AND GRAMMAR FOR POCKET PC
Author(s)	フィゲロア松本カルロス健司
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1958
Rights	
Description	Supervisor:白井 清昭，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

JAPANESE – ENGLISH – SPANISH DICTIONARY AND GRAMMAR FOR POCKET PC.[1]

フィゲロア松本カルロス健司(310202)

北陸先端科学技術大学院大学
情報科学研究科

2006 年 2 月 14 日

キーワード： 辞書, Pocket PC, 日本語 – 英語 – 日本語, 文法.

辞書とは、言葉や物事、漢字などを集め、その意味・背景（語源）・使用法（用例）などを解説した本である。辞書・字引とも言う。『辞典』『辞書』という単語は、主に言葉について書かれた辞典（国語辞典、英和辞典、和英辞典など）について使い、漢字について書かれた辞典を『字典』、事物について書かれた辞典（百科事典など）については『事典』という表記を用いる。

Pocket PC は Windows CE をベースとした PDA (携帯情報端末) 向けプラットフォームである。Pocket PC を使用した PDA (携帯情報端末) は、多忙な生活を送るビジネスユーザーを十分満足させ、いつでもどこでも必要な情報にアクセスし、利用するために最適なモバイル環境を提供する。Pocket PC にはビジネス、パーソナル、そしてエンターテイメント分野の多彩なアプリケーションが用意されている。また、業界標準のハードウェアおよびソフトウェアの拡張オプションが多数用意されており、急速に変化する顧客ニーズに対応したソリューションを簡単に提供することができる。

Windows Mobile 2003 software for Pocket PC 日本語版 (正式名称 : Windows Mobile™ 2003 software for Pocket PC 日本語版) は、PDA プラットフォームとして高い評価を受けている Pocket PC の最新バージョンである。

日本で生活している外国人に対して辞書と言うのは日用品になってると思われる。したがって、自分の Pocket PC の中に辞書を持ってれば便利である。実際に辞書がない時に困る場面は多い。外国で日本語を勉強している人々も同様に困っているし、漢字を勉強する為の辞書が必要である。したがって、Pocket PC の辞書を作れば日本語学習者にとってとても便利である。

ネット上にはたくさんの電子辞書があるが、ほとんどが書き言葉しか入っていない。本研究の目的は書き言葉と話し言葉の両方を含む日本語―英語―ス

[1] Copyright © 2006 by Carlos Kenlli Figueroa Matsumoto

ペイン語の辞書を作ることにある。また、その辞書に文法も入れる。文法は日本国際教育支援協会が外国人の為に毎年作るテスト用例文である。

この辞書は外国人の為のものであるため、漢字のリストも辞書に入れる。ユーザは漢字の音読みと訓読みの読み方を勉強することも出来る。

本研究の関連研究を以下に述べる。図書館情報大学の藤井らはネットを百科事典のように利用して、言葉の定義文の取得方法を提案した。彼らの方法は、ＷＥＢページから自動的に言葉や用語の定義文を取得するものである。用語説明獲得のために用語パターンと Html のテキスト構造を使った。情報通信研究機構の Yujie らは、ランキング候補翻訳方法を提案した。この方法は多くの情報から日本語―中国語の二カ国語の語彙を作った。 まず、一つの単語だけである英語の翻訳を集めて、英語―中国語辞書で調べて中国語を取得した。取得した中国語が日本語の翻訳の候補になる。東京大学の田中らは別の方法を提案している。この方法は辞書の構造を使って単語の類似度を測定する。その結果は、名詞のワード対ワード(word-to-word)の二カ国語の辞書になる。この辞書は、辞書に記載する単語を洗練する為に利用することが出来る。一方、KANJIDIC は漢字の情報を持ってるデータベースである。このテキストファイルの量は、6 千 355 行でできている。一つの行は一つの漢字の情報を持っている。このファイルは JIS X 0208-1990 セットになっており、EUC (Extended Unix Code) コーディング 方法で ASCII 文字とカナ/漢字文字が混ざっている。

本研究の辞書は日本語―英語―スペイン語の単語を含む。この情報は日本語―英語辞書ファイル(EDICT) から取得したものと、提案手法のプロセスで出来たものである。提案手法のプロセスとは、(EDICT) にスペイン語の単語を自動的に追加するものである。なお、EDICT はフリーの日本語―英語の辞書ファイルのことである。EDICT には、11 万 2 千 987 単語対が登録されている。

スペイン語の単語を獲得する為に別の英語―スペイン語の辞書を使った。この辞書のファイルには、2 万 4 千 245 単語が登録されている。EDICT の叙述はとても長いので色々な問題がある。本手法は、一つの日本語に対し一つのスペイン語の単語だけ取得する。 プログラムはマイクロソフトの Visual Basic .net と、データベースは SQL SERVER 2000 で作った。

日本国際教育支援協会が日本語を勉強している人々の為に毎年文法のリストを作っている。このリストはとても便利なものである。日本語を勉強をしている人が日本語能力試験のイメージをつかむために作られたものである。大体 1 つのレベルで、3 年間分の学習量に相当する文法がこの辞書に入っている。このシステムを使う人が自分で文法を入れることが出来るし、いらないものも消すことが出来る。これはとても便利な機能である。なぜなら、自分で勉強したい文法を入れて、いつでも何処でも勉強することが出来るからである。

本研究で作成した辞書が持つ情報は以下の通りである。日本語―英語で11万2千987単語対が入っている。日本語―英語―スペイン語で2万2千931単語対が入っている。 しかも、この辞書には漢字の文字が2千135個入っている。漢字の部首は214個入っている、ひらがなは79、カタカナは79文字が入っている。 文法ではレベル４で119行 、レベル３で131行、レベル２で108行とレベル１で108行が入っている。

このシステムのデータベースのマネジャーはSQL Server CE 2.0である。このマネジャーはPocket PCの為に作られているため、本研究の辞書開発に便利である。また、データベースに管理する為の便利なツールでもある。システムのレプリケーション管理はとても簡単に実現出来る。今のシステムに、まだレプリケーションは入っていないが、今後出来るようにするつもりである。

本システムの為にリレーショナルのデータベースを作った。このデータベースはテーブルを10個持っている。カタログ用に6個、単語リスト用に1個 、文法用に1個 、用語用に1個 と、最近検索した単語(履歴)用に1個 である.

本システムはC#のプログラミング言語で作った。.NETのPocket PC専用のコンパックとフレームワークで作成した。この言語はとても便利であり、計算能力も高いために選んだ。しかも、この言語はとても迅速に色々なことが出来る。

このシステムのメインメニューは二つの部分で別れている。

1. ビジュアル部
2. オペレーショナル部

1では、ユーザが勉強することが出来る。そこでは、カタカナやひらがなの文字の画面を表示出来る。漢字の文字の検索画面もある。その画面では、漢字の文字の読み方を勉強出来る。

2では、ユーザが単語や文法の検索画面を表示することが出来る。ここでは新たな情報を登録することが出来る。登録された情報の削除も出来る。

本研究における辞書は、外国人や日本語を勉強している人々にとって理想的な辞書になる為に、さらに多くの情報やオペレーションが必要だと考えられる。今後、辞書に追加するべき情報を以下に挙げる。

1. 日本語能力試験に出てくる動詞のリスト。
2. 助詞の説明。
3. 敬語や謙譲語。
4. 口語。
5. 方言（～弁）

6. サーバーのレプリケーション。（ネットでサーバーと情報を交換することが出来る）
7. システムの SQL Server 5.0 Mobile Manager への移植。(データベースの管理をする為や、スピードアップする為)