

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

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

Figueroa Matsumoto Carlos Kenlli (310202)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 14, 2006

Keywords - Dictionary, Pocket PC, Japanese – English – Spanish,
Japanese grammar.

A dictionary is a list of words with their definitions, a list of characters with their glyphs, or a list of words with corresponding words in other languages. In some languages, words can appear in many different forms, but only the lemma form appears as the main word or headword in most dictionaries. Many dictionaries also provide pronunciation information; grammatical information; word derivations, histories, or etymologies; illustrations; usage guidance; and examples in phrases or sentences. Dictionaries are most commonly found in the form of a book.

A Pocket PC is a computer in a handheld size that runs a variation of the operating system Windows CE. It has many capabilities of modern desktop PCs. Currently there are thousands of applications for Pocket PC, many free. Some of these devices, running Microsoft Windows Mobile 2003 Phone Edition, also include mobile phone features. Pocket PCs can also be used with many other add-ons like GPS receivers or barcode readers.

According to Microsoft, the Pocket PC is "a handheld device that enables you to store and retrieve e-mail, contacts, appointments, play multimedia files, games, exchange text messages with MSN Messenger, browse the Web, and more.

For instance, any device which is to be classified as a pocket PC must:

- Run Microsoft's Windows CE (PocketPC version) operating system
- Come bundled with a specific suite of applications in ROM
- Include a touchscreen
- Include a directional pad or touchpad
- Include a set of hardware application buttons

¹ Copyright © 2006 by Carlos Kenlli Figueroa Matsumoto

- Be based on a StrongARM compatible CPU (older Pocket PCs have a MIPS or SH3 processor)

Dictionaries and encyclopedias are indispensable tools for any professional. And since you never know when you're going to need one, they are ideal references to bring along with you on your Pocket PC. Since I know some of the problems that foreign people have when they study in Japan and also the people that are studying Japanese in other countries, I thought that develop a dictionary pocket dictionary is a good tool for all those people.

I know that there are a lot of Japanese Dictionaries in the web but all of them don't have speaking words or must the their words are written words. My goal is to develop a dictionary that has also written and spoken words. And since not also a dictionary is indispensable for the life, the dictionary will have some essential grammar that is produced each year for the Japan Foundation for the learners and teachers of Japanese living outside Japan. Also in the dictionary the user will be able to insert useful information for him.

Since my goal is mainly a dictionary of Japanese for foreign people, it will have a list of the most important kanji with their readings, and meanings.

I refer to some articles; some of them are included below.

- I refer to some articles of Professor Fujii and Professor Ishikawa of the University of Library and Information Science. They have some articles about utilizing the web like an encyclopedia to extract terms descriptions. They purpose a method used to automatically extract descriptions of technical terms from the Web pages in order to utilize the World Wide Web as an encyclopedia and summarize them.
- I refer to an article of Professor Yujie of the Communications Research Laboratory in Kyoto. She developed a method of ranking candidate translation by integrating multiple sources of information into one scoring scale, and applied it to construct a Japanese – Chinese bilingual lexicon.
- Professor Tanaka of the Division of Engineer of the University of Tokyo propose a method to discriminate equivalences from inappropriate words derived as a result of ambiguity in the third language, utilizing the structures of dictionaries to measured the nearness of the meanings of words. The resulting dictionary is a word-to-word bilingual dictionary of nouns and can be used to refine the entries and equivalences in published bilingual dictionaries.
- The KANJIDIC file contains comprehensive information about Japanese kanji. It is a text file currently 6,355 lines long, with one line for each kanji in the two levels of the characters specified in the JIS X 0208-1990 set.

The file contains a mixture of ASCII characters and kana/kanji encoded using the EUC (Extended Unix Code) coding.

Now I am going to talk about what my project is. On this project the information is included in triples, Japanese – English – Spanish. The information was obtained from a Japanese – English file dictionary (EDICT) and made a process that added the Spanish descriptions.

EDICT is a Japanese – English dictionary freely available file. It has 112,987 entries.

To get the Spanish descriptions I used a vocabulary of English – Spanish of 24,245 entries, and make matches to get the triples. Since EDICT English descriptions have different words for a certain Japanese words, I developed a process to make the matches and trying to get exactly correct Spanish translations only. The process was made in visual basic .net and the database in SQL SERVER 2000

The Japan Foundation generates yearly a list of grammar for Japanese language learners. This grammar help us to have a very rapidly image of the Japanese Language Proficiency Test.

Three years of this grammar per each level was inserted into the database of this system, I think that this is enough for the beginning, actually the system is able to add new grammar and delete the grammar that the user doesn't want. This is a good point for the user even that they have already the information for the Japanese test, they can add new grammar that the want to study whenever they want.

The information in this project is the most important part. As I said before the dictionary contains 112,987 entries with Japanese – English combinations and 22,931 entries of those contains the Japanese – English – Spanish descriptions.

Also the project contains 2,135 entries of kanji character with descriptions, 214 entries of radicals of kanji character, 79 entries of hiragana characters and 79 entries of katakana characters.

For the grammar it contains 119 entries for level 4, 131 entries for level 3, 108 entries for level 2 and 108 entries for level 1.

For this project I choose SQL Server CE because is a very good Database Manager, is very compact for the device and you can carry a very huge database without having problems with the space. Also is very simple to create a replication process with a server. This process will be enabled in the future.

I develop a relational database for the system that consists in 10 tables, 6 catalogs, 1 table for the words, 1 for the grammar, 1 for the dialects and 1 for the recent searched words.

This system was developed in C# programming language with a compact framework of .net for small devices. I choose this language because is powerful, easy to develop and since it is a member of the family of windows framework it has a very fast perform on the windows CE operating system.

I divided the project in two parts.

1. The only visual part
2. The operational part

On the 1st part the user can study the kanji, hiragana and katakana character. The kanji part is also an operational part, where the user can search for a certain kanji and read all its properties.

On the 2nd part the user can “Search”, “Add” or “Delete” information from the database, this means, an operational part where the user manage the information on their database.

Many things have to be added in this project to be the ideal dictionary for a foreign people; I think that the next items have to be added to this system to be the ideal dictionary.

1. List of verbs included on the JLPT,
2. Particles and the explanation of each one and its use
3. Words to talk to people that is older than us, or to people that is in higher level in the companies
4. More daily words
5. Dialects, etc...

Many times when I was on the street and I wanted to talk to older people they answered me with dialect of that region and I couldn't understand anything because that words we can't study in the school, so I think that making a huge vocabulary of different dialect of Japan will be helpful for the foreigners.

6. Replication to a central server, and exchange information with Pocket PC databases via web.
7. Migrate the system to the newest database manager for mobile devices, the SQL Server 5.0 Mobile Manager