

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.

By Carlos Kenlli Figueroa Matsumoto

A thesis submitted to

School of Information Science,
Japan Advanced Institute of Science and Technology,
In partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Kiyoaki Shirai

March, 2006

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

By Carlos Kenlli Figueroa Matsumoto (310202)

A thesis submitted to

School of Information Science,
Japan Advanced Institute of Science and Technology,
In partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Kiyoaki Shirai

And approved by
Associate Professor Kiyoaki Shirai
Professor Akira Shimazu
Professor Satoshi Tojo

February, 2006

¹ Copyright © 2006 by Carlos Kenlli Figueroa Matsumoto

Chapter 1

Introduction

1.1 BACKGROUND

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.[1]

As the previous definition of a dictionary say most of the dictionaries are on books. Nowadays there are some electronic dictionaries, one of them are the Pocket PC Dictionaries.

Microsoft® Windows® CE 5.0 is an open, scalable, 32-bit operating system (OS) that integrates reliable, real time capabilities with advanced Windows technologies. Windows CE allows you to build a wide range of innovative, small footprint devices. A typical Windows CE-based device is designed for a specific use, often runs disconnected from other computers, and requires a small OS that has a deterministic response to interrupts. Examples include enterprise tools, such as industrial controllers, communications hubs, and point-of-sale terminals, and consumer products, such as cameras, Internet appliances, and interactive televisions.

Windows CE also includes production-quality device drivers that use less hardware platform-specific code than previous drivers, abstract all resource assignments to the registry, and create reusable, hardware platform-independent libraries. These libraries work without modification for many of the common hardware platforms. This reduces the customization required for the drivers to function on custom hardware.

Windows CE brings enhanced productivity to application developers by providing a rich set of programming tools to create managed code and native code applications for Windows CE-based devices. Windows CE offers application developers the versatile environment of the Microsoft Win32® application programming interface (API) along with the ease and versatility of scripting languages. It also offers bundled support for multimedia, the Internet, local area networks (LANs), and mobile communications and security services.[2]

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.

From a technical standpoint, "Pocket PC" is a standard from Microsoft that sets various hardware and software requirements for mobile devices bearing the "Pocket PC" label.

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

Microsoft's current release is **Windows Mobile 5.0**, internally code-named *Magneto*. It was officially released by Microsoft on May 10, 2005. Based on Windows CE 5.0, Windows Mobile 5.0 incorporates features such as Direct3D Mobile, integrated Bluetooth and WiFi stacks, and the ability to be mounted as a USB Mass Storage device, bypassing ActiveSync. Pocket Word and Pocket Excel have been renamed Word Mobile and Excel Mobile, respectively, and both applications have gained more complete feature sets, including support for tables, ordered lists, and embedded graphics in Word Mobile, and support for charts in Excel Mobile. PowerPoint Mobile, a new addition to the Office Mobile suite, supports the ability to display and edit PowerPoint presentations on a Pocket PC. Windows Mobile 5.0 also improves support for one-handed usage

by inserting two, context-sensitive "soft" buttons at the bottom of the screen which can be mapped to hardware buttons on any specific device.[3]

Windows Mobile 5.0 marks the convergence of the Phone Edition and Professional Edition operating systems into one system that contains both phone and PDA capabilities. A 'Phone' application is now included in the OS, and all PIM applications have been updated to interface with it. Windows Mobile 5.0 is compatible with Microsoft's Smartphone operating system and is capable of running Smartphone applications.

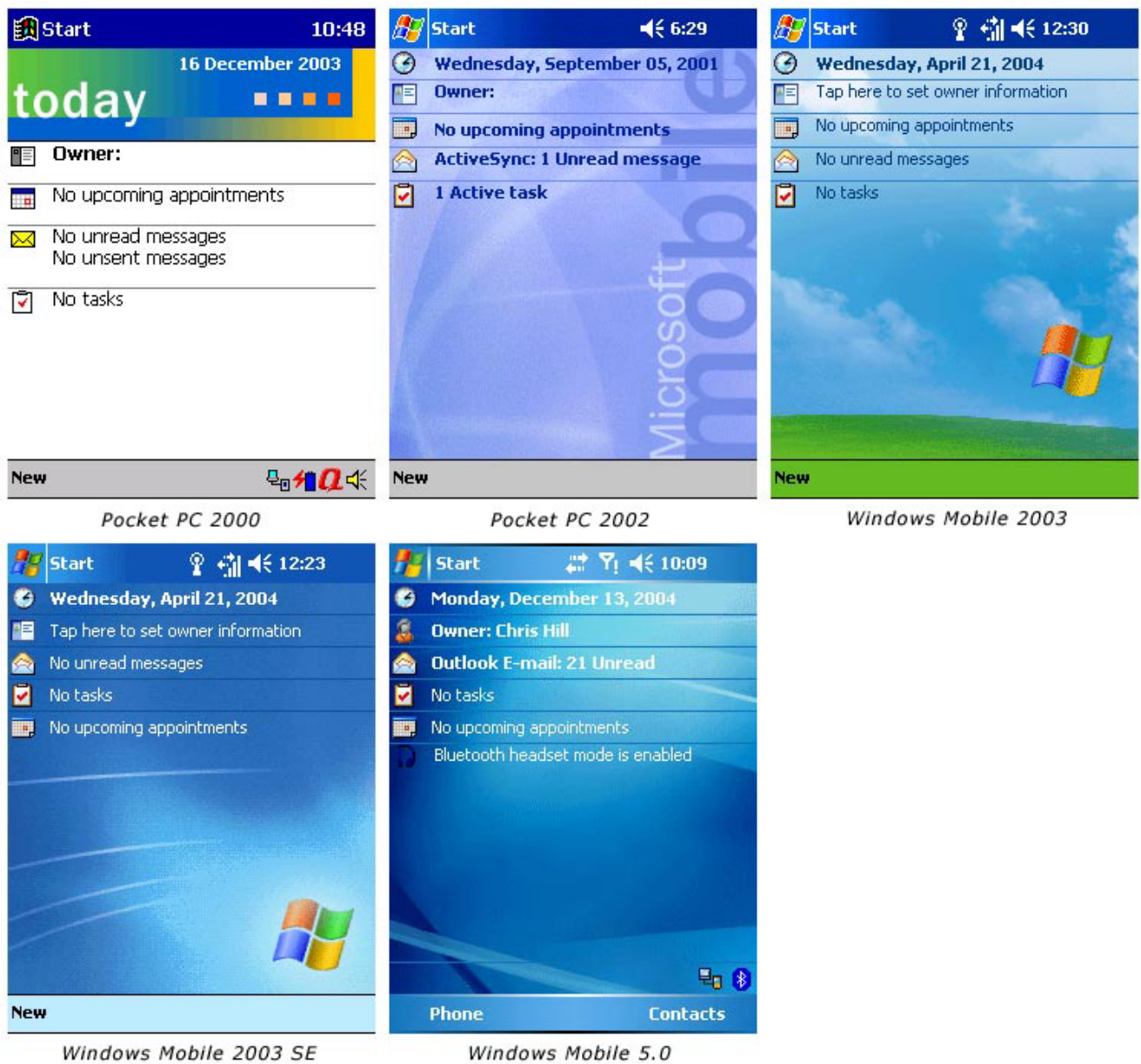


Figure 1: Screen shots the history of the OS for Pocket PC

Pocket PCs running previous versions of the operating system generally stored user-installed applications and data in RAM, which meant that if the battery was depleted the device would lose all of its data. Windows Mobile 5.0 solves this problem by storing all user data in the ROM, leaving the RAM to be used only for running applications, as it would be on a desktop computer. As a result, Windows Mobile 5.0 Pocket PCs generally have a greater amount of flash ROM, and a smaller amount of RAM, compared to earlier devices.

The previous release, **Windows Mobile 2003 Second Edition** added native landscape support as well as other fixes and changes to those features already present in the original release of Microsoft Windows Mobile 2003 for Pocket PC, also known as WM2003.

Microsoft Windows Mobile 2003 for Pocket PC consisted of the Windows CE.NET 4.2 operating system bundled with scaled-down versions of many popular desktop applications, including Microsoft Outlook, Internet Explorer, Word, Excel, Windows Media Player, and others. Past Pocket PC operating systems include Pocket PC 2002 (launched October 2001) and Pocket PC 2000 (launched April 2000), both running Windows CE 3.0 underneath.[3] The *Figure 1* shows the evolution of the Pocket PC operating system since 2000.

1.2 GOAL

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.

Chapter 2

Related Works

2.1 OTHER POCKET PC DICTIONARIES ON THE WEB

There are a bunch of dictionaries on the web for Pocket PC, I am going to mention some and compare them each other and then compare with my project.

1 SlovoEd.

SlovoEd is a series of best-selling translation and explanatory dictionaries for Pocket PC/Windows Mobile. SlovoEd covers more than 150 language pairs and provides the unique feature-rich dictionary solution powered with lexical databases from world-leading linguistic companies. [4]

One of the SlovoEd dictionaries related to my purpose is the Spanish-English-Spanish Pocket PC dictionary that contains 103 065 entries. Price: \$32.95

2 LingvoSoft (Ectaco [7]).

LingvoSoft is home to talented and experienced software engineers and linguists who create a variety of language-learning, dictionary, translation and localization software products.

Implementing superb word databases and supporting over 45 languages, LingvoSoft applications have become the de facto standard in working with languages. High quality, effectiveness, and use of the latest technologies set LingvoSoft software apart from the rest of the field. Moreover, we bring you top translation and localization software by other renowned developers. [5]

The LingvoSoft related software is:
LingvoSoft English <-> Spanish Talking Dictionary for Pocket PC. It has 400,000 (words/phrases) Price: \$49.95 and
LingvoSoft Talking Dictionary 2006 English <-> Japanese (Kanji). That also has 400,000 (words/phrases) Price: \$49.95

3 AbsoluteWord

AbsoluteWord has more than 300 multilingual and specialty dictionaries and encyclopedias for Palm OS® and Pocket PC. [6]

The related software is:

Japanese-English Dictionary has 97928 entries. Recommended for devices with expansion cards. Also you need to have Japanese localization system installed to display Japanese characters.

English-Japanese Dictionary has 78912 entries. This dictionary contains the most used words and it is perfect word set to fit in memory of 8MB devices, and devices with expansion cards. Also you need to have Japanese localization system installed to display Japanese characters. Price individually: \$19.95 and both: \$ 24.95

I think that those are the most important companies that make dictionaries for Pocket PC.

Most of the problem that I found on this dictionaries was that you have to install two dictionaries to have one meaning, I mean if I want to have a meaning in Japanese for a Spanish word I can not search for it immediately, you have to search first the word in English and then search that word in English – Japanese dictionary. That was very hard for me and I think is the same for a lot of people that speak Spanish and also in the opposite way (Japanese persons that want a meaning in Spanish).

Another problem is that not all the companies have Japanese – English dictionaries.

Looking to those problems I decide to include in one dictionary 3 languages (Japanese – English – Spanish) so with this the problem to install two dictionaries disappeared.

Another problem that I found is that sometimes is necessary the meaning or the reading of certain kanji and there is no dictionary that gives that advantage. So I include on this project a list of the most important kanji with the readings and some meanings of each one.

Nowadays there is no dictionary on the web that contains grammar.

My purpose on the grammar is give some important advice to people that is studying Japanese and want to get the proficiency test of Japanese.

2.2 ACQUIRING TERM DEFINITION

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. [8][9]

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. They use linguistic patterns and Html text structures to extract text fragments containing term descriptions.

The process consists in a search engine searches the Web pages for pages containing the terms listed in a lexicon. Then, fragments of retrieved Web pages are extracted based on linguistic and structural description patterns. However, in some cases the extraction process is not unsuccessful, and thus extracted fragments are not linguistically understandable. In view of this problem they perform a filtering to enhance the extraction accuracy.

The URLs of the Web pages from which descriptions were extracted are also indexed in the database, so that users can browse the full content. In the case where a number of descriptions are extracted for a single term, the resultant description is redundant, because it contains a number of similar descriptions. So they use a clustering method to divide descriptions for a single term into a certain number of clusters, and present only descriptions that are representative for each cluster.

To summarize the term descriptions they purpose a summarization method, which produces a single text from multiple descriptions. The resultant summary concisely describes a term from different viewpoints. Their purpose is to include as many viewpoints as possible in a concise description. The viewpoint-based summarization (VBS) method consists of the following steps:

- 1 Identification, which recognizes the language unit associated with a viewpoint.
- 2 Classification, which merges the identified units associated with the same viewpoint into a single group.
- 3 Selection, which determines one or more representative units for each group.
- 4 Presentation, which produces a summary in a specific format.

The implementation of each step varies depending on the application.

This method was pretty interesting but there are some differences with my project and it was not utilized because this project consists only in have definitions for words, I mean get only one pair of descriptions. And my project mainly gets triples of description, so the relation was not so close, and wasn't used.

2.3 CONSTRUCT A BILINGUAL DICTIONARY INTERMEDIATED BY A THIRD LANGUAGE

I refer to an article of Professor Yujie of the Communications Research Laboratory in Kyoto [10]. 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.

She used two bilingual lexicons.

- 1 Japanese – English lexicon that contains 364,430 records. Each record consists of the record number, Japanese word, part of the speech, concept identifier, English translations, that may be a single word, phrase or an explanatory sentence, and related information.
- 2 English – Chinese lexicon that contains 110,834 entries and 128,366 entries in the Chinese – English lexicon. In both each entry consists of only the source word and the corresponding target words.

The procedure of acquiring candidate Chinese translations was as follows. First, collect the English translations that are single words. Second, look up each collected English translation in the English – Chinese lexicon and obtain Chinese translations. Then, gather all the Chinese translations together and define them as the set of candidate Chinese translations for the Japanese record.

With this procedure she obtained sets of candidate Chinese translations for 144,002 records in the Japanese – English lexicon. Professor Yujie found that almost all of the sets of candidates contained correct Chinese translations, along with many incorrect translations. These incorrect translations came because although the English translations of the lexicon records are completely equivalent meaning to the corresponding Japanese words, the English words may also have other meanings of functions as other parts of speech. These ambiguities result in incorrect Chinese translations.

To select correct translations, we considered ranking candidates according to their possibilities of being correct translations. The possibilities can be viewed as measuring the similarity between a Japanese word and a candidate Chinese

translation in terms of several aspects, such as semantics, part of the speech, and composition.

She assumed that the characteristics of lexicon composition in one language would probably be more or less reflected in the corresponding translation in another language. Japanese words consist of kanji, hiragana and katakana. Chinese words consist of Chinese characters, so she exploited the correspondences between kanji of Japanese and Chinese characters in order to use the correspondences in ranking candidate Chinese translations. After obtained the correspondences between kanji and Chinese characters Professor Yujie utilized the correspondences to measure the similarity between Japanese words and candidate Chinese translations in meaning and added the similarity into the scoring function.

Professor Tanaka of the Division of Engineer of the University of Tokyo [11] 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. Looking up words in bilingual dictionaries intermediating the third language is a method often used by people who handle uncommon languages in a specific domain. If this process can be automated, bilingual dictionaries of any kind between any languages may be obtained as long as these concerned languages have dictionaries to common languages. One objective of her research was to establish a first step in automating process.

To construct a Japanese – French dictionary, she chose English as the intermediary language because Japanese – English and English – French dictionaries already exists in electronic forms and because published Japanese – French dictionaries provide enough vocabulary in comparison with the resulting dictionary.

Her method relies on inverse consultation. The most naïve way to use English to obtain French words corresponding to a Japanese word is to look up the Japanese word in a dictionary Japanese – English and the look up the resultant English words in an English – French dictionary. The resulting French words can be regarded as equivalence candidates of the original Japanese word. [12]

The following cases generate irrelevant equivalence candidates.

1. An English word with the same spelling but with different meanings is intermediated.
2. An English word with a wider meaning than that of original Japanese word is intermediated.
3. There are mistakes in dictionaries.

These papers were very useful to get the triples of my project, I used the same method of them, I mean get information from Japanese – English dictionary and then get Japanese – Spanish dictionary. I used English like intermediated dictionary to get the final information.

The difference between these projects and mine are that I can't use kanji characters to get the final information. Other one is that nowadays there are no triples between Japanese – English – Spanish.

2.5 KANJIDIC

The KANJIDIC [13] 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.

KANJIDIC is now rather large, and has information in it, which is not much use for people who are not studying and researching Japanese orthography. It is still appropriate to maintain it as a useful freely available compendium of such information.

KANJIDIC project was very related with my project, because I also show the information of each kanji, but the difference is that I don't use the same syntaxes as this project. They sometimes use txt files or XML files to read the information and I use SQL database to get the information.

Chapter 3

Compilation of the Dictionary

In this chapter, I will describe basically some points of the information, how did I get it, problems that I had and also I will talk about the grammar included in this project. This part of the project (information) wasn't developed complete so I will also talk about the future work of this part of my system.

3.1 WORD TRIPLES OF JAPANESE – ENGLISH – SPANISH

On this project the information is included in triples, Japanese – English – Spanish. The information was obtained from a Japanese – English file dictionary (EDICT) and with the process explained below 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 function below is the main function for the process.

1. *If (dsENGSP.Tables("ENGSP").Rows.Count > 0) Then*

```

2. For i = 0 To dsENGSP.Tables("ENGSP").Rows.Count - 1

3.     stresp = CStr(dsENGSP.Tables("ENGSP").Rows(i)("Spanish"))
4.     streng = CStr(dsENGSP.Tables("ENGSP").Rows(i)("English"))

5.     If (Me.Find(streng)) Then

6.         strqry = "SELECT * FROM TEMP WHERE English LIKE '%" & to " +
streng & ",%'"

7.         dsJES = con.ExecuteDSQuery(strqry, Conn, "JESP")

8.         If (dsJES.Tables("JESP").Rows.Count > 0) Then

9.             For z = 0 To dsJES.Tables("JESP").Rows.Count - 1

10.                If (IsDBNull(dsJES.Tables("JESP").Rows(z)("Spanish"))) Then
11.                    id = CStr(dsJES.Tables("JESP").Rows(z)("id"))

12.                    strqry = "UPDATE TEMP SET SPANISH = " & stresp & "
WHERE ID = " & id
13.                    dsResult = con.ExecuteDSQuery(strqry, Conn, "LAST")

14.                End If
15.            Next
16.            dsJES.Clear()
17.        End If
18.    End If
19. Next
20. End if

```

Before line 1, I get all the information of the English – Spanish dictionary, and then I will search the words on the Japanese – English dictionary to get the triple Japanese – English – Spanish.

At the line 2, I loop the entire English – Spanish dictionary. Then at line 3 I get one by one all the English and Spanish words and the search on the Japanese – English dictionary.

Some of the English words contains estranges characters so I make a function that finds that estrange characters and erase them, the call to this function is mentioned at line 5.

At line 7, the function searches on the Japanese – English dictionary the English word obtained in the English – Spanish dictionary.

If I found some results on the Japanese – English dictionary loop that results as shown in the line 9, then I will try to insert the Spanish word to make the triples.

If the Spanish field is blank (line10), this means that there is no Spanish meaning, I update the field and insert the new meaning (line 12). In the case that the Spanish field already got a meaning I discard that the new Spanish word and continue with the next word until finish the results.

Then I continue with the next word in the English – Spanish dictionary until I finish with this dictionary (line 19).

This function takes a word from the English – Spanish vocabulary and searches on the EDICT English description. This process was run many times, because I found some problems on the matches. For example in the English – Japanese vocabulary I didn't have the word "to" for the verbs, and EDICT puts the prefix "to" for the verbs. To solve this problem I search all the verbs in the English – Spanish vocabulary and add the prefix "to", then search and try to make the match with the descriptions of EDICT.

E.g.

EDICT: to become
E-S: (v) become

In this case the function didn't match because the word "to become" and the word "(v) become" was not equal.

The solution was:

EDICT: to become
E-S: (v) to become

And with this change I can get the correct match.

Another problem was that in the English description of EDICT there are many descriptions for a certain Japanese word, and sometimes different English words have the same meaning in Spanish, so I had the same Spanish word two or more times or sometimes the meaning in the Spanish position was not correct. After check it randomly and making the comparison of quantity of correct words between two or more times the same meaning and have only on meaning I decide to manage this problem just adding one description on the Spanish field. The function check if the word already has a Spanish description I continue with the next word.

EDICT: often, frequent
Spanish: frecuentemente, frecuentemente

The solution was to check if the word already has description do not write another description

EDICT: often, frequent
Spanish: frecuentemente

After getting all the information of the system, I had to check the information, randomly, it means I take information from the database and check it manually; I take around 100 random triples and check if the information was correct. The first time that I checked the information (when I got two or more words in the Spanish field) maybe a 70% of the information had errors. After get only one word in the Spanish field, I got maybe a 95% that the information was correct.

Maybe one of the advantages that I have on this part is that the Spanish speakers can understand the meaning of the English word with just one word in Spanish. And also for English speakers sometimes with only one word they can say a meaning that in Spanish we say with many words.

When the Japanese word has a large description in English some times the match most of the times gets incorrect matches so If the English description is more than 1 word and it doesn't match with the Spanish description It goes to the next word.

EDICT: divine message...
Spanish: divinidad, mensaje

The solution was that if the Spanish description doesn't match with the whole sentences do not write the description

EDICT: divine message...
Spanish: <NULL>

3.2 GRAMMAR AND RELATED INFORMATION

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.

What is the Japanese Language Proficiency Test [17]?

As the number of people studying Japanese is rapidly increasing worldwide, so are the opportunities to put their skills to use in a wide variety of careers. This has led to the demand among the students of Japanese for the establishment of a venue by which their language proficiency can be certified. The Japanese-Language Proficiency Test was devised to respond to this growing demand in 1984.

The Japanese-Language Proficiency Test is conducted both in Japan and outside Japan to evaluate and certify the language proficiency of primarily non-native speakers of Japanese.

The test is administered by Japan Education Exchanges and Services, inside Japan, and by the Japan Foundation outside Japan.

Contents of the Test.

The test has four different levels (*Table 1*); the examinee can choose the level that best matches his or her ability and training. Each test is made up of three sections: writing-vocabulary; listening; reading-grammar[18].

The contents and criteria of the test are as follows:

Level	Contents			Criteria
	Sections	Hours	Points	
1	Writing-vocabulary Listening Reading-grammar Total	45min. 45min. 90min. 180min.	100points 100points 200points 400points	The examinee has mastered grammar to a high level, knows around 2,000 kanji and 10,000 words, and has an integrated command of the language sufficient for life in Japanese society. This level is normally reached after studying Japanese for around 900 hours.
2	Writing-vocabulary Listening Reading-grammar Total	35min. 40min. 70min. 145min.	100points 100points 200points 400points	The examinee has mastered grammar to a relatively high level, knows around 1,000 kanji and 6,000 words, and has the ability to converse, read, and write about matters of a general nature. This level is normally reached after studying Japanese for around 600 hours and after completion of an intermediate course.
3	Writing-vocabulary	35min. 35min.	100points 100points	The examinee has mastered grammar to a limited level, knows

	Listening Reading- grammar Total	70min. 140min.	200points 400points	around 300 kanji and 1,500 words, has the ability to take part in everyday conversation and read and write simple sentences. This level is normally reached after studying Japanese for around 300 hours and after completion of an elementary course.
4	Writing- vocabulary Listening Reading- grammar Total	25min. 25min. 50min. 100min.	100points 100points 200points 400points	The examinee has mastered the basic elements of grammar, knows around 100 kanji and 800 words, has the ability to engage in simple conversation and read and write short, simple sentences. This level is normally reached after studying Japanese for around 150 hours and after completion of the first half of an elementary course.

Table 1: Requirements for each level of the JLPT

The Figure 2 shows the requirements of each level for the different parts of the examination.

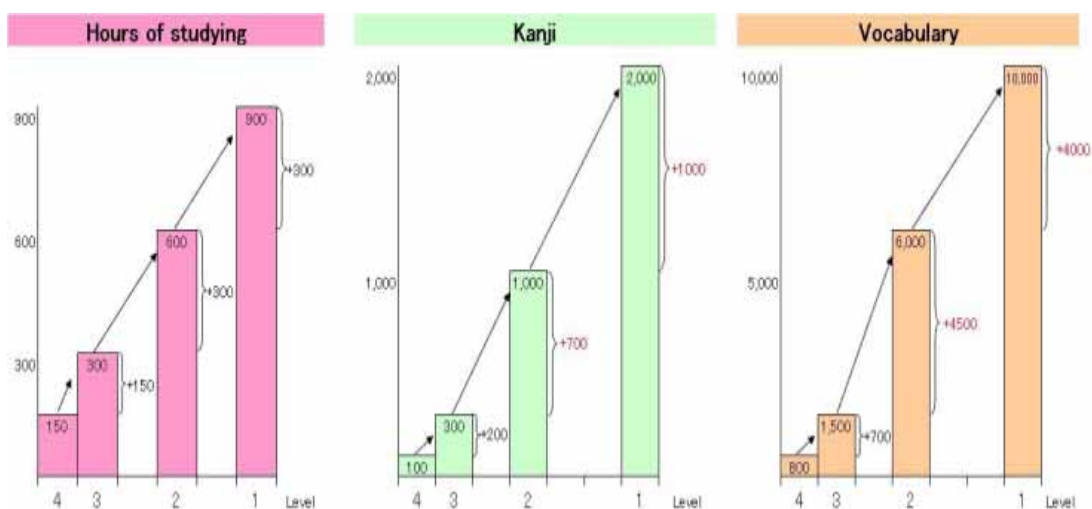


Figure 2: Chart the requirements for JLPT for the different levels.

All this information is also included on the dictionary so the user can check the level on where he or she is.

An example of the Grammar included on the project is mentioned below. As there are different things to study in a language on this part the user can insert grammar to the project.

For the level 4 “Basic Level” the grammar asked on the 2000-year examination was:

- 1.きのう ここに だれが 来ましたか。
- 2.私は 山田さんを ここで まちます。
- 3.1日に 2かい、はを みがきます。
- 4.テーブルの上に みかんや りんごが あります。
- 5.母は だいどころに います。

As you can see the level of the sentences are very poor, this is the level for the beginners of the studying of Japanese. What did we get from these sentences? Well foreign people study these kinds of sentences for the examination; most of the subjects studied in the classroom are contained in these sentences. I mean grammar and vocabulary. For example, one of the basics of Japanese is the particle (に, が, を, etc...) so the foreign people with sentences has to study that the particle に comes after a place, or that the particle が comes after the person that makes the action. I think these sentences are a very good way to train for the examination and that's the reason of including in the dictionary.

For the level 3 “Conversational Level” the grammar asked on the 2004-year examination was:

- 1.じしん 『で』 ビルがたおれました。
- 2.いっしょにコ - ヒー 『でも』 いかがですか。
- 3.来週ここでテニスの試合 『が』 行われます。
- 4.じしよをわすれたので、友だち 『に』 貸してもらった。
- 5.私は、中川 『と』 もうします。

Here in the level 3 of the test, you can see that the sentences have more kanji characters and that the level of the sentences is different from the level 4. Some of the grammar including in the level 4 are included here. Some of the foreign people that study on intensive courses, the used to apply level 3 examination instead of 4.

After the level 3 some times you can see the brackets (『 』) for us this means that we have to be careful on this point. Most of the sentences in level 3 that has this bracket mean that we have to be careful about the particles, that is one of the basics of the Japanese grammar. This point is extremely important for the Japanese learners, so the brackets remark it.

For the level 2 “Business Level “ the grammar asked on the 2004-year examination was:

- 1.戦争に『対して』]、批判の声が次第に高まっている。
- 2.関係者のみなさまのご協力に『よって』、無事この会を終了することができました。
- 3.思いつく『かぎり』のアイデアはすべて出したが、社長は認めてくれなかった。
- 4.外国人社員の労働条件を『めぐって』、会社側と労働者側が対立している。
- 5.これから旅行に『行くのなら』、海がいい。

And for the level 1 “Native Level” the grammar asked on the 2004-year examination was:

- 1.当社では学歴を『問わず』多くの優秀な人材を集めるため、履歴書の学歴欄を廃止した。
- 2.二酸化炭素を多く発生させている国の協力『ぬきには』、温暖化を防ぐことはできないだろう。
- 3.アパートの住人が突然追い出される『ことのないよう』保護する必要がある。
- 4.周囲の人が反対『しようとしまいと』、私の気持ちは変わらない。
- 5.彼は会社勤めの『かたわら』、福祉活動に積極的に取り組んでいる。

As you can see in the these two levels the quantity of kanji characters increase so much, so the step between level 3 and level 2 is to huge. Also the grammar increase and the level of the sentences is too difficult.

3.3 RESULTS

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.

3.4 FUTURE WORK

For this part I think that there are a lot of things to include.

1. List of verbs included on the JLPT,
The verbs for each level is divided by the Japan Foundation, and for the future I want to add the verb list for each level with all the tenses of the verb. It means for example.
食べます 食べました 食べる 食べた 食べよう。。。.
2. Particles and the explanation of each one and its use
As I said before one of the most important part of the Japanese grammar is the use of the particles, so I want to add an explanation of this part of the grammar, the most common uses, when to use the particle, after which verb we have to use that particle, etc...
3. Words to talk to people that is older than us, or to people that is in higher level in the companies
Add information about honorific words (尊敬語, 謙讓語), I mean when to use, to who use, etc...
4. More daily words
I want to add to the dictionary common words, more spoken words and some slang that most of the people use when they are talking with friends.
5. Dialects
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. So for the future I want to add information about the different regions dialects. Nowadays there is information about dialects but it is a few words only.

I think that this part has to grow a lot because it's a very important that the foreign people speak Japanese like native, and by this way the Japanese people start to be more comfortable when they are with them.

Chapter 4

System

In this Chapter, I will describe the entire system, I want to talk about each part of the system, explain the specifically the database and the functionality of the interface.

4.1 Database

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 consist 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 (*Diagram 1*).

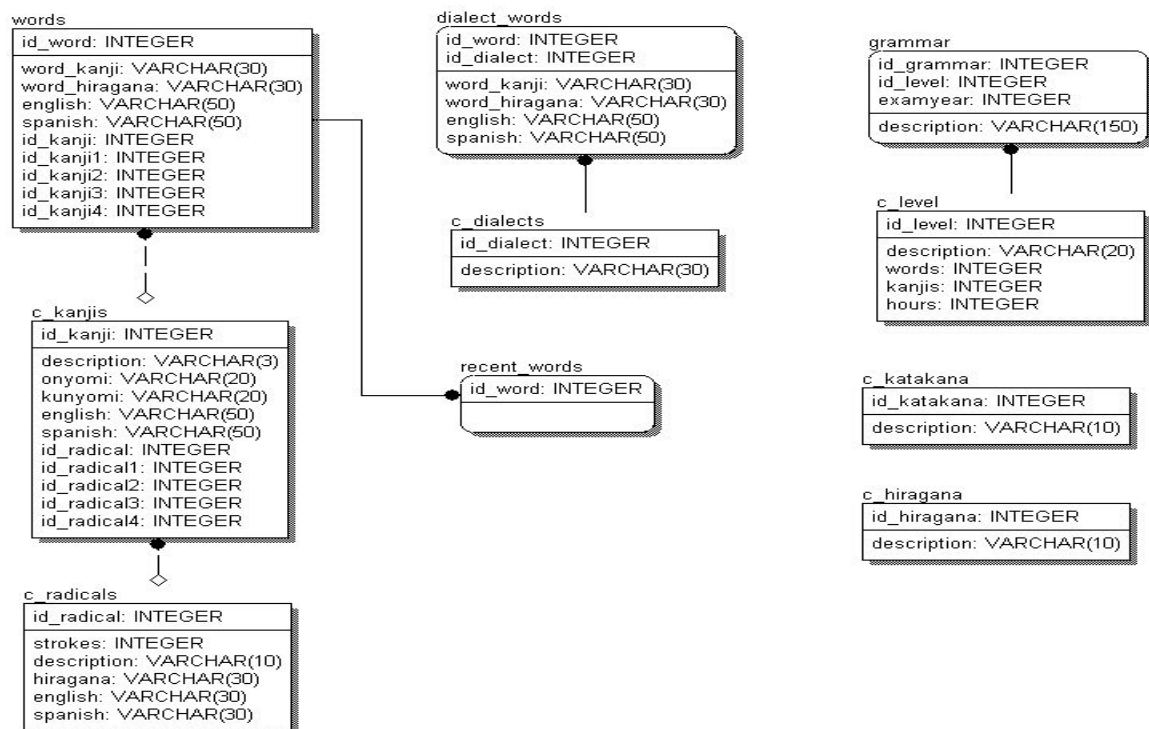


Diagram 1: Relational Database of the System.

TABLE NAME: c_hiragana

DESCRIPTION: This catalog contains all the Hiragana characters. These characters are shown in the system for their study.

	Field Name	Description
PK	Id_hiragana	Character id
	Description	Hiragana Character

TABLE NAME: c_katakana

DESCRIPTION: This catalog contains all the Katakana characters. These characters are shown in the system for their study.

	Field Name	Description
PK	Id_katakana	Character id

	Description	Katakana Character
--	-------------	--------------------

TABLE NAME: c_level

DESCRIPTION: This catalog contains all the levels for the grammar.

	Field Name	Description
PK	Id_level	Level id
	Description	Level
	Words	Number of words studied to apply to get this level
	Kanji	Number of kanji character that must know to get this level
	Hours	Number of hours that must studied to get this level

TABLE NAME: c_dialects

DESCRIPTION: This catalog contains all the dialects in Japan. This table is not complete yet; it has only a few dialects region names.

	Field Name	Description
PK	Id_dialect	Dialect id
	Description	Name of the region of the dialect

TABLE NAME: c_radicals

DESCRIPTION: This catalog contains all the radicals and their strokes of the kanji characters. This table contains 214 entries.

	Field Name	Description
PK	Id_radical	Radical id
	Strokes	Number of strokes of the

		radical
	Description	Radical character
	Hiragana	Name in hiragana
	English	Description in English
	Spanish	Description in Spanish

TABLE NAME: c_kanji

DESCRIPTION: This catalog contains the most important kanji character.
This table contains 2135 entries.

	Field Name	Description
PK	Id_kanji	Kanji id
	Description	Kanji character
	Onyomi	Onyomi lecture
	Kunyomi	Kunyomi lecture
	English	Description in English
	Spanish	Description in Spanish
FK	Id_radical1	Radical character 1
FK	Id_radical2	Radical character 2
FK	Id_radical3	Radical character 3
FK	Id_radical4	Radical character 4

TABLE NAME: grammar

DESCRIPTION: This table contains all the grammar. For the future work this table could change.

	Field Name	Description
PK	Id_grammar	Grammar id
PK	Id_level	Level id
	Examyyear	Year of examination
	Description	Grammar description.

TABLE NAME: dialect_words

DESCRIPTION: This table contains the words for the dialects. Actually this table is not used. In the future it will contains all the words and descriptions for the dialect words.

	Field Name	Description
PK	Id_word	Word id
PK	Id_dialect	Dialect id
	Word_kanji	Word in kanji characters
	Word_hiragana	Word in hiragana characters
	English	Description in English
	Spanish	Description in Spanish

TABLE NAME: words

DESCRIPTION: This table contains the words for the dictionary. It contains 112,987 entries with Japanese – English combinations and 22,931 entries of those contain the Japanese – English – Spanish descriptions.

	Field Name	Description
PK	Id_word	Word id
	Word_kanji	Word in kanji characters
	Word_hiragana	Word in hiragana characters
	English	Description in English
	Spanish	Description in Spanish
	Id_kanji1	Kanji id 1
	Id_kanji2	Kanji id 2
	Id_kanji3	Kanji id 3
	Id_kanji4	Kanji id 4

TABLE NAME: recent_words

DESCRIPTION: This table contains the last 10 searched words.

	Field Name	Description
PK	Id_word	Word id

4.2 SOFTWARE

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.

Some screenshots of the system running will be shown below. After each one I will write a briefly explanation of it.

4.2.1 Main page and Menu



Figure 3: Main page.

The *Figure 3* shows the main page of the system. In the left part of the bar is the Menu.



Figure 4: Menu of the system.

The menu (Figure 4) was divided 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.

4.2.2 Visual Modules



Figure 5: Studying Katakana.

The *Figure 5* shows the katakana module. The user can navigate in the Grid and study all the katakana characters. As you can see this module doesn't have menu bar, because this is a read only page. To return to the main menu, the user just clicks on the OK button and then return to the main page.



NO	HIRAGANA
1	あ
2	い
3	う
4	え
5	お
6	か
7	き
8	く

Figure 6: Studying Hiragana.

As in the previous *Figure 6* the user can study the hiragana characters. Also this page is for read only, and as the same as the katakana page to return to the main page just clicks on the OK button.

These two parts of the system are the only pages where the user can only read and study.

4.2.3 Operational Modules

- **Kanji Search**

In the next modules the user can search different information and navigate on the Grid to study or just to search for something.

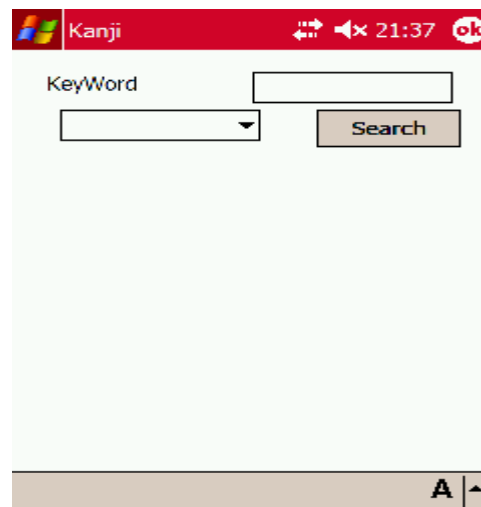


Figure 7: Kanji Search module.

Figure 7 shows the kanji search interface. At the beginning no kanji information is shown.

The user can select if want to search a specific kanji character or all the list of kanji. In this case the user just click the button “Search” and in a few seconds the information will appear.

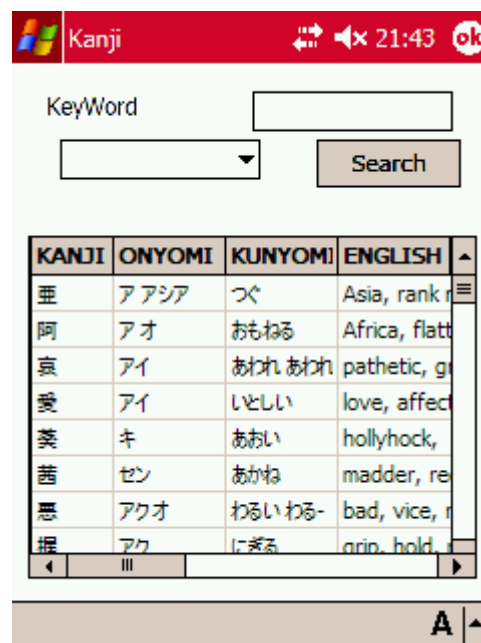


Figure 8: Kanji Search Result after all search.

Figure 8 shows the result of search all the information of the kanji characters. The user can navigate into the grid. On this module the user could not insert any information to the database.



Figure 9: Kanji Search menu.

The other form to search a certain kanji character is by selecting the way to search into the database (Figure 9). I mean, the user can select if the keyword is an onyomi part or is a specific kanji character, etc...



Figure 10: “鍵” Kanji Search.

After select the way to search the user has to insert a keyword to search. In this case I used the kanji character “鍵” (Figure 10). Then the user clicks the button “Search” to get the results.



Figure 11: "鍵" Kanji Search result.

After a few seconds the result will appear in the grid (Figure 11). As I said before, the user can navigate and read all the information on the grid, by using the scroll bar in both sides of the grid.

Now days the information on the kanji characters module is just the kanji character, the onyomi and kunyomi lectures and the description in English and Spanish.

In the future, the user would search by stroke a kanji character. Also the information showed in the grid will contain the 4 main radicals of each kanji character. So the ways to find a certain kanji character will be more.

In the next part of the system the user can insert and delete information to the database.

This kind of function is not available in other dictionaries. In most of the dictionaries we cannot modify that information.

Because each user use different information for your own purpose and also that is impossible to insert all the information for all the users, I think on this function, that everybody can manipulate your information.

- **Grammar**



Figure 12: Grammar Main Menu.

Figure 12 shows the menu to call the Grammar module. Inside of this module, the user can search, add and delete grammar information.

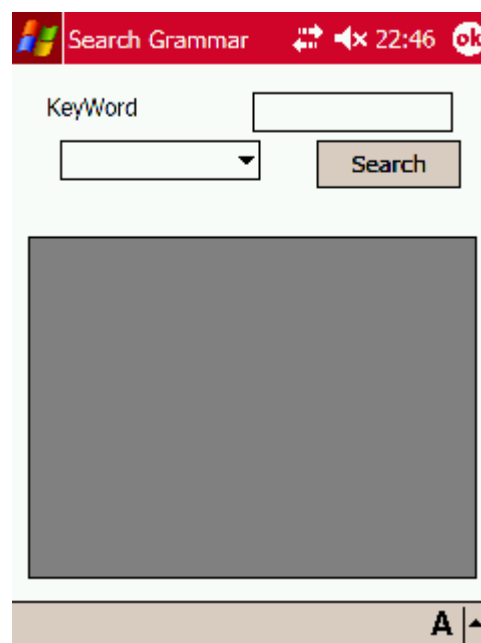


Figure 13: Search Grammar Module.

As in the kanji character search the grammar search always appears in blank (Figure 13). The user can choose if give some keywords to find something

specific or just search all the information on the database. On this module the user can select the level that want to search.

As I said before the Japan Foundation divide in four levels the Japanese language. Then the information is also divided in these four levels.



Figure 14: Search Grammar Result after search all.

On Figure 14 all the information of the grammar is shown. As you can see, the keyword and the level selection is not selected, and just click on the button "Search" and in a few seconds the information is shown on the grid.

The user could navigate on the grid, and read all the information with the help of the scroll bars.



Figure 15: Grammar Search Result after search by level 2.

Figure 15 shows how can the user select the level of the grammar that want to search. As you can see in the picture, any keyword was given. Also if you want a specific part on your grammar, giving the keyword you will have it.



Figure 16: Grammar Keyword Search Result.

On Figure 16 I selected the Level 2 “Business level” and give a key word “結果” and clicked the button “Search” to get only 2 results on the grid.

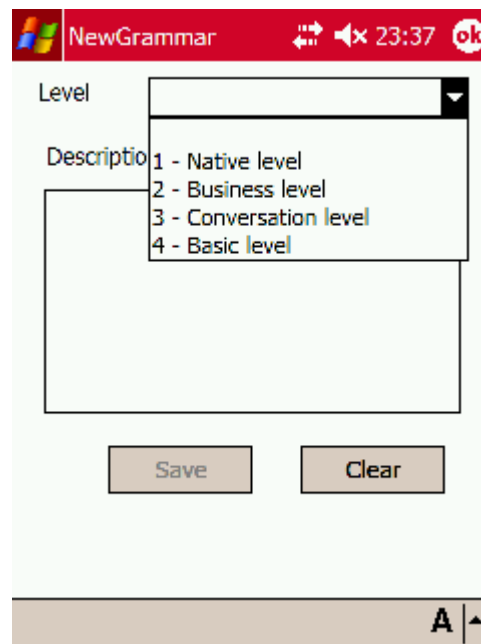


Figure 17: New Grammar Module menu.

On Figure 17 another part of the Grammar module is shown. On this the user can add grammar to the database, and continue studying all the important things that he or she consider.

This part is very simple, the user has to select which level want to insert and write the description on the blank space, then click the “Save” button and the information will be saved on the database.

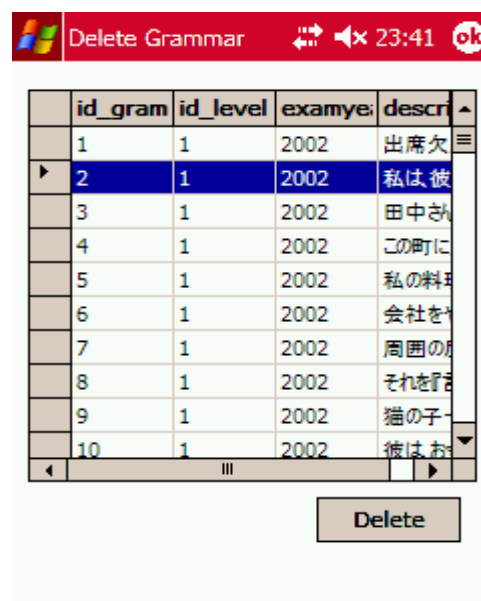


Figure 18: Delete Grammar Module.

Figure 18 shows the 3rd part of the grammar module, where the user can delete information from the database. On this module, the user has to select which record does want to delete and then click on the “Delete” button.

- **Dialects**



Figure 19: Dialect Main Menu.

Figure 19 shows the menu to search, add or delete information about the dialects. Now this module is useless because there’s no information on the database, of course, if the user want to add information on this module it will works.



Figure 20: New Dialect Word Module.

As the others modules, the dialect module also can insert new information to the database (*Figure 20*). The user has to select which kind of dialect does want to insert and then fill the boxes, with the appropriate information, then click on the “Save” button to insert in the database.

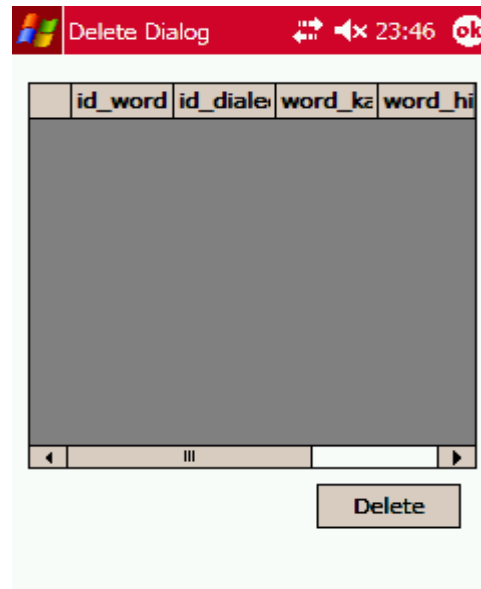


Figure 21: Delete Dialect Word Module.

Also the dialect module has its own delete part (*Figure 21*), where the user can delete information from the database. As I said before this module is not working yet.

Now I am going to explain the main part of the system, which is the dictionary. In this module, as the others, is divided in 3 parts, Search, New and Delete.

The main part of this module is the Search, where the user can search for all the words in the database.

In the New part the user can insert data to the database, this part is divided in two parts, the insertion of a new word and the insertion of a new kanji.

And the third part, the deletion, is where the user can delete words from the database.

- **Dictionary**



Figure 22: Dictionary Main Menu.

Figure 22 shows the menu of the dictionary module, as shown, the New part have a sub-menu where the Kanji part and Word part can be selected.

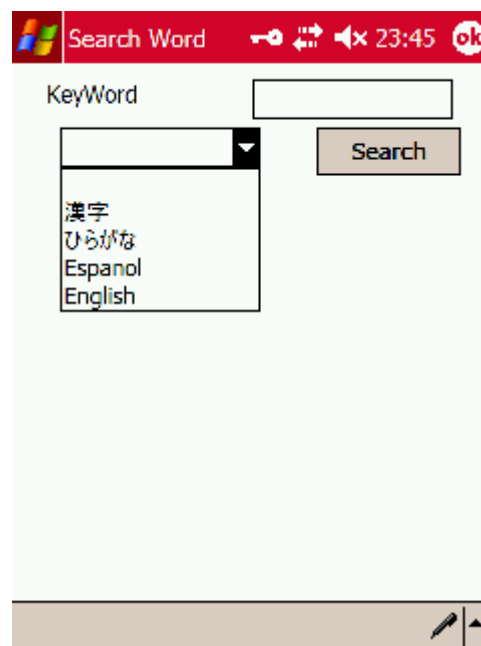


Figure 23: Dictionary Module Menu.

Figure 23 shows the first page of the search, here the user will select the field where wants to search a word, and then input the word that want to search. Some restrictions where taken for each one.

In the next images I will explain each one of them.

1. If the user selects “漢字” the system will search only for the words with exactly the same kanji as the keyword.



Figure 24: Dictionary result searching by “漢字”

For example, in the Figure 24 I searched the word “今年” and the system gave me all the words with the same keyword.

2. If the user selects “ひらがな” as option to select, the system will search for the words with the same beginning as the keyword.



Figure 25: Dictionary result searching by “ひらがな”.

In *Figure 25* I search for the word “こんねん” and the system gave me all the words with that beginning, as for this example the resulting words was “今年（こんねん）”, “今年度（こんねんど）”. As you can see in the first part, where the user select to search by “漢字” kanji characters, the system only search for exactly the same kanji character and the for example the character “今年度” will not appear. The reason is that the speed of the system will decrease a lot and the user will wait a lot of time to get one word. This problem will be solved in the future, actually a new database manager for Pocket PC was developed and it has more power on the searches.

3. If the user wants to search by the “English” or by the “Spanish” option the system will search for all the English sentences (for the English option) or Spanish sentences (for the Spanish option) that contain that keyword.



Figure 26: Dictionary result searching by “Spanish”.

In *Figure 26* I search for the word in “sistema” in Spanish, and the system gave me all the words that contain the keyword in the description, as the same as the English part that makes the search in all the sentences of the description.

As you can see in the image the system gave me 3 words, one with the meaning “system operator” that means “operador del sistema” in Spanish, the next with the meaning “series, system” and the third that means “order, system” all of them contains the word “system” in Spanish “sistema” in their descriptions.

4. If the user doesn’t select any of the options in the combo box the system will throw a message box asking to select an option.

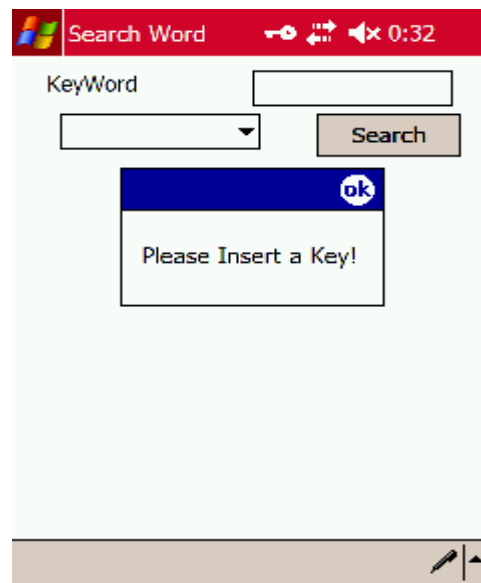


Figure 27: Dictionary Warning Message.

Figure 27 shows the warning message that the system throws when no key was entered.

5. When only one record is selected as a result, the system shows it in another form.



Figure 28: Dictionary Result after getting only one result.

In Figure 28 I search for the word “ことし” and the system gave me only one result. To get a better view of all the meanings and the readings of the word I choose to show the word in this way, individually.

As I said before the user can also insert your own words or kanji characters. In the next part I will explain the modules where the user can do it.



Figure 29: New Word Main Menu.

Figure 29 shows the menu where the user can select which module does want to enter.



Figure 30: New Word Module.

Figure 30 shows the module where the user can input new words to the database and get on it more information.

To do this, the user has to input the “漢字”, kanji character of the word, the “ひらがな”, hiragana character of the word, and the meanings of the word. Then click in the button “Save”, that will be enabled as soon as the user enter the fields.

The image shows a software window titled "New Kanji" with a red header bar. The header bar contains a Windows logo, the title "New Kanji", and several icons including a key, a double-headed arrow, a speaker, a timer showing "0:45", and an "ok" button. Below the header, there are five text input fields, each preceded by a label: "漢字:" (Kanji), "Onyomi:", "Kunyomi:", "Espanol:", and "English:". Below these fields are two buttons: "Save" and "Clear". At the bottom of the window, there is a status bar with a pencil icon and a vertical line.

Figure 31: New Kanji Module.

Figure 31 shows the module where the user can input to the database a new kanji character. To do this just has to fill the fields “漢字”, kanji character, the onyomi, and the kunyomi readings, and finally the English and Spanish descriptions for that kanji character, then by clicking in the button “Save”, that will be enabled as soon as the user fill the fields, the kanji character will be added to the database.

Finally I am going to explain the Delete module, where the user can delete information (in this case words) from the database.

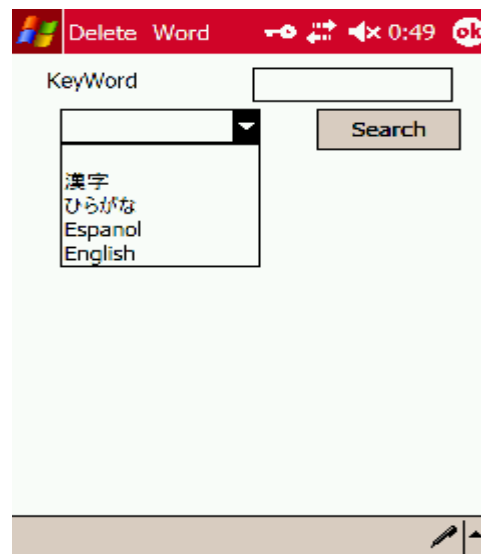


Figure 32: Delete Module Menu.

Because there are many words in this system and the memory of a pocket pc is not enough to show all the information of the table as in the deletion of the grammar, in this module the user has to search for the word that want to delete (Figure 32). The search is the same as for searching for meanings; the user has to select the field where want to search the word to delete and then the information will appear.



Figure 33: Delete Module Search.

Figure 33 shows how the information will appear as soon as the user search for the word to delete.

For example, I choose again the word “今年” and ask to search for the “漢字” kanji characters field. The system gave two words, and then I select which of them I want to delete, and click in the button “Delete” to delete the information. As the user could select a different word a confirmation message will be shown (Figure 34).



Figure 34: Delete Module Warning Message.

4.4 FUTURE WORK

1. As I mentioned before one of the modules information is not compiled yet. The dialect words I think is a very important part in the system, since many people in Japan nowadays use it. I want to develop a huge vocabulary of dialect words and compile all the prefectures dialects of Japan.
2. Another point that I think is important for this system is a replication server, where a main database will be developed. The people that use the system could update their own systems with new information that other people updated to the database. All the people in the word would access to the server by Internet and update the systems by a certain process that will be developed; actually 50% of the process has developed.

As the many people in the word would access to the database and update new information a check batch process has to be developed, this process will check the contents of the new information and update the database with only the correct information, with this I can be sure that all the information is correct.

3. Another part that I think important for this project is the speed of the system, nowadays the response of the system is good but in the kanji

characters search the response is too slow, so I want to migrate this system to SQL Server 5 that has developed a new manager for Pocket PC.

Chapter 5

Conclusion

This system was developed as a necessity of new things, as a necessity of better way to understand a new language. When I arrived for the first time to Japan and didn't know how to speak Japanese I try to get a good dictionary and try to speak with Japanese people, must of the laugh at me because I used to use the dictionary and speak with only words on it, of course my Japanese friends told me that the words that my dictionary contains was for written times only and they don't use for the casual way. Then I try to find some other dictionaries but all of them contain almost the same words. From this time I think in make a dictionary for the people like me that want to have new Japanese friends and sometimes is impossible to communicate because the words of the dictionary are so polite.

This necessity makes me the way to make this dictionary. Of course I think is not already finished but in a few time I think it will be helpful for many people inside Japan or outside.

Reference

- [1] <http://en.wikipedia.org>
- [2] <http://www.microsoft.com>
- [3] <http://msdn.microsoft.com>
- [4] <http://www.penreader.com>
- [5] <http://www.lingvosfot.com>
- [6] <http://www.absoluteword.com>
- [7] <http://www.ectaco.co.uk>
- [8] Fujii, Atsushi and Ishikawa, Tetsuya. *Utilizing the Word Wide Web as an Encyclopedia: Extracting Term Descriptions from Semi-Structured Texts*. University of Library and Information Science. Tsukuba, Japan.
- [9] Fujii, Atsushi and Ishikawa, Tetsuya. *Summarizing Encyclopedic Term Descriptions on the Web*. Graduate School of Library, Information and Media Studies. Tsukuba, Japan
- [10] Zhang, Yujie, Ma, Quing, Ishihara, Hitoshi, *Use of Kanji Information in Constructing a Japanese – Chinese Bilingual Lexicon*. Communications Research Laboratory. Kyoto, Japan.
- [11] Tanaka, Kumiko and Umemura, Kyoji. *Construction of a Bilingual Dictionary Intermediated by a Third Language*. Division of Engineering, University of Tokyo. Tokyo, Japan.
- [12] Bond, Francis, Yamazaki, Takefumi, Binti Sulong, Ruhaida and Ogura Kentaro. *Design and Construction of a machine-tractable Japanese – Malay Lexicon*. NTT Communication Science Laboratories. Tokyo, Japan.
- [13] <http://csse.monash.edu.au/~jab/kanjidic>

- [14] Shirai, Satoshi, Yamamoto, Kazuhide. *Linking English Words in Two Bilingual Dictionaries to Generate Another Language Pair Dictionary*. ATR Spoken Language Translation Research Laboratories. Kyoto, Japan.
- [15] Breen, James. *Multiple Indexing in an Electronic Kanji Dictionary*. Monash University. Clayton, Australia.
- [16] Breen, James. *JMdict: A Japanese – Multilingual Dictionary*. Monash University. Clayton Australia.
- [17] <http://www.jees.or.jp/jlpt/en/>
- [18] http://momo.jpf.go.jp/jlpt/e/about_e.html
- [19] GRATTAN, NICK. 2002. *Pocket PC, Handheld PC with Microsoft embedded Visual Basic*. Prentice Hall series on Microsoft Technologies.