

JITEN: Japanese Input for Text Exploration and Navigation

Ray Vichot
Georgia Institute of Technology
686 Cherry Street
Atlanta, GA 30332-0165
+1 646-270-7670
ray@stop-making-sense.net

ABSTRACT

JITEN is a program that allows for Kanji lookup and learning using the context of newspaper articles from the Asahi Shinbun, a major Japanese newspaper. The software is portable, easy to use on most of the Google platforms.

Keywords

Japanese, translation, literacy, dictionary, Kanji, RSS, language learning, pedagogy

1. INTRODUCTION

JITEN is a Google Gadget that allows a user to search Kanji (Chinese derived characters used in Japanese) as well as be able to read RSS news feeds from the Asahi Shinbun (the largest newspaper in Japan) in order to obtain a context with which to learn about Kanji. Thus the ultimate goal of the software is to achieve an increased literacy and fluency in the Japanese language.

The conceptual framework of the project began with a personal need. As an intermediate learner of Japanese, I am attempting to most difficult part of the language, which is gaining literacy. I was looking for a simple way to integrating reading with dictionary help, but could not find software of any sort that could fulfill this function adequately.

Another strategy I wished to employ was to create a solution that was simple in use and not dependent on a specific webpage. The solution was meant to be self working, integrating into a webpage like iGoogle, or extensible to a platform that is browser independent such as Google Desktop. Ideally it should be feasible for the software to be ported into a mobile device for on the fly lookup and translation in context outside of online sources.



Figure 1: JavaDict, a standalone Java application

2. CURRENT SOLUTIONS

There are obviously multiple solutions for the main problem of Kanji lookup. However, the current solutions have distinct drawbacks. The first and most obvious solution is of course paper-based dictionaries. While they have a utility in use and, some element of portability, dictionaries, especially Kanji dictionaries, are large and bulky. And the action of lookup requires the use of one of various complicated classification schemes such as Bushu or Radical Code, Nelson Index, or SKIP code.

The other solutions bypass the bulkiness of page and paste dictionaries. However, they suffer from a lack of portability. JavaDict, an application allows for Kanji lookup as well as dictionary lookup. It is relatively light in resource usage. However, being a Java application, its use is limited to the desktop/laptop platform. In addition JavaDict does not distribute installers, rather the writer require users to compile their own application, which makes the user group for this application (speakers of Japanese and people who can compile Java quickly and easily) even smaller.

There are several software options made for the Nintendo DS which address the portability and ease of use issues. However, because of the consoles limited connectivity as well as the fact that these software options are made for and licensed primarily for speakers in Japan, these options are not practical for those learning abroad.



Figure 2: Zaidanhoujin Nippon Kanji Nouryoku Kentei Kyoukai Kounin: KanKen DS, one of a series of DS software aimed at Kanji literacy among native speakers

Other widgets and gadgets that relate to Japanese language learning are focused mostly on flashcard type games, vocabulary lists, or simple word for word translation, which is of limited use for the speaker attempting to learn written Japanese.

JITEN: Japanese Input for Text Exploration and Navigation

My Data Model

JITEN deals with two specific data forms. First it deals with character data that is run through the search engine of the WWWJDIC, a website which has a detailed dictionary and Kanji Lookup Function. The Widget utilizes this site in doing the search. Secondly it is utilizing an RDF feed to display current stories from the Asahi Shinbun.

The Character input data, provided by the user is allowed to be inputted either through direct input (using internationalized keyboard standards), any of the various analog and digital classification schemes, or even by Romanized input.

The RSS data is processed by using a JavaScript function that organizes the data into a chart that displays the 5 most recent headlines.

Organization / Architecture

JITEN takes advantage of the sort of hypermediality afforded to Digital Media as described by Jay Bolter. The data is organized into two windows, each of which houses one of the two main functions. The left window displays the Kanji search window. The left window displays the Asahi RSS feed.

Originally, the layout was going to rely on the “tabbing” model of design, as used in web browsers such as Firefox and Opera and in certain Google Gadgets. However, there were technical and interface related issues. This was resolved by having the screens laid side by side.

feed does not have to stop looking at what she is doing to search and then look back. This also reduces extraneous clicking.

The buttons on the bottom of the screen allow a user to reload the starting pages of each window. This is not ideal, adding an extraneous layer of complexity to the multi-windowed widget. However, due to technical limitations, this was the best method for ensuring a consistent, repeatable user experience. This will hopefully be modified in future implementations.

The user interaction model is essentially a user moving from left to right to left alternating the lookup functionality with the news feed, allowing the user to quickly move from the text to lexical information and back. The end goal is to reduce the time between when a user does not recognize a character, performs a lookup, and returns to the text, having learned the appropriate meaning and reading of the character. Using the mouse and keyboard either with or without Japanese input capability (which most learners of Japanese at the level to benefit from JITEN would have), one can quickly copy and paste text, or input text.

A typical case narrative for JITEN would be along the lines of a user opening the widget when he or she is spending time practicing reading. He or she would open the widget and see the latest headlines on the right screen. They can then navigate through the stories and read them. If they come across a Kanji character in say a word that uses both of them, they can input it on the search screen on the left, look up the pertinent information (meaning, pronunciation, etc.) and simply look to the right screen and continue reading the headline. Alternately, they may see a Kanji in an article they know in one context (say as the main part of a verb such as (iku), to go), but do not know how it is pronounced when it is part of a multi-kanji compound (ginkou), bank). They can input the Kanji based on one reading, and can lookup alternate readings.

Another example is that of a translator who is working on a document and have come across an unfamiliar name, say in an article detailing a member of the Diet. In Japanese the pronouncing of names based on Kanji can be rather difficult, even for native speakers. The translator can then input the Kanji and see the possible ways in which it is used as part of a name.



Figure 2: The current implementation of JITEN

Interface / Interaction

JITEN's interface is unique because of the side by side display. The advantage of this interface and layout is that the dictionary tool is immediately available so the person who is reading the

3. FUTURE IMPLEMENTATIONS

Handwritten Kanji Input

The major future implementation I would like to add would be the ability for a user to use a pseudo-handwriting input for Kanji lookup. This could either be done via mouse if on a PC, a touch screen or touch pad, or with a stylus. The latter two inputs are “most natural” as demonstrating how a character is written user finger motions in the air is very common and a stylus input is similar to a pen or pencil.

This implementation I fell would add greatly to the application as this would allow users to not only gain knowledge in reading Kanji, but in writing as well. Since the rules for writing Kanji are standardized for the most part, inputting even an unfamiliar kanji is typically very easy.

Dictionary Lookup

Beyond simple character lookup, I would also like to make JITEN extensible to word lookup as well. The applications for this are numerous, but specifically there are times when Kanji are combined to form a word whose meaning is not obvious from knowing the meaning of each individual component in the

compound. Also a word lookup feature would be useful in the defining of verbs, nominal verbs, idiomatic expressions, and onomatopoeia, the latter two of which are used very often.

Site Independent Search

Currently the site lookup relies on the WWWJDIC, a site that relies on cgi forms. While the 10 year old site allows remote searches and is not in any danger of disappearing anytime soon, a goal would be to move the information to a separate server and have all the search functionality be under my control. This would allow for a slightly cleaner interface and possibly a sleeker visual design.

Multilingual Search

Since the KANJIDIC and EDICT that power the WWWJDIC have multilingual indexes (At least Spanish, Portuguese, German, and French), it would be good to implement a multilingual index in the search portion of JITEN as well. This would increase my user base to native speakers of those languages to use my tool in their pursuit of Japanese fluency.

Something that Might Add Value

I've considered possibly adding functionality like full blown translation in the vein of Google Translate or Babelfish, but ultimately decided that JITEN should be a language learning tool rather than a simple translator. I feel that in order to learn any foreign language, especially one so different to a native English speaker, one must really understand to some extent the linguistic components of the language. JITEN currently offers much of that structural knowledge, at least for a major hurdle of literacy for learners of Japanese. Turning JITEN into a translator would for one, be ultimately meaningless as machine translation is always imperfect and cannot convey connotation or simply do things or ignore word types such as onomatopoeia, which may need to be translated into a different word type in English, such as a verb.

4.REFERENCES

- [1] Bolter, Jay and Grusin, Richard. Remediation: Understanding New Media. MIT Press.

- [2] Howell, Lawrence J. and Morimoto, Hikaru. Etymological Dictionary of Chinese Characters. Kanji Networks.
<http://www.kanjinetworks.com/reference.html>