WWWJDIC - A Feature-rich WWW-based Japanese Dictionary

James Breen
Monash University, Australia

Abstract

The WWWJDIC WWW-based Japanese dictionary service is described. The service integrates many different dictionary functions into a single package, and contains many features aimed to assist language learners and readers of Japanese texts.

1. Introduction

The WWWJDIC WWW-based Japanese dictionary (Breen 2003) is an evolving multi-feature dictionary service based on free and public dictionary files. It is widely used in Japanese-language education, and has a number of functions specifically to aid language learners. The main server is at Monash University in Australia, and there are five mirror sites in Europe (2), North America (2) and Japan. Usage is currently at several hundred thousand accesses per day.

2. Design Goals

The dictionary service has been developed in an attempt to give expression to concepts of "tomorrow's dictionary"(Atkins 2002) in providing a wide range of configurable features and options which go well beyond the common commercial dictionary services based on accesses to copies of published bilingual dictionaries.(Kenkyusha 2009, NTT 2009, Yahoo 2009)

The dictionary files used by the server are:

a. the JMdict/EDICT Japanese-English dictionary (Breen 2004a), which has about 140,000 entries;
b. the ENAMDICT dictionary of named entities, which has over 700,000 entries
c. the KANJIDIC kanji (Chinese character) dictionary (Breen 2004b), which has detailed information on over 12,000 characters
d. a collection of glossary files in fields such as life sciences, law, engineering, Buddhism, business, etc.

Entries in the dictionaries can be accessed either by the Japanese headwords (either the kanji form or the reading/pronunciation) or by words in the glosses.

(See figure 1) The kanji dictionary can be accessed via a variety of methods including the traditional radical/stroke-count and four-corner techniques, the character pronunciations, the character meanings, various dictionary indices, etc. A multi-component index based on the visual elements in the characters is particularly effective and popular. An external handwriting interface can also be used. The dictionaries are integrated so that a user, having found a particular character, can display word entries containing that character, or having selected a word, can examine the details of the constituent characters.

```
Search Key: せいじ Current Dictionary: Jpn-Eng General (EDICT)
Options:[G]oogle search, [GI] Google images, [S]anseido dictionary, [A]LC dictionary (Eijiro),
[Ex]ample sentences, [V]erb conjugations, [F] Feedback, Japanese[W]ikipedia.

⊙ 政治 【せいじ】 (n) politics; government; (P) [Ex][G][GI][S][A][W]
     政治について討論しよう。 I suggest we discuss politics.[Amend]
○ セージ; セイジ (n,adj-no) sage [G][GI][S][A][W] [G][GI][S][A][W]
○ 青磁 【せいじ; あおじ】 (n) celadon porcelain [G][GI][S][A][W]
○ 正字 【せいじ】 (n) correct characters [G][GI][S][A][W]
○ 政事 【せいじ】 (n) political affairs [G][GI][S][A]
○ 盛時 【せいじ】 (n) prime of life; era of national prosperity [G][GI][S][A]
○ 盛事 【せいじ】 (n) prosperous undertaking [G][GI][S][A]
```

Figure 1: Example of dictionary word display

3. Text Glossing

One function of the service commonly used by translators is a text-glossing capability in which Japanese text is segmented and matched with dictionary entries. The segmentation and matching uses a combination of most of the dictionary files, and allows inflected forms of verbs and adjectives to be aligned with the dictionary forms. (Figure 2)

```
心停止の患者を応急処置で生かした。

  ◆ 心停止 【しんていし】 (n) cardiac arrest; ED
  ◆ 患者 【かんじゃ】 (n) (a) patient; (P); EP
  ◆ 応急処置 【おうきゅうしょち】 (n,adj-no) (1) emergency measure; emergency
     procedures; first-aid treatment; (2) temporary repairs; stop-gap treatment; ED
  ◆ Possible inflected verb or adjective: (plain, past)
     生かす 【いかす】 (v5s,vt) (1) to make (the best) use of; to leverage (skills,
     attributes, experience, etc.); to capitalise on (experience, etc); (2) to let
     live; to keep alive; (3) to revive; to resuscitate; (P); EP
```

Figure 2: Example of text-glossing function

4. Features for Language Learners

Aspects of WWWJDIC's service which are of particular interest in CALL are:

a. the option of displaying a table of conjugations for any of the verbs or verbal nouns in the dictionary (approx. 17,000 entries). (Figure 3)
b. animated stroke-order-diagrams for the 2,000 most common kanji. (Figure 4)
c. links at the entry level to the Tanaka Corpus of 150,000 Japanese-English sentence pairs. (Figure 5) The Corpus can also be searched independently.
d. sound clips of the Japanese pronunciation of almost all EDICT entries.

Conjugating 食べる: 一段 (type II) verb. (Supplementary Comments)

	Affirmative		Negative	
	Plain/Informal	Polite/Formal	Plain/Informal	Polite/Formal
Non-past	食べる	食べます	食べない 食べぬ(x) 食べず(に)(x)	食べません
Past	食べた	食べました	食べなかった	食べませんでした
Te-form	食べて	食べまして	食べなくて 食べないで	食べませんで
Conditional	食べたら	食べましたら	食べなかったら	食べませんでしたら
Provisional	食べれば	食べますなら(ば)	食べなければ	食べませんなら(ば)
Potential(#) & Passive	食べられる	食べられます	食べられない	食べられません
Causative	食べさせる 食べさす	食べさせます 食べさします	食べさせない 食べささない	食べさせません 食べさしません
Caus-Pass	食べさせられる	食べさせられます	食べさせられない	食べさせられません
Volitional/ Hortative	食べよう 食べる[よう/こと] にしよう	食べましょう 食べる[よう/こと] にしましょう	食べまい(+) 食べない[よう/こと]にしよう	食べますまい 食べない[よう/こと]にしましょう
Conjectural	食べるだろう	食べるでしょう	食べないだろう	食べないでしょう
Alternative	食べたり	食べましたり	食べなかったり	食べませんでしたり
Imperative	食べろ	食べなさい	食べるな	食べなさるな

Figure 3: Example of verb conjugation table

Figure 4: Example of animated stroke order display

Figure 5: Example sentences linked to the バス停 entry

Other features of the service are:

a. a configurable interface enabling users to structure the display and enable
    or disable options to suit their needs;

b. multi-lingual operation. At present the main operating pages are available in either English and Japanese. Other languages can be added by extending the catalogue files, and a French interface is in preparation.
c. a restricted interface tailored for use with Japanese mobile telephones.
d. links from each entry to a range of online dictionaries, search engines, Japanese Wikipedia entries, the Japanese WordNet, etc..
e. an edit interface enabling users to provide suggestions, amendments, etc. about dictionary entries or to propose new entries.
f. an API enabling access from software and servers.

Although most of the dictionary files used are Japanese-English, it also includes the major WaDokuJT Japanese-German dictionary and smaller Japanese-French, Japanese-Spanish, Japanese-Swedish, Japanese-Hungarian, Japanese-Slovenian and Japanese-Dutch files.

5. Conclusion

The WWWJDIC WWW-based Japanese dictionary server has developed to become an effective multifunction service which integrates many functions which are typically spread over several traditional printed dictionaries. It also has extensive integrated support for many language features.

References

ATKINS B.T.S. (2002). Bilingual Dictionaries - Past, Present and Future. Lexicography and Natural Language Processing - A Festschrift in Honour of B.T.S. Atkins, Eurolex 2002.
BREEN J. (2003). A WWW Japanese Dictionary. In "Language Teaching at the Crossroads", JSC Working paper No. 13, Monash Asia Institute, Monash University Press, 2003 http://www.wwwjdic.net/
BREEN J. (2004a). JMdict: a Japanese-Multilingual Dictionary. COLING Multilingual Linguistic Resources Workshop, Geneva, 2004.
BREEN J (2004b). Multiple Indexing in an Electronic Kanji Dictionary. COLING Enhancing and Using Electronic Dictionaries Workshop, Geneva, 2004.
KENKYUSHA LTD. (2009). Kenkyusha Online Dictionary http://kod.kenkyusha.co.jp/service/
NTT Resonant Inc. (2009). Goo Jisho http://dictionary.goo.ne.jp/
YAHOO Japan Corporation. (2009). Yahoo Jisho http://dic.yahoo.co.jp/