Addition of Search Function by Suffix to Hyper English Dictionary with Indo-European roots

Toshiyuki Oowada   s1080053

Supervised by Prof. Nobuyoshi Asai

Abstract

The Hyper English Dictionary is one type of Internet dictionary that is comprised of Indo-European roots. It enables learners of English to understand the roots and definitions of English words at the same time. It is very important to learn the roots of English word because all English words are branched to Indo-European roots. Some English words are branched to the same Indo-European roots. Learners of English can learn some English words at the same time by learning one of the Indo-European roots. In this research, a new search function by the suffixes of English words is added to the Hyper English Dictionary. This new search function by suffix of English words makes understanding and searching of English words and stems easier for learners. Understanding many stems assists learners in learning many English words because many words are comprised of some aggregations of stems.

1 Introduction

1.1 Background

1.1.1 What is good dictionary

A good language dictionary can help learners of English to greatly improve their English. English education is increasingly getting harder. However, many student's English level has still not improved. Many Japanese students use a dictionary to learn English words and sentence, but almost all students struggle to communicate well in English. One reason is that learners of English do not understand the accurate meanings of many English words. Most dictionaries contain many definitions of English words and example sentences, but they are not easy to use, in regard to organization. Many definitions and example sentences only confuse learners because learners are not good at searching. A good dictionary should not confuse learners because they are as important as a good English textbook or teacher.

1.1.2 Internet dictionary

Now, many people can access the Internet and almost all Internet dictionaries can be used at a small cost. In terms of number of words, Internet dictionaries cost less than normal dictionaries, and they include additional data and can be easily updated. An Internet dictionary has great potential to improve the English of many learners.

1.2 About Hyper English Dictionary

1.2.1 What is Hyper English Dictionary

The Hyper English Dictionary is an Internet dictionary with many functions. It is made up of Indo-European roots, so learners of English can understand roots and definitions at the same time. It is very important to learn the roots of English words because almost all English words are branched to Indo-European roots, and some English words are branched to the same Indo-European root. [?] [?]

1.2.2 What are Indo-European roots

Indo-European roots are an aggregation of some stem patterns of Indo-European and Proto-Indo-European languages. Indo-European and Proto-Indo-European are extinct languages and were created by linguists in the early part of the 19th century. Many languages trace their origins to Indo-European roots, for example, Indo-Iranian, Greek, Armenian, Balto-Slavic, Albanian, Celtic, Italic, and Germanic(see Figure1), and are referenced in the American Heritage Dictionary of The English Language. [?] [?]

1.2.3 What is American Heritage Dictionary of The English Language

The American Heritage Dictionary of The English Language is a long-standing favorite family dictionary and one of the prime dictionary references for magazines, newspapers, and dot.com content providers. For scads of writers and editors across the U.S., it sets the standard on matters of style and lexicographical authority. Its panel of special consultants includes authorities on anthropology, architecture, cinema, and law, plus military science, music, religion, and sports, and that is reflected in a comprehensive coverage of the arts, culture, and technology. [?]

1.3 Functions of Hyper English Dictionary

1.3.1 Functions of former Hyper English Dictionary

Until this year, the Hyper English Dictionary had three major alterations. The Hyper English Dictionary has some search functions using prefixes, Indo-European roots, synonyms, and word stems, which have made it easier to search and understand English words. The op-
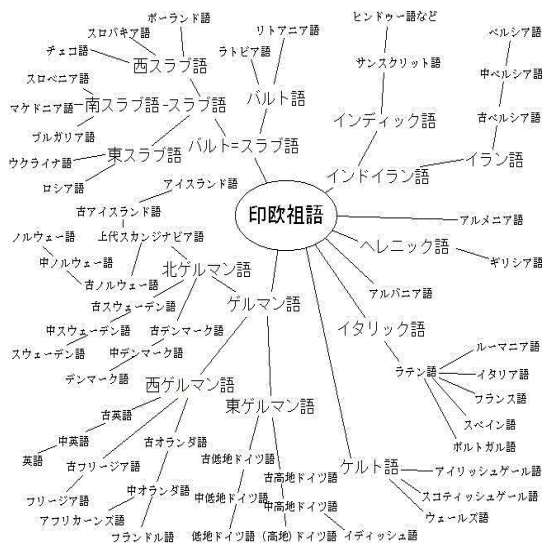
Figure 1: derivation

erating system was changed from Windows to Linux, the most popular operating system because it is free. The Hyper English Dictionary also has had an application made of JSP called a *Hyper-Tool* added. The *Hyper-Tool* is a quiz application with Indo-European roots. The *Hyper-Tool* gives provides quizzes of English word definitions and hints of origins. This application is useful for learners of English.

1.3.2   Additional functions

This year, the Hyper English Dictionary added four major features; (1) links were added from English words to example sentences; (2) sound data of English words; (3) groups of English words; and (4) search functions by suffixes. These alterations make it easier for learners to search and understand English words. In this thesis, I focus on adding search functions by suffixes.

Number of English words	20000
Number of stems	97
Number of example sentences	10669
Number of preffixes	72
Number of suffixes	345
Number of roots	1350

Table 1: Functions of Hyper English Dictionary

1.4   Purpose of research

1.4.1   Additional search function by suffix

A suffix is an affix added to the end of a word or stem, serving to form a new word or functioning as an inflectional ending, such as *-ness* in *gentleness*, *-ing* in *walking*, or *-s* in *sits*. It is very important for learners to understand suffixes when they learn English words because a suffix is one of the stems. The understanding of many stems can help learners to learn many English words because many words are comprised of some aggregations of stems. The search function by suffixes enables the search of English words from a suffix, and it lists English words with the same suffix. Learners of English can learn some English words simultaneously because some English words have the same suffix.

1.4.2   Additional English words

New English words were added to the Hyper English Dictionary because it contained fewer words than other Internet dictionaries. In 2003, about 7,000 new English words along with example sentences and about 700 new Indo-European roots were added to the Hyper English Dictionary. Project members added new English words, and I was put in charge of words beginning with the letters *P, I, J, K, M, N*. These English words were checked using many dictionaries and definitions and parts of speech of these words were added along with example sentences.

2   Discussion

2.1   System of Hyper English Dictionary

The Hyper English Dictionary was created using many kinds of software, for example, Tomcat, Apache MySQL and mm.mysql. All of these software are free and do not require a specific platform. [**?**] [**?**] [**?**] [**?**] [**?**]

OS	Red Hat 9.0
Java Environment	J2SDK 2.0.40
Servlet Container	Tomcat 4.1.27
RDB Server	MySQL 4.0.16
JDBC Driver	mm.mysql
Web Server	Apache 2.0.40

Table 2: The Hyper English Dictionary System

2.1.1   Explanation of Tomcat

Tomcat is a Servlet container that is needed in order to use the Java Servlet and JavaServer Pages technologies. Servlet is a program of Java that runs on a Web server. If there is a demand from the Web browser of a client,

the Java Servlet will generate the resource of HTML or others dynamically and will return a result to the Web browser. The Java Servlet and JavaServer Pages specifications are developed by Sun under the Java Community Process.

2.1.2 Explanation of Apache

Apache is a software that works on a Web server and offers a Web service and works in UNIX or Windows operating system. It is a free software and was created by many volunteers. Apache is the most popular Web server software in the world. The Hyper English Dictionary therefore runs by a combination of Tomcat and Apache.

2.1.3 Explanation of MySQL

The Hyper English Dictionary database was created by using MySQL, which enables easy addition and deletion of data. MySQL is *Relational DataBase Management System* that it is open source. *Relational DataBase Management System* is called RDBMS. MySQL does not choose a platform, for example, a UNIX and Windows operating system.

2.1.4 Explanation of mm.mysql

The mm.mysql is *Java DataBase Connectivity* Driver. *Java DataBase Connectivity* called JDBC. JDBC is *Application Program Interface* for access from JAVA program to RDMS. The mm.mysql can operate the MySQL database using *Structured Query Language*.

2.1.5 Explanation of JAVA

A trademark used for a programming language designed to develop applications, especially ones for the internet, that operate on different platforms, JAVA was originally develop by *Sun Microsystems*. *Sun Microsystems* calls JAVA a *Write Once, Run Anywhere Software*.

2.1.6 Attention of systems

The Linux OS initially has some software installed. But if you reinstall the software that was already installed, it will not run well. This phenomenon is called *conflict*. *Conflict* is the most frequently occurring error condition in Linux. If that software still does not run, one must check the setting of the environmental variable, which is very important to run different software simultaneously. The *.bashprofile* on the home directory must be rewritten for setting of the environmental variable.

2.2 Development of search function by suffix

2.2.1 Methods analysis

The Search function by suffix was created using MySQL and webmin and many kinds of dictionaries, for example, New English-Japanese Dictionary, Ge-

nius English-Japanese Dictionary and Random House English-Japanese Dictionary. Webmin is a web-based system administration tool for Unix servers and services. In order to use webmin, a web browser must start and access the port 10000 of PC by SSL. Then, the browser shows a display of the inputted user name and password. By inputting a user name and password, webmin can be used.

- http://localhost:10000 (In case of access from localhost)

- http://*Host name*:10000 (In case of access by input host name)

- http://*IP address*:10000 (In case of access by input IP address)

The suffix database was created by MySQL and is very easy to add and delete data using webmin. Webmin expresses a MySQL database graphically. Data can be added or deleted by a mouse click using webmin. Webmin is used to create a suffix database easy. Webmin is able to be used not only on Linux but also on many operating systems, for example, *FreeBSD*, *Mac OS X*, *Solaris* and *HP-UX* because it was made by a *CGI* program using *Perl 5*. The suffix database of definitions was cre-
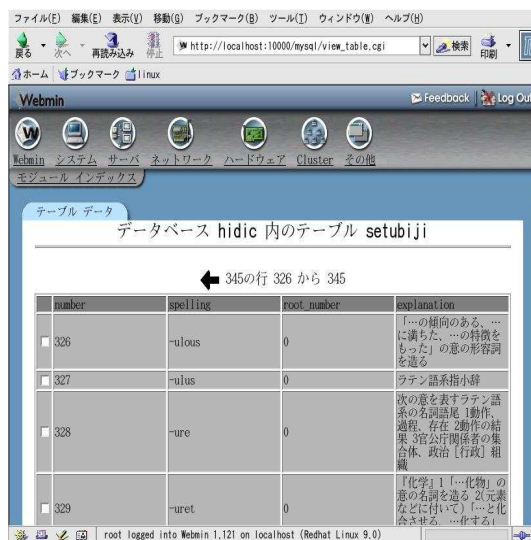

Figure 2: Display of Webmin

ated using many kinds of dictionaries, for example, New English-Japanese Dictionary, Genius English-Japanese Dictionary and Random House English-Japanese Dictionary, all of which are famous in Japan. The New English-Japanese Dictionary, published by Kenkyusha in

1998, contains many useful explanations of detailed etymology. This is very helpful in tracing the origin of a word. The Genius English-Japanese Dictionary, published by Taishukan in 2001, is a very intelligible dictionary. The Random House English-Japanese Dictionary, published by Shougakukan in 2002, contains the largest number of English words. I searched the meaning of suffixes using dictionaries, and added suffix data to the suffix database using webmin. [?] [**?**] [**?**] [**?**]
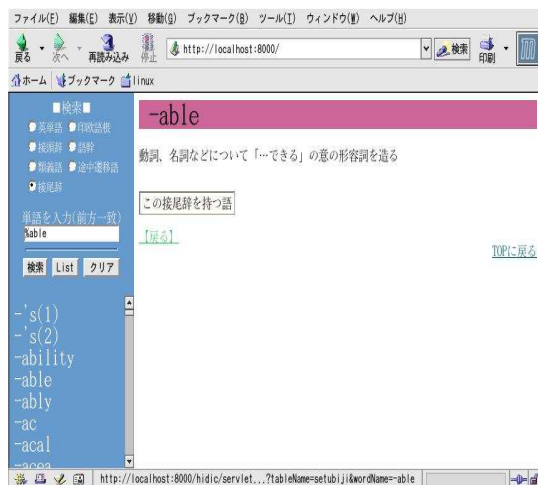

Figure 3: Search function by suffix

2.2.2 Evaluation

The search function by suffix was evaluated by ten participants when it was completed. Most participants said that the new search function was more convenient than the former one. However, a small number said that the suffix data was still insufficient. Therefore, some suffixes were added to reference other dictionaries, and some suffix words were changed in detail to enable a deeper understanding of words.

3 Conclusion

In this research, new functions were added to the Hyper English Dictionary, such as a new search function by suffix of English words. This new search function allows for a deeper understanding of words and searching for words has become easier for learners of English. Overall, excellent results were obtained to increasing the usability of the Hyper English Dictionary.

Acknowledgment

References

[1] N. Asai laboratory at University of Aizu, Hyper English Dictionary,
http://mfntsv1.u-aizu.ac.jp/hidic/

[2] Margery S. Berube, *The American Heritage Dictionary of The English Language 4th Edition*, Houghton Mifflin Company, 2000.

[3] C. Watkins, *The American Heritage Dictionary of Indo-European Roots Second Edition*, Houghton Mifflin Company, 2000.

[4] A. S. Hornby and Sally Wehmeier, *Oxford Advanced Learner's Dictionary*, Oxford University Press, 2002.

[5] K. Kihara, *GRAND CONCISE ENGLISH-JAPANESE DICTIONARY* SANSEIDO, 2001.

[6] T. Konishi and K. Minamide, *TAISHUKAN'S Unbridged Genius English-Japanese Dictionary* Taishukan Publishing Company, Tokyo, 2001.

[7] T. Konishi and M. Yasui and T. Kunihiro and K. Horiuchi, *Random House English-Japanese Dictionary* Shogakukan, 2002.

[8] H. Michihata, , ALC Inc., 2002.

[9] Linux Operating System,
"Linux Online Web Site"
http://www.linux.org/

[10] Apache Jakarta Project,
"The Jakarta Site — Apache Tomcat"
http://jakarta.apache.org/tomcat/index.html

[11] MySQL AB,
http://www.mysql.com/

[12] Sun microsystems,
"Site of Sun microsytems – JAVA"
http://java.sun.com/

[13] Java Boutique,
"WEB technology with java Site"
http://javaboutique.internet.com/tutorials/JSP/part01/page03.html

[14] S. Yoshimura,
"A study about the hyper-dictionary used Indo-European roots information Master's thesis in the department of Science and Engineering at University of Tsukuba, 1997."