Improving Portability and Usability of the Indo-European-Root-Based English-Japanese Hyper Dictionary

Kumi Miyata   s1061148   Supervised by Prof. Yasushi Kikuchi

Abstract

Proto-Indo-European is considered to be the ancestor language of Indo-European languages. Most Indo-European languages inherited their meanings from Proto-Indo-European. Indo-European-Roots were structured with some stem patterns groups of Proto-Indo-European. In previous research, it was proved that 70-80 percent of the modern English words have Indo-European-Roots. The modern plural English words come from singular Indo-European-Roots, so Indo-European-Roots are useful for learning the vocabularies. For example, it's possible to consider some modern English words as one common nuance. Each synonym has a different Indo-European-Root, so the learner can use the word in the right way. The first edition of Hyper Dictionary which allowed searching Indo-European-Roots was finished, and it is running [1].

The first edition has some problems (see 1.3.1): The stems and the prefixes could be listed, but they could not be searched. Words which contain the symbol, #, could not be searched. If the quantity of a word information was immensely large, the system outputs an error message. In this research, we improved the first edition using Servlets and JavaServer Pages a kind of Java technology. The second edition can be made at low cost and high portability using Java. The database server is MySQL which is free and speedy. Indo-European-Roots have some particular expressions which cannot be dealt with by a computer. In the first edition, these words were replaced with another expression like #e, $i. These expressions are replaced by images in second edition.

1 Introduction

1.1 About Indo-European-Roots

Proto-Indo-European which is the ancestor language of Indo-European languages, is an unreal language, but is supposed to be one language group existing from 7000 years ago to about 5000-6000 years ago. It is called Indo-European-Roots that was structured by some stem patterns groups of Proto-Indo-European. This language's derivatives are about 12 linguistic families; GERMANIC, CELTIC, ITALIC, INDO-IRANIA, BALTO-SLAVIC, and so on. The population using Indo-European-Roots occupies half of world, and forms the most important family. This family has about 70 languages.

William Jones, an Indian scholar and lawyer, was interested in the ancient language SANSKRIT during his administration of India in 1786. He found a marked similarity between SANSKRIT and GREEK and LATIN, classical European languages. Indo-European-Roots started when he presented a hypothesis that these languages specialized from common source, and later GOHTO, CELTIC, ancient PERSIAN were added to this family.

In previous research, it was proved that 70-80 percent of the modern English words have Indo-European-Roots. The Indo-European-Roots family and characteristics were clarify by Indo-European languages comparative linguistics. Now, research is moving to discover the social system of the people who used Indo-European-Roots.

1.1.1 The Reasons For Observing the Indo-European-Roots

The Method of studying Indo-European-Roots is useful in vocabulary training. For instance, *hero* is often seen in cinemas and novels. *Hero:* "someone who is admired for doing something very brave or good, someone, especially a man or boy, who is the main character of a book, play or movie" in dictionary. According to The American Heritage Dictionary of the English Language (AHD) [5], *hero*'s Indo-European-Root is *-ser*. According to The American Heritage Dictionary of Indo-European-Roots SECOND EDITION [4],

ser-: To protect

This words correspond these modern English [10];

conserve: to prevent something from being wasted, damaged, or destroyed.
observe: to watch someone or something carefully.

preserve: to keep something or someone from being harmed, destroyed, or changed too much.

reserve: an amount of something that is kept to be used in the future when it may be needed.

reservoir: a special lake where water is stored to be used by people in the city.

hero: someone who is admired for doing something very brave or good.

hero is a "protector".

Conserve's prefix *con-*, which was derived from Indo-European-Root *kom*, means "communal, common." *Conserve*'s stem is *serve*, which comes from *ser-*, means "to protect." *Conserve* comes from "to protect together."

Observe's prefix *ob-*, which was derived from Indo-European-Root *epi* and *wer-*, means "reverse, to turn toward." *Observe*'s stem is *serve*. *Observe* comes from "to protect against."

Preserve's prefix *pre*, which was derived from Indo-European-Root *per*, means "fast, forward, precede, spare." *Preserve*'s stem is *serve*. *Preserve*'s meaning "to keep" comes from "to protect the spare."

Reserve's prefix *re-*, which was derived from Indo-European-Root *re-*, means "second, newly, backward." *Reserve*'s stem is *serve*. *Reserve* comes from "keep in the back, protect back".

These words are joined by *-ser*'s meaning, "to protect."

The Indo-European-Roots are useful for distinguish words. *Rage* and *angry* have similarly meanings, for example. This problem can be solved using Indo-European-Roots. According to the AHD, the Indo-European-Root for *anger* is *angh-* and *rage* is *rebh-*.

angh-: Tight, painfully, constricted, painful

agnail: a piece of dead skin that has become loose near the bottom to your fingernail.

hangnail: a piece of dead skin that has become loosenear the bottom to your fingernail

anger: a strong feeling of wanting to hurt or criticize someone because he has done something bad to you or been unkind to you.

angst: strong feelings of anxiety and sadness because you are worried about your life.

anxious: very worried about something, or showing to hat you are worried.

anguish: very great pain or worry.

quinsy: tonsillitis.

angina: strep throat.

Angina, *quinsy*, *agnail*, *hangnail*, *anguish* come from "tight, painfully", *angst*, *anxious* come from "painful mind". Then, *anger* means "anger to the torment, or painfully constrict your body." On the other hand,

rebh-: violent, impetuous

rabid: suffering from *rabies*.

rabies: a disease that kills animals and people that are bitten by an infected animal.

rage: strong feeling of anger that you cannot control

Rage is "(bitterly, violently) fly into a fury" from these words. This shows why *road rage* "driver flares on the road and drives recklessly" and *air rage* "passenger flare on the airplane and fly into a fury" cannot be expressed as *road anger*, *air anger*. The difference between *anger* and *rage* was distinguished in this way [6].

1.2 About the Hyper Dictionary

1.2.1 What Is a Hyper Dictionary

In recent years, the information oriented society is popularizing rapidly, so common people can use the Internet easily. The number of useful Internet contents is increasing, so daily life is becoming more and more convenient. In addition, people can shop and get data from databases at home. It is possible for us to look up with the dictionary not only on paper but also on the Internet. We call the dictionary which makes use of the merit of Hyper Text Markup Language (HTML) a Hyper dictionary.

1.2.2 The Reasons for Making Dictionary of Indo-European-Roots not Using the Paper but Using the Internet

The greatest advantage of a Hyper Dictionary is speed in looking up words. If using Indo-European-Roots, the quantity of information increases, but the Hyper Dictionary system is unlimited virtually. In addition, it is possible to update quickly if information becomes old if using the Hyper Dictionary. A Hyper Dictionary can come up to these needs.

1.3 About the First Edition of the Hyper Dictionary
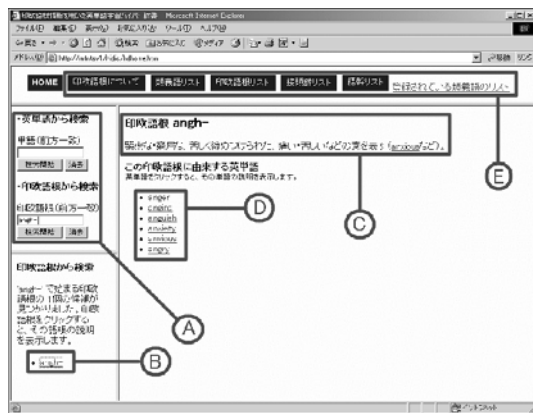
The first edition has the following features.

Figure 1: The First Edition

A A space to search for English words or Indo-European Roots (forward agreement).

B A space of for search results. This space shows candidates for applicable keywords.

C This space shows keywords and meanings.

D The words which have the same Indo-European-Roots.

E An explanation about Proto-Indo-European and Indo-European Roots, a list of synonyms, a list of Indo-European-Roots, a list of prefixes, and a list of stems. If you want to see these, click this image.

1.3.1   The Problems of the First Edition

The first edition had the following problems.

1. The stems and prefixes could be listed, but they could not be searched.

2. Words which contain the symbol, #, could not be searched.

3. If the quantity of the word's information immensely large, the system outputs an error message.

1.3.2   In Order to Make the First Edition more Useful

Indo-European-Roots have some particular expressions. A computer cannot deal with these words. In the first edition, these words were replaced with another expression like #e, $i, and so on. In second edition, in the case of particular expressions were replaced by images. In addition, the second edition outputs not only words which have the same Indo-European-Roots, the stems and the prefixes but also the word's meanings in order to compare each derivative easily.
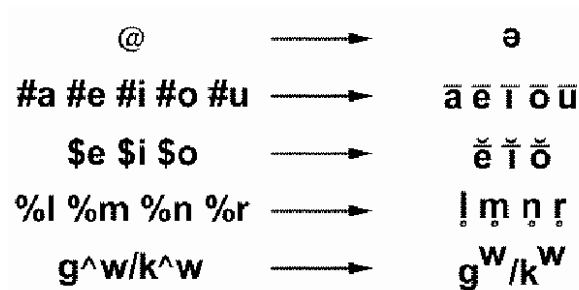

Figure 2: Particular Character

1.4   The Purposes of this Research

The purposes of this research are to fix the problems and to improve on the first edition. It should make it low cost and high portability for distributing in the future. The means which were selected to fulfill these conditions are Servlet and JavaServer Pages (JSP). These are Java technology. Java is a programming language. It was developed by SUN, and it isn't platform dependent. Using Servlet and JSP, the second edition can be managed more usefully and easily than first edition. The Relational DataBase (RDB) server for saving the information of words is MySQL. JDBC is used to access MySQL server from the Servlet. Servlet Container which is the environment for Servlet and JSP is Tomcat3.2.3.

OS	Windows2000
Java Environment	JSDK1.3.1
Servlet Container	Tomcat3.2.3
RDB Server	MySQL3.23.28
JDBC Driver	mmmysql
JDBC Driver type	Type4

2   About Servlet, JSP, and MySQL

This section explains how the second edition was developed and why this method was selected.

2.0.1   About Servlet

Servlet is one of the ServerSide Java technology, and has the same ability Common Gateway Interface (CGI). At first, Servlet reads the data from form tags of a web page or from a Java applet. It is possible to get the browser's ability and kind, information about cockies, and client user data. Second, Servlet processes these data. The results are created by HTML. Finally, Servlet sends HTML to the client, and the result is showed on browser.

2.0.2   About JSP

Using Servlet, HTML must be written in source code, so the programmer has to compile all source code if he want to change HTML a little. JSP is made to this Servlet's
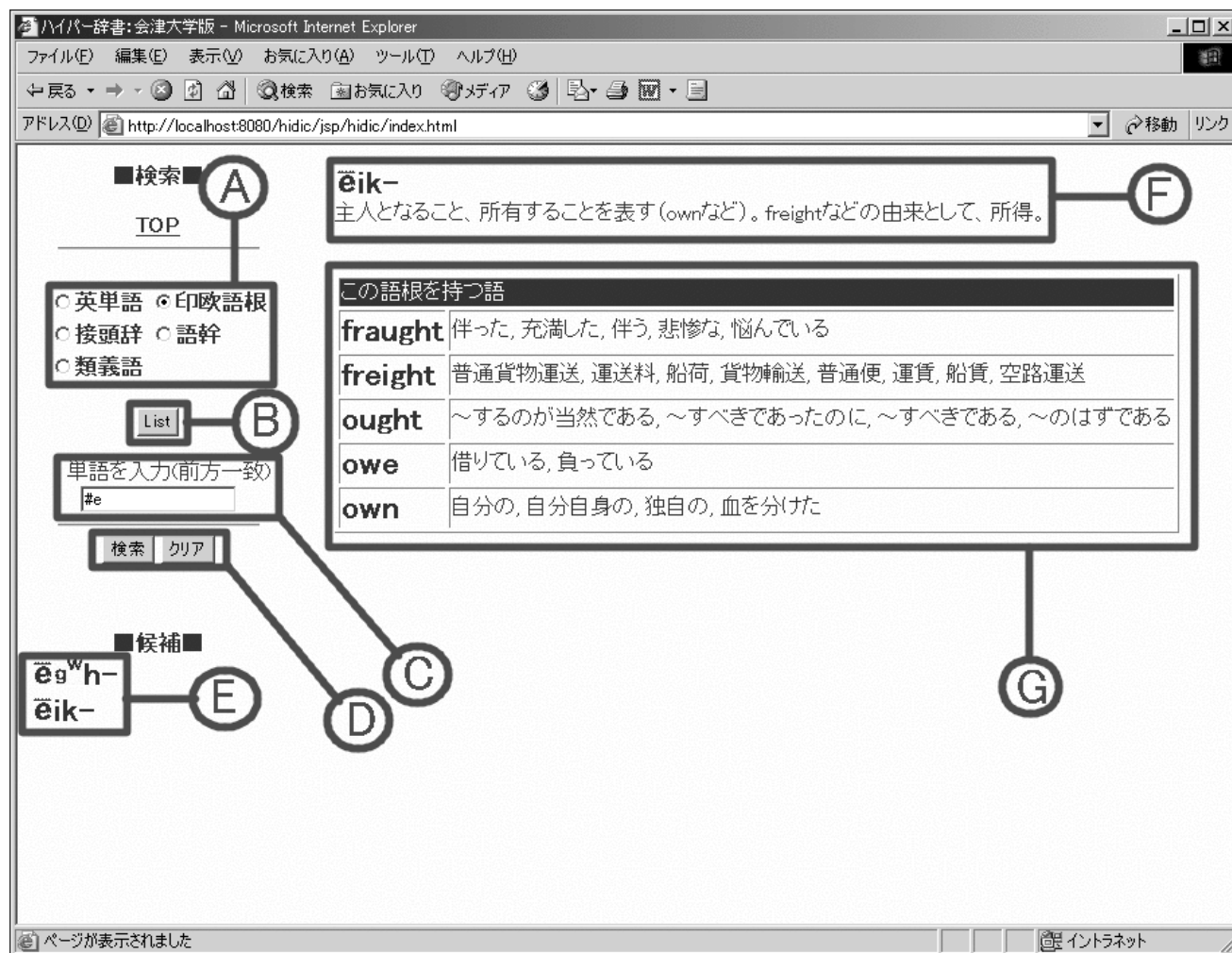
Figure 3: The Second Edition

weak point. JSP has the following characteristics:

- Using JSP with Servlet, a programmer can use both of static and dynamic HTML.

- JSP is written in HTML basically. Java source code are written in tags which are called Scriptlet.

- If JSP is called by Servlet, the source is compiled. Then, JSP get the data and create HTML.

2.0.3 Why it is not CGI but Servlet and JSP

Servlet has great advantages;

1. Portability
   Java is platform independent, so Servlet can be uses on all web server.

2. Security
   CGI uses functions depending on OS, so a programmer must pay attention to back quotes, semicolons, and so on. Even now, the famous library of CGI has bugs.

   Most CGI are compiled by languages which don't check the border of array or strings. If an array's max number of elements is 100, the programmer can write the 105th element even if it is a used memory.

3. Using Servlet and JSP, a program can be divided into a design area and a logic area. For instance, Servlet is in charge of a logic, and JSP is in charge of design. However, designer and programmer can use their ability.

2.0.4 About MySQL

MySQL is a free RDB server. MySQL's function is simple but its search speed is faster than any other RDB server. MySQL adopts GNU GENERAL PUBLIC LICENSE.

2.0.5 About JDBC

JDBC, which was developed by JavaSoft, is a tool to access RDB from Java program. JDBC is a trademark, so it isn't an abbreviation. First, JDBC accesses the RDB. Then, it sends SQL sentences to RDB and gets the results.

3 Research Contents

3.1 About the Second Edition

This section explains about the second edition in detail.

3.1.1 The Second Edition's Functions

The functions illustrated in Figure 3 and listed bellow:

A The User can choose the kind of word what he want to search.

B The list button.

C The space for the search keyword.

D The search button and the reset button. If the user clicks the search button, the system searches for the keyword. If the user wants search space to be reset, he clicks reset button.

E The search results space. This space lists candidate for applicable keywords.

F This space shows the word name and its meanings.

G This space shows words and meanings which have the same Indo-European-Roots.

3.1.2 Solved Problems

The problems in 1.3.1 were all solved.

1: Using checkbox, the second edition can change the kind of word.

2: It can search the words which contains the symbol, #.

3: The second edition can output words which could not output by first edition.

The particular expressions were replaced by images (see E in Figure3), so the user can see the accurate representation of Indo-European-Roots. Now, the second edition shows not only the word's name but also its meanings in order to confirm differences or similarity of meaning all at once (see G in Figure3).

3.1.3 Future Work

The particular expressions were replaced by images (see E in Figure3), so the user can see the accurate representation of Indo-European-Roots. The user can resize the word's size with browser setting, but the size of images cannot change. In the future, it is a goal that both words and images sizes can be changed in the same time.

4 Conclusion

In this research, problems of the first edition (see 1.3.1) were all solved. The second edition can be searched by English word, Indo-European-Root, stem, or prefix. The accurate representation shows by image. Now, the second edition has became more useful and user friendly.

5 Acknowledgment

References

[1] Y. Ikebe, "Hyper Dictionary for English Word Study", http://mfntsv1.u-aizu.ac.jp:8080/hidic/hdhome.htm.

[2] M. Agnes, "Webster's New World Collegiate Dictionary Forth Edition", *IDG Books Worldwide Ink*, 1999.

[3] Y. Terasawa, "The Kenkyusha Dictionary Of English Etymology", *IDG Kenkyusha*, 1997 (Japanese).

[4] C. Watkins, "The American Heritage Dictionary of Indo-European-Roots SECOND EDITION", *HOUGHTON MIFFLIN COMPANY*, 2000.

[5] "The American Heritage Dictionary Forth Edition", *Houghton Mifflin Company*, 2000.

[6] Y. Ikebe, "The Construction of Hyper Dictionary Database For The Study Of English Words Second Edition", *International Symposium on the East and West Language Cultures*, July 25-27, 2001

[7] Y. Harada, "JAVA Server Side Programming", *Gijutsi-Hyoron-Sha*, 2001 (Japanese).

[8] H. Tanaka, "Hajimete-no JSP & Servlet", *SOFT BANK*, 2001 (Japanese).

[9] M. Tomita, "MySQL Tettei-Nyumon", *Shoei-Sha*, 2001 (Japanese).

[10] "LONGMAN DICTIONARY of AMERICAN ENGLISH", *Obunsha*, 1997.