WEB BROWSER WORD BASED TRANSLATOR

Recep Ayaz, Fatih Dalgıç, Banu Diri
Yıldız Technical University, Department of Computer Engineering, Istanbul
recepayaz@hotmail.com, fatihdalgic@yahoo.com, banu@ce.yildiz.edu.tr

Özet. *Bu çalışmada; sıradan bir web tarayıcısının sahip olması gereken özellikleri üzerinde barındıran ve bu özelliklere ek olarak gezilen yabancı sayfaların kullanıcının diline kelime tabanlı olarak çevrilmesine yardımcı olan çevirmen modülünü ve bu tarayıcıyı kullananların kendi aralarında sohbet etmelerini sağlayacak modülü barındıran bir web tarayıcısı geliştirilmiştir. Gerçekleştirilen web tarayıcısı, mevcut web tarayıcıları içerisinde ilk olma özelliğine sahiptir. Yabancı dili yeterli olmayan (kelime haznesi az olan) kullanıcılar bu çalışmanın hedef kitlesini teşkil etmektedir.*

Abstract. *The goal of the study is to design a web browser that has all unique properties of a normal web browser and to help the user who is not good a foreign language by translating the web sites based on word translation. Also the users will be able to chat each other by the chat module. The destination group of the study is the users that has no enough foreign language ability. The interface of the web browser is very user-friendly so any user can use the program easily.*

1. INTRODUCTION

The Internet is a world wide communication network that is continuously growing. Today, the most effective way to reach information in an easy, inexpensive, fast and reliable manner and to share it is to use the Internet.

People can now carry out their works, shopping and even education from the Internet and communicate with each other through it. The number of registered e-mails has increased rapidly in the recent years and the Internet has also generated its own culture. The applications are required as the Internet comes into our life. Chat environments and software constitute a very good example for this. Today there are many different chat environments and web browsers; however there is not a common usage yet.

There are various problems that prevent it from becoming a common platform, despite is wide usage. For example, when you conduct a search based on a keyword, you are faced with sites in languages other than your native language. However, unless we know foreign language very well, we shall probably not be able to use these sites. Even if we assume that a person may know more than one language, it may not be possible for him/her to know all of them very well.

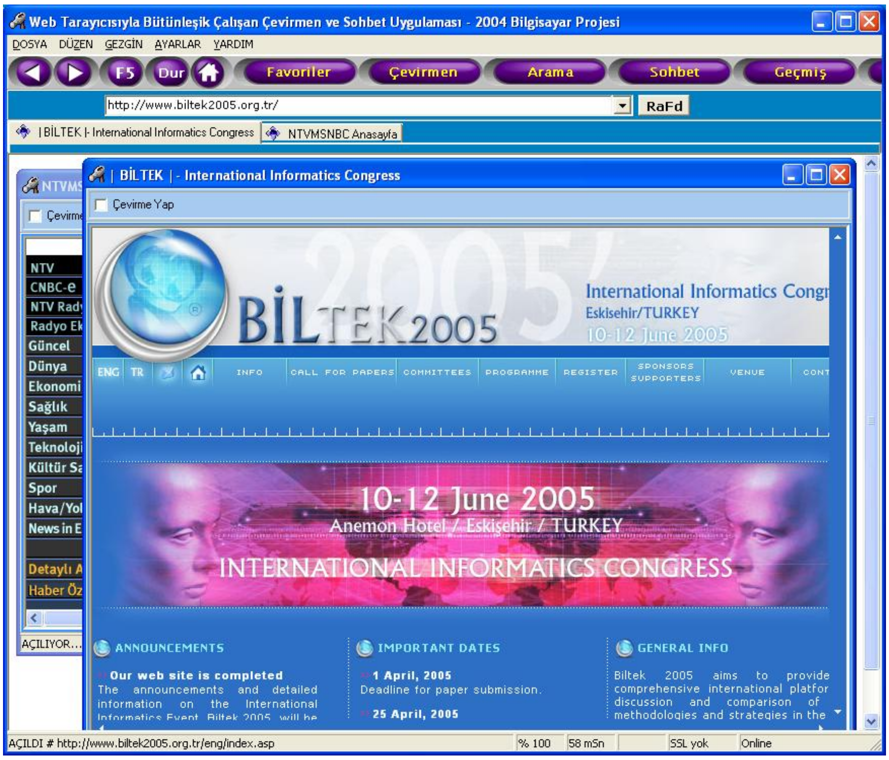
We have attempted to bring a solution to this problem with the software that we developed. We have enabled users to realize on a single platform all that they need on the Internet. We have developed an easy to use and comprehensible browser that includes an integrated chat application, a word based translator module that can be adjusted according to the foreign language level of the user. Alternative for each module of this application exists. MSN Messenger can be shown as an example to the "Chat Module" and Babylon and Systran[1] for the "Translation Module".

The modular structure of this developed application has been explained in the paper and a general comment and evaluation have been made in the last section.

2. GENERAL STRUCTURE OF THE BROWSER

The developed web browser is a comprehensive and easy to use program with its menus. Its structure and general appearance has been given in Figure 1 and Figure 2.

[1] http://www.dejavu.og, "The history of the web"



Web Tarayıcısıyla Bütünleşik Çalışan Çevirmen ve Sohbet Uygulaması - 2004 Bilgisayar Projesi

http://www.biltek2005.org.tr/

| BİLTEK |- International Informatics Congress

BİLTEK2005 International Informatics Congr Eskisehir/TURKEY 10-12 June 2005

10-12 June 2005 Anemon Hotel / Eskişehir / TURKEY INTERNATIONAL INFORMATICS CONGRESS

ANNOUNCEMENTS

Our web site is completed
The announcements and detailed information on the International Informatics Event Biltek 2005 will be

IMPORTANT DATES

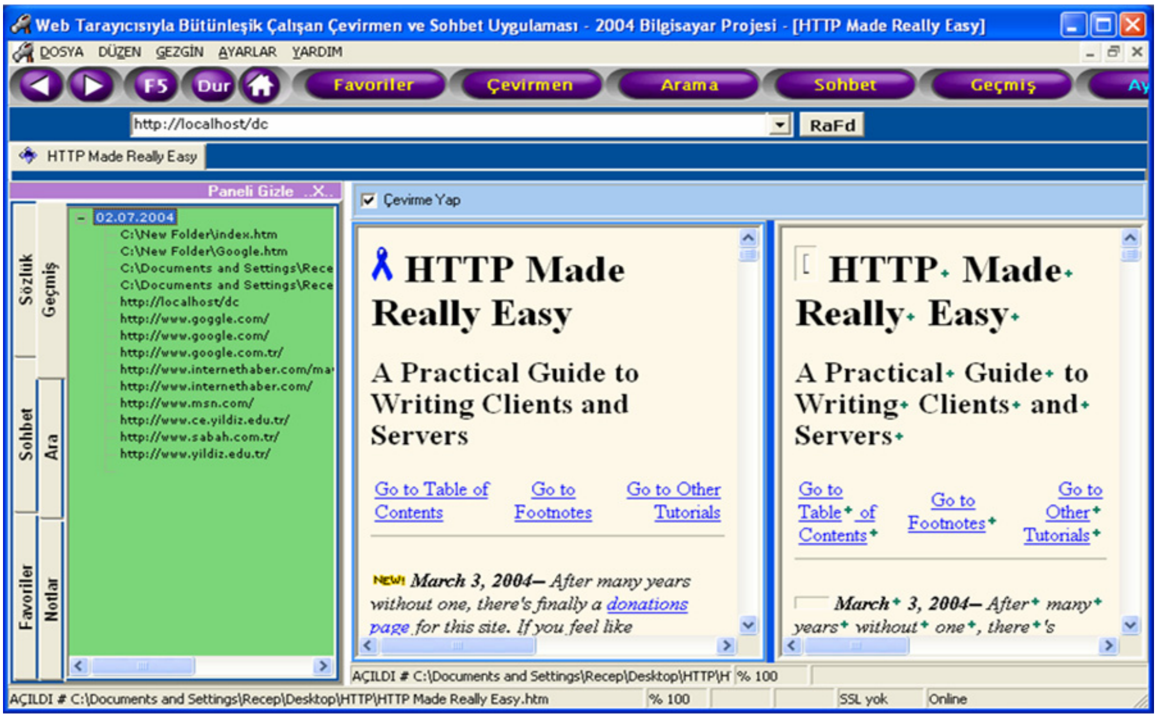
1 April, 2005
Deadline for paper submission.
25 April, 2005

GENERAL INFO

Biltek 2005 aims to provide comprehensive international platfor discussion and comparison of methodologies and strategies in the

AÇILDI # http://www.biltek2005.org.tr/eng/index.asp % 100 58 mSn SSL yok Online

Figure 1. Word based translator and chat application integrated with the web browser



Web Tarayıcısıyla Bütünleşik Çalışan Çevirmen ve Sohbet Uygulaması - 2004 Bilgisayar Projesi - [HTTP Made Really Easy]

http://localhost/dc

02.07.2004
C:\New Folder\index.htm
C:\New Folder\Google.htm
C:\Documents and Settings\Rece
C:\Documents and Settings\Rece
http://localhost/dc
http://www.goggle.com/
http://www.google.com/
http://www.google.com.tr/
http://www.internethaber.com/ma
http://www.internethaber.com/
http://www.msn.com/
http://www.ce.yildiz.edu.tr/
http://www.sabah.com.tr/
http://www.yildiz.edu.tr/

HTTP Made Really Easy

A Practical Guide to Writing Clients and Servers

Go to Table of Contents Go to Footnotes Go to Other Tutorials

NEW! *March 3, 2004— After many years without one, there's finally a donations page for this site. If you feel like*

HTTP• Made• Really• Easy•

A Practical• Guide• to Writing• Clients• and• Servers•

Go to Table• of Contents• Go to Footnotes• Go to Other• Tutorials•

March• 3, 2004— After• many• years• without• one•, there•'s

AÇILDI # C:\Documents and Settings\Recep\Desktop\HTTP\H % 100

AÇILDI # C:\Documents and Settings\Recep\Desktop\HTTP\HTTP Made Really Easy.htm % 100 SSL yok Online

Figure 2. Word based translator module

The developed application consists of "Web Browser Module", "Chat Module" and "Word Based Translator Module".

2.1. The Web Browser Module

The web browser module has been designed such as not to required the user to start more than one program. The possibility has been provided to the user within the program to open more than one browser. The web browser supports JavaScript, Flash and other plug-ins[2]. It uses the same libraries as the Internet Explorer.

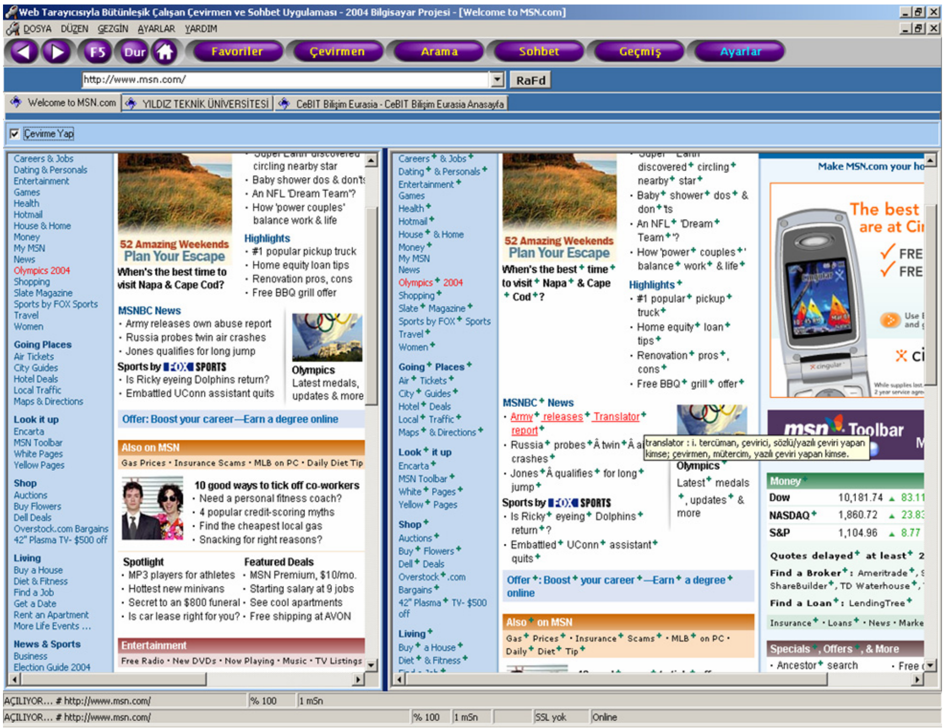
2.2. The Chat Module

The purpose of the module is to provide the opportunity to the users to establish a common platform and communicate with each other. The users can also send offline messages to each other in the chat module.

2.3. The Translator Module

The aim of the Translator Module is to ensure that Internet pages other than the mother language of the user are more comprehendible. To enable the user to understand an Internet site in any language is the aim of this work. The application conducts a word based translation of a page in a foreign language. After the page that user opens is loaded, its word based translation is shown next to it.

Certain formal problems were encountered in the Translator Module of the work. When we translate the page we encounter as word based in other words, instead of words themselves if the translation of those words is written, a text is created that has no meaning. Also, the layout of the page is deformed. In order to prevent these complexities, a different displaying method has been used in the application. A symbol has been placed beside the translated words. When you bring the mouse pointer ever the word in order to lean its meaning, the meaning of the word is displayed. Thus the original layout of the page is not distorted. The user shall see a screen such as the one in Figure 3.


Web Tarayıcısıyla Bütünleşik Çalışan Çevirmen ve Sohbet Uygulaması - 2004 Bilgisayar Projesi - [Welcome to MSN.com]
DOSYA
DÜZEN
GEZGİN
AYARLAR
YARDIM
F5
Dur
Favoriler
Çevirmen
Arama
Sohbet
Geçmiş
Ayarlar
http://www.msn.com/
RaFd
Welcome to MSN.com
YILDIZ TEKNİK ÜNİVERSİTESİ
CeBIT Bilişim Eurasia - CeBIT Bilişim Eurasia Anasayfa
☑ Çevirme Yap
AÇILIYOR... # http://www.msn.com/
% 100
1 mSn
SSL yok
Online

translator : i. tercüman, çevirici, sözlü/yazılı çeviri yapan kimse; çevirmen, mütercim, yazılı çeviri yapan kimse.

Figure 3. The word based translator and chat program integrated into the web browser. By means of this there is no complexity, and the whole page can be read from the beginning to the end without pressing any extra key or operation other than moving the mouse.

[2] http://en.wikipedia.org/wiki/Web_browser, "Web Browser-Wikipedia, The Free Encyclopedia"

There is a menu with three options in the Translation Module for the user. With this menu, one is chosen from the options "beginner, intermediate and advanced" and thus the translation of all the words is prevented and time is saved.

The Translator module has not been written especially for or dependent on any specific language. The purpose was to have it support all languages. Its current version supports only English; however it has been designed so as to support other languages once the word database is loaded. Also, it has been designed to store the word based translated (processed) text in a directory (temp). This way, there is a chance to return to the page later and continue reading.

Other than automatic translation, the user can also do manual word search. The system also has a dictionary menu and additionally, the meaning of any selected word can be displayed by pressing "Ctrl + T".

3. STRUCTURE AND ALGORITHM OF THE WORD BASED TRANSLATION MODULE

The translation structure of the application is word based and words are translated one by one from the HTML (Marshall, 1977) code. As mentioned above, certain restrictions have been put in order to prevent the web browser from translating all the words on the page. First, a list has been made of the words most used in that foreign language. Since Turkish-English word based translation has been taken as example in this study, there is a menu with three options in the Translation Module. With this menu, users can select one of "beginner, intermediate and advanced" language levels from the menu prevent the translation of all the words on the site and thus save time. These words are absolutely not used for automatic word based translation (pronouns, and, or, …). For other words, the words are selected and translated depending on the foreign language level of the user.

Usage rate curves have been drawn out for 25 foreign sites (all the study has been conducted for English) in order to determine this foreign language level criterion. The curve drawn out for a single site has been shown in Figure 4. It is observed that these curves comply with the Zipf Law (Manning and Schütze, 2002) just like all natural languages. When we interpret these curves, we obtain the usage frequency on the page of each word. It also gives the knowledge or familiarity rate among people who do not know the language (or are new learners). By interpreting these ratios, we left 3 language levels for the user in our application: "beginner", "intermediate", and "advanced":

- beginner: Meanings of all words can be seen other than ones known as frequently used, which are stored in the buffer database.
- intermediate: Words other than ones with usage frequency higher than 7 and higher can be seen.
- advanced: Meanings of words whose usage frequencies are 1, 2 and 3 can be seen. These are words used on the page with least frequency.

The time required by the Translation Module for determining words for which meanings are to be displayed to the user and reloading of the page, is approximately twice as much as the loading time of the page. Since it works in a word based manner, the Translation Module is designed so as to work with any language for which a word database is loaded.

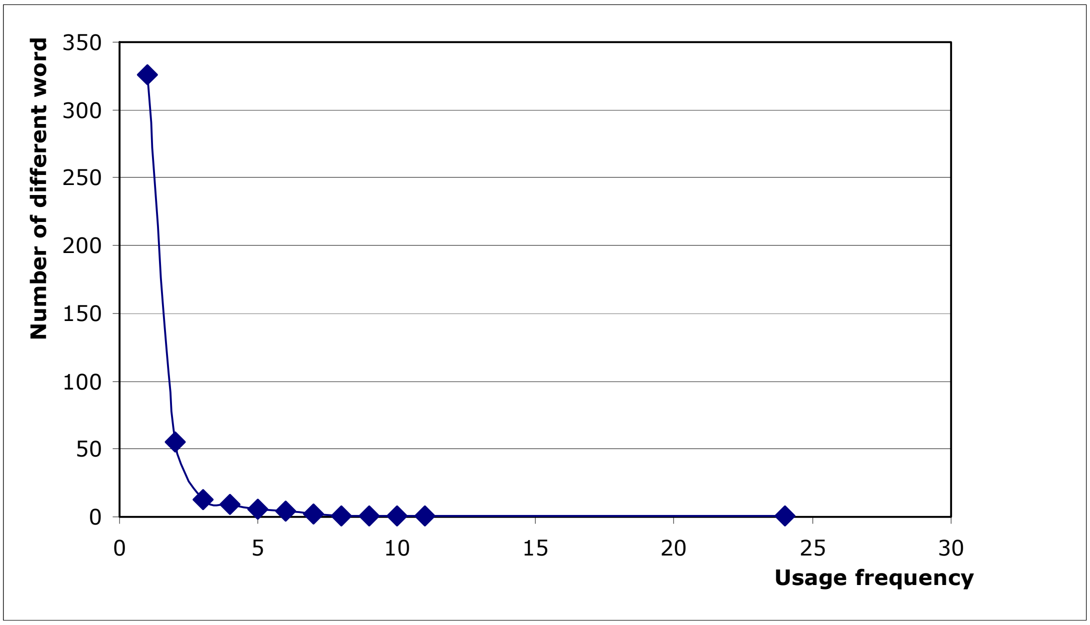


Figure 4. Usage frequency curve of different number of words for the site "http://www.msn.com"

Table 1. Distribution of words according to usage frequency in 25 different sites

Row	**Usage frequency**	**Number of different word**	**ratio**
1	1	326	%77,62
2	2	55	%13,10
3	3	13	%3,10
4	4	9	%2,14
5	5	6	%1,43
6	6	4	%0,95
7	7	2	%0,48
8	8	1	%0,24
9	9	1	%0,24
10	10	1	%0,24
11	11	1	%0,24
12	24	1	%0,24

Table 1 allows for a better interpretation of Figure 4. It shows that the 326 different words in row one have been used only once in 25 sites, and similarly, that the 55 different words have been used only twice in 25 sites.

4. CONCLUSION

contains the ordinary features of existing browser, it also includes a word based translator that helps the translation of a browsed page. Users do not have to use a separate dictionary program while browsing foreign sites in order to find the meaning of any word. The structure of the application supports all languages. The only necessary procedure is to introduce to the system the alphabet and dictionary of the language. The user interface of the web browser has been designed as simple and comprehendible. The user shall not have any problem using the web browser and shall have the opportunity to chat with persons using this web browser.

References

(Bell and Parr, 2001)	D.Bell and M. Parr, Java, Prentice Hall, 2001
(Marshall, 1997)	James Marshall , HTTP Made Really Easy , 1997
(Manning and Schütze, 2002)	C.D.Manning ve H.Schütze, Foundations of Statistical Natural Language Processing, MIT, 2002
(Borland, 1997)	Borlan Delphi User's Guide

Autobiography

Recep Ayaz. He was born in Istanbul on 1983. He graduated from Fatih Erkek High School. He is currently continuing his undergraduate education in Computer Engineering in Yıldız Technical University.

Fatih Dalgıç. He was in Bursa on 1981. He graduated from Çapa Anatolian Teachers High School. He is currently continuing his undergraduate education in Computer Engineering in Yıldız Technical University.

Banu Diri. She was born in Istanbul on 1966. After graduated from Şehremini High School, she completed her undergraduate, graduate and PhD degrees in Computer Engineer at Yıldız Technical University. She is currently working as assistant professor in the Department of Computer Engineering at the Yıldız Technical University. She has publications in data compression and natural language processing.