

# A System for Bangla Community to Enhance English Capability through Web Browsing

Md. Nazim Uddin, Md. Abdus Salim Mollah and M. A. H. Akhand

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna-9203, Bangladesh

nazim.cse.kuet@gmail.com, salim9326@yahoo.com, akhand@cse.kuet.ac.bd

**Abstract**—Internet is the single largest source of information and it grows day by day. Since most of websites are in English language; so it is difficult to understand for a person when he is weak in English. To understand information by other languages community, there existed ‘websites translator’ facilities provided by Google, Microsoft and other companies. Unfortunately the translation facility from English to Bangla is not adequate although it is fifth most spoken language. The number of informative websites in Bangla are very few with respect to English and a major portion of less educated Bangla community are away to gather information from internet due to weakness in English. The aim of this project is to (1) develop a system that will provide English website in a transformed from containing Bangla meaning of all or specific English words and (2) provide facility to enhance knowledge on English language when browse English website. This developed ‘Browse English Website with Bangla Words (BEWBW)’ will analyze the contents of source English website and provide the transformed website with Bangla words corresponding to English with the help of a dictionary. This system will also track which English website he/she visited and for which words bangle meaning provided. The BEWBW (web and desktop based) will be helpful for the people who are weak in English in Bangladesh as well as Bangla community anywhere in the world.

**Keywords**— *Bangla Browsing; Website Translator; Learning; Bilingual Dictionary;*

## I. INTRODUCTION

The internet is the single largest source of any kind of information and it grows every day. Information available in internet website sometimes are very helpful for various purposes like health care, education, research and other daily life matters. Since most of websites are in English [1], [2], it is difficult to realize for a person when he/she is weak in English. To understand information by other languages community, there are some translator facilities given by Google, Microsoft and other companies. Unfortunately none of the translator service provides in Bangla although it is fifth most spoken language. Bangla is the first language of Bangladesh, West Bengal and Tripura (two states in India) and is spoken by a population that now exceeds 250 million [2], [3], [4]. Moreover, the number of informative websites in Bangla is very few with respect to English; and a major portion of Bangla community who are weak in English are away to gather information from the English Website in the Internet.

This paper presents a system ‘Browse English Website with Bangla Words (BEWBW)’ to provide facility to Bangla community to acquire information from English website as well as to enhance English capability. The aim of this research is to develop a system that will provide English website into a transformed format having Bangla meaning of specific English words. The BEWBW will analyze the contents of source English website and provide the transformed website with Bangla words corresponding to English word with the help of a dictionary for word translation. The system (web and desktop based) will be helpful for less English educated people in Bangladesh as well as anywhere in the world.

The rest of the paper is organized as follows. In section II the paper discussed about the Methodology of the proposed System; development of the system, output and access of the project is discussed in section III. Section IV illustrates the concluding remarks and scope.

## II. METHODOLOGY

The main goal of this research project is to facilitate comparatively less English educated Bangla community by providing web facility with Bangla meaning of the selected (hard and harder) English words as well as the English learning facilities. So the focused goals are: (i) develop Web browsing facility with Bangla words (ii) Provide English learning facilities. To reach the goal, the specific objectives are as follows:

- Analyze the requested English website.
- Identify the words for which Bangla meaning will be provided with the help of the dictionary.
- Prepare the transformed website with Bangla words and provide to internet users.
- Develop the mechanism to build and/or update the dictionary for translation service.

The system will be hosted on the Internet web server and accessible from anywhere through Internet. A user will browse English website through specific way of the system. The system will analyze the original requested English website and provide the transformed website with Bangla words with the help of the dictionary.

The structure of the whole proces has been shown in Fig. 1 and the major sequential steps are as follows:

- Receive request for an English webpage from a user.
- The BEWBW load the requested website from origin Server via Internet.
- Identify the words for which Bangla service will be provided based on user status and dictionary content.
- Prepare the transformed website for the user with Bangla words in bracket for the corresponding selected English words.
- Provide the transformed website to the user.
- Finally, update the user status by maintaining how much words he/she has already learnt. For those words he/she has already learnt, will not appear in the transformed webpage next time.

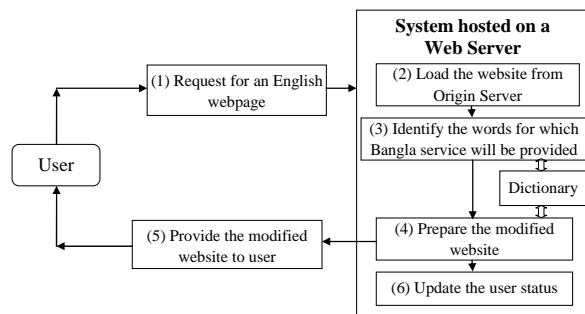


Fig1. Proposed System Structure

### 1. Receiving Request from User

The system must have a user interface to facilitate the user of website browsing. By using the BEWBW's user interface, the user can enter the URL of the English website and get the page with Bangla words as an output. The BEWBW will have two type of user: (i) Unregistered user identified as anonymous user and (ii) Registered user. Unregistered user will enjoy the basic browsing facilities and registered user will get the English learning facilities.

### 2. Loading Requested Website from Origin Server

The system then gets the requested page of the website from the web server and advances to the further processing steps. In that case only the HTML page of the website is downloaded not the whole page, i.e. website resources: images, videos, flash animation, java applet, audio files etc. URL of the resources of the page is kept unchanged. Normally the downloaded HTML page is not well formed and contains lots of scripting language code (JavaScript, AJAX, jQuery), style code (CSS), META etc. This is very hard to perform any processing operation on this unformed HTML page. First of all, it is required to make the HTML pages well formed. Generally content of the page are in the HTML tags mainly the <div></div>, <p></p> tags. If tag is not well formed then it will be very hard to replace the content of the pages.

### 3. Identifying Words for Bangla Trnslation

To function properly, the system required bilingual English to Bangla Dictionary in the database. The system

reads the words from the HTML/ web page and replaces the English words with corresponding Bangla words/ meaning depending on the user status that is user level and preferences set by the user earlier. Some HTML tags or CSS, JavaScript, jQuery syntax and there attributes are like English words so it is not be possible to translate the whole page in Bangla. And the replacement is very slow process, this may reduce the speed of processing the page significantly. To speed up the process, we have used the HTML Agility Pack (XML node search has some limitation, it required the well formed HTML page) to search the node element and replace.

### 4. Preparing Transformed Website with Bangla

The system will replace the English words with the English words and the Bangla meaning in first bracket with it e.g. "right" will be replaced with "right (অধিকার, সঠিক)". Once the replacement is finished, the page is being saved in the current server and redirected to the user end. Dictionary formation and Server setup process is discussed in details in the further section.

### 5. Providing Transformed Website to User

The recent transformed web pages is saved to the current server and sent to the user. The end user then gets the transformed webpage that contains the Bangla meaning of the specific English words.

### 6. Updating User Status

The registered user learns the English words and marked as learned. These learned words will not be displayed with Bangla further. The words that user has already learnt will not appear in the transformed webpage for the next time.

## III. DEVELOPMENT OF THE SYTEM

To create and run the BEWBW there is some elements and components are required. The key component of the system is Bilingual Dictionary to generate a useable Dictionary which is very challenging task. The other components are HTTP/ Web Server, Software (Web and Desktop based) to browse the website etc. These are described in details in the letter sub sections.

### A. Software Development

After analyzing the system the required features are identified. The required features of the developed software are enlisted below:

1. New user registration facilities
2. Website browsing facilities
3. Bangla meaning of the English Words
4. Word Level Selection facilities
5. Word Learning facilities
6. Marking Learned Words (not be displayed later)
7. Deleting unlearned words
8. Quick Dictionary
9. Searching words meaning in other dictionaries.

The software of the system has been developed in both web and desktop versions. The web version is hosted in a web server [9] and publicly accessible through internet. Desktop version is downloadable and installable to the local machine although internet connection is needed for browsing website. Desktop version provides the English learning facilities without any network connection because the database is compacted in it. We have used the .NET framework to develop the software. In web based version we have used C#, ASP .net, HTML, CSS, jQuery as User Interface (UI) and SQL Server as Database. In the desktop based version we have used C# windows form and MS Access as database.

### B. Dictionary Preparation

To translate from one language to another one must need Bilingual Dictionary. English words are collected to prepare Dictionary with used “Frequency” and “Parts of Speech” of the words from “Word frequency data” [5], [6]. The data contains the English Word list (100000 words approximately with some useless words and repetition), Parts of speech, Word Used Frequency (relative frequency that is calculated using books and human utilization of words) and Lemma, the more frequency means the word use more frequently, and the dispersion of the words [7], [8]. All these information’s are useful for us but we need some additional information (Bangla meaning and Level of the words) to make the useful bilingual dictionary.

TABLE I. SAMPLE ENGLISH WORD LIST WITH FREQUENCY

Sl.	Word	Lemma	PoS	Frequency
1.	include	include	vv0	65764
2.	knowledge	knowledge	nn1	57974
3.	caused	cause	vvn	17794
4.	devices	device	nn2	14373
5.	wake	wake	vv0	10755
6.	conversations	conversation	nn2	9488
7.	citizenship	citizenship	nn1	6344
8.	occupied	occupy	vvn	5813
9.	boost	boost	vv0	5438
10.	confirmed	confirm	vvd	5321
11.	wrapped	wrap	vvd	4477
12.	cooperate	cooperate	vv0	4385
...	...	...	...	...

TABLE I illustrates the English Word List with Frequency that we have collected from the internet resources [5]. This contains 100000 (1 Lac) English words (Word) with Word Used Frequency (Frequency), Part of Speech (PoS), and Lemma (Lemma) but still the Data is not in Bilingual Dictionary format. These Data is helpful to prepare the Bilingual Dictionary. To generate the Bangla meaning of the English Words we have used the “Google translator tool kit” and finalized the bilingual Dictionary. The “Google translator tool kit” is not error free; lots of Bengali meaning is incorrect.

With the Data mining we have corrected the words meaning by using Dictionary books. Among the 100000 words there are some repeated words as well as unused useless words. To make the search faster we have removed those repeated and useless words and over 40000 words are kept that can fully supports our requirements.

### C. Data Mining and Leveling

We have prepared the Dictionary with Bangla meaning and classified the words in 3 Level. Description of the three level of word is illustrated and enlisted in TABLE II.

TABLE II. WORDS LEVELING

Sl.	Level	Type	Definition
1.	Level 1	Normal	All Words in the ditionary except preposition
2.	Level 2	Hard	Hard and Harder Words
3.	Level 3	Harder	Harder Words only

Leveling is performed depending on the Word used Frequency and the length of the Word. Word Leveling procedure has no standard rules and regulations. Leveling is varies depending on human education. This implementation use the Frequency as the approximate Leveling criteria. Leveling will be more apporoprite if this is performed by the human voting system. There is much scope to works in this area. We are not fully success in this area.

TABLE III. SAMPLE FINAL ENGLISH TO BANGLA DICTIONARY

Frequency	Word Level	Main Word	Word Type	Bengali Meaning
748	3	maintenance	Noun	পরিচর্যা
148	3	metabolism	Noun	বিপাক বা সংশ্লেষণ
4879	2	survive	Transitive verb	টেকা
686	2	caution	Noun, Transitive verb	সাবধানতা
3447	1	loss	Noun	ক্ষতি
4932	2	suspect	Verb, Noun, Adjective	সন্দেহ করা
518	1	loot	Noun, Verb	অপহরণ
15	3	sweetening	Noun	মধুরতা - স্বধারকারী
100	3	bureaucrat	Noun	আমলাতন্ত্রবাদী
1803	2	specific	Adjective, Noun	নির্দিষ্ট
...	...	...	...	...

The implementation categorize the English words in three level, is a novel idea. Level 1 includes the all words except the prepositions and nouns, Level 2 includes hard words that is flaged by the users. Level 3 includes harder words only.

There is no boundary of hard and harder words; its depends on the user. The final generated bilingual Dictionary with leveling is shown in TABLE III. The Dictionary has Word Id as a unique Primary key where each word has a unique Id, Word level, main English word, part of speech/ word type and last of all Bangla meaning of the word. TABLE III illustrates the Dictionary structure.

#### D. System Setup

We have developed both desktop and web version. Desktop version of the project does not require any server; it can be used like a web browser. The desktop version of the software contains all the English words in Microsoft Access database with it; so database server is not needed to be installed. The web based version of the developed software is based on IIS web server. The system may be placed in any IIS based web server and may be accessed from anywhere through internet. At present the system is placed KUET web portal [9]. The desktop version of the project is also available to download in the same place.

### IV. EXPERIMENTAL OUTCOME OF THE DEVELOPED SYSTEM

This section explains the outcome as well proficiency of the developed system on basis practical testing.

#### A. System Performance

The developed system has been tested on a large number of international and national famous websites with different levels. Fig. 2 shows the level wise word translation measure for homepage of four selected websites that are KUET, BBC, CNN and Microsoft. For all the cases, Level 1 translates most of the word because system provides translation facility for all the words in the dictionary except preposition for this level. On the other hand, Level 2 translates hard and harder words and Level 3 translates only the harder words. The performance of the system is found similar for other tested websites including AOL, Yahoo, Google, BUET and LinkedIn.

Fig. 3 illustrates the website conversion with the proposed system using Level 3 for a sample webpage <http://www.kuet.ac.bd/index.php/welcome/welcomereadmore>. It is observed that only the hardest word is transformed that was marked as “Level 3” in the system.

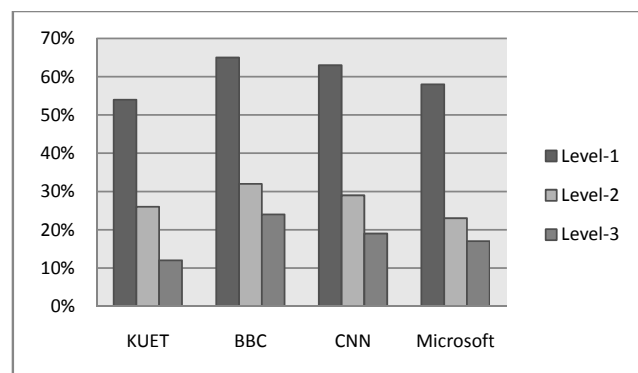


Fig 2. Level wise Word Translation for selected websites.

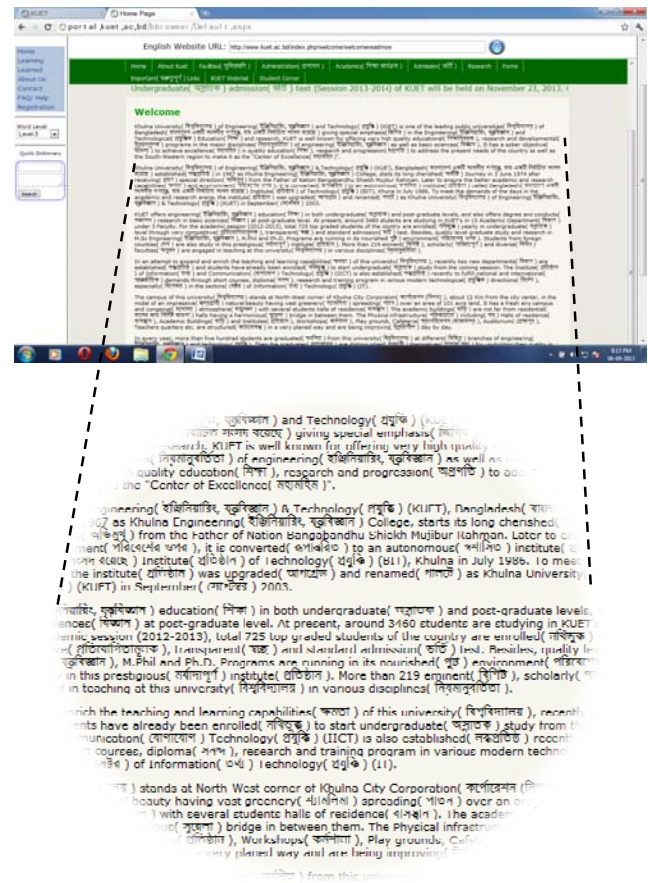


Fig 3. Level-3 (Harder words only) output of KUET website.

#### B. Learning Facility through Web Browsing

Learning by website browsing is another major concern of the proposed system. When user browse any page then transformed words will be automatically cached in user learning table.

Display Word: All This Week This Month This Year							
Learned Words List							
Sl.	English	Bangla	Word_Type	Website	Date	Time	Del
1	about	উপরে	Preposition, Adverb, Noun, Adjective	www.yahoo.com	2013-04-22		[Del]
2	account	বিশেষ	Noun, Transitive verb	www.aol.com	2013-04-09	11:50:34 AM	[Del]
3	access	গেয়ে		www.yahoo.com	2013-04-22	11:58:21 AM	[Del]
4	Access	গেয়ে		www.yahoo.com	2013-04-22	12:04:38 AM	[Del]
5	afghanistan	আফগানিস্তান	Noun	www.aol.com	2013-04-09	11:50:04 AM	[Del]
6	affairs	বিষয়		www.yahoo.com	2013-04-22		[Del]
7	amazon	আমাজন	Adjective	www.aol.com	2013-04-22	11:50:28 AM	[Del]
8	amazon	আমাজন	Adjective	www.aol.com	2013-04-22	12:06:06 AM	[Del]
9	after	পরে	Preposition, Adverb, Conjunction, Adjective	www.aol.com	2013-04-09	11:50:03 AM	[Del]
10	access	আমাজন	Preposition, Adverb	www.yahoo.com	2013-04-22	11:58:34 AM	[Del]
11	access	আমাজন	Preposition, Adverb	www.yahoo.com	2013-04-22	12:04:44 AM	[Del]
12	activity	কার্যক্রম	Noun	www.yahoo.com	2013-04-22	11:57:57 AM	[Del]
13	activity	কার্যক্রম	Noun	www.yahoo.com	2013-04-22	12:04:31 AM	[Del]
14	affected	প্রভাবিত	Adjective	www.aol.com	2013-04-09	11:50:12 AM	[Del]
15	about	সম্পর্কে	Preposition, Adverb, Adjective	www.aol.com	2013-04-09	11:50:34 AM	[Del]
16	Against	বিরুদ্ধে	Preposition	www.aol.com	2013-04-22	11:50:26 AM	[Del]
17	Against	বিরুদ্ধে	Preposition	www.aol.com	2013-04-22	12:06:04 AM	[Del]
18	Against	বিরুদ্ধে	Preposition	www.yahoo.com	2013-04-22	12:04:33 AM	[Del]
19	afghan	আফগানিস্তানের আফগান	Noun	www.aol.com	2013-04-09	11:50:01 AM	[Del]

Fig 4. Learning words.

User can mark the word as learned after learning and the word further will be stored at the learned table. The learned word will not further be translated since the word is marked as learned. User is able enough to unmark the learned word for further translation. Fig. 4 illustrates the user learning process. Word will be stored with source website link, browsing time and date. Easy manageable interface contains the deleting and shorting facilities. English words are linked with other online dictionaries for more details.

### C. Comparison with other tools

Google website translator translates the whole page in Bangla. The conversion seems word by word translation only and not matures enough for Bangla Grammar. Since, the grammatical structure of the Bangla sentences is not the equivalent structure of the English sentences. Bangla Grammar implementation is very challenging, Google yet unable to implement the Bangla Grammar properly. Our project works in different way than Google or Bing. We show English Words with Bangla Words in Bracket as well as we have provided the English learning facilities that Google have not. English learning facilities during website browsing is a novel technique.

## V. CONCLUSION

Now-a-days internet is the easy accessible source of information for all aspects of life. Most of the websites are in English; and informative websites are very few in Bangla though it is a major language. This study investigated a technique and developed the system based on it to provide any English website with Bangla word meaning of the select English words based on users preferences. The system can be very handy to the mass people in Bangladesh to acquire huge of knowledge from English website in the internet. An additional benefit of the system is the English learning facility. The English learning facility through web browsing seems an interesting feature of the system.

Still there are some scopes to improve the system's performance. The developed system has some limitations: (i) Unable to convert and display some websites and web pages, (ii) Slow conversion and loading process, and (iii) Lack of Word leveling procedure etc. Bad website design i.e. using local link address seem the main reason not to convert the pages, since the converted page is stored at different server so the local link will not be reachable. A special technique is required for such badly designed websites. And local web

resource's (CSS file, JS/ jQuery file, Images etc.) link problem solving technique is still in immature stage and require more attention. There are also some scopes to work in this searching and replacement technique as it is seemed slows in some cases. Moreover, the service might be better if the dictionary is rich and defined word level more precisely.

## ACKNOWLEDGEMENT

The work is partially financed by the Ministry of Science and Technology (MoST), Government of the Peoples Republic of Bangladesh as a research grant for the financial year 2012-2013. The authors also acknowledge KUET authority for the technical support and to host the developed system in the server.

## REFERENCES

- [1] Rhaman, M.K.; Tarannum, N., "A Rule Based Approach for Implementation of Bangla to English Translation," Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on , vol., no., pp.13,18, 26-28 Nov. 2012 doi: 10.1109/ACSAT.2012.98
- [2] Hasan, M.S.; Mondal, A.; Saha, A., "A context free grammar and its predictive parser for bangla grammar recognition," Computer and Information Technology (ICCIT), 2010 13th International Conference on , vol., no., pp.87,91, 23-25 Dec. 2010 doi: 10.1109/ICCITECHN.2010.5723834
- [3] Mridha, M.F.; Huda, M.N.; Rahman, M.S.; Rahman, C.M., "Structure of Dictionary Entries of Bangla morphemes for morphological rule generation for Universal Networking Language," Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on , vol., no., pp.454,459, 8-10 Oct. 2010 doi: 10.1109/CISIM.2010.5643498
- [4] Ali, M.N.Y.; Al-Mamun, S.M.A.; Das, J.K.; Nurannabi, A.M., "Morphological analysis of Bangla words for Universal Networking Language," Digital Information Management, 2008. ICDIM 2008. Third International Conference on , vol., no., pp.532,537, 13-16 Nov. 2008 doi: 10.1109/ICDIM.2008.4746734
- [5] Word frequency data - Corpus of Contemporary American English Available: <http://www.wordfrequency.info/100k.asp>
- [6] Khan, N.; and Khan, M., "Developing a Computational Grammar for Bengali Using the HPSG Formalism", Center for Research on Bangla Language Processing, BRAC University, Dhaka, Bangladesh.
- [7] Dasgupta, S.; Khan, N.; Sarkar, A.I.; Pavel, D.S.H and Khan, M., "Morphological Analysis of Inflecting Compound Words in Bangla", <http://hdl.handle.net/10361/615>.
- [8] Sengupta, P.; Chaudhuri, B.B., "Morphological processing of Indian languages for lexical interaction with application to spelling error correction", Sadhana, Vol. 21, Part. 3, pp. 363-380, 1996.
- [9] Portal of Khulna University of Engineering & Technology Available: <http://portal.kuet.ac.bd/bbrowser/>.