

Multilingual Bengali Electronic Dictionary Using Sequential Search Technique

Md. Saiful Islam¹, Bipul Syam Purkayastha²

Research Scholar, Department of Computer Science, Assam University, Silchar, Assam, India¹

Professor, Department of Computer Science, Assam University, Silchar, Assam, India²

ABSTRACT: A language is one of the most important and effective modes of communication between the people belonging to different communities and cultures. The language acts as a bridge among us and helps in creating a bond among our cultures. Therefore, to learn mother language as well as other new languages is very important for us. The dictionary is one of the important tools that can be used for learning new languages. Electronic dictionaries are very popular nowadays and many users can be accessed it simultaneously on online. This paper describes the development of Bengali to Assamese, English and Hindi (B-AEH) Multilingual Electronic Dictionary using Sequential Search Technique. The B-AEH dictionary is a user friendly dictionary and user can easily look up the meaning of word and other related information of the word like word Id, POS, synonyms and examples from Bengali language to Assamese, English and Hindi languages on online. This dictionary will be beneficial for Bengali people as well as other people living in India.

KEYWORDS: Electronic Dictionary, Languages, Natural Language Processing, Sequential Search Technique.

I. INTRODUCTION

A. Electronic Dictionary

A dictionary is a very important component of any Natural Language Processing system. The word 'Dictionary' is derived from Latin word 'Dictionarium'. A dictionary is a book of words with one or more specific languages and the words are listed alphabetically with their meaning, synonyms, phonetics, POS, and examples [5][6]. It is one of the important tools to assist students in understanding as well as enlightening the skill of reading. There are mainly two types of dictionary, namely (1) *Paper Dictionary*: This dictionary is also known as hard or printed dictionary, and (2) *Electronic Dictionary*: This dictionary is also known as digital or Internet dictionary

An Electronic Dictionary (E-Dictionary) is one kind of dictionary whose data exists in digital form and can be accessed through a number of different media. The E-Dictionary is a very important and powerful tool for any person who is learning a new language using computer on both online and offline. It has the advantage of providing the user to access much larger database than a single book. The most important advantage of an E-Dictionary is that it is very convenient to use. In modern electronic form, electronic dictionaries have tremendous potential.

According to the languages involve, the dictionaries are found in three categories as below:

1. **Monolingual Dictionary:** In this dictionary, user can search the meaning of word and other related information of the word from one language to same language. English-English and Bengali-Bengali are some of the examples of monolingual dictionary.
2. **Bilingual Dictionary:** In this dictionary, user can search the meaning of word and other related information of the word from one language to another language. Bengali-English and Bengali-Assamese are some of the examples of bilingual dictionary.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

3. **Multilingual Dictionary:** In this dictionary, user can search the meanings of words and other related information of the words from one language to several languages. Bengali-Assamese, English and Hindi is the example of multilingual dictionary.

According to Al-Rabi'i, the E-Dictionary can be divided into two different types [5] as follows:

1. **Online E-Dictionary:** This type of dictionary is provided on the World Wide Web and is also known as Internet dictionary. It is directly used in digital form through the Internet using web browsers from anywhere place in the world. The advantage of this dictionary is that many users can be accessed it simultaneously on online [5].
2. **Offline E-Dictionary:** This type of dictionary can be used in digital computer, PDA and Mobile Phone. This dictionary is also known as portable digital dictionary. We can carry and backup it using CD, DVD, HD and pen drive. The advantage of this dictionary is that no need of Internet connection [5].

B. Natural Language Processing

Natural Language Processing (NLP) is a field of Computer Science and linguistics concerned with the interactions between computers and natural languages [4]. Some examples of natural languages are Assamese, Bengali, English, Hindi, and so on. The major goal of the NLP group is to design and build software that will analyze, understand, and generate languages that humans use naturally. NLP is an area of research and application that explores how computer can be used to understand and manipulate natural language text or speech to do useful things. Some of the most commonly researched tasks in NLP include Machine Translation, Electronic Dictionary, Morphological Segmentation, Natural Language Generation, Natural Language Understanding, Optical Character Recognition, Part of Speech (POS) Tagging, Parsing, Question Answering, Speech Recognition, Information Retrieval, and Speech Segmentation [6].

C. Language and Its Impact

The language can be considered as one of the most important ethno-cultural links between the people belonging to two different communities. No community can survive without a language. We use language to express our inner thoughts and emotions, to communicate with other people, and to establish rules and maintain our culture.

This paper describes about the four languages, namely Bengali, Assamese, English and Hindi which are used in B-AEH dictionary as follows:

1. **Bengali Language:** Bengali language is an Indo-Aryan language spoken mostly in the East Indian subcontinent. It is also known as Bangla language. It has evolved from the Magadhi Prakrit and Sanskrit language. The Bengali language is spoken mainly by the people of Indian states like West Bengal, Tripura and some people of Assam. It is the seventh most spoken language in the world and is the second most spoken language in India. The Bengali language is written using Bengali scripts and is the 6th most widely used writing system in the world. The script with minor variations is shared by Assamese and is the basis for the other languages like Manipuri and Bishnupriya Manipuri [6]. The Bengali language is one of the recognised languages of India. It is the official language of West Bengal and Tripura [15].
2. **Assamese Language:** Assamese is an Eastern Indo-Aryan language used mainly in the state of Assam. It is the state language as well as official language of Assam. The Assamese language is also known as Asamiya (Axomiya). It is the mother tongue/language of Assamese people. The Assamese language is spoken mainly by the people of Assam and by the some people of other North-Eastern states. Nearly 15 to 20 million people speak the Assamese language. It is one of the recognized languages of India [6][7]. The Assamese language is written using Assamese scripts that are developed from the Gupta alphabets around 1200 AD and which closely resemble with the Bengali alphabets.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

3. **English Language:** English is the West Germanic language that was first spoken in early medieval England and is now a global lingua franca. The English language is spoken mainly by the people of Canada, Australia, United Kingdom, United States, Ireland, and New Zealand. It is an official language of almost sixty (60) sovereign states. It is the third most common native language in the world. English was introduced in India in 1830 during the rule of the East India Company. At the time of Independence of India in 1947, English was the only functional lingua franca in the country. The Constitution of India (1951) declared Hindi as the primary official language and English as the associate official language of India [6].

4. **Hindi Language:** Hindi is the fourth most widely spoken language in the world. The Hindi language is spoken widely by the people of Indian states, including Delhi, Madhya Pradesh, Bihar, Uttar Pradesh, Chhattisgarh, Himachal Pradesh, Chandigarh, Haryana, and Rajasthan. It is the primary spoken language of Madhya Pradesh and Uttar Pradesh [6]. The Hindi language is also spoken in the other neighbouring countries of India, such as Bangladesh, Bhutan and Nepal. The Indian constitution, adopted in 1950, declared Hindi would be written in the Devanagari script and would be the primary official language of the Government of India.

D. Objectives

The primary objectives of this paper can be summarized as follows:

1. To develop user-friendly B-AEH dictionary where the user can easily look up the meaning of word and other related information of the word like word Id, lexical item (POS), synonyms and example from Bengali language to Assamese, English and Hindi languages.
2. To type Bengali words using English hard Keyboard as well as Virtual Keyboard from any computer on online.
3. To ensure B-AEH dictionary will be a valuable dictionary and to be beneficial for young learners, students, researcher scholars and teachers.
4. To improve the knowledge of Bengali, Assamese, English and Hindi languages basically for people of North-East India.

II. REVIEW OF RELATED LITERATURE

Lots of Bengali hard (or paper) dictionaries have been compiled by many lexicographers in different times. The first Bengali dictionary was compiled by Portuguese missionary Manuel da Assumpcao in 1734 [15]. There are also a very few numbers of Bengali-Assamese, Bengali-Hindi and Bengali-English paper dictionaries available in the market nowadays. In recent years, due to expansion of computer and Internet, there are a few number of Bengali-English, Bengali-Hindi and English-Bengali electronic dictionaries available on both offline and online [17]. The bilingual online electronic dictionaries like English-Bengali (WordNet) is developed by Dr. Pushpak Bhattacharyya, IIT Bombay [8]. The bilingual offline electronic dictionaries like English-Bengali and Bengali-Hindi are developed by C-DAC and TDIL for android.

Some examples of Bengali dictionaries with their lexicographer names are mentioned below:

- a. Samsad Bengali-English Dictionary, compiled by Sailendra Biswas in 2000 [12].
- b. The Modern Anglo-Bengali Dictionary, compiled by Charuchandra Guha in 2007 [14].
- c. Samsad English to Bengali Dictionary, compiled by Birendramohan Dasgupta in 2013 [12].
- d. The English-English-Bengali Dictionary, edited by Moitreyee Mitra and Dipendranath Mitra in 2013 [13].
- e. Oxford English-English-Bengali Dictionary, published by Oxford University Press in 2013[16].
- f. Concise English-Bengali Dictionary, compiled by Jyoti Bhusan Chaki in 2014 [1].

III. DATAFLOW DIAGRAM OF B-AEH DICTIONARY

A Data Flow Diagram (DFD) is a pictorial representation of information flows in a system. The DFD is often used as a preliminary step to create an overview of the system. It is an attractive technique because it provides what users do rather than what computers do. We have used two types of DFD for the development of B-AEH dictionary which are as below:

A. Level 0 DFD: The Level 0 DFD is also known as Context Diagram (CD). A CD is the most basic form of the DFD. Its aim to show how the entire system works at a glance. The CD demonstrates the interactions between the process and external entities. The CD of B-AEH dictionary is shown in fig.1.

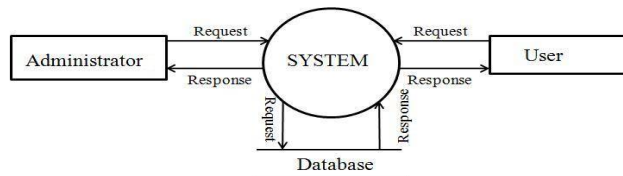


Fig.1: Context Diagram of B-AEH dictionary

B. Level 1 DFD: Level 1 DFD is the next level of CD that shows the overview of the full system of the B-AEH dictionary. Level 1 DFD is used to describe more details on how the data are processed and what type of data is needed in the system. The Level 1 DFD of the B-AEH dictionary is shown in fig. 2.

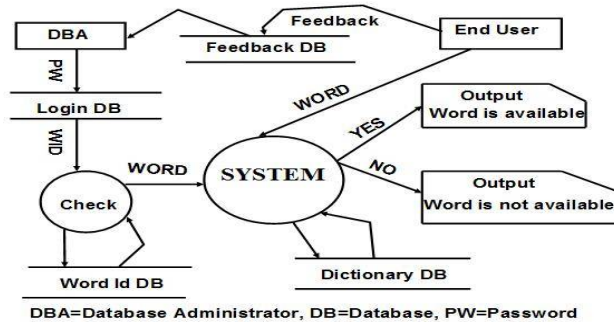


Fig. 2: Level 1 DFD of B-AEH dictionary

In Level 1 DFD, the Administrator and End-user are two external entities. The Administrator needs to login first; if the login is successful, then the Administrator can enter data into the B-AEH dictionary. The End-user can search the meaning of word.

IV. IMPLEMENTATION

The implementation part of B-AEH dictionary contains three phases which are discuss as below:

A. Software Requirements

We have used PHP, HTML, CSS and JavaScript as Front-End and MySQL as Back-End for the development of B-AEH dictionary [10][18].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

B. Data (or word) Entry

In B-AEH dictionary, only the Administrator can enter data (or word). The Administrator needs to login first with proper username and password. If the login is successful, then he/she can be able to enter words into the dictionary based on the following word entry algorithm.

1. Enter word Id

If (found)

```
{  
  Print- word Id already exists in the dictionary;  
  Stop  
}
```

Else

Go to next step;

2. Search headword with its POS

If (found)

```
{  
  Print- headword already exists in the dictionary;  
  Stop  
}
```

Else

Go to next step;

3. Enter new word Id, headword and other related information of the headword (POS, synonyms and examples) of Bengali, Assamese, English and Hindi languages.

4. Submit.

According to this algorithm, suppose, an Administrator wants to enter a word (headword) into this dictionary. The Administrator needs to check desired word Id for the headword first. If the word Id is not available in the dictionary, then the Administrator needs to also check the headword with its POS in the dictionary. If the headword and its corresponding POS are not available in the system, then the Administrator can enter the desired word Id, the headword and other related information of the word like word meaning, POS, synonyms and example in the dictionary.

C. Word Search (or look up)

There are many word search techniques available for E-Dictionary. We have used Sequential Search Technique to look up (or search) the meaning of the word quickly and easily in B-AEH dictionary. Sequential Search Technique (SST) is the simplest and most popular word search technique for electronic dictionaries. It is a very useful and efficient technique to look up the words easily and quickly. If we want to search a particular word in a database table using SST, then the SST checks each word one by one in sequence until the desired word is found in the table. It starts to compare with each word from the beginning of the database table. In SST, the database table need not be sorted. The average number of comparisons in SST is $(N+1)/2$, where N is the size of the row in the table. Its worst case cost is proportional to the number of elements in the list. The searching time for SST is $O(n)$ [9][11].

1. **Architecture of SST:** Suppose, a user wants to look up (or search) a Bengali word **ভালো** from the database table of B-AEH dictionary using SST, then the SST will work as below fig. 3.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016



Word Id	Assamese	Bengali	English	Hindi	POS
1001	কাম	কাজ	Work	काम	Verb
1002	ভাল	ভালো	Good	अच्छा	Adjective
1003	তেওঁ	তিনি	He	वह	Pronoun

Fig. 3: Architecture of SST

In this database table, the given word will compare with each word existing in Bengali field (or column) one by one in sequence until the desired word is found in the table. If the given word is found in the table, then the given word and other related information of the word would be displayed. Otherwise, the word is not available in the database table.

2. **Algorithm of SST:** An algorithm is a self contained step by step set of operations to be performed. The algorithm of SST in the B-AEH dictionary is shown as below:

- Step 1: Initialize sarray, kword, len;
- Step 2: Initialize pos=0;
- Step 3: Repeat step 4 until pos<=len
- Step 4: if (sarray[pos])==kword)
return pos (Print kword is found);
else pos=pos+1;
- Step 5: if (pos>len)
Print kword is not found;
- Step 6: Stop

In this algorithm,

- sarray= Search array (specific column or field of language)
- kword=Keyword (headword or root word or searching word)
- len=Length (number of words in specific language field of database table)
- pos=Position (position of word in a specific language field of database table)

V. RESULT AND DISCUSSION

A. B-AEH Dictionary

We have developed Bengali to Assamese, English and Hindi (B-AEH) multilingual online electronic dictionary. In this dictionary, user can search the meaning of Bengali word in corresponding Assamese, English and Hindi words. The user can also find word Id, word class (POS), synonyms and example of the headword (or search word) in Bengali, Assamese, English and Hindi languages. One example of Bengali word and its corresponding meanings in Assamese, English and Hindi words is shown in fig. 4.

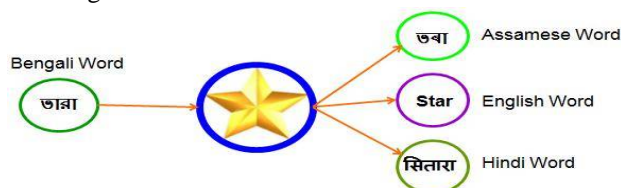


Fig.4: Example of word meaning in B-AEH dictionary

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

B. User Interface of B-AEH Dictionary

The following figure (Fig. 5) is the output snapshot of user interface of B-AEH dictionary. In B-AEH dictionary, user can type Bengali word in the text box using English hard Keyboard as well as Virtual Keyboard from any computer on online. After typing the desired search word, the user can press the Enter key or Click on Look Up button, then the Bengali word and its corresponding meaning in Assamese, English and Hindi words and other related information of the headword will be displayed on the web page as shown in fig. 5.

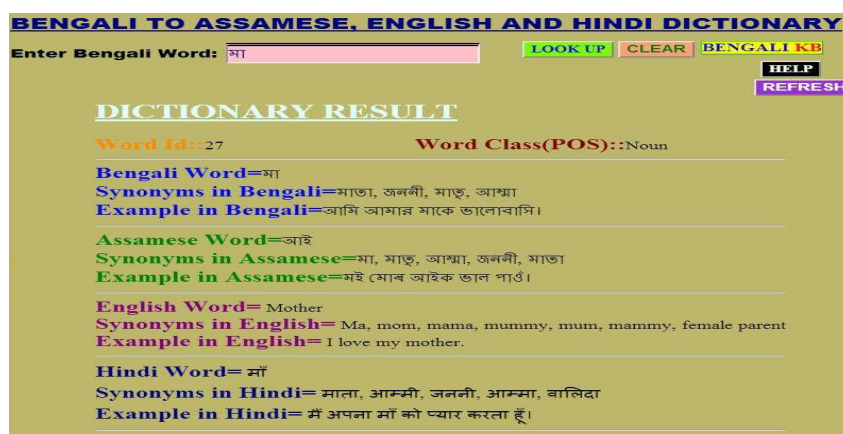


Fig.5: Snapshot of web page of B-AEH dictionary

C. Keyboard

The Keyboard is the most essential component of E-Dictionary. We have developed two types of Keyboard using Unicode (UTF8) for Bengali language in B-AEH dictionary which are: Bengali General Keyboard and Bengali Virtual Keyboard

1. Bengali General Keyboard: The Bengali general Keyboard is used to type Bengali letters, words and sentences in the B-AEH dictionary with the help of English hard Keyboard. When a user wants to type Bengali words in text box using an English hard Keyboard, then the English hard Keyboard will automatically convert into Bengali general Keyboard. The typing instructions of Bengali words using English hard Keyboard are given in online B-AEH dictionary (at HELP button).

2. Bengali Virtual Keyboard: The Bengali Virtual Keyboard is used to type Bengali letters, words and sentences in the B-AEH dictionary. The snapshot of Bengali Virtual Keyboard is shown in fig. 6.



Fig. 6: Snapshot of Bengali Virtual Keyboard

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

VI. CONCLUSIONS

The paper entitled “Multilingual Bengali Electronic Dictionary using Sequential Search Technique” is carried out to look up the meanings of words and other related information of the words from Bengali language to Assamese, English and Hindi languages on online. This dictionary is developed in such a way that it makes the concepts of knowledge and provides to look up the accurate meaning of the Bengali words in corresponding Assamese, English and Hindi words. It plays an important role to improve the knowledge of Assamese, Bengali, English and Hindi languages. This dictionary will save time and money in comparison to traditional paper dictionaries. It will be beneficial for students, research scholars, teachers, travelers and other people. The B-AEH dictionary is developed both for people of North-East India as well as for other people living in other parts of India. Since, India is a multilingual country, therefore we hope, B-AEH dictionary will be helpful for Indian people and for the educational institutions.

REFERENCES

- [1] Jyoti Bhusan Chaki, “ English to Bengali Dictionary”, 2014.
- [2] Dr. Hiren Gohain, “Anglo-Assamese (English- Assamese) Dictionary, 2005”.
- [3] Dr. Hardev Bahri, “English-Hindi Dictionary”, 2010.
- [4] Ela Kumar, “ Natural Language Processing”, 2011.
- [5] Al-Rabi'i, S. M. et. al. (2001), “Al-Mukjam al-Syamil li Mustalahaat al-Hasib al-Aali wa al- Internet [A complete dictionary for terminologies in computers and internet]. Riyadh: Maktabah al Abikan”.
- [6] Saiful Islam and Bipul Syam Purkayastha, “Development of Multilingual Assamese Electronic Dictionary”, 2015.
- [7] Shikhar Kr. Sarma, Himadri Bharali, Mayashree Mahanta, Utpal Saikia, Dibyajyoti Sarmah, “An Analytical Study of Synonymy in Assamese Language Using WorldNet: Classification and Structure”.
- [8] Dr. Pushpak Bhattacharyya, “An Indo WordNet (A WordNet of Indian Languages)”, IIT Bombay.
- [9] Robert Lew, “Online dictionary skills”, Adam Mickiewicz University”, 2013.
- [10] Chris Lea, “PHP MySQL Website Programming: Problem - Design – Solution”, Wrox Press Ltd. Paperback- 24 March, 2003.
- [11] Paul M. Dorfman, “Array Look up Techniques: From Sequential Search to Key Indexing, Citibank Universal Card Services”.
- [12] URL:[http://www.amazon.in/Samsad-Bengali-English-Dictionary-Sailendra- Biswas/dp/8186806865](http://www.amazon.in/Samsad-Bengali-English-Dictionary-Sailendra-Biswas/dp/8186806865).
- [13] URL:[http://www.amazon.in/English-English-Bengali-Dictionary-Moitreyee- Mitra/dp/019568964X](http://www.amazon.in/English-English-Bengali-Dictionary-Moitreyee-Mitra/dp/019568964X).
- [14] URL:[http://www.amazon.com/The-Modern-Anglo-Bengali-Dictionary Volumes/dp](http://www.amazon.com/The-Modern-Anglo-Bengali-Dictionary-Volumes/dp).
- [15] URL: https://en.wikipedia.org/wiki/Bengali_language.
- [16] URL: <http://dictionary.tamilcube.com/bengali-dictionary.aspx>
- [17] URL: <http://translate.google.co.in>
- [18] URL: [https:// www.w3schools.com](https://www.w3schools.com)