AN ENGLISH TO ASSAMESE, BENGALI AND HINDI MULTILINGUAL E-DICTIONARY

Md. Saiful Islam
Department of Computer Science
Assam University, Silchar, Assam, India
E-mail:sislam.mca@gmail.com

Abstract

Dictionary is a very demandable components of Natural Language Processing system nowadays. A dictionary is one of the important tools that can be used for learning new languages. A word is basically an association of linguistic sound and meaning. The spelling does not always easily correlate with the sound of a word. A dictionary helps us both with the spelling and pronunciation of such words. Electronic dictionaries are very popular nowadays. It can be accessed by many users simultaneously on online. The main objective of this paper is to develop an English to Assamese, Bengali and Hindi (E-ABH) multilingual electronic dictionary in such a way that it is 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 English to Assamese, Bengali and Hindi languages. This dictionary will be beneficial and must be improved the knowledge of Assamese, Bengali, English and Hindi languages basically for people of North-East India.

Keywords: **Electronic Dictionary, Languages, Natural Language Processing, Sequential Search Technique**

I. INTRODUCTION

A. Electronic Dictionary

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 two types of dictionary, namely Paper dictionary which is also known as hard or printed dictionary and Electronic dictionary which is also known as digital or Internet dictionary.

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:** Here, 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:** Here, user can search the meaning of word and other related

information of the word from one language to another language. Assamese-English and English-Bengali are some of the examples of bilingual dictionary.

3. **Multilingual Dictionary:** Here, user can search the meanings of words and other related information of the words from one language to several languages. English-Assamese, Bengali and Hindi is an 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 dictionary is directly used in digital form through Internet using web browsers from anywhere place in the world. It is also known as Internet dictionary. Many users can be accessed it simultaneously on online.

2. **Offline E-Dictionary:** This dictionary can be used in digital computer, PDA (Personal Data Assistant), and mobile phone. It is also known as portable digital dictionary. We can carry and backup Offline E-Dictionary using CD, DVD, HD and pen drive. We can also download this type of dictionary from Internet and can be installed in our own computer or other devices.

B. **Natural Language Processing**
Natural languages are most commonly used by humans for communication purposes naturally. Natural Language Processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and natural languages[4]. NLP deals with computer programs to understand human languages both in written and oral form. 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 common research tasks in NLP are Machine Translation, Electronic Dictionary, Morphological Segmentation, Natural Language Generation, Optical Character Recognition, Part of Speech (POS) Tagging, Question Answering, Speech Recognition, Information Retrieval (IR), and Speech Segmentation[6].

C. **Languages**
In this section, we discuss briefly about the Assamese, Bengali, English and Hindi languages as follows:

1. 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. 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. Assamese is one of the recognized languages of India [6][7]. It is evolved in the 7th century AD having its roots from the Sanskrit language. However, its vocabulary, phonology and grammar have been substantially influenced by the original inhabitants of Assam, such as the Boros and the Kacharis. Assamese script is derived from Brahmi script. The Assamese language is written using Assamese scripts that are developed from the Gupta alphabets around 1200 AD and which closely resemble the Mithilakshar and Bengali alphabets.

2. 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. Bengali is one of the recognised languages of India. It is the official language of West Bengal and Tripura. It is also a major language in the Indian Union Territory of Andaman and Nicobar Islands. The Bengali is mainly spoken by the people of Indian states like West Bengal, Tripura and Assam. It is the seventh most spoken language in the world and 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].

3. English Language: English is the West Germanic language that was first spoken in early medieval England. English 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 sovereign states. It is the third most common native language in the world. It has become the leading language of international discourse [6]. 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 English as the associate official language of India. It has various dialects in India due to the influence of local languages.

4. Hindi Language: Hindi is the fourth most widely spoken language in the world. It is spoken widely by the people of Indian states like Delhi, Madhya Pradesh, Bihar, Uttar Pradesh, Chhattisgarh, Haryana. Himachal Pradesh, Chandigarh, and Rajasthan. It is the primary spoken language of Madhya Pradesh and Uttar Pradesh [6]. In the 2001 census of India, 258 million people is reported Hindi to be their native language. Hindi is also spoken in the other neighbouring countries of India, such as Bangladesh, Bhutan and Nepal. Hindi derives its vocabulary from several major sources like Sanskrit, Persian and Arabic.

II. REVIEW OF RELATED LITERATURE

Lots of English paper dictionaries have been compiled by many lexicographers in different times. The first English dictionary was compiled by Robert Cawdrey in 1604 [17]. It contains about 2,543 words. The first electronic version of Oxford English Dictionary (OED) was made available in 1988 [14]. The digital OED was developed by Tony Smith and published by Oxford University Press in 1999. The online version of OED has been available since 2000. Presently, there are many English-Assamese[1], English-Bengali[2], English-Hindi[3] and English-English paper dictionaries available in market. There are also a few number of English-Assamese, English-Bengali[8][19], English-Hindi and English-English electronic dictionaries available on both online and offline nowadays.
Some examples of English dictionaries with their lexicographer names are mentioned as below:
 a. A Dictionary of the English Language, compiled by Samuel Johnson in 1755 [14].
 b. The Oxford English Dictionary, published by Oxford University Press in 1989.

 c. The Compact Oxford English Dictionary, edited by J. A. Simpson and E. S. C. Weiner in 1991[15].
 d. The Oxford Dictionary of Current English, compiled by Catherine Soanes in 2006.
 e. The Concise Oxford English Dictionary, edited by Angus Stevenson and Maurice Waite in 2011 [16].

III. DATAFLOW DIAGRAM OF E-ABH 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 [12]. It is an attractive technique because it provides what users do rather than what computers do. The DFD technique is very popular, because it is very simple to understand and use. We have used two types of DFD to implement the E-ABH 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. It aims to show how the entire system works at a glance. CD demonstrates the interactions between the process and external entities. The CD of E-ABH dictionary is shown in figure1.
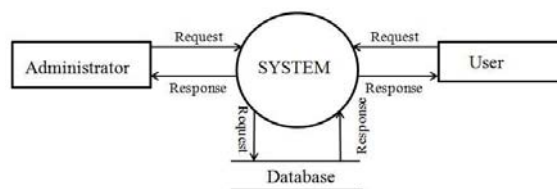

Fig.1: Context Diagram of E-ABH dictionary

In CD, the Administrator and User are two external entities. The Administrator can enter data into the database of the system, whereas the User can search data from the database of the system.

B. Level 1 DFD
Level 1 DFD is the next level of CD that shows the overview of the full system of the E-ABH dictionary. It is used to describe more details on

how the data are processed and what type of data is needed in the system. Level 1 DFD of the E-ABH dictionary is shown in figure 2.
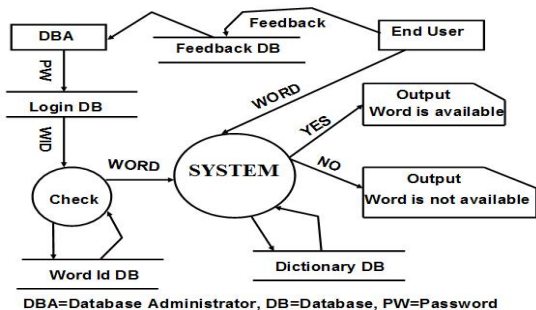

Fig. 2: Level 1 DFD of E-ABH dictionary

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 E-ABH dictionary. The End-user can search the meaning of word. In addition, the End-user can also give feedback to the Administrator about the performance of the E-ABH dictionary.

IV. IMPLEMENTATION

The implementation part of E-ABH dictionary contains three phases which are:

A. Necessary Software

We have used PHP, HTML, CSS and JavaScript as Front-End and MySQL as Back-End for the development of E-ABH dictionary [10][11][20].

B. Data (or word) Entry

In E-ABH 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;
2. Enter new word Id, headword and other related information of the headword (POS, synonyms and examples) of Assamese, Bengali, English and Hindi languages.

3. 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 lots of 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 E-ABH 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][13]..

1. Architecture of SST: Suppose, a user wants to look up (or search) an English word "Good" from the database table of E-ABH dictionary using SST, then the SST will work as below figure 3.

Fig. 3: Architecture of SST

In this table, the given word will compare with each word existing in English 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 E-ABH 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)

3. Flowchart of SST: A flowchart is a graphical representation of an algorithm. The flowchart is used to easily understand and implement the algorithm of SST. The flowchart of the SST algorithm in E-ABH dictionary is shown in figure 4.
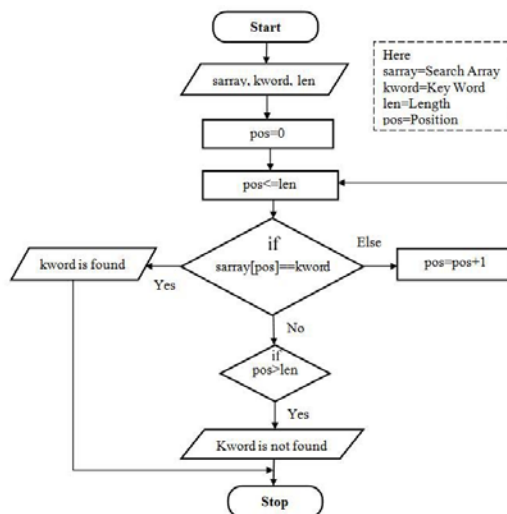

Fig. 4: Flowchart of SST

4. *Advantages of SST:* The advantages of Sequential Search Technique are as follows:

a. The primary advantage of SST is its simplicity.
b. The SST is very simple to implement.
c. It is very easy to understand and is straightforward.
d. It provides good performance in small database.
e. In SST, database table need not be sorted as ascending or descending ordered.

V. RESULT AND DISCUSSION

A. E-ABH Dictionary

We have developed English to Assamese, Bengali, English and Hindi multilingual electronic dictionary. In E-ABH dictionary, user can search the meaning of English word in corresponding Assamese, Bengali and Hindi words. The user can also find word Id, word class (POS), synonyms and example of the headword (or search word) in English, Assamese, Bengali and Hindi languages. One example of English word and its corresponding meanings in Assamese, English and Hindi words is shown in figure 5.

Fig.5: Example of word meaning in E-ABH dictionary

B. **User Interface of E-ABH Dictionary**

The following figure (Fig. 6) is the output snapshot of user interface of E-ABH dictionary. In E-ABH dictionary, user can type English word in the text box using English hard Keyboard as well as English Virtual Keyboard from any computer. After typing the desired search word, the user can press the Enter key or Click on Look Up button, then the English word and its corresponding meaning in Assamese, Bengali and Hindi words and other related information of the headword will be displayed on the web page as shown in figure 6.
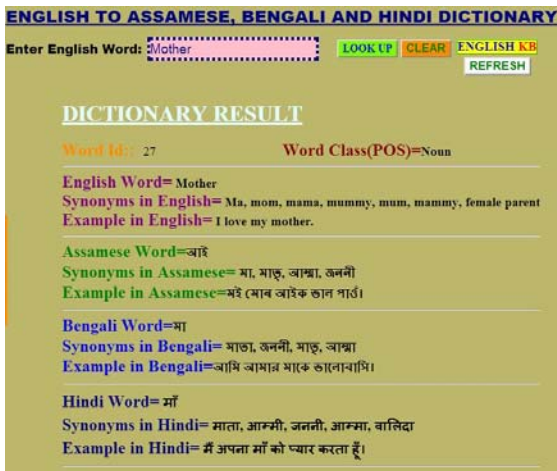

Fig.6: Snapshot of web page of E-ABH dictionary

C. **Keyboard**

Keyboard is the most essential component of E-Dictionary. We have developed Unicode (UTF 8) based Virtual Keyboard for English language in E-ABH dictionary. The snapshot of English Virtual Keyboard is shown in figure 7.

Fig. 7: Snapshot of English Virtual Keyboard

VI. **CONCLUSIONS**

The paper entitled "An English to Assamese, Bengali and Hindi Multilingual Electronic Dictionary" is carried out to look up the meanings of words and other related information of the words from English to Assamese, Bengali and Hindi languages. 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 English words in corresponding Assamese, Bengali 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 E-ABH 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, the E-ABH multilingual electronic dictionary will be helpful for Indian people.

REFERENCES
[1] Dr. Hiren Gohain, "Anglo-Assamese (English- Assamese) Dictionary", 2005.
[2] Jyoti Bhusan Chaki, " English to Bengali Dictionary", 2014.
[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] Elmasri & Navathe, "Fundamentals of Database Systems", 2010.

[12] Donald S. Le Vie Jr, "Understanding Data Flow Diagrams".

[13] Paul M. Dorfman, "Array Look up Techniques: From Sequential Search to Key Indexing, Citibank Universal Card Services".

[14] URL:http://www.bl.uk/learning/langlit/dic/johnson/1755johnsonsdictionary.html

[15] URL: http://ukcatalogue.oup.com/product/9780198612582.do

[16] URL:https://en.wikipedia.org/wiki/Concise_Oxford_English_Dictionary.

[17] URL:http://press.uchicago.edu/ucp/books/book/distributed/F/bo5456677.html

[18] URL: http://wordnet.princeton.edu

[19] URL: http://translate.google.co.in

[20] URL: https:// www.w3schools.com