Implementation of an Assamese-English Bilingual E-Dictionary

Saiful Islam1, Bipul Syam Purkayastha2

1M.Phil, Research Scholar, Assam University, Silchar, Assam, India.
2Professor, Department of Computer Science, Assam University, Silchar, Assam, India.

Abstract—Assamese is the state language of Assam and English is the associate language of India. A Dictionary is one of the most important tools for language learning, in education and to know the relative information in everyday life. An Electronic Dictionary (E-Dictionary) is a powerful tool that enables students to improve their learning ability and achievements. An Assamese-English Bilingual Electronic Dictionary (AEBED) is an important part of Natural Language Processing tasks and linguistic work. This paper describes the implementation of an Assamese-English Bilingual Electronic (Online) Dictionary. The AEBED is user friendly and users can easily search or look up the meaning of Assamese words and relative information into English words and similarly English words to Assamese words. This dictionary will be beneficial for Assamese people as well as other people living in India.

Keywords—Assamese Language, Electronic Dictionary, English Language, Index Based Search Technique, Natural Language Processing.

I. INTRODUCTION

A. Natural Language Processing: Natural Languages are the languages which have naturally evolved and used by human beings for communication purposes. The some examples of Natural Languages are Assamese, Bengali, English and Hindi. Natural Language Processing (NLP) is the scientific study of languages from computational perspective. NLP is a field of Computer science, artificial intelligence and linguistics concerned with the interactions between Computers and Natural languages. NLP is a very attractive method of human-computer interaction. Natural language processing is the ability of a Computer program to understand human speech as it is spoken. The goal of the Natural Language Processing group is to design and build software that will analyze, understand, and generate languages that humans use naturally. Application areas within NLP include automatic (Machine) translation between languages, Natural language generation, Natural language understanding, Optical Character Recognition (OCR), Part-of-Speech Tagging (POST), Parsing, Speech Recognition (SR), Speech Processing (SP), Information Retrieval (IR), and Speech Segmentation (SS).

B. Assamese Language: Assamese is an Eastern Indo-Aryan language used mainly in the state of Assam. Assamese language is the state and official language of Assam and spoken by nearly 30 millions of people in India. Assamese Language is written with Assamese script. Assamese language is also known as Asamiya (অসমীয়া Oxomiya). Assamese language is also spoken by the some other people of the North-East states. The Assamese language has been declared as one of the recognized languages of the Republic of India by the Indian Constitution. Assamese script is derived from the Brahmi script. It played a vital role in the evolution of the Indian script. There are 52 alphabets in Assamese language, among them 11 alphabets are vowels and 41 alphabets are consonants. In this paper, we have mentioned some examples of Assamese words and sentences as below

Example of Assamese Words are as below

মহ, উপকাব, মবম, ধুনীয়া, ইতাদি।

Example of Assamese Sentences are as below

মহ অসমীয়া ভাষা তোলকে কব পাবে।

সিহতে ফুটবল খেলি আছে।

C. English Language: English is a West Germanic language that was first spoken in early medieval England and is now a global lingua franca. It is spoken as a first language by the majority populations of several sovereign states, including the United Kingdom, the United States, Canada, Australia, Ireland, New Zealand and a number of Caribbean nations. It is an official language of almost 60 sovereign states. It is the third most common native language in the world.

Indian English is the group of English dialects or regional language varieties, spoken primarily in the Indian subcontinent.
English language public instruction began 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. In 1951, the constitution of India declared Hindi is the official language and English is the associate official language of the new republic. English is the official language of the state Arunachal Pradesh, Nagaland and Meghalaya in North-East. There are 26 alphabets in English language, among them 5 alphabets are vowels and 21 alphabets are consonants. In this paper, we have mentioned some examples of English words and sentences as below:

Examples of English words are as below

I know English language.

They are playing cricket.

She loves music.

D. Electronic Dictionary: A Dictionary is a collection of words in one or more specific languages, often listed alphabetically with usage of information, definitions, etymologies, phonetics, pronunciations, translation, and other information. An Electronic Dictionary (E-Dictionary) is a very important and powerful tool for any person who is learning a new language on online. Electronic dictionaries have the advantage of providing the user with almost instant access to a database much larger than a single book. The most important advantage of an E-Dictionary is that it is very convenient to use. Whenever you meet new words or expressions, you can know the meaning quickly. You don’t need to waste much time turning the page seeking for the new words. The use of Electronic dictionaries has many advantages over the traditional paper dictionary. A Bilingual Electronic Dictionary (BED) is a specialized dictionary used to translate words or phrases from one language to another language. A BED can be unidirectional, meaning that the words of one language can be translate in another language or can be bidirectional meaning that allowing translation to and from both languages.

The Electronic Dictionary can be divided into two different types as below:

1. An Online Electronic Dictionary: – This Dictionary is provided on the World Wide Web also known as Internet Dictionary. This Dictionary is directly used in digital form through the Internet using web browsers from any where place in the world.

2. An Offline Electronic Dictionary: – This type of Dictionary can be used in a Computer or a Personal Data Assistant (PDA). This Dictionary is also known as Digital Dictionary (DD). We can carry and backup DD using Compact Disk (CD), Digital Video Disk (DVD), Hard Disk (HD) and Pen Drive.

II. LITERATURE REVIEW

There are many hardcopy or paper Assamese Dictionary developed by many Authors. Miles Bronson, an American missionary was the first to compile a Dictionary of the Assamese language. His Dictionary published in 1867 at American Baptist Mission Press Sibsagar is out of print now. The first Anglo-Assamese Dictionary was compiled by a student of Cotton College in the year 1910. Makhan Lal Chaliha of Chiring Chapori, which was found in the British Library by a researcher of the Jatiya Sikh Samanay Parishad. The third Assamese Dictionary Chandrakanta Abhidhan, a comprehensive Bilingual Dictionary with words and their meanings in Assamese and English, originally compiled and published by Assam Sahitya Sabha in 1933, 32 years after the publication of the Hem Kosh.

In recent year a wealth of digital Dictionaries have appeared because of expansion of the Internet and new media technologies. Some Authors developed bilingual electronic English to Assamese and Assamese to English Dictionary but most of them are Offline Electronic Dictionary or Digital Dictionary.

III. OBJECTIVE OF THIS PAPER

The main objective of this paper is to implement an Assamese-English Bilingual Online Electronic Dictionary to translate Assamese words to and from English words from any where place in the world.

The Assamese-English Bilingual Electronic Dictionary (AEBED) focuses on the following points:

1. To develop AEBED that will be user friendly and Students or People can easily search or look up the meaning of Assamese words to English words and English words to Assamese words within a short period of time through internet.

2. To use Word Id, Lexical items, Synonyms and Examples of Assamese and English words for easily understand the words.

3. To directly type Assamese words in AEBED from any Computer using General Keyboard or Virtual Keyboard.
4. To ensure AEBED is a valuable Dictionary of Assamese and English language and to be beneficial for both young learners and scholars.

IV. DATA FLOW DIAGRAM OF AEBED

Data flow diagram (DFD) represents the flows of data among different processes in a system. It is a graphical technique that shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated. The DFD can also be used for the visualization of data processing (structured design). The DFDs are attractive technique because they provide what users do rather than what computers do. A DFD is intended to serve as a communication tool among system analysts, end users, database designer and system programmers. DFD is also called bubble chart. The DFD technique is so popular because it is simple to understand and use. The DFD for implementation of Assamese-English Bilingual Electronic Dictionary (AEBRD) is as below

V. METHODOLOGY USED IN AEBED

A. Introduction: In this system, we used Index Based Search Technique (IBST). An index is a data structure that stores the values for a specific column in a table. A database index is a data structure that improves the speed of data retrieval operations on a database table. Indexes are used to quickly locate data without having to search every row in a database table. An index improves the performance of lookup, because an index is basically used to store column values, looking up those values becomes much faster. A very good analogy is to think of a database index as an index in a book.

In Index Based Search Technique, we used basically Index and Sequential (Index Sequential) search technique. Indexed sequential search is very useful where records are often retrieved randomly and is also processed in key order. Index sequential search is simple and it is primary used for unordered list of data in database table.

B. Architecture of IBST: Suppose, we want to search an Assamese word “кал” from the following database table “Edictionary” using this technique, then the architecture of IBST in this Dictionary is as below

If the given word is found in the table then its corresponding words in English will be display, otherwise the given word is not available in the database table or dictionary.

C. Algorithm of IBST: The algorithm of IBST in this Dictionary is as below

Step1: Initialize table_row, searchword, length.
Step2: Initialize index=0.
Step3: Repeat step 4 till index<=length.
Step4: if (table_row[index]==searchword) Searchword is found
Else Increment index by 1.
Step5: Stop

D. Advantages and Disadvantages of IBST of this dictionary: The advantages of IBST are as below

1. The use of indexes in SQL queries usually the results are in much better performance and allow records to be accessed directly or sequentially.
2. Quite easy to process and Index can be used to sorting the data in database table.
3. With proper selection of a key field, records in the database table can be searched and accessed in very quickly.
The Disadvantages of IBST are as below
1. Indexes increase the disk space in database table.
2. Indexes decrease the performance of Data Manipulation Languages (i.e. Insert, Update and Delete).
3. Index sequential search is very ineffective in some cases, especially for large database.

VI. EXPERIMENTAL RESULT AND SNAPSHOT

Using IBST, we have developed the following dictionaries
A. Assamese to English Electronic Dictionary
B. English to Assamese Electronic Dictionary

A. Assamese to English Electronic Dictionary: In this Dictionary, user can search or look up the meaning of any Assamese word into English word. The users also get other information like Word ID, Word Class (Lexical items or POS), Synonyms and Example of both languages. In this Dictionary, user can directly types Assamese word in the search text box using General keyboard or Virtual keyboard as shown in Fig.2. We have given here, an example of how we can search an Assamese word and its corresponding meaning in English word and other relative information with snapshot as shown below in Fig.1

B. English to Assamese Electronic Dictionary: In this Dictionary, user can search or look up the meaning of any English word into Assamese word. The users also get other information like Word ID, Word Class (Lexical items or POS), Synonyms and Example of both languages. In this dictionary, user can directly types English word in the search text box using General keyboard or Virtual keyboard as shown in Fig.4. We have given here, an example of how we can search an English word and its corresponding word in Assamese word and other relative information with snapshot as shown below in Fig.3

VII. CONCLUSION

Dictionary plays an important role in learning a new language and enlightening the skill of reading. An Electronic Dictionary is a powerful tool that enables students to improve their learning ability and achievements. The use of dictionary does not only improve students’ proficiency in second language but also helps to increase the command of their mother language.
The Assamese-English Bilingual Electronic Dictionary (AEBED) is one kind of Dictionary that people directly can type Assamese and English words using General Keyboard and will get the information about Word Id, Lexical items Words, their meaning, Synonyms, Examples and other relative information from Assamese language to English language and English Language to Assamese Language through Online (Internet). This paper must be beneficial basically for North-East students and also other people.

REFERENCES

Author’s Profile
Saiful Islam, M.Phil, Research Scholar, Department of Computer Science, Assam University, Silchar, Assam, India.

Prof. Bipul Syam Purkayastha, Head of the Department Computer Science, Dean of Physical Sciences, Assam University, Silchar, Assam, India. Specialization: Computational Linear Algebra, Computer Programming, Natural Language Processing, Internet Programming, Website Designing.