

Word Search in a WWW Manipuri-English Electronic Dictionary

Sagolsem Poireiton Meitei
Department of Computer Science
Assam University, Silchar

Shantikumar Ningombam
Department of Computer Science
Assam University, Silchar

Bipul Syam Purkayastha
Department of Computer Science
Assam University, Silchar

ABSTRACT

The dictionary is one of the important tools for language learning in everyday life, education and learning of many things. With recent developments of information technologies a wide range of new software tools been developed, especially for natural language processing. This paper presents an online bilingual dictionary as a technological tool, and describes a Manipuri-English online dictionary. It focuses especially on how word is search using this dictionary. The word search is analysed into three categories namely (1) simple word search, (2) wild card search and (3) search by lexical item. This paper also describes about the development and database design of the electronic dictionary.

Keywords:

Manipuri, Tibeto-Burman, Natural language Processing

1. INTRODUCTION

Electronic dictionaries (EDRs) have been seen as likely a source of information in NLP (Natural Language Processing) because they contain an enormous amount of lexical and semantic knowledge collected together over years of effort by lexicographers. The electronic dictionary is not simply a machine readable dictionary; it is a dictionary containing all the information necessary for computers to understand natural language. Both people and computers must know the meaning of word to understand natural language text, the meaning of words, i.e. the concepts expressed by words, the grammatical characteristics of words when they express concepts, and knowledge necessary for understanding concepts, is contained in the electronic dictionary. Early results in EDRs research were promising and led many to feel that large knowledge bases could easily be derived automatically from EDRs. EDR research has contributed significantly to our understanding of the nature, kinds, and role of semantic information in the processing of natural languages. It has also led to an understanding and formal description of the structure of dictionaries and lexical information in general. In addition, the analysis of the information in dictionaries from the point of view of what is needed for NLP has contributed to lexicographers' own understanding of dictionaries and the dictionary-making process. Some electronic dictionaries contain only corpora for a single language (monolingual), but others are bilingual dictionaries and translation dictionaries and may also include medical or legal dictionaries, thesauri, travel dictionaries, dictionaries of idioms, a guide to pronunciation, a grammar reference, common phrases and collocations, and a dictionary of foreign loan words.

2. MANIPUR LANGUAGE

Manipuri, locally known as Meiteilon (the Meitei+lon language), is spoken basically in the state of Manipur which is in the north eastern India. It is also spoken in Assam, Tripura, Mizoram, Burma and Bangladesh. Manipuri is the only medium of com-

munication among 29 different mother tongues. Hence it is regarded as lingua franca of Manipur state. Since 20 August 1992 Manipuri become the first TB (Tibeto Burman) language to receive recognition as a schedule VIII language of India. Before this every date, Manipuri has adopted as the medium of instruction and examination. More and above, Manipuri is taught up to Ph.D. level in the Indian Universities. It is also the medium of education in Manipuri Schools. This language has been a major adhesive in the integration of all ethnic groups of Manipur. In the classification of Sino-Tibetan family, Manipuri is placed in Kuki-Chin Proper separated by Grierson-Konow (1903-1928), in Meitei branch of Kukish Section by Safer (1955, 1966, 1974). Manipuri share genetic features of TB (Tibeto Burman) features including phonemic tones, SOV (Subject Object Verb) word order, agglutinative verb morphology and tendency to reduce disyllabic forms to monosyllabic ones (Delancy 1987)[4]. Grierson (1903)[5] in his LSI, Manipuri is briefly described: vowels, consonants, word categories syllable, sentence type etc. however neither of this word is exhaustive in its description. Meitei had its original script named Meitei-Mayek which was in use up to the 18th century. However, later the Eastern Nagari or Bengali script adopted for scripting Manipuri. It is also the script used till now. However, efforts to revive the Meitei-Mayek are still going on.

3. STRUCTURE OF THE LANGUAGE

The writing style of Manipuri Language is from left to right. SOV (Subject Object Verb) word order is obeyed in Manipuri Language. There are two types of roots-the free roots and bound roots. All nouns (except the kinship terms), in Manipuri, are free roots. There is also no grammatical gender in Manipuri, the human and animal. Meitei had its original script named Meitei-Mayek which was in use up to the 18th century. However, later the Eastern Nagari or Bengali script adopted for scripting Manipuri. It is also the script used till now. However, efforts to revive the Meitei-Mayek are still on. The Grammar of the Manipuri Language is interesting yet simple; the phonological system of the Manipuri Language can be basically divided into three levels- Vowels, Consonants and Tones. In Manipuri nouns are addressed according to the masculine or feminine on the basis of the natural sex.

4. COLLECTION OF DATA

The data for the electronic dictionary are collected from many sources. Some of them are as given below.

4.1 Daily Newspapers

Words collected from the daily local newspapers become one of the sources of data collection for the Dictionary. Some of local newspapers published in Manipur as

- (1) The Sangai Express.
- (2) The Poknapham.

Table 1. Table structure of the Dictionary database in MYSQL

field	data type	comment
root_word	vchar	Root word in the Dictionary.
p_o_s	vchar	part of speech(lexical item).
trans	vchar	Transliteration of the root word in English.
e_meaning	vchar	English meaning of the word.
man_ex	vchar	Example in Manipuri.
eng_ex	vchar	Example in English.

- (3) The Naharolgi Thoudang.
- (4) The Yakairol.
- (5) The Huiyen Lanpao etc

With the development of internet and Information technology, we can download and read some of the newspapers daily directly through the internet.

4.2 Weekly or monthly journals

We also collect the words from the journals which published weekly or monthly. Some of the journals areas.

- (1) Sanathong
- (2) Kangla Pao.
- (3) Eastern Frontier etc

4.3 Manipuri Text Corpus

Manipur corpus is the backbone for researchers of Manipuri language. Some of the words are also picked up from the Manipuri Corpus which is created in Manipuri University under the aegis of Prof. Ningomba and Dr. N.Pramodini.

4.4 Manipuri lexicons (paper dictionaries)

- (1) *Manipuri to Manipuri and English dictionary*
Compiled by Sri Khelchandra Singh
- (2) *Manipuri to English Dictionary*
Compiled by Dr. Imoba
- (3) *Learners Manipuri-English Dictionary*
Compiled by Dr. Surmangol Sharma etc

5. IMPLEMENTATION OF SOFTWARE AND DICTIONARY DATABASE

The softwares used for designing the Electronic Dictionary are as:-

Frontend: - PHP (Hypertext Preprocessor)

Backend: - MYSQL SERVER

WEB SERVER: - UNISERVER (Orion)

The structure of the dictionary (lexical) database is represented as:

- (1) *Root word:* -the main word in the dictionary.
- (2) *Lexical item:* -the part of speech category of the word.
- (3) *Transliteration:* -transliteration of the word in (target language) English.
- (4) *Meaning in English:* - meaning of the word in English .
- (5) *Example in Manipuri .*
- (6) *Example in English.*

We are using Binary search for our dictionary. Algorithm is quite simple. It can be done either recursively or iteratively.

- (1) Get the middle element;
- (2) If the middle element equals to the searched value, the algorithm stops;

Table 2. Wild Card search from Dictionary

Search string with wild card	Words from the dictionary according to the wild card search
ক??? (single search)	ককপা, ককলা, কংখু, কংবা, কংলা, কতলা, কদাষetc
ক*র (group search)	কমপোর , কর্পূর, কলাকার, কারetc

- (3) Otherwise, two cases are possible:
 - 3.(a). Searched value is less, than the middle element. In this case, Go to the step 1, before middle element.
 - 3.(b). Searched value is greater, than the middle element. In this case, Go to the step 1, after middle element.

6. WORD SEARCH IN THE ONLINE DICTIONARY

The word search is divided into three sections they are:

- (1) Simple word search
- (2) Wild card search
- (3) Search by lexical items

6.1 Simple Word Search

The algorithm of word search in the dictionary is given as:

- (1) START.
- (2) Search of the word in the Dictionary (BINARY SEARCH).
- (3) If word is found.
Print the word and its corresponding information: STOP.
- (4) If word is not found:
Do you want to Entry it in DICTIONARY ENTRY?
4.(a). If (Option=yes)
Dictionary entry is done: STOP.
4.(b). If (Option=no)
Display a message: Word is not found: STOP.

6.2 Wild Card search

The wild card search is of two types:

- (1) Single wild card search (using ?).
- (2) Group wild card search (using *).

The algorithm of wild card search in the dictionary is given as:

- (1) scan for wild card in the input string.
If any wild card is not found then display message: wild card is not in the given string STOP.
Else (here 3 cases arise: according to the position of the wild card)
- (2) If the wild card is found at the beginning of the string.
The comparison is done from the right to left until the wild card is not found.
- (3) If the wild card is found at the last of the string.
The comparison is done from the left to right until the wild card is not found.
- (4) If the wild card is found within the string.
Then 2 & 3 steps are to be use till the wild card is not found.
- (5) Display the output

6.3 Search by lexical items

Here searching is done through the lexical item. It is mainly to know the numbers of a particular lexical item in the dictionary. The user has to select a lexical item from the list of lexical item and push the search button. then we will find the output .(as in figure 3)

7. CONCLUSION AND FUTURE WORKS

This paper is to show that the Online Electronic Dictionary so developed contribution to the quality of education in Manipuri language corresponding with the development of education technology. Further research should be carried out either directly or indirectly by those who involved in the study of Manipuri language. It is to ensure that Manipuri language is highly valued and continues to be developed in future. And finally we need young researchers and linguistics to develop CLIR (cross lingual information retrieval systems) and MT (machine translation) for Manipuri Language as future work.

8. REFERENCES

[1]. Bhat, D. N. S. and Ningomba **Manipuri grammar** Munich: Lincom Europa, 1997.

[2]. Chelliah, Shobhana L. **A grammar of Meithei** , Berlin: Mouton de Gruyter. ISBN 0-19-564331-3, 1986

[3]. Ch. Yashwanta Singh **Manipuri Grammer**, Rajesh Publications, New Delhi,2000.

[4]. DeLancey, Scott **The Sino Tibetan languages** in B. Comrie (ed.). *The Worlds Major Languages*, New York:Oxford University Press, 1987.

[5]. G.A.Grierson **LSI, Vol III. Part III** .1st Edition Motilal Banarasidas, Delhi,1903.

[7]. S. Poireiton Meitei ,Shantikumar Ningombam, Prof. Bipul Syam Purkayastha, **An Analysis towards the Development of Electronic Bilingual Dictionary (Manipuri-English) -a report**. International Journal of Computer Science and Information Technology (IJCSIT)ISSN: 0975-9646, Vol. 3 (2) , 2012.

[7]. S. Poireiton Meitei ,Shantikumar Ningombam, H.Mamata Devi and Prof. Bipul Syam Purkayastha, **A MANIPURI-ENGLISH BILINGUAL ELECTRONIC DICTIONARY-Design and Implementation**, (IJEIT)ISSN:2277-3754, Vol. 1 (5) , 2012.

[8]. Singh Imoba **Manipuri to English Dictionary**,S.Ibetombi, Imphal 2004.

[9]. Sharma H. Surmangol **Learners Manipuri Dictionary**, Sangam Book Store, Paona Bazar, Imphal, 2006 .

[10]. T. C. Hudson **The meiteis**, Neeraj Publication, House, New Delhi, 1975.

9. EXPERIMENTAL RESULT AND DICTIONARY OUTPUT



Fig. 1. Simple word search



Fig. 2. Wild Card Search

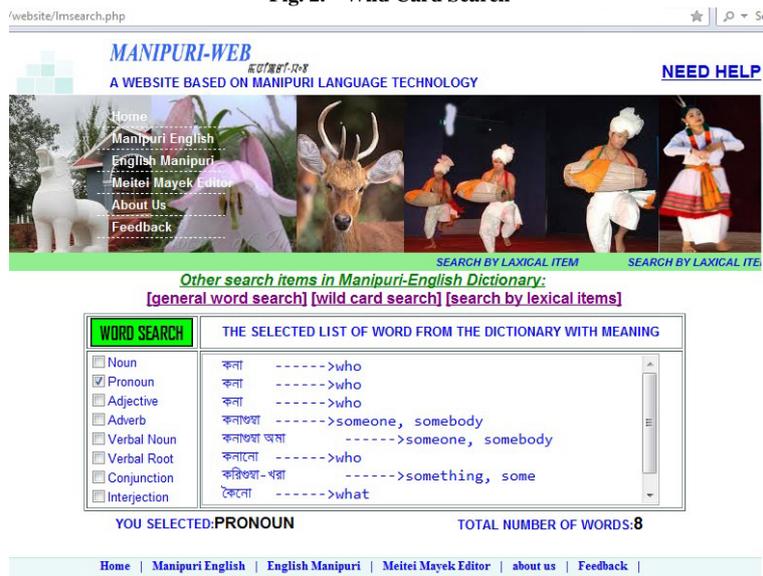


Fig. 3. Search By Lexical Item