ミャンマー語電子辞書の検索インターフェース

イェ　チョウ　トゥ　　浦野　義頼
早稲田大学大学院　国際情報通信研究科

他の発展途上諸国同様、ミャンマー語辞書は、紙ベースから CD-ROM へ、そしてウェブベースへ、さらに現在では携帯電子辞書へと変化を遂げてきた。しかしながら、電子辞書上でのミャンマー語の単語検索は未だ難しく不便である。本論文では、考えられる検索方法を数点取り上げる。中でも、ミャンマー語のためのワイルドカード、文字計数、句読点による分類は、ミャンマー語電子辞書での効率的な検索方法を生み出す新たな 3 つのアプローチである。各検索方法について複数の実験を実施し、その結果に基づいて有用性の比較を行った。

Lookup Interfaces for Myanmar Electronic Dictionaries

YE KYAW THU and YOSHIYORI URANO
Graduate School of Global Information and Telecommunication Studies,
Waseda University

Like in other developing countries, Myanmar dictionaries have been transforming from paper based to CD-ROM, then to web based, and now to portable electronic dictionaries. However, searching Myanmar words in electronic dictionaries is still difficult and inconvenient. In this paper, several possible lookup methods are discussed. Among them, Wildcard for Myanmar (WFM), Character Counting for Myanmar (CCM) and Punctuation Marks Grouping (PMG) lookup methods are three new approaches to create efficient lookup methods for Myanmar language electronic dictionaries. Several experiments were conducted for each lookup method, and usability comparisons were made based on the experiments result.

1. Introduction

1.1 Myanmar Language

Myanmar language is the official language of Myanmar. It belongs to the Tibeto-Burman language family and derives from Sino-Tibetan. Myanmar language alphabet is recognized as containing 33 consonants, vowels (containing dependent and independent vowels) and some conjunction alphabets or abbreviations. Unlike other Southeast Asian languages like Thai and Khmer, Myanmar language adopted words primarily from Pali rather than from Sanskrit. The Myanmar language writing direction is from left to right. The word order of the Myanmar language is Subject-Object-Verb. But 'က' (the verb "to be") is the only exception that is placed directly after the subject.

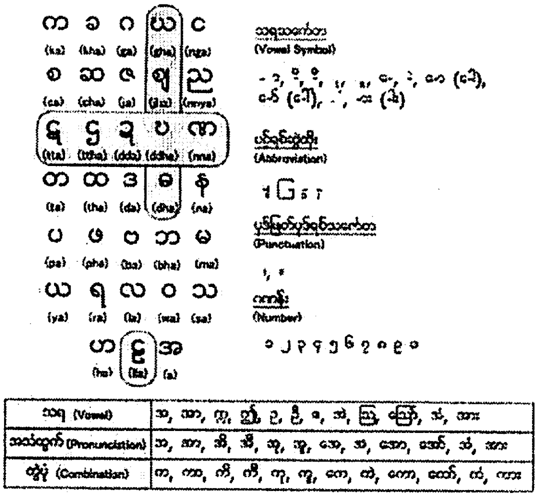

(ka) (kha) (ga) (gha) (nga)
(ca) (cha) (ja) (jha) (nnya)
(tta) (ttha) (dda) (ddha) (nna)
(ta) (tha) (da) (dha) (na)
(pa) (pha) (ba) (bha) (ma)
(ya) (ra) (la) (wa) (sa)
(ha) (lla) (a)
(Vowel Symbol)
(Abbreviation)
(Punctuation)
(Number)
(Vowel)
(Pronunciation)
(Combination)

Fig.1 Characters of Myanmar Language (Highlighted Characters Are Come from Pali Language)

1.2 Myanmar Dictionaries

Throughout history, the first "Burmese-English Dictionary" was compiled and published by an American Baptist "Adoniram Judson" (1788-1850). U Tun Nyein Dictionary (published in 1953) and "Tet Toe Dictionary" are famous among Myanmar students. In 1979, "Biruma Jiten" (Burmese-Japanese Dictionary) written by Masaharu HARADA and Toru OHNO was published by the Myanmar-Japan Cultural Association. U Hla Pe (retired Burmese Professor, University of London, UK in 1948-1980) worked for Burmese-English dictionary project which was transferred to Myanmar University. His small group completed 5 volumes of this dictionary, each 80 pages.

The "Myanmar Language Commission" has a very important role for developing and preserving the Myanmar Language. It published, Myanmar-Myanmar, Myanmar-English, English-Myanmar and Myanmar-other languages such as "Burmese Travelers' Dictionary" (1999), "Burmese Dictionary" (1991), "Concise Burmese Dictionary" (1978-80) [5 volumes], "Myanmar-English Dictionary" (1993), "English-Burmese Dictionary" (2001) etc. 1990s, several CD-ROM based dictionaries were developed by some companies such as "English-Myanmar dictionary" by World Peace IT Co., Ltd., and "Cherry Myanmar-Japanese/Japanese-English dictionary" in 2004 etc.

"E.R.A Reality Network Company Limited" distributed the very first English-Myanmar portable electronic dictionary named "Smart e212", in April 2005 and "English-Myanmar/Myanmar-English" electronic dictionary "Smart e212s" in 2006. Keyboard layout for inputting Myanmar language is similar to other Myanmar keyboard layouts in PCs.

2. Current Lookup Methods

Basically, current Myanmar electronic dictionaries (including CD-ROM based and WWW based) use 2 types of lookup methods. One is the "alphabetically ordered buttons method" and the other is the "type and check method". In the first method, alphabetically sorted buttons (only consonants) are used as User Interface. When the user clicks a button, the possible Myanmar word combinations list started with the clicked character will be shown and the user has to click the first part of word combinations of the search word inside the list. Then the user will see the word list started with chosen combination word. For example, if the user wants to find the word "မေမေ" (mother or mom), first the user has to click the button "မ" then the possible combinations of "မ" word will be shown such as "မ", "မာ", "မား" etc. in the list and the user has to click "မေ". Then, the user has to find the word "မေမေ" in the list. (see Fig.2) This lookup method is very easy to use and there is no need to type Myanmar words. But usually, the users have to scroll down until lookup word is found. For the second method, although the background idea is the same as first one, the difference is that users have to type their lookup word instead of clicking character buttons. When the user is typing, the word list will be updated starting with typed characters.

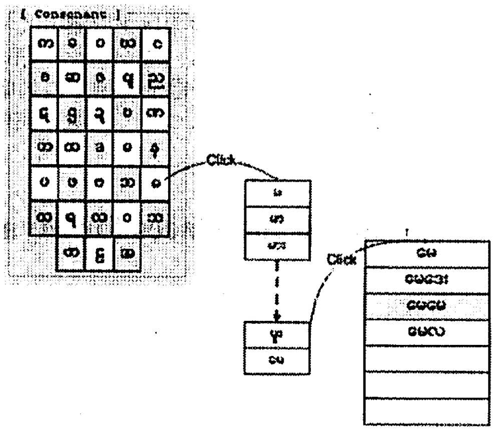


Fig.2 Alphabetically Ordered Lookup Method

Here the problem is that typing Myanmar characters with keyboard is still difficult for users. This is not only for portable electronic dictionaries but also for computers. In the current situation, there is no standard keyboard layout for Myanmar language and thus we have several keyboard layouts with propriety input methods such as "WinMyanmar", "GeoComp", "Myazedi" and "MyMyanmar" etc.

According to Myanmar language nature, keys on the PC keyboard are not enough for inputting Myanmar characters. That is why current Myanmar Keyboards use (Ctrl + Alt + Key), (Shift + Ctrl+ Alt + Key) and (Alt + Numeric Keys) etc for Pali words and other conjunction symbols. For example, typing (Ctrl + Alt + Q) for Myanmar character "ဿ" and (Alt + 0142) for Myanmar character "ဋ္ဌ". These kinds of solutions are not good user interface and need many hours of practice to become familiar. One of the other difficulties is Myanmar language encoding. We do not have a standard encoding system yet. And thus, we cannot make text processing like sorting, searching, word breaking and typing together with other languages properly. Some of the characters drop out when we copy from one application to another.

Many Myanmar Natural Language Processing (NLP) researchers are now trying to get Myanmar Unicode standard and International Organization for Standardization (ISO) standard for Myanmar language. Current Myanmar Unicode Keyboards are not pure Unicode standard and still use code points in the Unicode Private User Areas. For these reasons, developing dictionary lookup methods for Myanmar language is still difficult, and therefore, this development is a necessary research area for Myanmar language information retrieving.

3. Lookup Methods

Here, we will discuss possible lookup methods for Myanmar language. These lookup interfaces are considered for all types of Myanmar electronic dictionaries such as CD-ROM based, World Wide Web based and portable electronic dictionaries etc.

3.1 Visual Lookup

3.1.1 Writing System of Myanmar Language

Myanmar language has various types of characters comparing with English, i.e. consonants, dependent vowels, independent vowels, medials, finals or killers, visarga, stacking characters, conjunction alphabets etc. And Myanmar language contains many Pali words especially for religious things such as praying.

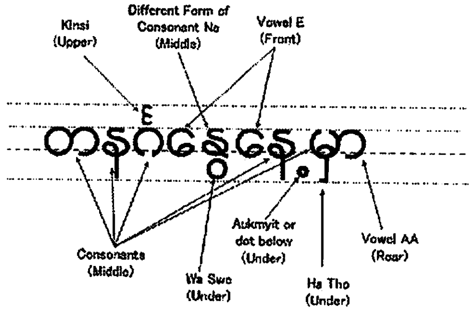


Fig.3 An example of Myanmar phrase (on Sunday)

When we make an analysis on Myanmar sentences, basically we can consider that Myanmar characters are written in three levels (up, normal, under) and most of the characters have their defined positions (e.g. ေ have to be written as front vowel, ု, ူ, ု, ူ, ွ, ွှ, ှ etc. should be written as a lower vowel and ိ, ီ and ံ should be written as an upper vowel and so on). But there are several exceptional cases, e.g. although consonant က (ka) are usually written in the normal level, when it becomes 'subscripted ka' it becomes smaller than normal ka and written in the lower level as 'က္က' in 'ဥက္ကဋ္ဌ' (the chief officer of a society). The basic typing characteristic of Myanmar language is that we type initial consonant character at first, and then add many symbols (medials, finals, visarga etc.) under, over or round it. (see Fig.3) Myanmar language learners as a foreign language have difficulties like "Which come first in a Myanmar word like "ခွေး" (dog) or "မြွေ" (snake)?

3.1.2 Lookup Method by Visual Input

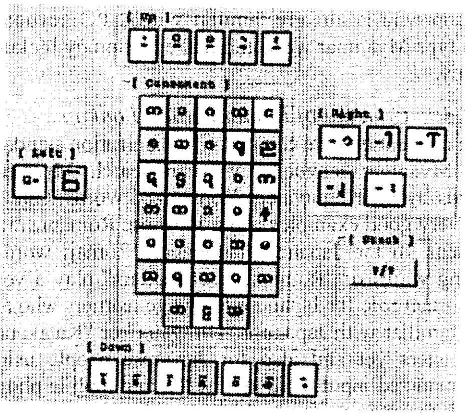


Fig.4 Visual Lookup Method

This method uses the visual input interface for Myanmar word searching. The user can type a Myanmar word according to visual order (i.e. we can neglect the Myanmar language hand writing order). The layout of keys is according to visual orders of Myanmar characters. (see Fig.4)

3.2 Romanization Lookup

3.2.1 Romanization Standard

There is no official standard for Myanmar language Romanization yet. There have been attempts to make standard, but none has been successful. In Myanmar language, many words are spelled differently from the way they are pronounced. For example, the word for "thief" is pronounced tha-kou "သခိုး" but spelled thu-kou "သူခိုး" and the word for "ox" is pronounced na-thou "နသိုး" but spelled nwa-thou "နွားသိုး". Replicating Myanmar sounds in the Roman script is difficult. In this research, we use Romanization system of Myanmar language commission for representing Myanmar words but not use for lookup experiments. (see Table 1)

3.2.2 Romanization Input Method

Lookup Method by Romanization in our demo program use our proposed "Partial Romanization input method" or "Assign and Combine Method". In "Assign and Combine Method", English alphabet(s) are assigned to each Myanmar letter based on its pronunciation instead of assigning a whole word pronunciation. (see Fig.5)

ေ = twt, က = ka, ျ = yp, ာ = yc, င = nga, ် = at, း = wsp

twtkaypycngaatwsp → ကျောင်း

Fig.5 Assign and Combine Method

By using a Romanization interface, every Myanmar person who is already familiar with a PC keyboard can type Myanmar word and make dictionary lookups easily.

3.2.3 Concept of Romanization Lookup

The concept is if we can make Romanization standard for Myanmar language, we can search Myanmar words by their equivalent Romanized words. One of the very good examples is the Japanese Romanization system. In the Japanese language, "Romaji words" (slang word) or "Romanized characters" play a very important role for Japanese language learners who are not familiar with Japanese "Hiragana" or "Katakana" characters yet and also for Japanese people using "Romanized input method" for PC and mobile phone SMS messaging. From this example, Romanization lookup method might be useful for Myanmar language learners as a foreign language.

3.3 Wildcard Lookup

Wildcards allow you to find words using patterns for a set of words (replacing single or multiple characters) and to broaden your search or capture different spellings or endings of a word. The two most commonly used wildcards are:

1) The question mark '?' may be used to represent a single alphanumeric character. For example, searching for the term 'car?' will find 'card', 'care', 'cars' etc.

2) An asterisk '*' may be used to specify zero or more alphanumeric characters. For example, searching for the term 'colo*r' will find both 'color' and 'colour'.

Based on the English wildcard searching method, we consider a simple pattern searching algorithm for Myanmar language that we called 'Wildcard for Myanmar Language' (WFM). Although there are some more symbols in wildcard such as '[ ]' (brackets enclose a set of characters, any one of which may match a single character at that position) and '-'(a hyphen used within [ ] denotes a range of characters), as a very first logic we only used '?' and '*' for WFM. Here, as we all know character breaking and word breaking for English is very simple. But not for the Myanmar language that used various types of characters and various combinations. And thus, we used the pronunciation level word breaking method as preprocess for wildcard grouping. Then we consider 'Consonant Level Matching' (CLM). By using this method, the user only needs to type main consonant characters such as 'သခသ' for 'သခွားသီး' (cucumber) and 'သဘသ' for 'သင်္ဘောသီး' (papaya).

က →	က
ကခ →	ကချင်၊ ကာချင်း၊ ကာချယ်၊ ကားချပ်၊ ကူးခပ် ...
ကခ* →	ကချလာ၊ ကချော်ကချွတ်၊ ကချင်၊ ကာချင်း၊ ကာချယ်၊ ...
က* →	က၊ ကကုသန်၊ ကကြိုကကြောင်လုပ်၊ ကချလာ၊ ...
က*ခ →	ကချော်ကချွတ်၊ ကချင်၊ ကစော့ခါး၊ ကဆစ်သခွား ...
*က →	က၊ ကကြီး၊ ကကြိုကကြောင်၊ ကစားကွင်း ...
က? →	ကချင်၊ ကာချင်း၊ ကာချယ်၊ ကားချပ်၊ ကူးခပ် ...
မ?? →	မင်္ဂကြက်၊ မဂိုကြက်၊ မဟာည်မွေး၊ မတောင်းမှု ...
?တ? →	ကတီတည်၊ ကတီထား၊ ကတိပေး၊ ကတိပျက် ...
??အ →	ကလအော်၊ ကုလားအော်၊ ကုလားအုတ် ...
???? →	ကဖျက်ယဖျက်၊ ကကြိုးတန်ဆာ၊ ကချော်ကချွတ် ...
တ?ရ* →	တစ်တစ်ရစ်ရစ်၊ တင်းတင်းရင်းရင်း၊ တကိုယ်ရေတကာယ ...

Fig.6 Wildcard Lookup Method

Here, the user does not need to type 'front vowel', 'lower vowel', 'upper vowel' and 'visarga' etc. By using wildcard symbols '?' and '*' the user can make pattern searching to require Myanmar words. CLM will be very useful for foreigners and even for Myanmar users who are not familiar with the Myanmar keyboard layout.

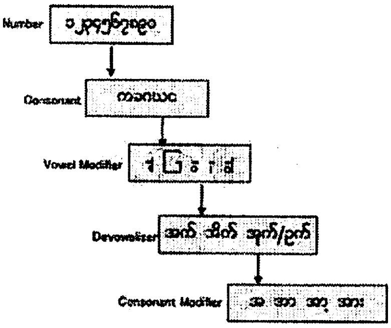


Fig.7 Sorting Order in Dictionary Index File

In the wildcard lookup method, we assume that a normal user will not be searching for broken characters such as 'ို*', '*္ဂ', '?ြ' etc and WFM algorithm will automatically be filtered. Excepts from a broken character search, others pronounceable Myanmar characters such as 'ကျ', 'ကြွ', 'ကွ', 'ပါ', 'မေ', 'ပြ' etc. can make searching together with wildcards. In this method, word by word online matching can reduce word retrieving efficiency and appropriate indexing methods can apply to a dictionary database. In this method, we used one index file that contains sorted Myanmar characters. (see Fig.7)

3.4 Character Count Lookup

Although counting characters and words in English is very simple, it is not so in Myanmar language. And there is no space between Myanmar words. In a Myanmar sentence, spaces are used to mark phrases. Traditionally, there is no indexing method by character count in Myanmar dictionaries. Our proposed idea is if we can define the appropriate counting method for Myanmar language then we can consider as new lookup or indexing method. When we analyze the Myanmar sentence, the general form of sentences contain phrases, phrases contain words, words contain syllables, syllables contain graphemes, and graphemes contain features. Note that even some features can have their separate meaning in Myanmar language (e.g. "တာ" means "is or from", "ည" means "night" etc.). In this method, we count 1 for one pronunciation part of a Myanmar words. We assume this counting method is meaningful for indexing Myanmar words. And we can apply this counting way for wildcard lookup as well. For example searching "???" will show three pronunciation words such as "အမျိုးသား" (men), "ဆရာမ" (female teacher), "စာရေးသူ" (author) etc. The following is an example of possible counting steps for the word "Myanmar citizen" in Myanmar language.

When we develop a demo function for counting lookup, we face some programming difficulties, for example, a string compare function "strcmp(string1, string2)" cannot be used properly for Myanmar sentences. We have to make Unicode values or Byte value comparison.

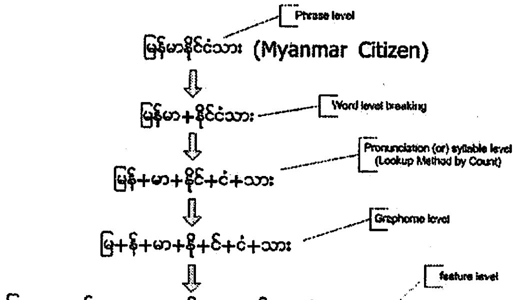


Fig.8 Phrase to Feature Level Breaking Steps

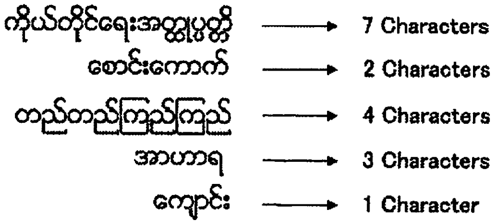


Fig.9 An Example of Character Counting

3.5 Meaning or Cross-Language Lookup

Meaning Lookup is the most traditional and simplest indexing methods for dictionaries. For example, in an English-Myanmar dictionary, the word "car" will be indexed with the equivalent Myanmar word "ကား", and in a Japan-Myanmar dictionary, the Kanji word "車", Hiragana "くるま" and Katakana "クルマ" will be indexed. "Cross-Language Searching" not only for one language to another but also for multiple languages can possible with today's electronic dictionary technologies. Non-Myanmar users, such as learners, and even native speakers, can greatly benefit from this kind of cross-language searching. Cross-Language Lookup has the additional benefit of enabling users without using Myanmar language input methods to retrieve Myanmar words. But in the real world, we need to build reliable lexicons such as English-Myanmar, Japanese-Myanmar, German-Myanmar, Thai-Myanmar etc to use the Cross-Language Information Retrieval (CLIR) technologies effectively.

Table.1 Meaning or Cross Language Lookup

Search Word	Search Result	Romanization
Math	သင်္ချာ	thin char
地図	မြေပုံ	mjei boun
ပညာရေး	教育	pjin nja jei

3.6 Pronunciation Lookup

Another possible lookup or indexing method is "Pronunciation Lookup". Almost all of the Myanmar dictionaries are not using pronunciation symbols (some dictionaries used Romanization). The phonetic assignment that we used in our experiments is referred from Myanmar-Japan dictionary [7] published in 1990 by Japan-Myanmar Cultural Association. This lookup method consideration is not for the ordinary users who are not familiar with phonetic symbols. We assume this lookup method might be useful for the advanced users such as linguists.

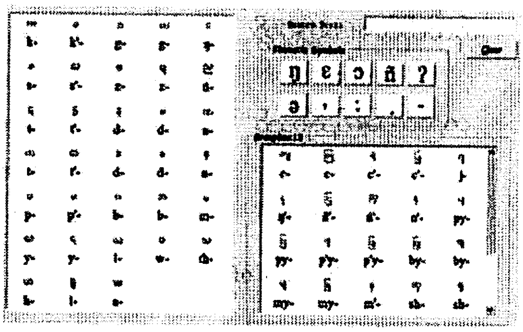


Fig.10 Pronunciation Lookup Interface

3.7 Pattern Lookup

The Pattern Lookup Method is based on an idea of direct identification of geometrical patterns of letters. Like Jack Halpern's "System of Kanji Indexing by Patterns (SKIP)" [9], we are grouping Myanmar characters according to their pattern similarity.

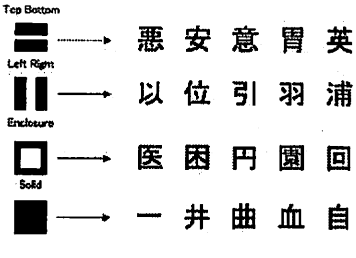
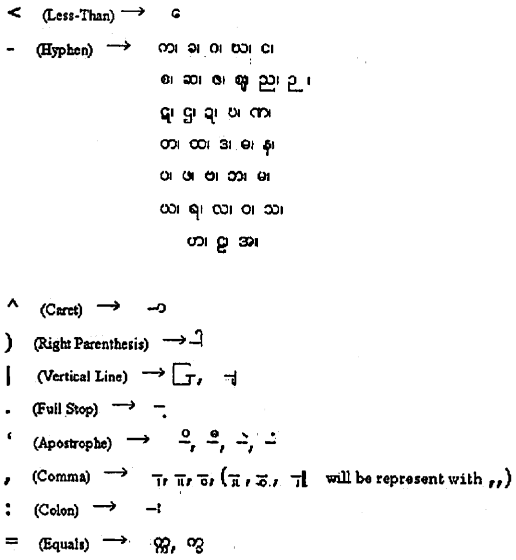


Fig.11 Four Patterns of SKIP



< (Less-Than) →
- (Hyphen) →
^ (Caret) →
) (Right Parenthesis) →
| (Vertical Line) →
. (Full Stop) →
' (Apostrophe) →
, (Comma) →
will be represent with ,,)
: (Colon) →
= (Equals) →

Fig.12 Myanmar Characters Representations in PMG

However, Myanmar letters are not stroke-based but have different characteristics compared to Japanese Kanji or Chinese characters, and thus, a lot of pattern groups come out. After making analysis on SKIP and Four Corner Code Coding System of classifying Kanji, we propose "Punctuation Marks Grouping (PMG) lookup" method for Myanmar language.

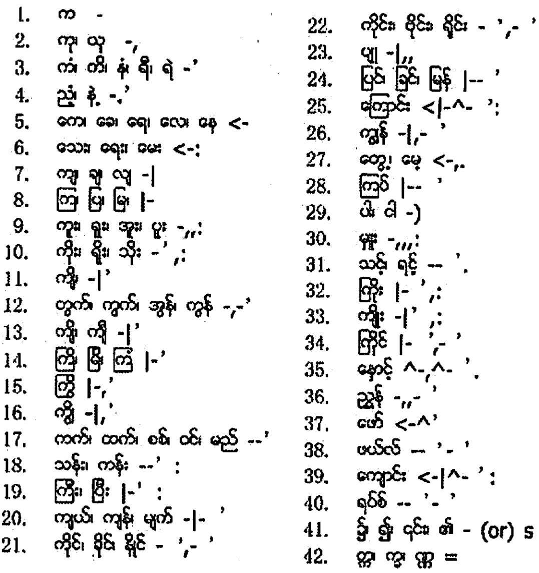


Fig.13 Myanmar Words Grouping with PMG

PMG uses English punctuation marks such as "-" (hyphen), "," (comma) and ":" (colon) etc. for representation similar pattern of Myanmar characters. For example, "-" (hyphen) will represent Myanmar consonants such as "တ", "ခ", "ပ", "ဟ", "င", to "အ" and it will also represents some Myanmar conjunction symbols such as "ၡ", "ၦး", "ၥ" etc. Similarly "<" (less than) will represents Myanmar front vowel "ေ" (E), ")" (right parenthesis) will represents Myanmar rear vowel "ာ" (AA), ":" (colon) will represents Myanmar character "း" (visarga) respectively. (see Fig.13)

4. Experiments and Discussions

We developed a simulation program by using Microsoft Visual Basic Programming to make experiments on all possible lookup methods. We also created a small Myanmar-Japan dictionary that contains 3000 Myanmar words. In this section, we discuss more on some of our proposed lookup methods.

4.1 Experiments on Wildcard Lookup

In WFM experiments, users have to know 'meaningful' and 'meaningless' Myanmar characters combinations. Refer to the following example.

In the word: 'ခို' (pigeon)

Meaningful characters: {ခ, ခိ, ခု, ခို}

Meaningless characters: {ု, ိ, ို}

Although users can make searching with meaningful characters like 'ခ*' or 'ခိ*', the WFM algorithm will not allow searching with meaningless characters like 'ို*','?ိ-?' etc. The meaningful characters and meaningless characters combination can be represented by the following permutation calculation:

$$\sum_{i=1}^{m} {}_mC_i = \sum_{i=1}^{m} \frac{m(m-1)\times(m-2)\cdots(m-i+1)}{i(i-1)(i-2)\cdots 1}$$

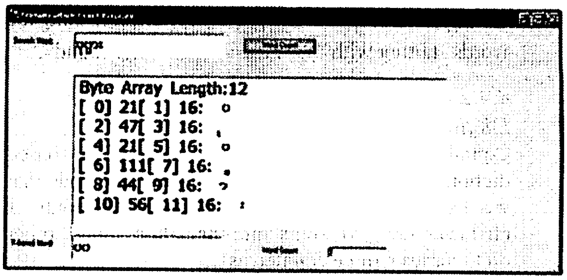
$$\boxed{mf + ml = \sum_{i=1}^{m} {}_mC_i}$$

Here,
m = total number of Myanmar characters (level 4)
mf = number of meaningful combination patterns
ml = number of meaningless combination patterns

4.2 Experiment on Character Count Lookup

For the "Character Count Lookup", we considered a general character breaking algorithm for Myanmar language. The algorithm used the "pronunciation level breaking method". The brief description of the pronunciation level breaking algorithm will be as follows:

```
Algorithm PronunciationBreak
        Input: Myanmar Language String S
        Output: Total Number of Characters in S
Count ← 0
for (i= 0 to length of S; i= i+1)
        select case(S[i])
                case 0: //Check for Ka character
                        if not S[i+1] = 57 // not killer
                          Count = Count + 1
                \\Check for all possible break point
                                else if not S[i+1] =
                        ....
                        end if
                case 1: //Check for Kha character
                to
                case 33: //Check end for consonants
                case 35:
                to
                case 79: // Check end for I, II, etc.
                case 65: To 64 // Check for numbers
//Check for Tha Gyi and other special characters etc.
                case 137: 141: 120: 138: 139: 121
        //Check for Nya Lay
        //(Note: Nya Lay is 2 characters)
                case 9:
        end select
end for
return Count
```



```
Byte Array Length:12
[ 0] 21[ 1] 16:
[ 2] 47[ 3] 16:
[ 4] 21[ 5] 16:
[ 6] 111[ 7] 16:
[ 8] 44[ 9] 16:
[ 10] 56[ 11] 16:
```

Fig.14 Pronunciation Level Breaking by Demo Program

4.3 Experiments on Pattern Lookup

Note that "စျ" (Sa+Yayit) and "ဩ" (O) characters are different from each other in Myanmar language. Similarly, in "၎င်း" (Le Gaung) character, there is not combination of "၄" (number 4). And thus, "စျ" (Sa+Yayit) will be represented with "|-" and "ဩ" (O) will be represented with "-" symbols in PMG

Table.2 Lookup Results for Pattern "-^:"

Myanmar Word	Japanese	English
ကား	車	car
စား	食べる	eat
ဆား	塩	salt
ဓား	刀、ほうちょう	sword, kitchen knife
နား	休む	rest

4.4 Experiments with other Dictionaries

Here, we made lookup experiments on 2 other electronic dictionaries.

4.4.1 Experiments on Smart e212s

Like other CD-ROM based Myanmar dictionaries, the lookup interface used by Smart e212s is the "type and check" method. It is based on font type face instead of actual Myanmar characters. Current Myanmar fonts assign many code points for "ျ" and "ြ-" characters. According to Myanmar language nature, we need to draw different sizes of Medials. For example big “ြ” (Yayit) for “ကြ”, “”, a little small “ြ” (Yayit) for “ဓြ” and so on. Although humans can adjust in handwriting, intelligent justification is needed for computers. Today Myanmar font cannot give automatic justification facility and thus, there will be several type faces only for one Yayit medial like “ြြြြြြ” etc. The program will be seen as a different character. The Smart e212s lookup system has the same problem, and it compares words with their type face. And thus, if a user tries to find the words starting with “ပြာ”, it will show “not found”.

4.4.2 Experiments on CD-ROM based Dictionaries

Counting methods on today’s CD-ROM based dictionaries are different from counting methods that we used in this research. Although we count 1 character for 1 Myanmar pronunciation sound, other dictionaries will be counted as 2.

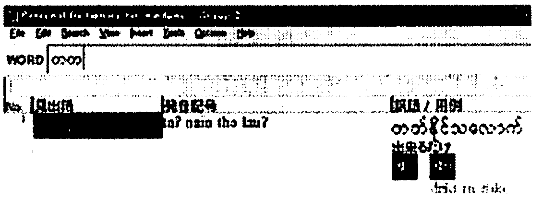


Fig.15 Lookup in Cherry’s Myanmar-Japan Dictionary

4.4.3 Usability Comparison for Lookup Methods

We conducted a small survey on the usability of propose seven lookup methods with 10 native users. Their response was as follows:

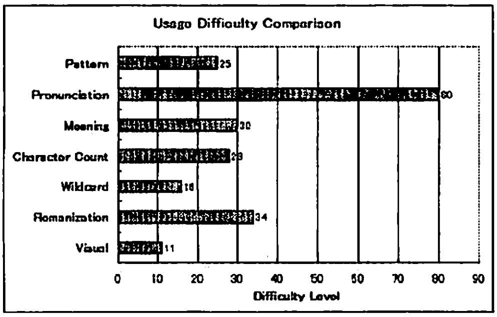


Fig.16 Users’ Responses on Lookup Methods

The graph represents the usage difficulty comparison among propose seven lookup methods. Here, X axis represents the difficulty level and Y axis represents lookup methods. Usage difficulty is divided from level 0 to level 10. And thus, for 10 users 100 in total will be the highest difficulty level, and we found that Pronunciation Lookup is the most difficult and Visual Lookup is the easiest interface for users.

5. Conclusion

The purpose of this research is to develop efficient lookup methods for Myanmar language, based on today’s Myanmar and other languages electronic dictionary lookup technologies. We propose seven possible lookup methods for Myanmar electronic dictionaries. The efficiency of the proposed lookup methods highly depend on input methods, encoding standard, Romanization, phonetic representations and indexing methods of Myanmar language. Wildcard and Pattern Lookup methods are also dependant on local Myanmar people’s acceptances. We believe that all of the proposed lookup methods can apply to on Myanmar language information retrieving. According to the survey, tested users are very interested in Wildcard and Pattern lookup methods. For the future, we plan to create a WWW based Myanmar dictionary and want to continue an analysis on lookup methods.

6. Acknowledgements

This research was supported by Prof. URANO Lab., Graduate School of Global Information and Telecommunication Studies (GITS), Waseda University, Tokyo, Japan. We would like to give special thanks to ‘Mr. Ye Kyaw Thein’ (my younger brother) who helped us gather information and participated in some experiments from the Myanmar side.

7. References

[1] Cherry’s Japan-Myanmar/Myanmar-Japan Dictionary, CDROM

[2] John Okell, *A Guide To The Romanization Of Burmese*, The Royal Asiatic Society of Great Britain and Ireland, 1971

[3] John Okell, “*Alphabetical Order In Burmese*”, JBRS, LI, ii, December 1968

[4] J.E. Bridges, *The Burmese Manual*, British Burma Press, Rangoon, 1906

[5] Jan Goyvaerts, *Regular Expressions (The Complete Tutorial)*, Printed in United States: Jan Goyvaerts, 2006

[6] Myanmar Language Commission, *Myanmar-English Dictionary*, Myanmar Language Commission Press, No.27, Pyi Road, Yangon, 1993

[7] Masaharu HARADA and Toru OHNO, *Myanmar-Japan Dictionary*, Printed in Japan: Japan Myanmar Culture Association, 1990

[8] Robert C. Stevenson, Rev. F. H. Eveleth, *Judson's Burmese-English Dictionary*, Baptist Board of Publications, 1953

[9] System of Kanji Indexing by Pattern (SKIP)
http://www.csse.monash.edu.au/%7Ejwb/SKIP.html