Standardization and Implementations of Thai Language

Theppitak Karoonboonyanan
National Electronics and Computer Technology Center,
National Science and Technology Development Agency,
Ministry of Science Technology and Environment, THAILAND.
thep@links.nectec.or.th

1 Thai Language

Thai, spoken by almost 100 percent of the population, is the official national language of Thailand. It is a tonal, uninflected, and predominantly monosyllabic language. Most polysyllabic words in the vocabulary have been borrowed, mainly from Khmer, Pali or Sanskrit. [1]

The ancient Thai language itself belongs to the Tai language family, which includes language spoken in Assam, northern Myanmar, Thailand, Laos, northern Vietnam and the Chinese provinces of Yunnan, Guizhou and Guangxi. [2]

The oldest Thai script found in the history dated over 700 years ago, in Sukhothai age. Over years, Thai script has been gradually changed. The contemporary Thai alphabet comprises 44 consonants, 32 vowels, 4 tone marks, 2 diacritics, and 10 decimal digits. These cover only the minimal set of symbols to write Thai words. There are additional 3 symbols particularly used for writing Pali and Sanskrit words, and 6 typographical symbols.

Consonants

Thai consonants include 34 of 35 Sanskrit equivalences, plus 10 invented ones. Thai consonants are classified into three classes—namely, high, middle, and low consonants—which can affect the syllable tone when functioning as initial sound.

Table 1 is the consonant chart laid after the grammar of Sanskrit. The first 5 rows are classified according to their articulators. The last row is the remainder of the classification. The parenthesized consonants are invented by Thai. Within the first 5 rows, according to Thai grammar, the first column contains *middle consonants,* the second contains *high consonants,* the next two columns contain *paired low consonants* (low consonant with the same sound as the high consonant in the same row), and the last column contains the *single low consonants*, whose sounds are nasal.

middle	high	low		
		paired		single
ก [k]	ข (ฃ) [kh]	ค (ฅ) [kh]	ฆ [kh]	ง [ŋ]
จ [tɕ]	ฉ [tɕh]	ช (ซ) [tɕh] [s]	ฌ [tɕh]	ญ [j]
(ฎ) ฏ [d] [t]	ฐ [th]	ฑ [th]	ฒ [th]	ณ [n]
(ด) ต [d] [t]	ถ [th]	ท [th]	ธ [th]	น [n]
(บ) ป [b] [p]	ผ (ฝ) [ph] [f]	พ (ฟ) [ph] [f]	ภ [ph]	ม [m]
ย [j] ร [r] ล [l] ว [w] ศ [s] ษ [s] ส [s] ห [h] ฬ [l] (อ) [ʔ] (ฮ) [h]				

Table 1—Thai consonant chart inherited from Sanskrit

Vowels

There are 32 vowel forms in Thai, which could be written by combinations of 18 vowel symbols and 3 consonants. The 32 vowel forms are:

-ะ -า ◌ิ ◌ี ◌ึ ◌ื ◌ุ ◌ู เ-ะ เ- แ-ะ แ- โ-ะ โ-
[a] [aː] [i] [iː] [ɯ] [ɯː] [u] [uː] [e] [eː] [ɛ] [ɛː] [o] [oː]

เ-าะ -อ เ-อะ เ-อ ◌ัวะ ◌ัว เ◌ียะ เ◌ีย เ◌ือะ เ◌ือ ◌ำ ใ- ไ- เ-า
[ɔ] [ɔː] [ɤ] [ɤː] [uə] [uəː] [iə] [iəː] [ɯə] [ɯəː] [am] [ai] [ai] [au]

ฤ ฤๅ ฦ ฦๅ
[rɯ] [rɯː] [lɯ] [lɯː]

The 18 vowel symbols are:

ะ ◌ั า ◌ำ ◌ิ ◌ี ◌ึ ◌ื ◌ุ ◌ู เ แ โ ใ ไ ฤ ฦ ๅ

And the 3 consonants used to compose vowels are ย ว and อ.

Tones

Thai is a tonal language. There are 5 tones in Thai syllables, namely Saman (middle), Ek (low), Tho (falling), Tri (high) and Chattawa (rising).

There are 4 tone marks in Thai script, namely Mai Ek (◌่), Mai Tho (◌้), Mai Tri (◌๊) and Mai Chattawa (◌๋). Despite their names, the tone marks do not directly imply the actual tone of the syllalbles in which they appear. The initial consonant and the syllable ending sound are also factors which determine the actual tone.

Diacritics

There are 2 diacritics for writing Thai words: Mai Taikhu (◌็) or sound shortener, and Thanthakhat (◌์) or sound killer; and additional 3 marks for Pali/Sanskrit words: Nikhahit (◌ํ), Pinthu (◌ฺ), and Yamakkan (◌๎).

Digits

Thai has ten decimal digits (0-9): ๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙ .

Typographical Symbols

Paiyan Noi (ฯ) or omission sign, and Mai Yamok (ๆ) or repetition sign, are contemporary punctuation marks generally used in documents. Thai Baht currency sign is written as ฿.

In addition, there are three ancient typographical symbols used in some documents: Fongman (๏) or sentence/stanza/paragraph beginner, Angkhandeaw (ฯ) or sentence/stanza terminator (normally written using the same glyph as Paiyan Noi), Angkhankhu (๚) or chapter/episode terminator, and Khomut (๛) or story terminator. [24]

2 Character Sets

The *only* standard for Thai character set for information interchange is **TIS 620-2533 (1990),** defined by the Thai Industrial Standard Institute (TISI), Ministry of Industry. It defines two eight-bit character code sets by extending ISO 646-1983 and IBM GX20-1850-4 (EBCDIC), respectively. TIS 620-2533 (1990) is indeed a minor correction of the previous one, **TIS 620-2529 (1986).** The character set table remains the same. Only some detailed descriptions are added for compliance with international standards.

TIS 620 extends the 7-bit character set of ISO 646 by using 8-bit code table and defining 87 Thai characters in the GR area, namely A1-DA and DF-FB, as shown in Table 2. And it has also been submitted to the European Computing Manufacturers' Association (ECMA) for registering in the ISO 2375 repertoire. ECMA registered and assigned it ISO-IR-166. Therefore, it can be used with ISO/IEC 2022, and an example of such implementation is GNU Emacs.

Around 1996, a TISI working group had drafted the Latin/Thai part (part 11) of ISO/IEC 8859 based on TIS 620, but, due to the prohibition of combining characters in the sense of ISO/IEC 2022, it did not pass, and the work was suspended.

In Thailand, most Thai documents are encoded using the extension of ISO 646 in TIS 620. (The use of the EBCDIC extension part is as rare as the use of EBCDIC itself.) In the internet, there has been a confusion of MIME character sets for Thai e-mails and web pages among the ad hoc solutions, such as using "iso-8859-1", "x-user-defined" and "windows-874", until, in 1998, Trin Tantsetthi [21] requested for "tis-620" character set registration with the Internet Assigned Number Authority (IANA), according to RFC 2278, for information interchange on the internet. IANA registered it in the same year. Now the MIME character set is under usage campaign. Mozilla project [30] has already supported it.

Thai part in ISO/IEC 10646-1 and Unicode 2.1 (U+0E00 – U+0E5F) is the code-by-code mapping with the characters A0-FF of TIS 620.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	■	■	■	■	■	■	■	■				ฐ	ภ	◌ะ	เ	๐
1	■	■	■	■	■	■	■	■			ก	ฑ	ม	◌ั	แ	๑
2	■	■	■	■	■	■	■	■			ข	ฒ	ย	า	โ	๒
3	■	■	■	■	■	■	■	■			ฃ	ณ	ร	◌ำ	ใ	๓
4	■	■	■	■	■	■	■	■			ค	ด	ฤ	◌ิ	ไ	๔
5	■	■	■	■	■	■	■	■			ฅ	ต	ล	◌ี	◌ๅ	๕
6	■	■	■	■	■	■	■	■			ฆ	ถ	ฦ	◌ึ	◌ๆ	๖
7	■	■	■	■	■	■	■	■			ง	ท	ว	◌ื	◌็	๗
8	■	■	■	■	■	■	■	■			จ	ธ	ศ	◌ุ	◌่	๘
9	■	■	■	■	■	■	■	■			ฉ	น	ษ	◌ู	◌้	๙
A	■	■	■	■	■	■	■	■			ช	บ	ส	◌ฺ	◌๊	๚
B	■	■	■	■	■	■	■	■			ซ	ป	ห		◌๋	๛
C	■	■	■	■	■	■	■	■			ฌ	ผ	ฬ		◌์	
D	■	■	■	■	■	■	■	■			ญ	ฝ	อ		◌ํ	
E	■	■	■	■	■	■	■	■			ฎ	พ	ฮ		◌๎	
F	■	■	■	■	■	■	■	■			ฏ	ฟ	ฯ	฿	๏	

■ = same as ISO 646

shaded = unspecified

Table 2—TIS 620-2533 character set (ISO 646 extension)

3 Input Method

Thai software industry had suffered a diversity of Thai language implementations until a group of professionals had joined together to form the Thai API Consortium (TAPIC), headed by Thaweesak Koanantakool and sponsored by National Electronics and Computer Technology Center (NECTEC), for drafting a proposal for Thai API standard called WTT 2.0 [14] in 1991. (The name WTT comes from its Thai abbreviation "วทท", or "Wor Tor Tor", of which the full name is "วิ่งทุกที่", meaning "runs everywhere".) The proposal comprises 3 parts, defining the general facilities, Thai input/output methods, and printer identification number, respectively. The proposal was submitted to the TISI/TC536/SC2 for adopting as a standard in 1991.

WTT 2.0 was a consequence of WTT 1.0 system developed by Thaweesak Koanantakool and his team at Thammasat University. Although not proclaimed as a standard, due to the wide cooperation among industrial companies, such as Digital, Sun, Microsoft and IBM, WTT 2.0 has been adopted in many systems, such as Digital UNIX, Thai Language Environment (TLE) for Solaris [16], Microsoft Windows 95/98 and Windows NT [17]. This makes WTT 2.0 become de facto.

According to WTT 2.0, Thai words are input and stored letter-by-letter from left to right. For the cells which contain symbols above or below baseline, the base character will be stored first, followed by the lower or upper vowel, and then the tone mark or diacritic at the top-most level. For example, the phrase 'ไปทั่วคุ้งน้ำ' will be input and stored as 'ไ-ป-ท-ั-ว-ค-ุ-้-ง-น-้-ำ'. A WTT 2.0 compliant input method can provide a sequence checking mechanism to ensure the validity of the input sequence in one of the three levels of strictness: level 0: pass through, level 1: basic check, level 2: strict check.

WTT 2.0 has also specified the cursor movements and editing behavior of Thai text editors and word processors. Cursor must be moved from cells to cells, that is, all characters in other levels than the base line must be skipped. Text deletion using the "Delete" key must also remove all characters in the current cell, including the above, below and top symbols. Meanwhile, character-by-character, right-to-left, removal is still possible by using the "Backspace" key, where the order of removal is considered by the order they are stored.

Thai computer keyboard layout is specified by **TIS 820-2538 (1995),** as shown in Figure 1. TIS 820-2538 is a modification of the previous one, TIS 820-2531 (1988). TIS 820 are modified from Ketmanee, a famous layout for typewriters. There was, however, another keyboard layout resulted from a character frequency distribution research, named Pattachote. But it is not as familiar for typists as Ketmanee.
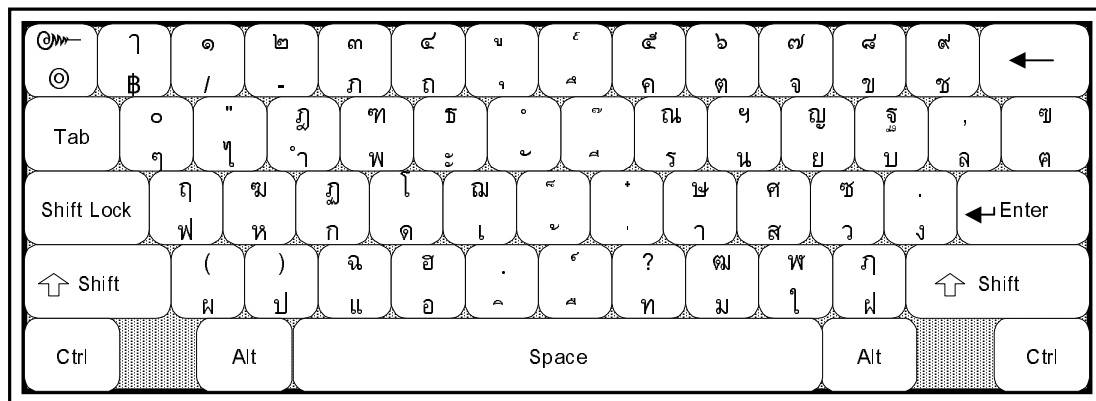

Figure 1—Layout of Thai Computer Keyboards (TIS 820-2538)

4 Output Method

Thai is written in 4 levels, as shown in Figure 2.

ไปทั่วคุ้งน้ำ

— top level
— above level
— base line
— below level

Figure 2—Thai 4-level writing system

The multiple levels of Thai texts are rendered by means of *combining characters*. WTT 2.0 classifies printable characters in TIS 620 as *forward characters* and *dead characters*. Forward characters are those characters written on the base line and occupy horizontal space, while dead characters are those written below or above base line and whose widths are zero.

Thai quality display and printing has been enabled by the font technologies. Among these, TrueType is the most widely used. New codes for presentation purpose have been added. For example, since some Thai consonants with long tails occupy a little space in the upper level, new codes for upper vowels and tone marks shifted left have been added. Another need is the adjustment of tone mark position when upper vowel is absent. Thus, new codes for lower tone marks are also added. (See Figure 3 for other adjustments.)

ปี่ ปี่ ป่ บ่

high, left
high, right
low, left
low, right
left
right

(a) Vowels and tone marks adjustment

ญ ญ ฐ จุ กุ ฎุ
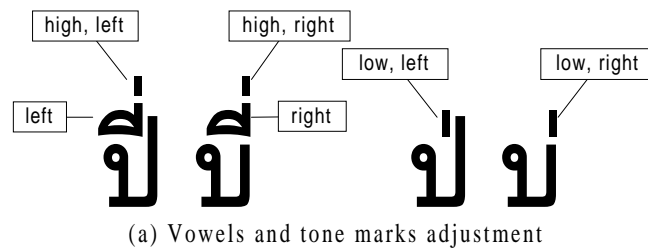
(b) Base removal when combined with below vowel   (c) Lowered below vowel

Figure 3—Added glyphs for quality text representations

However, there are two major different code tables for Thai TrueType fonts. One is defined for Mac OS Thai, another is for Microsoft Windows. The one for Mac OS is based on MacThai character set [19], while that for Microsoft is an extension of code page 874 [18]. This makes the two kinds of TrueType fonts unable to be used together.

5 Lexicographical Ordering

The widely accepted standard for Thai lexicographical ordering is defined in the Royal Institute Dictionary 2525 B.E. Edition [23], the official standard Thai dictionary. And TIS 620 code table has been carefully specified according to this principle. The ordering principle is:

1. Words are ordered alphabetically, not by sounds. Consonants are ordered according to traditional order. TIS 620 has defined codes for 44 consonants and 2 vowels ฤ, ฦ according to this order in the range A1-CE.

2. Vowels are also ordered by written forms, not by sounds. TIS 620 has defined codes for vowels according to this order in the range D0-D9 and E0-E4.

3. Consonants always precede vowels. Word comparison is processed from left to right by considering initial consonants before vowels in the same syllable.

4. Tones and diacritics are normally ignored, unless all other parts are equal, in which case the order is as traditional order. TIS 620 defines order for these marks in the range E7-EB.

The principle is sufficient for use in the dictionary. And TIS 620 code specification almost enables C standard library function *strcmp()* to work with Thai strings, with two exceptions [26] :

1. **Leading vowels** (E0-E4 in TIS 620), which are written before the consonants, must be considered after the consonants. Therefore a reordering is needed before the actual comparison. This is also described as *rearrangement* in the draft of Unicode Technical Report #10 (Unicode Collation Algorithm) [25].

2. **Tone marks and diacritics** (E7-EB in TIS 620) must be ignored in the first pass, and be considered at later pass if the first pass yields equality.

Several Thai developers have designed sorting algorithms after this consideration, some of which have been summarized in [26]. However, the Royal Institute principle does not cover all cases in TIS 620 table. Therefore, a group of developers [27] have worked out the non-covered part and have proposed a generic string collation principle for Thai, based on ISO/IEC 14651 sorting model. Table 3 summarizes the proposed principle.

Table 3—Thai lexicographical ordering proposed by T. Karoonboonyanan et. al. [27]

Level	Order	TIS 620	ISO/IEC 10646	Remark
1	digits (language-insensitive),	30-39	U0030-U0039	• U0030=$_1$U0E50, ... U0039=$_1$U0E59
		F0-F9	U0E50-U0E59	
	English alphabets (case-insensitive),	41-5A	U0041-U005A	• U0041=$_1$U0061, ... U005A=$_1$U007A
		61-7A	U0061-U007A	
	Thai consonants,	A1-CE	U0E01-U0E2E	
	Nikkhahit ˙ ,	ED	U0E4D	
	Thai vowels	D0-D9	U0E30-U0E39	• U0E32=$_1$U0E45
		E0-E5	U0E40-U0E45	• rearrange U0E40–U0E44
2	Latin-script property,	30-39	U0030-U0039	• U0030<$_2$U0E50, ... U0039<$_2$U0E59
	Thai-script property,	F0-F9	U0E50-U0E59	
	Yamakkan ˟ ,	EE	U0E4E	
	Pinthu ˳ ,	DA	U0E3A	
	Thanthakhat ʻ ,	EC	U0E4C	
	Mai Taikhu ˮ ,	E7	U0E47	
	Thai tone marks	E8-EB	U0E48-U0E4B	
3	English-lower-case property,	61-7A	U0061-U007A	• U0041>$_3$U0061, ... U005A>$_3$U007A
	English-upper-case property,	41-5A	U0041-U005A	
	Thai-extra property	E5	U0E45	• U0E32<$_3$U0E45
4	space,	20	U0020	
	non-breaking space,	A0	U00A0	
	low line _ ,	5F	U005F	
	hyphen - ,	2D	U002D	
	comma , ,	2C	U002C	
	semicolon ; ,	3B	U003B	
	colon : ,	3A	U003A	

	exclamation ! ,	21	U0021	
	question ? ,	3F	U003F	
	solidus / ,	2F	U002F	
	full stop . ,	2E	U002E	
	Paiyan Noi ๆ ,	CF	U0E2F	••abbreviation
	Mai Yamok ๆ ,	E6	U0E46	••repetition
	grave ` ,	60	U0060	
	circumflex ^ ,	5E	U005E	
	tilde ~ ,	7E	U007E	
	apostrophe ' ,	27	U0027	
	quotation " ,	22	U0022	
	left parenthesis (,	28	U0028	
	left bracket [,	5B	U005B	
	left brace { ,	7B	U007B	
	right brace } ,	7D	U007D	
	right bracket] ,	5D	U005D	
	right parenthesis) ,	29	U0029	
	at @ ,	40	U0040	
	Baht ฿ ,	DF	U0E3F	••currency
	dollar $,	24	U0024	
	Fongman ๏ ,	EF	U0E4F	••text beginner
	Angkhankhu ๚ ,	FA	U0E5A	••chapter terminator
	Khomut ๛ ,	FB	U0E5B	••story terminator
	asterisk * ,	2A	U002A	
	back solidus \ ,	5C	U005C	
	ampersand & ,	26	U0026	
	number # ,	23	U0023	
	percent % ,	25	U0025	
	plus + ,	2B	U002B	
	less than < ,	3C	U003C	
	equal = ,	3D	U003D	
	greater than > ,	3E	U003E	
	virtical line |	7C	U007C	

Here are the explanations of some cases [27] [24] :

Paiyan Noi (U0E2F) and Mai Yamok (U0E46)

Paiyan Noi (ๆ) is an abbreviation sign, representing omitted text. Its function is like ellipsis, although the more correspondent Thai punctuation mark for ellipsis is Paiyan Yai (ฯลฯ). Therefore, it makes sense to order it after ellipsis category in the Common Template, after Devanagari abbreviation sign <U0970>.

Mai Yamok (ๆ) is a duplication sign, representing a repeat of the preceding word. Therefore, it deserves to be sorted in the duplication marks category in the Common Template.

Nikhahit (U0E4D)

According to traditional Thai principle, Nikhahit (ํ) is a name of a vowel symbol. It, with Pin I (ิ) compose Sara Ue (ึ), with Lakkhang (า) compose Sara Am (ำ).

However, in Thai document, such function is ineffective, because TIS 620 has defined codes for such combined vowels, and they were input with a single keystroke. Instead, Nikhahit is normally used as Pali-Sanskrit-inherited consonant. For example, following Pali, พุทธํ is pronounced [phútthaŋ]; and following Sanskrit, จุํณ is pronounced [tɕhumnum]. Therefore,

Nikhahit arguably deserves to be classified as a consonant, ordered after the last Thai consonant (ฮ).

Yamakkan (U0E4E), Pinthu (U0E3A), Thanthakhat (U0E4C), Mai Taikhu (U0E47) and Tone marks (U0E48-U0E4B)

Yamakkan (๎) is an ancient punctuation mark used to mark cluster, such as พ๎ราห๎มณ [phraːmanǎ]. Pinthu (ฺ) has two functions, one is the same as Yamakkan, such as a Sanskrit word พราหมณ [phraːmaná], the other is to mark final consonant in Pali writing system, such as วิชฺชา [wíttɕhaː].

Thanthakhat (์) is used for killing letter's sounds, such as พราหมณ์ [phraːm], in which the consonant ณ is killed (not pronounced), ลักษมณ์ [lák], in which the three consonants ษมณ are killed, ปาฏิหาริย์ [paːtihǎːn], in which the two letters ิย are killed.

The three marks can be considered as having less effect in changing words than tone marks. For example, the meaning of พราหมณ and พราหมณ์ are not different, while บน [bon] and บ่น [bòn] are. Therefore, they deserved to be ordered before tone marks.

Pinthu may come between Yamakkan and Thanthakhat, because another function of Thanthakhat is to mark final consonant in ancient writing of Pali words, which is in common with Pinthu.

In summary, the order should be Yamakkan, Pinthu, Thanthakhat, Mai Taikhu, Mai Ek, Mai Tho, Mai Tri, and Mai Chattawa.

Lakkhang Yao (U0E45) and Sara Aa (U0E32)

Lakkhang Yao (ๅ) is always written after Ru (ฤ) and Lu (ฦ), to produce new symbols Rue (ฤๅ) and Lue (ฦๅ), respectively. Traditional Thai principle treats each of these two new symbols as an atomic entity. Therefore, in general, Lakkhang Yao is never written after any letters other than Ru (ฤ) and Lu (ฦ). Meanwhile, the two symbols are never followed by a vowel. Hence, the occurrence of Lakkhang Yao (ๅ) and Sara Aa (า —a vowel) are mutually exclusive, and assigning order between them seems not necessary. However, due to their similar shapes, they are occasionally confused (in a reasonable way). So, they should be treated as weakly identical, that is, their weights should be different at the last level of comparison.

Fongman (U0E4F), Angkhankhu (U0E5A), and Khomut (U0E5B)

As described earlier, Fongman (๏) is used in ancient books as paragraph, sentence, or poem stanza begin marker. Angkhandeaw (ๆ) is used to end a sentence or a stanza. Angkhankhu (๚) is used to end a chapter or episode. Khomut (๛) ends a story.

These marks can be classified as typographic symbols in the Common Template, where Fongman functions like a bullet. Unfortunately, there is no specific code for Angkhandeaw (ๆ) and Paiyan Noi is always used instead. Therefore, we should ignore the existence of Angkhandeaw and the similarity between Paiyan Noi (ๆ) and Angkhankhu (๚) when considering their orders.

6 Word Boundary

There is no word delimiter in Thai writing system. Therefore, a computer program that processes Thai documents needs to be able to tokenize Thai strings into words, using certain kinds of linguistic knowledge.

There have been several approaches to this famous problem of Thai language. The first approaches were rule-based. This needed human knowledge to find out rules. But there were too much exceptions and an accurate rule set seems impossible. Dictionary-based approaches were thus proposed. Several Thai word processors nowadays are using these approaches and users are happy at some degree. However, it is still too coarse for more fine-grained applications like machine translation. Therefore, modern research are now focusing on statistical and machine learning approaches.

However, it still lacks standard API for Thai word tokenization, although there was awareness of requirement among software developers to mediate the diversity of the existing word break API's, so that the word boundary finding task of an internationalized software can be split and left to the pluggable libraries.

Existing word break APIs include:

1. Line wrapping within hot zone

    Example

    > *int LineWrap(const char* text, const int charWidths[], int rightMargin);*
    >
    > returns last word position that is not greater than the *rightMargin* for the given *text* string, of which the width of each character is given in *charWidths[]* correspondingly,

2. Word delimiters insertion

    Example

    > *int InsertWBR(const char* text, char* output, const char* delimiter);*
    >
    > reads *text,* copies each word into *output* buffer, putting *delimiter* at the end of each word, and returns the number of words found.

3. Filling array of pointers to word break positions

    Example

    > *int FindWBR(const char* text, int* output);*
    >
    > reads *text,* finds the word boundaries within it, fills the word break positions in *output* buffer, and returns the number of words found.

4. Finite state machine [28]

    Example

    > *int Reset();*
    > *int AddChar(char c);*
    >
    > *Reset()* initializes the internal abstract state machine. *AddChar(c)* feeds *c* to the internal machine, returns word length if reaching a word break position, or zero if the word is not complete yet.

5. Word break iterator [29]

    Example

    > *class ThBreakIterator {*

```
public:
    // Operation
    virtual void setText(const char* begin, const char* end=NULL) = 0;
    virtual void appendText(const char* begin, const char* end=NULL)
    = 0;
    virtual void clear() = 0;

    // Iteration
    virtual int first() = 0;
    virtual int last() = 0;
    virtual int next() = 0;
    virtual int prev() = 0;
    virtual int current() = 0;
};
```

setText() and *appendText()* feeds the input text. Then, each word can be retrieved with iteration operations: *first()* initializes the iteration and return the first word boundary; *last()* returns the word boundary; *next()* and *prev()* steps the iteration forward and backward word by word.

7 Minority Scripts in Thailand

There is no thorough research about contemporary scripts uses in Thailand. It is clear that almost 100 percent of Thai population are using Thai script, as they are taught in public schools and it is the only official national language.

However, in the four Muslim provinces of southern Thailand, Pattani dialect of Jawi with script is being used in everyday life by the local people. And the script is different from Malay Jawi script used in Malaysia.

In the northern and north-eastern regions, there are two ancient scripts used in old writing materials like stones and palm leaves. One is Fagkham script, the old Tham script closely blended with Sukhothai script. The other is Muang script, which looks closer to Tham script used in Laos [31] [32]. Nowadays, Muang script is found being used for religious study materials.

However, the need of the computerization of the scripts, including what other scripts being used, need further research and survey.

References

[1]   Secretariat of the Prime Minister, Thai Government House. **Thailand at a Glance.** *http://www.thaigov.go.th/general/bgthai/bgthailand.htm.*
[2]   Thomas J. Hudak. **Some Historical Background of Thai Language.** *http://thaiarc.tu.ac.th/thai/thai.htm.*
[3]   Thai Industrial Standard Institute. **TIS 620-2533 Standard for Thai Character Codes for Computers.** Bangkok. 1990.
[4]   Thai Industrial Standard Institute. **TIS 820-2538 Layout of Thai Character Keys on Computer Keyboards.** Bangkok. 1995.
[5]   **ISO/IEC 646 Information technology – ISO 7-bit Coded Character Set of Information Interchange.**
[6]   ISO/IEC JTC1/SC2. **ISO/IEC 2022 Information Technology – Character Code Structure and Extension Techniques.**
[7]   **ISO 2375 Data processing – Procedure for Registration of Escape Sequences.**
[8]   ISO/IEC JTC1/SC2/WG3, **ISO International Register of Coded Character Sets To Be Used With Escape Sequences.**

[9]   ISO TC97/SC2. **ISO 8859 Information processing – 8-bit Single-Byte Coded Graphic Character Sets.**

[10]   ISO/IEC JTC1/SC2. **ISO/IEC 8859 Information technology – 8-bit Single-Byte Coded Graphic Character Sets.**

[11]   ISO/IEC JTC1/SC2/WG2. **ISO/IEC 10646-1 Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architectural and Basic Multilingual Plane.**

[12]   ISO/IEC JTC1/SC22/WG20. **ISO/IEC FCD 14651 – International String Ordering – Method for comparing Character Strings and Description of the Common Template Tailorable Ordering.**

[13]   The Unicode Consortium. **Unicode 2.1 Character Charts.** *http://charts.unicode.org/charts.html.*

[14]   Thaweesak Koanantakool. **Computer and Thai Language: Development of Fundamental Standards for Thai Information Technology.** National Electronics and Computer Technology Center. Bangkok. 1991. (in Thai)

[15]   Thaweesak Koanantakool. **Information Technology and Open System Industrial Standards.** *http://www.nectec.or.th/it-standards/ITStaNew.htm.* (in Thai)

[16]   Sun Microsystem, Inc. **TLE 2.2 Release Overview.** 1994.

[17]   Microsoft Corporation. **Thai Character Sequence Checking Method.** *http://www.microsoft.com/globaldev/fareast/thai.asp.*

[18]   Microsoft Corporation. **Microsoft Windows Codepage : 874 (Thai).** *http://www.microsoft.com/globaldev/reference/sbcs/874.htm.*

[19]   Peter Edberg, Julio Gonzalez, and John Jenkins. **Map (external version) from Mac OS Thai character set to Unicode 2.0.** *ftp://ftp.unicode.org/Public/MAPPINGS/VENDORS/APPLE/THAI.TXT.*

[20]   Trin Tantsetthi. **An annotated reference to the Thai implementations.** *http://www.inet.co.th/cyberclub/trin/thairef/.*

[21]   Trin Tantsetthi. **Legitimate Thai Charset on the Internet.** *http://software.thai.net/alerts/tis-620/index.html.*

[22]   N. Freed and J. Postel. **IANA Charset Registration Procedures.** BCP 19, RFC 2278. January 1998.

[23]   The Royal Institute. **Royal Institute Dictionary 2525 B.E. Edition.** Aksorn Charoen That, Bangkok. 1995. (in Thai)

[24]   The Royal Institute. **The Principle of Punctuation Marks and Other Symbols, The Principle of Spacing, and The Principle of Abbreviation Writing.** Bangkok. 1990. (in Thai)

[25]   Mark Davis and Ken Whistler. **Draft Technical Report #10 : Unicode Collation Algorithm.** *http://www.unicode.org/unicode/reports/tr10-2.html.*

[26]   Theppitak Karoonboonyanan. **Thai Sorting Algorithms.** *http://www.links.nectec.or.th/~thep/tsort.html.*

[27]   Theppitak Karoonboonyanan, Samphan Raruenrom, and Pruet Boonma. **Thai-English Bilingual Sorting.** *http://www.links.nectec.or.th/~thep/blsort.html.*

[28]   Boonrueng Tanasuntonpaisan. **Design and Development of Interface for Thai Word Separation and Separator Insertion.** Master Thesis, Chulalongkorn University. 1990.

[29]   Samphan Raruenrom. **LibInThai : Platform Independent Thai word-break Library.** *http://www.thai.net/libinthai/.*

[30]   Samphan Raruenrom. **Mozilla Thai Enabling.** *http://www.thai.net/mozilla/.*

[31]   Kannikar Wimolkasem. *Influence of Sukhothai script in Lanna Thai Kingdom.* **Articles about Thai Language and Thai Scripts.** The Fine Arts Department. Bangkok. 1984. (in Thai)

[32]   Prasert Na Nakhon. *Lai Sue Thai.* **700 Years of Lai Sue Thai.** Dhurakit Bhundit. 1983. (in Thai)