

Winkler, Arnold F

From: Suwit Srivilairith [suwits@th.ibm.com]
Sent: Sunday, July 06, 2003 10:44 PM
To: Kent Karlsson
Cc: 'Kenneth Whistler'; 'Michael Everson'; 'Mark Davis (jtcsv)'; Winkler, Arnold F; Ranat Thopunya; Alain LaBonté
Subject: Fw: Thai ordering

Ken,

As I know, Theppitak, who wrote *Thai ordering in ISO/IEC 14651*, quitted from NECTEC. His e-mail at NECTEC may be invalid. Anyway, please let me share my idea on Thai ordering issue.

The standard of Thai ordering currently used in Thailand is **the Royal Institute Dictionary 2525 B.E. Edition**. I looked at the Thai ordering in ISO/IEC 14651:1999 (page 110-113) and I found that it uses the **same standard** (the Royal Institute Dictionary 2525 B.E.). In the ISO 14651, it also provides thai sorting implementation in the Algorithmic Aspect section. Furthermore, ISO 14651 suggests how to handle Thai decimal digits and Thai punctuation marks in Thai ordering (The Royal Institute Dictionary does not cover these characters).

Regarding to vowel ordering, I checked the Royal Institute dictionary. There are a list of 14 non-composites vowels in order (see Figure 1).

Figure 1: Vowel ordering list in the Royal Institute dictionary.

Also, there is a example of vowel combination in the Royal Institute dictionary (see Figure 2). This sample list contains 26 patterns. It may be useful sample for the users of the dictionary to look for a Thai word in dictionary.
 For the programmers who write Thai sorting function, A vowel list in the Figure 1 is sufficient to implement Thai sorting logic.
 The complete list of vowel combination is not necessary for Thai sorting logic.

1) ะ	10) ุ	19) เ ^๕ (เสื่อ)
2) ั (กัณ)	11) ู	20) เ ^๕ ะ (เกือะ)
3) ็ ะ (พัระ)	12) เ	21) แ
4) ิ	13) เ ะ (เกะ)	22) แ ะ (แพะ)
5) ิ ^๐	14) เ ิ (เขา)	23) โ
6) ิ ^๑	15) เ ิ ะ (เจาะ)	24) โ ะ (ไ้ะ)
7) ิ ^๒	16) เ ิ (เกิน)	25) ใ
8) ิ ^๓	17) เ ิ ^๔ (เสี่ย)	26) ใ
9) ิ ^๔	18) เ ิ ^๕ ะ (เพียะ)	

Figure 2: sample of vowel combination in the Royal Institute Dictionary

From the figure 2, the words between left and right parenthesis are samples of word pattern.

According to your question about dash, I am not sure because I do not have MacFarland dictionary. but I think that it can have more than one character per dash. One dash may be double consonants or consonant with a tonal mark.

Hope this helps!

Thanks and Have a nice day,
 Suwit Srivilairith
 National Language Development & Translation Services Center,
 GCoC-Bangkok, IBM Thailand
 Phone: (66-2)-273-4178, Fax: (66-2)-273-0686

----- Forwarded by Suwit Srivilairith/Thailand/IBM on 04/07/2003 16:40 -----

Ranat Thopunya/Thailand/IBM

To Suwit Srivilairith/Thailand/IBM@IBMTH

cc

04/07/2003 09:05

bcc

Subject Re: Thai ordering

Is Khun Theppitak still at NECTEC ?

regards,
 -ranat.

Manager, National Language Development & Translation Services Center,
 GCoC-CTL (Globalization Center of Competency - Complex Text Languages)
 Phone : (66-2)-273-4231, Fax : (66-2)-273-0686
 Infonet : 980-4231, e-mail : ranat@th.ibm.com

----- Forwarded by Ranat Thopunya/Thailand/IBM on 04/07/2003 09:04 -----

7/7/2003

Alain LaBonté

<alb@sct1.gouv.qc.ca>

To: "Kent Karlsson" <kent@cs.chalmers.se>, Theppitak Karoonboonyanan <thep@links.nectec.or.th>
 cc: "Kenneth Whistler" <kenw@sybase.com>, "Michael Everson" <everson@evertype.com>, "Mark
 Davis (jtcsv)" <mark.davis@jtcsv.com>, "Winkler, Arnold F" <Arnold.Winkler@unisys.com>, Ranat

03/07/2003 20:55

Thopunya/Thailand/IBM@IBMTH

Subject: Re: Thai ordering

Kent, the author of the text on Thai ordering in ISO/IEC 14651 is Theppitak Karoonboonyanan, from Thailand.

I send him your question, hoping that the email address I have is still valid (the email I have dates back to 1999).

Alain LaBonté
 Project Editor, ISO/IEC 14651

Secrétariat du Conseil du trésor du Québec
 Direction de l'architecture et de l'infrastructure
 1500-H, rue Jean-Talon Nord
 1er étage, porte 1.26, boîte n° 1
 Ste-Foy (Québec) G1N 4T5 Canada
 Téléphone : +1 418 528 6144
 Télécopieur : +1 418 646 3571

A 12:45 2003-07-03 +0200, Kent Karlsson a écrit :

>Hi!

>

> I'm still looking at Thai (and Lao) ordering issues (for a tailoring,
 >at least), even though the UTC did not want to consider changing the CTT.

>I checked "Thai-English student's dictionary" (Mary Haas, Stanford University

>Press, 1964) and "Thai-English dictionary" (George Bradley MacFarland, 1944).

>Ok, not the very latest dictionaries...

>

> Anyhow, both dictionaries claim that vowels are ordered according
 >to certain combinations. Each have their own list... though the latter

>claims

>to follow the "government dictionary" ordering. While the resumé in 14651

>only gives a list of 15 vowels (one of which is a composite), MacFarland's

>list the 14 non-composites as a "simple key to the complete list", and then

>lists 42 composite and non-composite vowels (in order). It is not quite

>clear if the dashes in that list always stand for one consonant or if there

>may be more than one per dash (such as double consonants [page vii]),

>but some of them seem to spread over two consonants (v-c-v-c-v).

>

> The Haas dictionary has a somewhat shorter list. An infamous

>"extended subset". In particular, it includes Thai r and Thai rr as vowels.

>

> And Lao?

>

> So has this been simplified after 1944 (and 1964)? Or is the list in

>the resumé in 14651 just a "simple key to a complete list"? If there has

>been

>a simplification, for who's benefit? Programmers writing collation programs

>for Thai? Or for the users of the dictionary? I guess the actual readings of

>these combinations haven't changed (much).

>

> I don't recall who wrote the text on Thai in 14651. Do any of you?

>Anyone of you knowing the answers to my questions (and implied questions)

>here?

>

> Kind regards

> /kent k

>