

กก	กำ	เกลียว
กรรม	กิน	เกา
กรรม	กี	เกาะ
-กระแฉ่ง	กิ้น	เกี้ยว
กราบ	กุน	เกียะ
กะเกณฑ์	กูด	เกือก
กัก	เกง	แกง
ก้าว	เกล้า	แกะ

โกน	แข่งขัน	ปาน
โกรน	แขน	พัด
ไกล	ครรภ-	ฯพณฯ
ไก	ครรภ	พนิชย์
ไกล	จุมพล	ย่อง
ขน	จุพล	รอง
ขนาน	ขาย	ฤทธิ์
ข้าง	เดา	ฤษี
ข้างๆ	เณร	ฤษี
ข้างกระดาน	ตลาด	ลลิตา
ข้างขึ้น	ทูลเกล้า	ภาษา
ข้างควาย	ทูลเกล้าฯ	วก
ข้างๆ คูๆ	ทูลเกล้าทูลกระหม่อม	ศาล
ข้างเงิน	ม	หริภุชชัย
ข้างออก	นา	หฤทัย
แขน	น้ำ	หลง
เข็น	นี้	แห่ง
แขน	บุญหลง	แห่ง
เข็ด	บุญ-หลง	แหนม
แข็ง	ป่า	แหนหวง
แข่ง	ป่า	แหบ
แข่ง	ป่า	แหม
แข่งขวา	ป่า	อาน
แข่งขัน	ป่า	ฮา

2 Algorithmic Aspect

The above principle, with appropriate character code assignment such as TIS-620 and ISO/IEC 10646, almost allows C standard library function **strcmp()** to collate Thai strings without much more complication, except:

1. Leading vowels (เ- แ- โ- ใ- ไ-), which are written before consonants, must be considered after the initial consonant. Therefore, the rearrangement is needed before comparison.
2. Diacritics and tone marks (̂ ̄ ̆ ̇ ̈ ̉ ̊) must be ignored in the first pass, and be considered at later pass if the first pass yields equality

And these are the only two mandatory requirements for Thai string collation algorithms. No syllable structure nor word boundary analysis is required, as Thai lexicons are ordered alphabetically, not phonetically.

2.1 Leading Vowel Rearrangement

To fulfill this requirement, either a preprocessing or collating-element grouping is required. The preprocessor scans the string once and swaps every leading vowel with its succeeding letter. The preprocessed string is then passed to the normal weight calculation process. Another way to manage this is by means of collating-element formation. Every possible pair of leading vowel and consonant is defined as a collating-element, whose weight equals to the weight of the rearranged substring.

Note that the rearrangement of a leading vowel is performed with its immediate succeeding consonant only. No consonant cluster analysis is needed. Instead, doing so would either face ambiguity problem or yield different order from the Royal Institute Dictionary.

For example, if consonant clusters were concerned, “เพลา” could be rearranged as either “เพลา” (two-syllable word เพล-ลา) or “พลเา” (one-syllable word, with consonant cluster “พล”).

To illustrate the different ordering caused by consonant cluster analysis, consider this conforming order: (เพล, เพลง, เพลศ). If consonant cluster were analyzed, it would be rearranged to (พลเ, พลง, พลศ), and would yield a different order: (เพลง, เพล, เพลศ) (if the consonant cluster “พล” were not grouped as a single collating element); or (เพล, เพลศ, เพลง) (if the consonant cluster “พล” were treated as a single collating element)

2.2 The Multiple Levels of Character Weights

The second requirement of diacritics and tone marks treatment implies multiple levels of weights. Tone marks and diacritics must be ignored in the first level, and weigh more than consonants and vowels in the second level.

There are ten Thai decimal digits (๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙), each semantically equivalent to Arabic digit 0-9, respectively. Their weights are then equal to their corresponding Arabic digits in the first level, and are different in the second level, to distinguish languages.

When punctuation marks (ฯ ๑ ๐ ๑๐ ๐๐) are concerned, another level of weights is required for them. This is correspondent to the fourth level in the Common Template Table. In string ordering, punctuation marks are less significant than any tone marks and diacritics, and must be ignored in all the first three levels.

For example, (ข้างฯ, ข้างกบ, ข้างฯ คุฯ, ข้างจัน) is a valid order in the Royal Institute Dictionary. In the first level, the considered weights are ขาง, ขางกบ, ขางคุ, ขางจัน respectively.