Thai Document Styles and Layouts

Virach Sornlertlamvanich
Thai Computational Linguistics Laboratory
Communications Research Laboratory (CRL)
Thailand
virach@crl-asia.org

Document Types

- Newspaper
- Textbook
- Magazine
- Manual
- Newsletter-Article-Letter
Newspaper

- Column
- Rare hyphenation
- Bold in the first line of text body
- Tab from the second paragraph
Textbook

- Preface, table of contents, book guide, chapters, authorized issue, editing committee, publishing information

- Tab from the first paragraph
Textbook
Magazine

- Decorated font
- Hyphenation
- Bold and italic fonts for emphasizing
Manual

- Header of the book title
- Session
- Punctuation
Newsletter/Article/Letter

- Header, logo
- Issue number, affiliation, date
- Subject, addressee, attachment
- Postscript, signature, printed name, title
- Contact person
Writing Style

- Deletion
- Hyphenation
- Caption
- Emphasizing mark
- Listing
- Indentation
- Alignment
Deletion

- Horizontal line across the whole word
- Horizontal line across each character
- Slash over each character
Hyphenation

- Syllable separator
  - line breaking and pronunciation
- Alternation “and” or “or”
- Duration “to”
- Detailing
- Listing
- Group separator
Caption

- Figure
- Table
- Picture

a) รูปที่ 1 กราฟแสดงรายได้ตลอดปี
   [Figure 1 Graph shows the annual income]

b) ตารางที่ 12.2 รายได้ประชาชาติประจำปี 2545
   [Table 12.2 GNP of the year 2002]

c) ภาพที่ 9 การเจริญเติบโตของทารกในครรภ์
   [Figure 9 Growth of the infant]
Emphasizing Mark

- Font style: bold, italic
- Underlining: single or double underline
- Quote: single or double quote
1. Sub-topic can be divided into “section” and each section can be sub-divided into 4 levels by using number and “.” as
   1
     1.1
       1.1.1
         1.1.1.1

2. If the second section should be divided, it can be also divided into 4 levels, and ( ) is introduced to identify each level. For example,
   1.1.1.1(1)   1.1.1.1(2)   1.1.1.1(3)
   1.1.1.1(1.1) 1.1.1.1(1.2) 1.1.1.1(1.3)
   1.1.1.1(1.1.1) 1.1.1.1(1.1.2) 1.1.1.1(1.1.3)
   1.1.1.1(1.1.1.1) 1.1.1.1(1.1.1.2) 1.1.1.1(1.1.1.3)

มหาสาระ (1) dddddddddddddddddddddd
  (2) dddddddddddddddddd
  (3) dddddddddddddddddd
  (4) dddddddddddddddddd
Indentation

1. dddddddddddddddddddddddddd
dddddddddddddddddddddddddd
1.1 dddddddddddddddddddddddddd
dddddddddddddddddddddd
1.1.1 dddddddddddddddddddddddddd
dddddddddddddddddd

2. dddddddddddddddddddddddddd
Alignment/Justification

- In *official letter*, the components are aligned restrictively. For example, date is center alignment, salutation is left alignment, and so on.
- *Caption* can be left, right or center alignment.
- *Sub-topic* is always aligned left.
- In the *table of content*, page number is right alignment.

- *Table*
  - Table header is usually center alignment.
  - Number is center alignment.
  - Text is left alignment.
  - Number (money) is right alignment.
Interesting Issues for Non-Segmenting Language

- Line breaking
  - Possible position:
    - End of word, then end of syllable (pronunciation unit)
  - Hyphenation:
    - Word unit is not definite. Hyphen is used to reduced ambiguity.

- Justification
  - Space position
  - Equally between characters
XML-based Document Style Service
XML-based Document Style Service
(External Message)