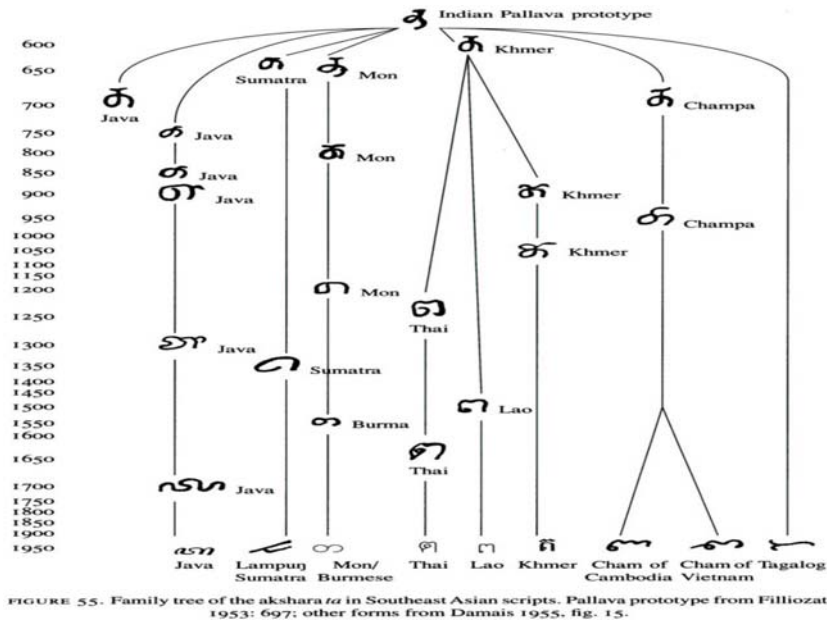


# Gap Analysis of HTML for Khmer

## 1. Introduction

The **Khmer script** is used to write the Khmer language . It derived from Indian Pallava. (Figure 5.5 Family tree of the akshara ta in Southeast Asian scripts; Pallava prototype from Filliozat 1953: 697; other forms from Damais 1995. fig. 15. See in document below). The Constitution of the Kingdom of Cambodia, article 5: Official Language and Script are Khmer Language and Khmer script.



The Khmer scripts consist of:

- 35 consonants (Khmer Unicode Script)
- 34 subscripts
- 13 independent vowels
- 20 dependent vowels
- 22 various signs in Khmer Unicode but only 19 various signs in Khmer Official Dictionary
- 2 consonant shifters
- Ligatures
- Numerals

**Consonants:**

There are 35 Khmer Unicode Consonants symbols but 2 consonants were obsolete (\*). They are:

កខគយង

ចឆជណញ

ដបឧធរណ

តថទធន

បផពភម

យរលវ គ\*ម\* សហឡ័អ

**Subscripts:**

Subscript Consonants are special figure used to form consonant clusters. Also sometimes referred to as "sub-consonants", subscript consonants often resemble the corresponding consonant symbol, only smaller. In Khmer, they are known as 'Cheung' (ទើង) meaning the *foot of a letter*. They are:

ក ខ គ ឃ ង

ច ឆ ជ ណ ញ

ត ថ ទ ធន

ប ផ ព ភ ម

យ រ ល វ

គ\* ម\* ស ហ ឡ័ អ

ឃ ង

ក ខ



័	17CC	KHMER SIGN ROBAT
័	17CD	KHMER SIGN TOANDAKHIAT
+	17CE	KHMER SIGN KAKABAT
័	17CF	KHMER SIGN AHSDA
័	17D0	KHMER SIGN SAMYOK SANNYA
-	17D1	KHMER SIGN VIRIAM
+	17D2	KHMER SIGN COENG
័	17D3	KHMER SIGN BATHAMASAT
័	17D4	KHMER SIGN KHAN
័	17D5	KHMER SIGN BARIYOOSAN
័	17D6	KHMER SIGN CAMNUC PII KUUH
័	17D7	KHMER SIGN LEK TOO
័	17D8	KHMER SIGN BEYYAL
័	17D9	KHMER SIGN PHNAEK MUAN
័	17DA	KHMER SIGN KOOMUUT
័	17DB	KHMER CURRENCY SYMBOL RIEL
័	17DC	KHMER SIGN AVAKRAHASANYA
័	17DD	KHMER SIGN ATTHACAN

## Consonant Shifters

These signs shift the base consonant between registers.

◌̣ ◌̤

## Ligatures

Most consonants, including a few of the subscripts, form ligatures with all dependent vowels. Many of these ligatures are easily recognizable but a few may not be recognized due to the shape transformation when it is combined with the vowel *a* (ា), *ao* (ៅ) and *au* (ៅ).

Examples of ligatured transform:

ក + ា = កា (kaô)

ផ + ៅ = ផៅ (pau)

ច + ា = ចា (cha)

ប + ា = បា (ba)

Ligatured consonant subscript and vowel combination:

ច + ្រ + ា = ច្រា (chba)

ម + ្រ + ៅ = ម្រៅ (msau)

## Numerals

The numerals of the Khmer script, similar to other civilizations in Southeast Asia. Arabic numerals are also used, but to a lesser extent.

<b>Khmer numerals</b>	០	១	២	៣	៤	៥	៦	៧	៨	៩
<b>Arabic numerals</b>	0	1	2	3	4	5	6	7	8	9

## 2 Analysis of Hand Written:

### 2.1 Text Selection

#### Khmer Script in hand written:

- Written in Bold Text <b> (missing title and author, only some pages available, 8, 9, 33, see 2.2.1)
- New line <br> (missing title and author, see 2.2.1)
- Horizontal rule <hr> (missing title and author, page 33, see 2.2.1)
- Defines centered text <center> (missing title and author, only some pages available 8, 9, see 2.2.1)
- Defines a long quotation <blockquote> (missing author and title, page 8,9 see 2.2.1)
- Defines an ordered list <ol> (missing author and title, page 57 see 2.2.1)

#### Older Published Documents:

- New line <br> (author: ពោល សាមី (Pov Samy); title: Theory of Khmer Language all pages, see 2.2.1)
- Horizontal rule <hr> History Book posted 1987( title: History Book, Author: Ministry of Education, page 8 see 2.2.1)
- Defines centered text <center> Khmer Language Book (author: ខ្មែរ សុខុម្ពុ (Khuon Sokphou, title: Khmer Language, page 32 see 2.2.1)
- Defines an ordered list<ol> (author: ស្រីសង្ហា សារិន (Sreng Sarin), title: Cambodian History Book posted in 1988, page 2 see 2.2.1)
- Underlined Text <u> in Khmer Language Book (author: ស្រីសង្ហា សារិន (Sreng Sarin), title: Cambodian History, page 32 see 2.2.1)

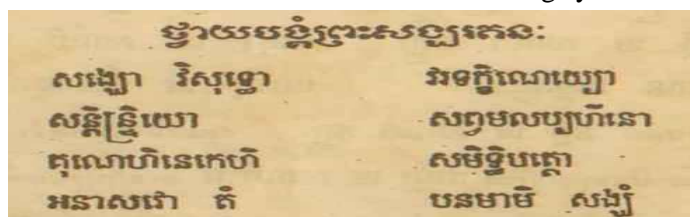
## 2.2 Analysis

Discuss on all the formatting conventions of Khmer Language the study of the selected text as follow:

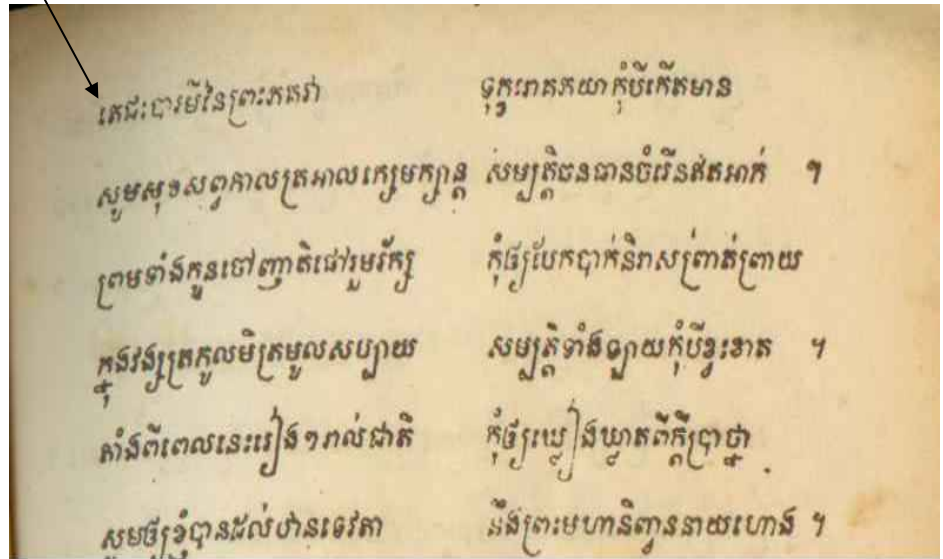
### 2.2.1 Format of Feature:

#### Hand written documents

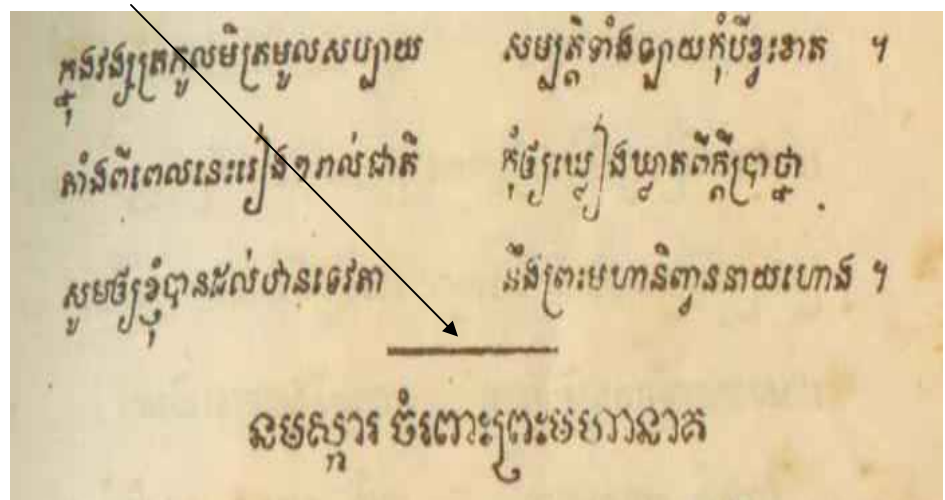
Bold <b> there were bold texts in Khmer writing system.



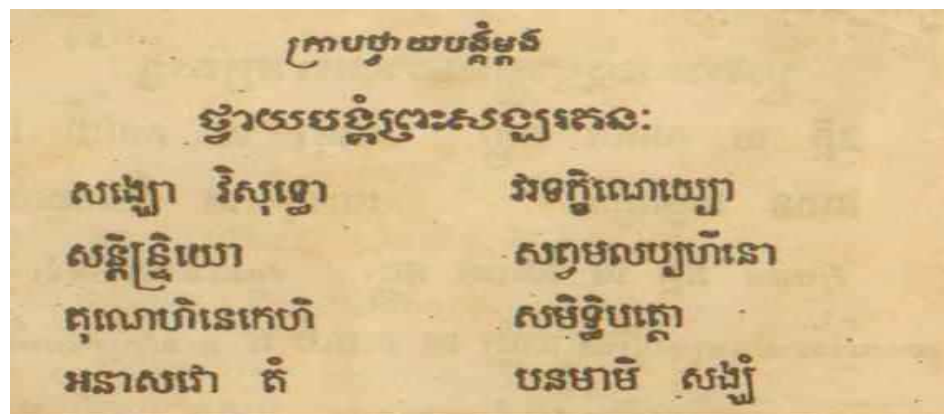
New line <br> :



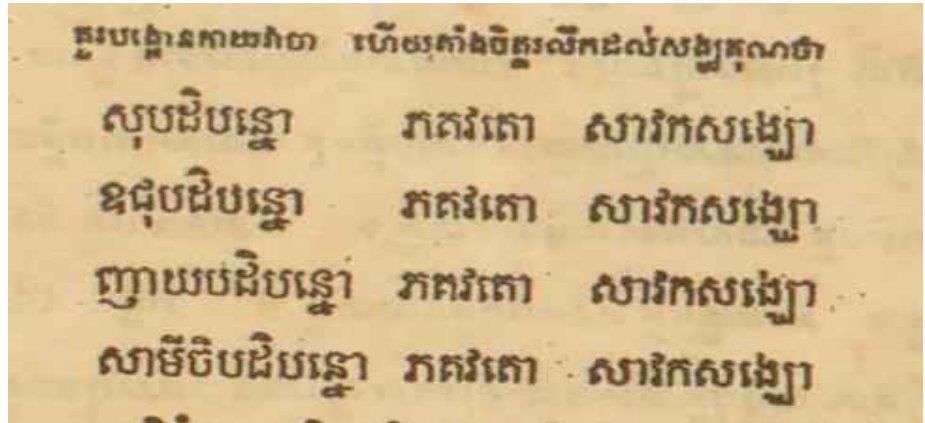
Horizontal Rule <hr> there was horizontal rule in Khmer writing system.



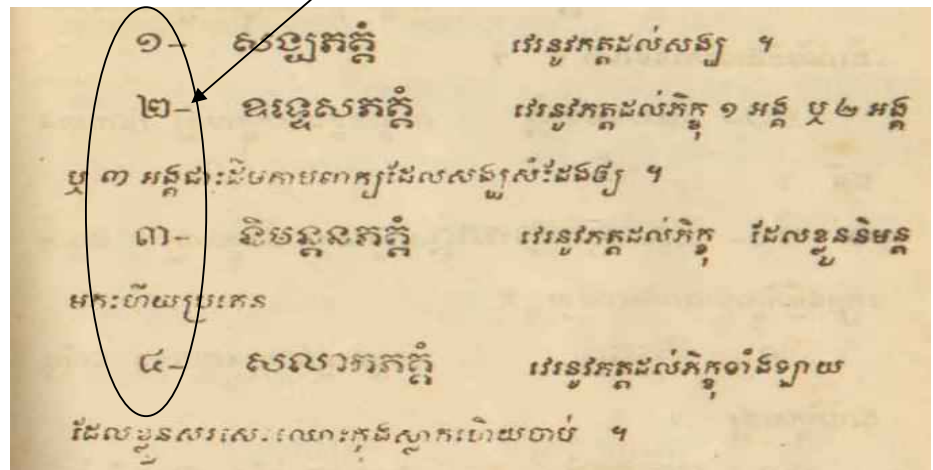
Centered text<center>



Long quotation <blockquote> it had a long quotation text in Khmer writing system.

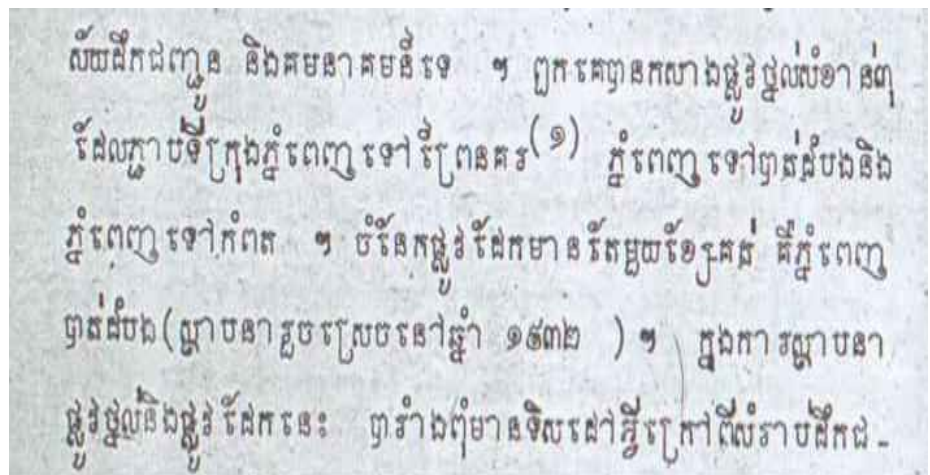


Ordered list <ol> Bullets (order list) were used in Khmer writing system.



### In Older Published Documents

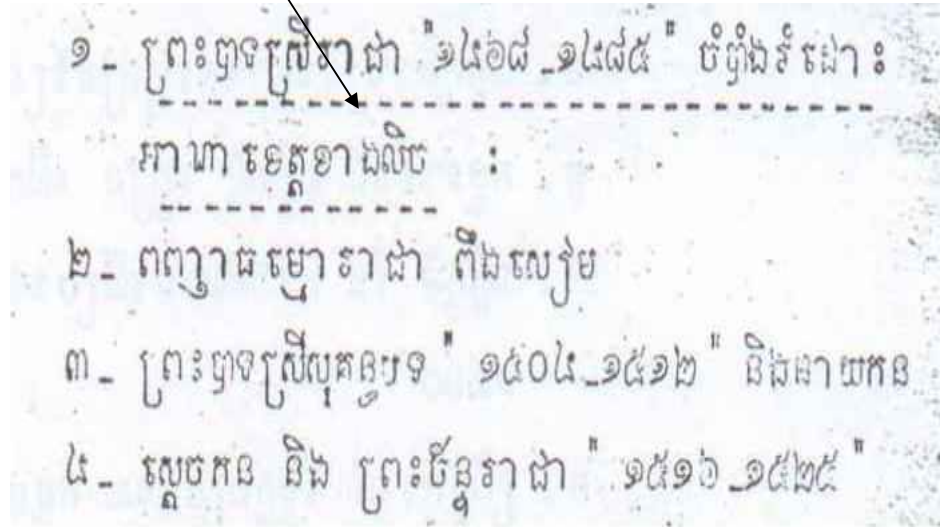
New line <br>







Underlined Text<u>



### 3 Comparison with HTML Tag:

List of all formatting conventions which are already supported in HTML in Hand Written Documents and Older Published Documents

HTML Tag	Khmer Convention	Already Supported
<b>Basic Tags</b>		
<h1> to <h6>	Defines header 1 to header 6	✓
 	Inserts a single line break	✓
<hr>	Defines a horizontal rule	✓
<b>Char Format</b>		
<b>	Defines bold text	✓
<font>	Deprecated. Defines text font, size, and color	✓
<em>	Defines emphasized text	✓
<big>	Defines big text	✓
<strong>	Defines strong text	✓
<small>	Defines small text	✓
<sup>	Defines superscripted text	✓
<sub>	Defines subscripted text	✓
<u>	Deprecated. Defines underlined text	✓
<b>Output</b>		
<pre>	Defines preformatted text	✓
<code>	Defines computer code text	✓
<tt>	Defines teletype text	✓
<kbd>	Defines keyboard text	✓
<var>	Defines a variable	✓

<u>&lt;dfn&gt;</u>	Defines a definition term	✓
<u>&lt;samp&gt;</u>	Defines sample computer code	✓
<b>Blocks</b>		
<u>&lt;blockquote&gt;</u>	Defines a long quotation	✓
<u>&lt;center&gt;</u>	Deprecated. Defines centered text	✓

List of all tags which are not found in Khmer conventions and also not relevant to Khmer Writing System

HTML Tag	Purpose	NN	IE
<b>Basic Tags</b>			
<u>&lt;!DOCTYPE&gt;</u>	Defines the document type		
<u>&lt;html&gt;</u>	Defines an html document	3.0	3.0
<u>&lt;body&gt;</u>	Defines the body element	3.0	3.0
<u>&lt;p&gt;</u>	Defines a paragraph	3.0	3.0
<u>&lt;!--...--&gt;</u>	Defines a comment	3.0	3.0
<b>Char Format</b>			
<u>&lt;i&gt;</u>	Defines italic text	3.0	3.0
<u>&lt;bdo&gt;</u>	Defines the direction of text display	6.2	5.0
<b>Blocks</b>			
<u>&lt;acronym&gt;</u>	Defines an acronym	6.2	4.0
<u>&lt;abbr&gt;</u>	Defines an abbreviation	6.2	
<u>&lt;address&gt;</u>	Defines an address element	4.0	4.0
<u>&lt;q&gt;</u>	Defines a short quotation	6.2	4.0
<u>&lt;cite&gt;</u>	Defines a citation	3.0	3.0
<u>&lt;ins&gt;</u>	Defines inserted text	6.2	4.0
<u>&lt;del&gt;</u>	Defines deleted text	6.2	4.0
<u>&lt;s&gt;</u>	Deprecated. Defines strikethrough text	3.0	3.0
<u>&lt;strike&gt;</u>	Deprecated. Defines strikethrough text	3.0	3.0
<b>Links</b>			
<u>&lt;a&gt;</u>	Defines an anchor	3.0	3.0
<u>&lt;link&gt;</u>	Defines a resource reference	4.0	3.0
<b>Frames</b>			
<u>&lt;frame&gt;</u>	Defines a sub window (a frame)	3.0	3.0
<u>&lt;frameset&gt;</u>	Defines a set of frames	3.0	3.0
<u>&lt;noframes&gt;</u>	Defines a noframe section	3.0	3.0
<u>&lt;iframe&gt;</u>	Defines an inline sub window (frame)	6.0	4.0
<b>Input</b>			

<u>&lt;form&gt;</u>	Defines a form	3.0	3.0
<u>&lt;input&gt;</u>	Defines an input field	3.0	3.0
<u>&lt;textarea&gt;</u>	Defines a text area	3.0	3.0
<u>&lt;button&gt;</u>	Defines a push button	6.2	4.0
<u>&lt;select&gt;</u>	Defines a selectable list	3.0	3.0
<u>&lt;optgroup&gt;</u>	Defines an option group	6.0	6.0
<u>&lt;option&gt;</u>	Defines an item in a list box	3.0	3.0
<u>&lt;label&gt;</u>	Defines a label for a form control	6.2	4.0
<u>&lt;fieldset&gt;</u>	Defines a fieldset	6.2	4.0
<u>&lt;legend&gt;</u>	Defines a title in a fieldset	6.2	4.0
<u>&lt;isindex&gt;</u>	Deprecated. Defines a single-line input field	3.0	3.0
<b>Lists</b>			
<u>&lt;ul&gt;</u>	Defines an unordered list	3.0	3.0
<u>&lt;dir&gt;</u>	Deprecated. Defines a directory list	3.0	3.0
<u>&lt;dl&gt;</u>	Defines a definition list	3.0	3.0
<u>&lt;dt&gt;</u>	Defines a definition term	3.0	3.0
<u>&lt;dd&gt;</u>	Defines a definition description	3.0	3.0
<b>Images</b>			
<u>&lt;img&gt;</u>	Defines an image	3.0	3.0
<u>&lt;map&gt;</u>	Defines an image map	3.0	3.0
<u>&lt;area&gt;</u>	Defines an area inside an image map	3.0	3.0
<b>Tables</b>			
<u>&lt;table&gt;</u>	Defines a table	3.0	3.0
<u>&lt;caption&gt;</u>	Defines a table caption	3.0	3.0
<u>&lt;th&gt;</u>	Defines a table header	3.0	3.0
<u>&lt;tr&gt;</u>	Defines a table row	3.0	3.0
<u>&lt;td&gt;</u>	Defines a table cell	3.0	3.0
<u>&lt;thead&gt;</u>	Defines a table header		4.0
<u>&lt;tbody&gt;</u>	Defines a table body		4.0
<u>&lt;tfoot&gt;</u>	Defines a table footer		4.0
<u>&lt;col&gt;</u>	Defines attributes for table columns		3.0
<u>&lt;colgroup&gt;</u>	Defines groups of table columns		3.0
<b>Styles</b>			
<u>&lt;style&gt;</u>	Defines a style definition	4.0	3.0
<u>&lt;div&gt;</u>	Defines a section in a document	3.0	3.0
<u>&lt;span&gt;</u>	Defines a section in a document	4.0	3.0
<b>Meta Info</b>			
<u>&lt;head&gt;</u>	Defines information about the document	3.0	3.0
<u>&lt;title&gt;</u>	Defines the document title	3.0	3.0
<u>&lt;base&gt;</u>	Defines a base URL for all the links in a page	3.0	3.0

List of all tags which are not supported in HTML and Require New Tags to be introduced for Khmer Writing System

<select>	Defines a selectable list
----------	---------------------------

For this <select> HTML tag it support with Khmer Unicode when using Mozilla Brower, but it not work with Internet Explorer Browser.

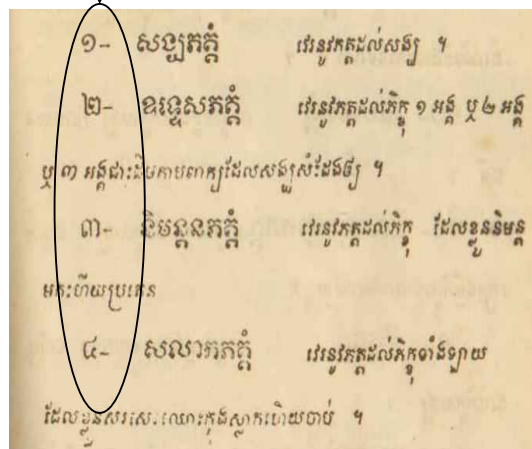
<ol>	Defines an ordered list	Not Support
------	-------------------------	-------------

## 1. Ordered list

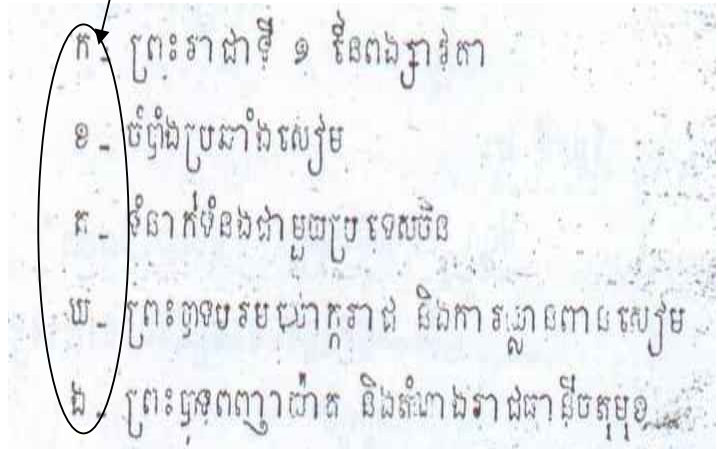
Ordered lists or Bullet in Khmer Writing System; starting with "១, ២, ៣" or ក,ខ,គ.... it should auto display numerically or alphabetically order, but it not support with Khmer Unicode.

<OL Type="list type">

Type="១" - Display in numerals

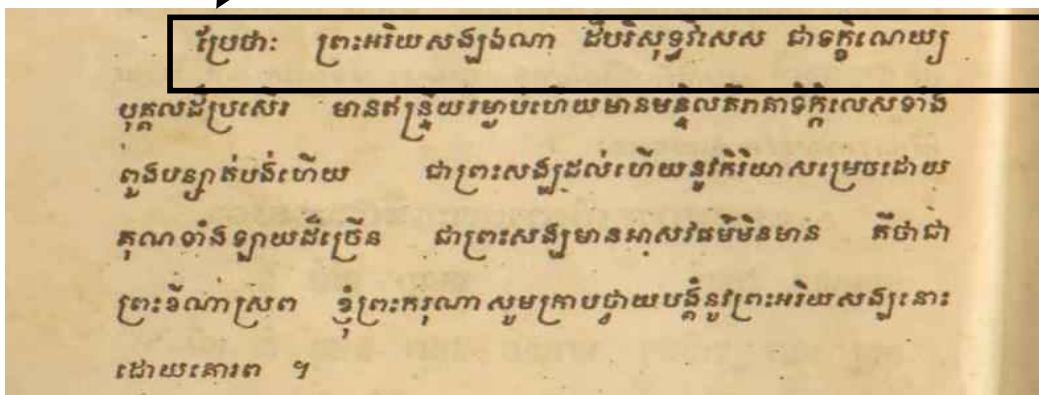


Type="ក" - Display in Khmer Consonant

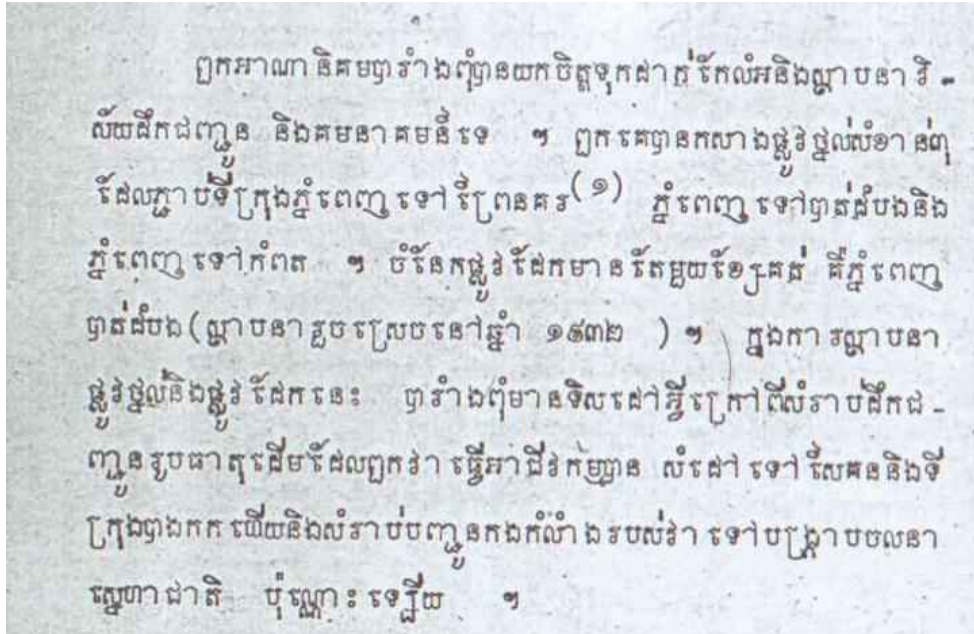


## 2. New Line

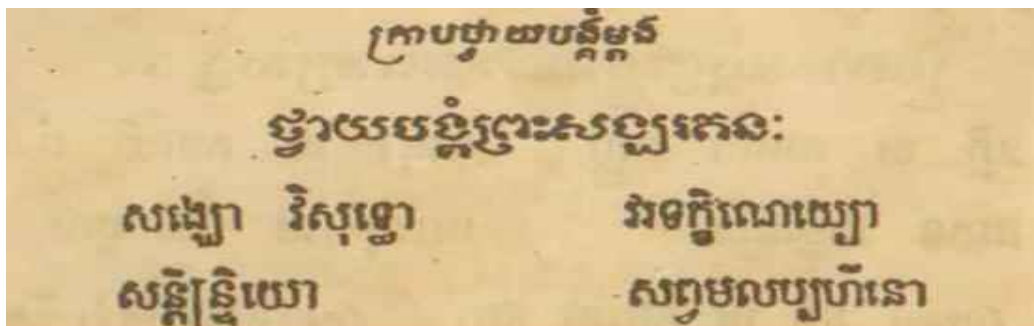
In Khmer Writing System, INDENT is always applied for new lines, HTML tag is not support this function.



3. Justify Khmer Unicode Writing, we would request HTML tag to support it.



4. In Khmer POETRY Writing System, in order to write poem, it used 2 columns and all functions above. We would request HTML tag to support these functions also.



Reference:

<http://www.unicode.org/charts/PDF/Unicode-4.1> (Range: 1780-17FF for charts showing only the characters added in Unicode 4.1)

<http://www.w3schools.com> (showing all html tag)

DICTIONNAIRE CAMBODGIEN (EDITION DE L'INSTITUT BOUDDHIQUE 1967)

Author: សម្តេចប្រះសង្ឃនាថ ជ.ណ. ជោតញ្ញាណ (Samdech Choun Nat)