Hope for the Lotus

Concluding Remark for the
EBTI / **CBETA** Conference 2008

Hope for the Lotus:

A Technical Perspective on Digital Buddhist Systems

Ching Chun Hsieh, Ph.D.
Chair Professor, Ming Chuan University

February 17, 2008

Outline

- On the interoperability of digital buddhist texts
- Content Markup
  - A way of processing the meanings of texts
- Conclusions

On the interoperability of digital buddhist texts

Digital Buddhist Text
vs.
Traditional Buddhist Text

In creating digital text, the followings are added to the traditional text:

- Link Information
  - Indexing words, subjective headings, keywords, hyperlinks …
- Context Information
  - Meta-data, Historical Calendar, GIS …
- Content Information
  - Explanations, interpretations, translations, comments …

Why add those Information?

- For text processing in computer
  - Without those information, the text is completely useless.
- More information, more powerful the processing
  - The text and those added information should properly organized and integrated in computer to facilitate the processing / applications of the text.

After those information are added…

- The digital text is no more the tradition text.
  - You may consider the definition of text changes while text becomes “digital”.
- So, the Digital Buddhist Text is different from the traditional Buddhist Text.
- **How?**

The representation of digital text

<table>
<tr><th colspan="2">Digital Text</th><th>Expression system</th></tr>
<tr><td colspan="2">Text itself</td><td>Natural Language</td></tr>
<tr><td rowspan="3">Text to the outside world</td><td>Link Information</td><td rowspan="3">Meta-Language</td></tr>
<tr><td>Context Information</td></tr>
<tr><td>Content Information</td></tr>
</table>

Digital Text

- Digital Text is bi-lingual in nature.
  - A Natural language and a meta-language work together to represent a digital text.
  - Meta-language is a kind of artificial language which can be processed by computer.
    - ***Therefore, the information described by meta-language should be interoperable / compatible among different systems.***

Interoperability of digital Buddhist text

- The link information, the context information and the content information of digital buddhist text should be and could be make them compatible / interoperable.
  - Although the those information may use different natural language for their “values”, but their forms can be make compatible.
    - *It includes the compatibility of syntactic and related semantic specifications of forms described by meta-language.*

Examples

- Indexing words, subjective headings, keywords, hyperlinks …
- All metadata sets
- Any markup specifications and rules
- Any specific context information
  - Context about author, translator, culture, social …
- Explanations, interpretations, translations, comments …

A few words about context

- Once a text being published, all its context has been fixed forever.
- Meaning is context sensitive. Once context has been fixed, the original meaning intended/expressed by the author also been fixed.
- Meaning can not be pin-pointed unless the relevant context has been properly and sufficiently described and understood.

A few words about context

- Digital Buddhist Context information should cross the boarders of nations, races, parties…
- Also, it should be interoperable among various hardware/software/networking systems.

A review of markups being used

Markups being used

- Metadata usually reveals facts or attributes of an object. Therefore, metadata can be verified, but can not be interpreted.
  - We can not apply Hermeneutics to explain an item of metadata.
  - The meaning of a metadata item is determined by the semantic specification of that field.

Markups being used

- They are all a selected set of context information aiming for a specific purpose for that a metadata set is designed.
- The tag set (or the fields) of a metadata reveals some common phenomena (共相) of a specific class of objects collected.
  - Metadata is not suitable to provide unique phenomena (别相) of an object.

Content Markup

What is content markup?

- Content markup is designed to explicitly denoted the meanings of a piece of text.
  - The explanation, interpretation, feeling, comparison, criticize, comment …of a text can be denoted by content markup.
  - Content markup is a tool to record the results of understanding of text by human beings.
  - So, content markup is a tool designed for Humanities Studies, such as History, Literature, Philosophy, Aesthetics, Sociology, etc.

The Significant of Content Markup

- Content markup is suitable to provide unique phenomena (別相) of a text (or an object).
- Content markup suggests a man-machine co-operation system that allows computer to have semantic processing capability by inputting the meaning of text markup by human expert.

Examples of Content Markup

- The Mandarin punctuation symbols
  - Classic writings in Chinese language do not use punctuation symbols.
  - Modern Chinese writings use punctuation symbols to clarify the meaning of text.
  - Punctuation symbols help people to understand text.
  - Punctuation symbols also helps writers to explicitly express the meanings of text.

A classic sentence as an example

- ***Without punctuation symbols***

  民可使由之不可使知之

- ***With proper punctuation symbols***

  民可，使由之；不可，使知之。

- ***With improper punctuation symbols***

  民，可使由之，不可使知之。

Examples of Content Markup

- Comments symbols used in classical Chinese writing (句讀符號)
  - A set of comments symbols is used to markup good or bad rhetorical wordings, impressive or nonsense parts of text, etc.
  - These symbols helps reader understand the content of text. They also helps reader to put down his/her understanding/opinion about the text.

Conclusions

Context Info. vs. Content Info.

- Context information and content information are different kind of information. They are complements each of the other.
  - ***Context information describes the relevant background information external to the text.***
  - ***Content information explains the meaning of the text.***

The Feasibility of “Meaning Processing”

- Meanings of texts can be processed by computer, if people will work together with computer:
  - Computer alone can not do “meaning processing”.
  - People can understands the meanings of texts, even there are ambiguities. Content markup tools will convey the meanings understood by people and put them into computer for further processing.
- The ability of representing sufficient context information of texts is crucial to solve the ambiguity of meanings of texts.

Hope for the Lotus

- As Professor Lancaster addressed in his opening keynote:

  "***One of the themes of this conference will be to discuss ways in which we can join forces for the sake of mutual support as well as action for interoperability of data sets***."

- ***I believe we all share these points of view.***

Hope for the Lotus

- But, what should the next-generation digital buddhist systems look like?
- When shall we start? by which way? and, how?
- These are open problems for us at this moment.

Hope for the Lotus

- As interoperability of digital buddhist text is concerned:
  - All the context information described by any meta-language, such as *SGML*, *HTML*, *XML* etc., could and should be make them sharable among all of us.
  - All technical information of applying meta-language on text, such as markup rules (including both syntactic and semantic specifications), tag sets, etc., should be open and accessible.

Hope for the Lotus

- The semantic processing capability is the keynote for the next-generation digital buddhist systems.
  - ***Content mark tools are indispensable for future digital buddhist text.***
- And lastly, I have the confident that:

We can do it !

Thank you !

The
End