

A Study on Creation and Application of Electronic Chinese Buddhist Texts: With the Version Comparison and the Commentaries of Yogācārabhūmi as a Case Study

*Huimin Bhikkhu, Christian Wittern, Aming Tu,

Li Juan Guo, Ray Chou

*Professor, Taipei National University of the Arts, Taiwan

Summary

The project of the first year is an attempt to create and markup electronic data. We try to use tag sets of TEI and markup the documents (including contextual features such as <monogr>, <series>, <analytic>; structural features such as <div>, and nonstructural features such as <item>) based on the various translations, and the outlines book of the *Yogācārabhūmi-śāstra*.

Second, the project described above will be cross-referenced with structural features of the electronic data, providing, for example <linkGrp><link targets="A1 B1 C1"/><linktargets="A2 B2 C2"/></linkGrp>, <link tagsets="n1579DD065 n1581DB002 n1582DB001"/>).

Interfaces will be developed, such as using the xml2dir.bat program on Sutra XML with IE5, xml parser, and DHTML and JavaScript to provide menus. The Link will provide cross references through html.bat. This will be useful for comparing different translations and outlines.

The project of the second year takes the *Yogācārabhūmi-śāstra* (YBh) and its Commentaries to explore the problems and solutions in processing electronic Buddhist texts-- the creation of markups, the link between the text and its Commentaries, full text search and web development.

First of all, we use the tag sets of TEI (*the Text Encoding Initiative*) as the foundation to develop the tag sets for Buddhist text and its Commentaries,. Examples of using tag sets include (1) `<cit><prname></prname><q></q></cit>` for the citation of the commentator, (2) `<cit><bibl><title type="lun"></title><biblScope></biblScope>` for the citation form the other Buddhist texts, (3) `<q type="unknown"></q>` for the unknown citation, (4) `<xref doc="" targets=""></xref>` for the cross-referenced link between the citation and its original texts, (5) `<term id="" type="yinyi"></term>`, `term id="" type="yiyi" ></term>`, `<gloss targets="" ></gloss>` for the Chinese explainaiton of Sanskrit term, (6) `<q type="proof "></q>` for the Buddhist logic argument, (7) `<pp`

id="T1828xp0001"></pp> and <pp targets=" "></pp> for the pinpoint link between YBh and its Commentaries, and (8)<sic resp="lijuan" cf1="T1579" corr=""></sic> for the correction for printing errors.

Secondly, we display the XML file of the sutra with xml2dir.bat and use IE5, xml parser, DHTML as well as JavaScript to display the table of content. Upon user selection, the contrast among various translated versions and its Commentaries is then displayed with html.bat Link function. Although the function of cookie is not yet provided, but will be further developed. Moreover, the toolbar is completed to provide resources and functions such as the introduction, the ordered hierarchy, searching of Sanskrit-Chinese-Tibetan term, index for searching proper nouns, and the bibliography. At last, we convert the marked-up electronic files into web files and put them on line for public use (<http://ybh.chibs.edu.tw>) .

Keywords : Electronic Chinese Buddhist Texts, the Commentaries of *Yogācārabhūmi*, Markup Language, TEI(*the Text Encoding Initiative*), tag sets

I. Background and Objectives of this Research

Buddhist texts feature a tremendous amount of valuable cultural treasure shared by all human beings. However, the enormous corpus of Buddhist texts as well as their secondary sources usually overwhelms interested researchers, students, or even teachers. Hence, it shall be a significant contribution to Buddhism Studies and an innovative move in the field of human science to employ the digital and electronic media and tools of contemporary information technology, and to utilize the characteristics of multivalent texts as well as standardized markup language (such as SGML, HTML, XML) in the management of vast Buddhist information so as to facilitate more efficient teaching, researching and information service.

The autobiography of Xuanzang (602-664) indicated that one of his goals to study in India was to learn and translate *Yogācārabhūmi-śāstra* (*YBh*). The *Yogācārabhūmi* records different levels of practice and achievements of a yogi (a practitioner or a Meditation practitioner). Encompassing the psychological, philosophical, and religious dimensions of the rich experiences of explorers who pursue the subtle spiritual life of enlightenment, the *YBh* can also serve as an encyclopedia for Mahayana Buddhism. Hence, this research shall focus on the *YBh* and aim to achieve

the following goals.

- 1 Take the *YBh*, its outline books, and various translations as an example and explore the problems and solutions in processing electronic Buddhist texts, the treatment of rare characters, proofreading, and the creation of markups, full text search and web system.
- 2 Explore possibilities for the application of the digitalized data obtained via the previous process in the development of user interface, research methodology and teaching activities.

II. Achievements of the Research and Discussions (1st years project)

1. Provide an actual sample case for the development of electronic Chinese Buddhist Texts

After one year of experiment and practice, with the *YBh*, its various translations, commentaries and the outlines book as an example in exploring the possibilities to benefit from the characteristics of electronic texts, a three-step process to create and markup the texts proves to be feasible and effective. It can be described as follows,

(1) Create and markup the contextual features of the document:

Take the compilation of the bibliography of the *YBh* for example. Its various references can be marked up as <monogr> (single volume books), <series> (books in series), <analytic>(periodical papers) with TEI tag sets. And meanwhile, these marked-up references can also be linked to annotations, the full text of the *YBh* and other reference books.

(2) Create and markup the structural features of the documents:

Treat the various translations of the *YBh* and the content of their outline books as ordered hierarchy instead of linear texts and thus structure the original chapters into 20 divisions, then mark them with <div> and check them with HTMLHelp functions.

(3) Create and markup the non-structural features of the documents:

Use the dictionary entries in *Yogacam Dictionary* as an index for searching proper nouns, keywords in the *YBh* and mark these entries with <item>.

2. Sample Application of Chinese Buddhist Texts: Comparing Various Translations and Contrasting the Outline Book

The digitalized and marked-up electronic texts thus created can be applied to the comparison and contrast among various translations and the outline book. This can be done in three steps.

(1) Create cross-reference markups among related texts according to their structural features:

YBh (n1579), various translations (n1581-4), commentaries (n1580, 1828-9, 2259, 2802-3, S0794) and the outlines book (n1602) are marked up as div1=DA, div2=DB, div3=DC, div4=DD, ...etc.). And add linking id, for instance, div1=n1579DA, div2=n1579DB, etc. The numbering can be done with a program that can add series numbers. Next, to add links markup. (See TEI, 14.4.3:A Three-way Alignment, pp. 433) :

```
<linkGrp><link targets="A1 B1 C1"/> <link targets="A2 B2 C2"/>
```

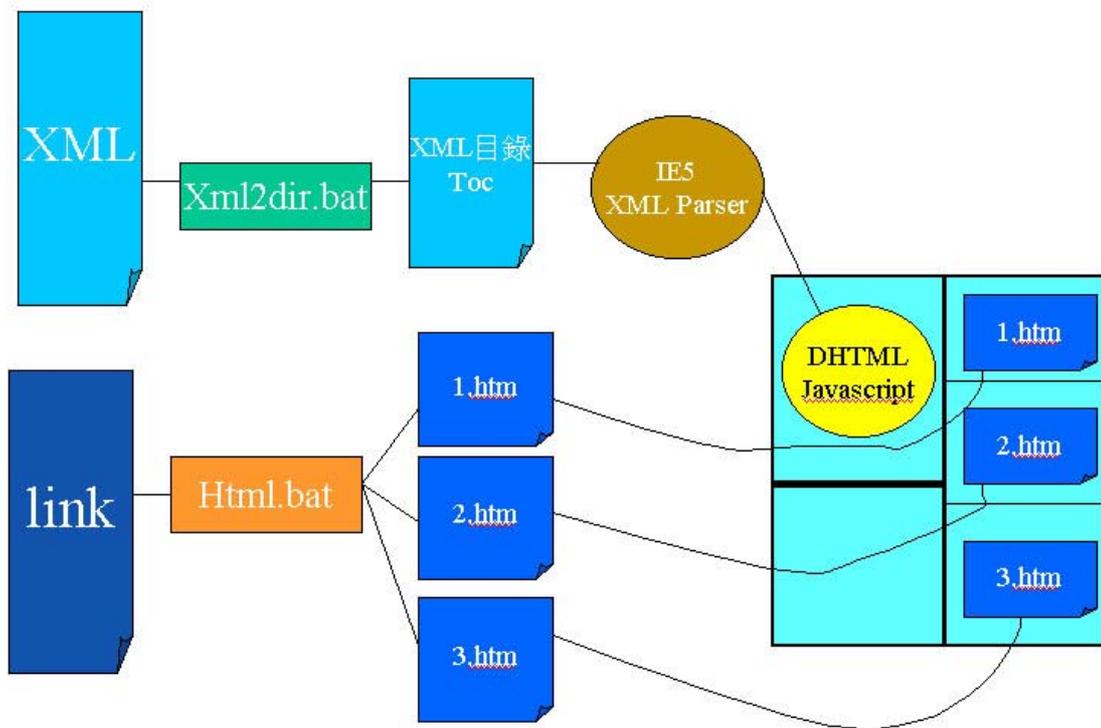
```
</linkGrp>. Related various translations and the outline book can be linked according to their ids. (e.g.:<link tagsets="n1579DD065 n1581DB002 n1582DB001 "/> )
```

(2) Developing a User Interface

Display the XML file of the sutra with xml2dir.bat and use IE5, xml

parser, DHTML as well as JavaScript to display the table of content.

Display the contrast among various translated versions with html.bat Link function, as illustrated in the following:



(3) Put the electronic files on the Internet

Convert the marked-up electronic files into web files and put them on line for public use (<http://ybh.chibs.edu.tw>).



The marked up electronic texts can be used for comparing and contrasting various translations and the outline book.



III. Research Purpose of 2nd year project

This study, conducted from 90/08/01 to 91/07/31, is the 2nd year of

research on creation and application of electronic Buddhist texts of *Yogācārabhūmi*. It takes the following 7 Commentaries of *Yogācārabhūmi* as examples to study the input, treatment of rare characters, proofreading, the creation of markups, the link between the text and its Commentaries, full-text search and web development.

1. Yuqieshidilun shi 《瑜伽師地論釋》 by Jinaputra, translated by Xuanzang, 1 volume (T30, 1580);
2. Yuqielun ji 《瑜伽論記》 Ed. by Dunlun, 48 volumes (T42, 1828);
3. Yuqieshidilun Luezwan 《瑜伽師地論略纂》 by Kuiji, 16 volumes (T43, 1829);
4. Yuqielun Q&A 《瑜伽論問答》 by Zenho, 7 volumes (T65, 2259);
5. Yuqieshidilun Fenmen ji 《瑜伽師地論分門記》 by Facheng, 智慧山記, 6 volumes (T85, 2801);
6. Yuqieshidilun So ji 《瑜伽師地論手記》 by Facheng, 福慧記, 4 volumes (T85, 2802);
7. Yuqieshidilun Jechang Song 《瑜伽師地論劫章頌》 by Kuiji, 1 volume (卅 47).

IV. Result and discussion (2nd years project) :

1. Basic procedure – markups with the TEI tag sets

The TEI (*the Text Encoding Initiative*) tag sets were used to mark up the following three kinds of features of the documents: (1) contextual features, (2) structural features—using the “div” tag to mark up the 54 levels of content structure in *Yuqielun ji* 《瑜伽論記》 (T1828), E.g. `<div54 n="1" type="L54"><head type="added">1 第一、不供三寶戒</head>`, 參 T42.1828.538b23), (3) non-structural features for cross-references. The above-mentioned features are linked via the file “link.xml” which links the id of documents. E.g. `<link targets="T1579D02_001 T1828D06_017 T1829D09_001T1580D03_002 T2801D7001"/>` links documents for “The stage of the five consciousnesses associated with the body *-pa~nca-vij~naana -kaaya- sa.mprayukta- bhumi*” (「1 五識身相應地」). The user-interface is window-based which consists of two columns (frames) – the column for directory and the column for texts (cross-referenced documents). The directory column shows the table of content by using `xml2dir.bat`, IE5 xml parser, DHTML, and Javascript. Upon user selection, the contrast among various translated versions and its Commentaries is shown in the text column via program `html.bat`.

2. The use of tag sets for Commentary – taking *Yuqielun ji* 《瑜伽論記》

(T1828) as an example

This study develops the following tag sets for the 48 volumes of *Yuqielun ji* 《瑜伽論記》 by Dunlun.

- (1) Using `<cit><prname></prname><q></q></cit>` to mark up citation of the commentator. For example,

```
<cit><prname key="zh-WB-x-0648AD">備</prname>云：<q>於五蘊中，隨何<br><lb n="0353b29"/>等蘊計爲實我，故云等隨。</q></cit>
```

- (2) Using `<cit><bibl><title type="lun"></title><biblScope></biblScope>` (or `<title type="jing"></title>`) `</bibl><quote></quote></cit>` to mark up citation of other Buddhist texts. For example,

```
<cit><bibl><title type="lun">對法</title><biblScope>第六</biblScope></bibl>說：<quote>不寂靜是諸煩惱共<br><lb n="0352c28"/>相。</quote></cit>
```

- (3) Using `<q type="unknown"></q>` to mark up the unknown citation (without reference), i.e., “it is interpreted as”, “it means”, “it is said”, “some said”, “it’s mentioned somewhere” (「有釋」、「有義」、「有說」、「有處說」、「有云」). For example,

```
有釋：<q type="unknown">此說佛爲一切智者，非外道師
```

```
<lb n="0347c21"/>也。 </q>
```

```
<cit>有說： <quote type="paraphrase">無此第六  
</quote></cit>
```

- (4) Using `<xref doc="" targets=""></xref>` to mark up the cross-referenced link between the citation and its original texts. For example, the link between T1828 and Jushelun 《俱舍論》,

```
T1828
<p type="inline"><pp targets="T1579xp0250">量等三千大
<lb n="0328b25"/>千世界</pp>者，若<xref doc="T29n1558"
targets="0057a08:1-0057a13:10"><cit><bibl><title
type="lun">俱舍</title><biblScope>頌</biblScope></bibl>
云：<quote><lg><l>安立器世間，</l><l>風輪最
<lb n="0328b26"/>居下，</l><l>其量廣無數，</l><l>厚十六
洛叉。</l><l>次上水輪深，</l>
<lb n="0328b27"/><l>十一億二萬，</l><l>下八洛叉水，
</l><l>餘凝結成金。</l><l>此水
<lb n="0328b28"/>金輪廣，</l><l>徑十二洛叉，</l><l>三千
四百半，</l><l>周圍此三
<lb n="0328b29"/>倍。</l></lg></quote></cit></xref></p>
```

- (5) Using `<term id="" type="yinyi"></term>`, `<term id="" type="yiyi"></term>`, `<gloss targets=""></gloss>` to mark up the Chinese explanation of Sanskrit terms. For example,

```
<div5 n="3" type="L5"><head type="added">3 地</head>
<div6 n="1" type="L6"><head type="added">1 步彌</head>
<p type="inline">梵云<term id="tmT42p0312c161112"
```

type="yinyi">步彌</term>，此云<term id="tmT42p0312c1615" type="yiyi">地</term>
<lb n="0312c17"/>者，<gloss targets="tmT42p0312c161112 tmT42p0312c1615">生成住持義也。</gloss>始自五識終暨無餘，凡聖
<lb n="0312c18"/>因果統無不備，皆具能生成住持，有類於地
<lb n="0312c19"/>故云也。</p></div6>

(6) Using <q type="proof"></q> to mark up the Buddhist logic

arguments, i.e., 「量云」、「總立量」、「別立量」、「應立量」、「外宗量」、「唯識比量」、「應破量」。 For example,

量云：<q type="proof">見色等起，我覺應非如實
<lb n="0346b29"/>見，此起餘覺故，如見杌爲人覺等。</q>

外宗量云：<q type="proof">去來定有，世所攝
<lb n="0344c07"/>故，猶如現在。</q>

唯識比量云：<q type="proof">色等諸法，皆不離識
<note type="inline">宗</note>，
<lb n="0595a12"/>是所識故<note type="inline">因
</note>，猶如於識<note type="inline">喻</note>，既
色等諸法，不離於
<lb n="0595a13"/>識，從識而現，無識之時，此境應
無，且如入無
<lb n="0595a14"/>心定或唯緣涅槃等境，爾時色等無
能緣之
<lb n="0595a15"/>識，眼等諸根及外器等須應斷絕。
</q>

應破量云：<q type="proof">第六意識無
 <pb ed="T" id="T42.1828.0596a" n="0596a"/>
 <lb n="0596a01"/>想定等，諸位中應無<note
 type="inline">宗</note>，在此位中於六識內隨
 <lb n="0596a02"/>一攝故<note type="inline">因
 </note>，如五識等<note type="inline">喻</note>。
 </q>

- (7) Using <pp id="T1828xp0001"></pp> and <pp targets=" "></pp>
 ([pinpoint](#)) to mark up the pinpoint link between T1828 and the YBh
 (T1579) itself which including other six Commentaries.

There are about 5000 pinpoint links in Bendifen 「 1 本地分 」, 3000 in
 Shejuezenfen 「 2 攝決擇分 」, 231 in Sheshifen 「 3 攝釋分 」, 181 in
 Sheyimenfen 「 4 攝異門分 」, and 1637 in Sheshifen 「 5 攝事分 」. In total,
 there are more than 8000 pinpoint links. The characteristic of pinpoint
 link is to display different comments from various Commentaries. For
 example, the link between T1828 and T1579, the links of terms in T2802,

<p>T1579 <lb n="0463b11"/><pp id="T1579xp0730">鄔波尼殺曇 分</pp>、不及此一。<pp id="T1579xp0731">餘如前說 </pp>。</p></p>
<p>T1828 <p type="inline"><pp targets="T1579xp0730">鄔波尼殺 曇分</pp>者，數中最多最 <lb n="0469a15"/>極無名可翻，餘如前說者。</p></p>

(8) Using `<sic resp="lijuan" cf1="T1579" corr=""></sic>` to mark up the correction for printing errors. For example,

```

<lb n="0336b13"/>初問，次略答。<pp
targets="T1579xp0402">[05]<sic resp="lijuan"
cf1="T30n1579p0294b19" corr="云">復</sic>何界施設建立
</pp>下，別釋

```

3. The development of multi-functional user interface

(1) Seven Commentaries for version-comparison were added for this 2nd year of research. Users could select to view the sutras of his/her interest. Figure 1 shows the example.

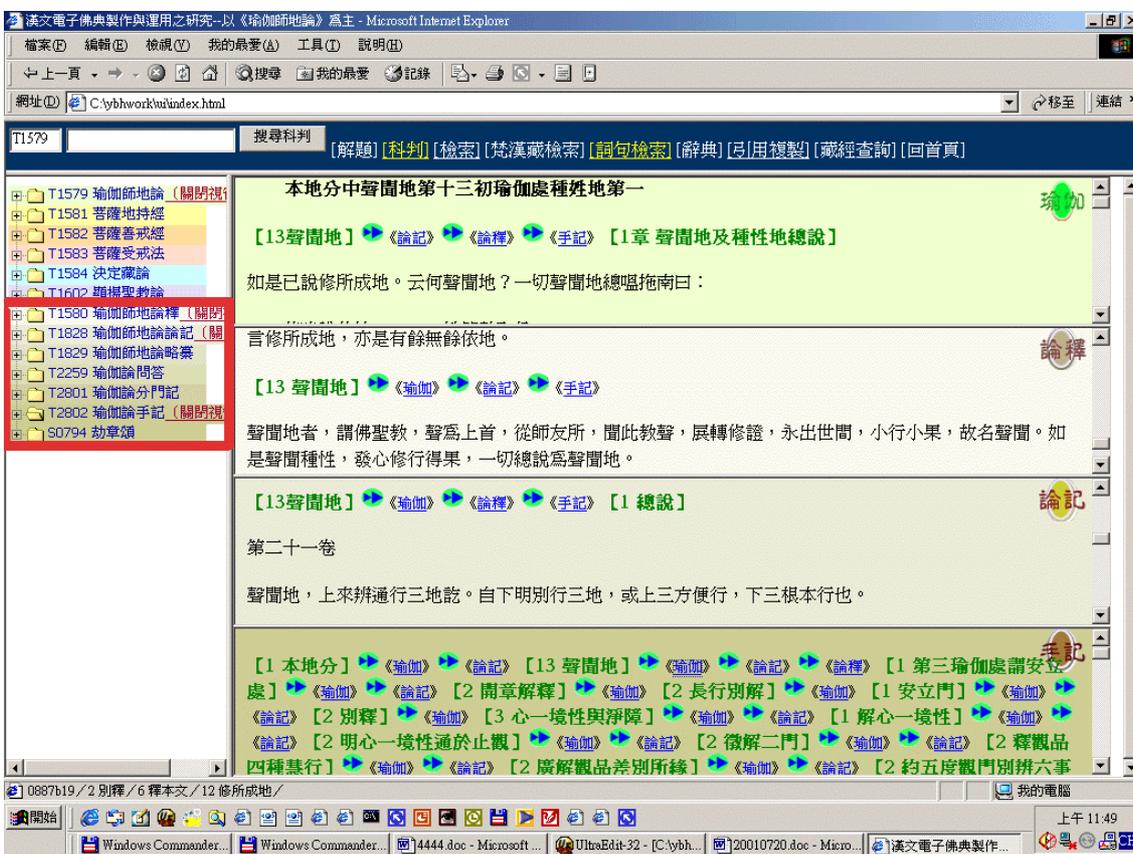


Figure 1. Seven Commentaries added for version-comparison

(2) The toolbar provides various resources and functions such as “Introduction”, “Ordered hierarchy”, “Searching of Sanskrit-Chinese-Tibetan terms”, “Index for searching proper nouns”, “Dictionaries”, “Bibliography”, and “Tripitaka” that links to the Chinese Tripitaka web site (「解題」、「科判」、「梵漢藏檢索」、「詞句檢索」、「辭典」、「參考書目」、「藏經查詢」) .

(3) To address the needs of researchers, the search results of a term are listed in the directory column in the order of sutra number. The term in search is highlighted in color in the text column. The status line at the bottom displays the “where-about” of cursor position (the red dot in Figure 2).

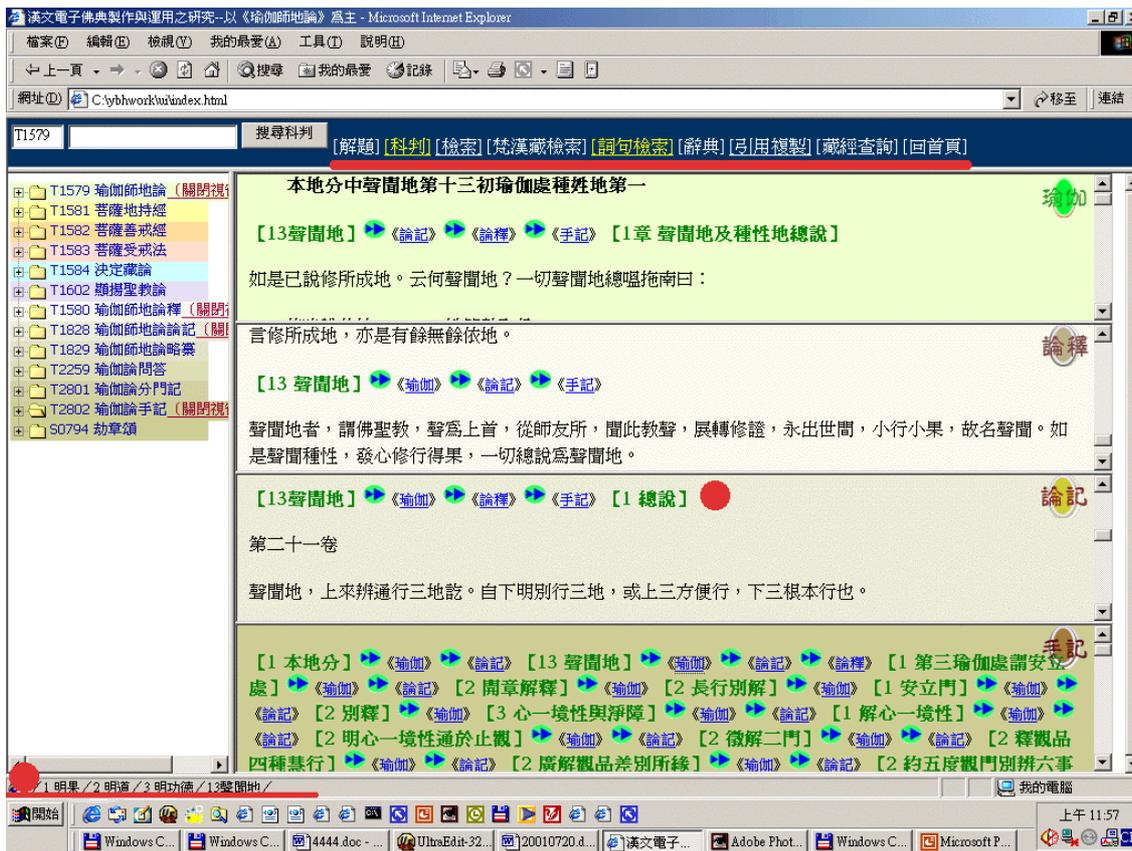


Figure 2 Toolbar and status line

V. Conclusion

This study provides an example of using electronic tag sets to mark up ancient Chinese texts and Commentaries. We found that the TEI tag sets are not satisfactory to use for ancient Chinese texts and Commentaries as TEI tag sets are designed for the western texts. Regarding user-interface design, the available functions now include the comparisons among various versions, multi-functional toolbar, and the display of search results that suits researchers' needs. Further developments such as the design of cookie that allows user selection of text and Commentaries for version-comparison

displayed as default, multiple selections for pinpoint links, etc. are planned.

Reference:

Abbreviations

CHumNN: *Computers and the Humanities Official Journal of the Association for Computers and the Humanities*, Dordrecht: Kluwer, 1966ff. NN is the sequential number of the volume.

RHComN: *Research in Humanities Computing*, Oxford: Clarendon, 1991ff. N is the sequential number of the volume.

1. Barnard, David, Ron Hayter, Maria Karaba, George Logan and John McFadden "SGML-Based Markup for Literary Texts: Two Problems and Some Solutions", in: [Chum22], p256-76.
2. Calzolari, Nicola and Zampolli, Antonio "Lexical Databases and Textual Corpora: A Trend of Convergence between Computational Linguistics and Literary and Linguistic Computing", in: Cole R.A., J.Mariani, H.Uszkoreit, A. Zaenen, V.Zue (Hrsg.) *Survey of the State of the Art in Human Language Technology*, Cambridge 1996
3. Giordano, Richard, Goble, Carole and Kiillgren, Gunnel "Problems of Multidatabase Construction for Literary and Linguistic Research", in: [RHCom5], p143-164.
4. Huitfeldt, Claus "Multi-Dimensional Texts in a One-Dimensional Medium", in: [CHum28], p235-241.1995
5. Lancashire, Ian (Ed.) *The Humanities Computing Yearbook 1989-90 A Comprehensive Guide to Software and other Resources*, Oxford: Clarendon Press, 1991.
6. Latz, Hans-Walter *Entwurf eines Modells der Verarbeitung von SGML-Dokumenten in versionsorientierten Hypertext-Systemen Das HyperSGMI, Konzept*, Diss. Berlin 1992.
7. Simons, Gary F. "implementing the TEI's Feature-Structure Markup by Direct Mapping to the Objects and Attributes of an Object-Oriented Database System", in: [RHCom5], p220-242.
8. Sperberg-McQueen, C. Michael and Burnard, Lou (Eds.) *Guidelines for Electronic Text Encoding and Interchange*, Chicago and Oxford, 1994.
9. Wilson, Eve "Hypertext Libraries: The Automated Production of Hypertext Documents", in [RHCom2], p232-246. 1994
10. Wittern, Christian "Chinese character codes: an update", in: *The Electronic Bodhidharma*, Nr. 4, June 1995, S.63-65.
11. Wiaern, Christian "Minimal Markup and More - Some Requirements for Public Texts", Conference talk held at the 3rd EBTI meeting on April 7th 1996 in Taipei, Taiwan.