

## Entrance Through the Scriptures: Catalogues and Electronic Text as a New Gate to the Buddhist Tradition

Christian Wittern

Associate Professor

Institute for Research in Humanities, Kyoto University

### Abstract

Almost from the beginning of the Buddhist tradition in China, cataloguing and recording of the scriptures has been an important activity in the attempt of coming to terms with the vast universe of the Buddhist teachings.

In this paper, I will recount the history of the WWW Database of Chinese Buddhist texts (available at <http://www.kanji.zinbun.kyoto-u.ac.jp/~wittern/canwww>), which was started with a similar purpose to make use of the electronic medium to get a handle on and a window to explore the scriptures. The problems that have been encountered during the ten years of its ongoing development, the many changes in format, methodology and outlook, its present state, as well as prospects for future development will be the main topic of this paper. This narrative will be framed by the underlying questions of what constitutes a scripture in the Chinese cultural sphere, how different translations of the “same” scriptures are to be treated and how they relate to the “original” of Indian or Central Asian origin.

This paper will however go beyond what a traditional scholarly paper is supposed to do and experiment with how a new electronic version of the scriptural catalogues, together with electronic versions of the texts themselves, can be combined to enhance our understanding of the scriptures and the times and people that produced.

Keywords: 1. Buddhist Canon 2. Cataloging 3. Bibliographical Database  
4. Modeling 5. Topic Maps

## Contents

The Birth of CANWWW: Experiments with the Electronic Medium

What is the Content of a Catalog?

The WWW Database of Chinese Buddhist Texts: A Bibliographic Database?

WWW Database of Chinese Buddhist Texts as a Topic Map?

Divide et Conquere

Using the Database

Combining the Database with the Electronic Version of the Scriptures

Conclusions

References

## The Birth of CANWWW: Experiments with the Electronic Medium

Work on what became the WWW Database of Chinese Buddhist texts (CANWWW) began some time in the early 1990s when I was using separate indices and reference books to locate texts in the different versions of the Chinese Buddhist Canon, comparing and collating titles and associated information while I went along. At that time, I was busy doing research for a Ph. D. thesis and could not spend the time necessary to systematically explore this topic, but I planned to come back to it. The chance to do this arrived in early 1996, when I started to complete the list of texts to include all texts contained in the *Taishō shinshō daizōkyō*,<sup>1</sup> cross-referenced them to those that are also included in the *Tripitaka Koreana*<sup>2</sup> and also added entries for texts from the *Dainippon zokuzōkyō*.<sup>3</sup> I also decided to add the titles mentioned in Nanjiō Bunyō's catalogue<sup>4</sup> that were not included in any of the other collections. Later, the opportunity arose to also include the *Zhonghua dazangjing*.<sup>5</sup> On top of that, I also added readings of the Chinese characters using the Hanyu pinyin 漢語拼音 system. The resource had thus already grown to include about 4000 texts and associated information.

Not surprisingly, the question arose of how to maintain such a resource. This might seem like a mere technical question, which bears no relation to the more interesting scholarly questions for the answering of which this resource was being created. However, this is far from the case; to the contrary, the way this resource is set up and maintained has a very deep impact on the way it can be used. The interpretive categories and methods that can be used depend quite literally on the architectural decisions made when creating the resource. One of the aims of this paper is to clarify some of these assumptions and discuss the implications of it. A significant part of this paper will thus be concerned with issues of how to model, create and present a digital resource in the Humanities, or more precisely, in Buddhist Studies. To make it more understandable, I will frame this discussion in the story of the development of the WWW Database of Chinese Buddhist texts.

The most natural choice might have been a simple database management system

---

<sup>1</sup> *Taishō shinshū daizōkyō* 大正新修大藏經, 85 vols, Tōkyō 1924–1932.

<sup>2</sup> *Tripitaka Koreana* 高麗大藏經, 46 vols, Seoul 1965.

<sup>3</sup> *Dainippon zokuzōkyō* 大日本續藏經, 150 vols, reprint Taipei n. y.

<sup>4</sup> *A Catalogue of the Chinese Translation of the Buddhist Tripitaka* by Bunyiu Nanjio, Oxford 1883, which is based on the *Daming sanzang shengjiao mulu* 大明三藏聖教目錄.

<sup>5</sup> *Zhonghua dazangjing* 中華大藏經, 106 vols, Beijing 1984–1996. In the published version, only the data up to vol. 90 has been prepared.

like dBase or FileMaker. They provide convenient and fast access to large bodies of data, allow for reporting and create sorted lists of many different types. However, the underlying data model of such programs is the table. While this might be linked to other tables, thus creating a more complex relational database, it is still the basic model of tabular data that lies underneath. The important question to ask here thus is: “Are we dealing with tabular data here?” A first look at existing catalogues might suggest a positive answer to this question. Most of the traditional catalogs of Buddhist texts simply list the works, sometimes in different categories, and give a fixed amount of information for each text. If there were ever tabular data, than we certainly have them here.

### What is the Content of a Catalog?

In order to decide what representation to use for the WWW Database of Chinese Buddhist texts, it turned out to be necessary to investigate more thoroughly exactly what kind of data are needed for a catalog. As an example, here is an entry from the *Kaiyuan Shijiao lu* 開元釋教錄 (Record of Buddhist scriptures, compiled during the Kaiyuan period), by Zhisheng 智昇:

大方廣佛華嚴經八十卷（第二出與東晉覺賢譯者同本證聖元年三月十四日於東都大內大遍空寺譯天后親受筆削至聖曆二年十月八日於佛授記寺功畢）<sup>6</sup>

As can be seen, the catalog gives information on the following items:

- Title of the work: 大方廣佛華嚴經
- Extent: 八十卷
- Additional information:
  - Reference to a previous translation: 第二出與東晉覺賢譯者
  - Date and place of the beginning and end of the translation: 本證聖元年三月十四日於東都大內大遍空寺譯天后親受筆削至聖曆二年十月八日於佛授記寺功畢

As can be seen, it does not give the name of the author or translator of this scripture, but instead gives some additional information about its creation. The name of the translator and additional information can be found at the end of this section at T55, p566a10ff.

Since some information has to be derived from the context of the individual catalog entry and the additional information is in a free text format, using a pure

---

<sup>6</sup> T55.2154, p565c.

tabular format would be not sufficient to capture the information. At the time of the creation of the WWW Database of Chinese Buddhist texts, the decision was thus made to use a format that would allow to flexibly recreate this structure, which was the bibliographic encoding as recommended by the Text Encoding Initiative Guidelines for Electronic Text Encoding and Interchange. This format was then used during the first years of the maintenance of this database and formed the base for the first publication in late 1996.

In addition to the items contained in traditional scriptural catalogues mentioned above, there was some more information that needed to be taken into account and added where available:

- Alternate names of the text
- Authors, translators and/or other persons responsible for the content of the text
- Dates of the text
- Canonical categories (divisions) under which the text has been subsumed
- References to locations in existing canonical collections (both modern and pre-modern)
- References to other catalogues
- References to related texts, for example other translations, commentaries etc.
- Structural division of the text

An example of how this would look for the text cited above is given here:

WWW Database of Chinese Buddhist texts

Div. Authors Periods Taisho (Nr. Vol.) Trip.Kor. (Nr. Vol.) 中華 (Nr. Vol.) 卍Nanjio Other

T10N0279 大方廣佛華嚴經 da4 fang1 guang3 fo2 hua1 yan2 jing1

Short title: 華嚴經 hua1 yan2 jing1

華嚴部 唐

實叉難陀 shi2 cha1 nan2 tuo2 [譯] 80 卷

Canonical Sources:

Canon	Vol.	Nr.	Page	Source Texts
大正	10	T0279	1	
高麗	08	k0080	0425	
中華	012	H0087	0629-0892 H0130001-0452	金藏廣勝寺本 明永樂北藏 本 金藏大寶集寺本
Nanjio	--	N0088	--	
Other	縮-天1-4, 卍-7.6-9, 國-華1-4			

Related Texts:

1. T09N0278 大方廣佛華嚴經
2. T35N1735 大方廣佛華嚴經疏
3. T36N1736 大方廣佛華嚴經隨疏演義鈔
4. T36N1737 大華嚴經略策
5. T36N1738 新譯華嚴經七處九會頌釋章
6. T36N1739 新華嚴經論
7. T36N1740 大方廣佛華嚴經中卷卷大意略敘
8. T36N1741 略釋新華嚴經修行次第決疑論
9. T36N1742 大方廣佛華嚴經願行觀門骨目
10. T36N1743 皇帝降誕日於麟德殿講大方廣佛華嚴經玄義一部
11. T84N2706
12. T85N2755 華嚴經疏

Structural Division:

1. 世主妙嚴品 **shi4 zhu3 miao4 yan2 pin3**  
→ T10N0281 佛說菩薩本業經
2. 如來現相品 **ru2 lai2 xian4 xiang4 pin3**
3. 普賢三昧品 **pu3 xian2 san1 mei4 pin3**
4. 世界成就品 **shi4 jie4 cheng2 jiu4 pin3**
5. 華藏世界品 **hua1 zang4 shi4 jie4 pin3**
6. 毘盧遮那品 **pi2 lu2 zhe1 na3 pin3**
7. 如來名號品 **ru2 lai2 ming2 hao4 pin3**  
→ T10N0280 佛說兜沙經
8. 四聖諦品 **si4 sheng4 di4 pin3**
9. 光明覺品 **guang1 ming2 jue2 pin3**  
→ T10N0280 佛說兜沙經
10. 菩薩問明品 **pu2 sa4 wen4 ming2 pin3**
11. 淨行品 **jing4 xing2 pin3**  
→ T10N0282 佛說菩薩求佛本業經
12. 賢首品 **xian2 hou3 pin3**
13. 昇須彌山頂品 **sheng1 xu1 mi2 shan1 ding3 pin3**
14. 須彌頂上偈讚品 **xu1 mi2 ding3 shang4 ji4 zan4 pin3**
15. 十住品 **shi2 zhu4 pin3**  
→ T10N0283 菩薩十住行道品  
→ T10N0284 佛說菩薩十住經
16. 梵行品 **fan4 xing2 pin3**
17. 初發心功德品 **chu1 fa1 xin1 gong1 de2 pin3**
18. 明法品 **ming2 fa3 pin3**
19. 昇夜摩天宮品 **sheng1 ye4 mo2 tian1 gong1 pin3**
20. 夜摩天宮中偈讚品 **ye4 mo2 tian1 gong1 zhong1 ji4 zan4 pin3**

21. 十行品 shi2 xing2 pin3  
 22. 十無盡藏品 shi2 wu2 jin4 zang4 pin3  
 23. 昇兜率天宮品 sheng1 dou1 shuai4 tian1 gong1 pin3  
 24. 兜率宮中偈讚品 dou1 shuai4 gong1 zhong1 ji4 zan4 pin3  
 25. 十迴向品 shi2 hui2 xiang4 pin3  
 26. 十地品 shi2 di4 pin3  
     → [T10N0285 漸備一切智德經](#)  
     → [T10N0286 十住經](#)  
     → [T10N0287 佛說十地經](#)  
 27. 十定品 shi2 ding4 pin3  
     → [T10N0288 等目菩薩所問三昧經](#)  
 28. 十通品 shi2 tong1 pin3  
 29. 十忍品 shi2 ren3 pin3  
 30. 阿僧祇品 a1 seng1 qi2 pin3  
 31. 壽量品 shou4 liang4 pin3  
     → [T10N0289 顯無邊佛土功德經](#)  
     → [T10N0290 佛說較量一切佛功德經](#)  
 32. 諸菩薩住處品 zhu1 pu2 sa4 zhu4 chu3 pin3  
 33. 佛不思議法品 fo2 bu4 si1 yi4 fa3 pin3  
 34. 如來十身相海品 ru2 lai2 shi2 shen1 xiang4 hai3 pin3  
 35. 如來隨好光明功德品 ru2 lai2 sui2 hao3 guang1 ming2 gong1 de2 pin3  
 36. 普賢品 pu3 xian2 pin3  
     → [T10N0293 大方廣佛華嚴經](#)  
     → [T10N0296 文殊師利發願經](#)  
     → [T10N0297 普賢菩薩行願讚](#)  
 37. 如來出現品 ru2 lai2 chu1 xian4 pin3  
     → [T10N0291 佛說如來興顯經](#)  
 38. 離世間品 li2 shi4 jian1 pin3  
     → [T10N0292 度世品經](#)  
 39. 入法界品 ru4 fa3 jie4 pin3  
     → [T10N0293 大方廣佛華嚴經](#)  
     → [T10N0294 佛說羅摩伽經](#)  
     → [T10N0295 大方廣佛華嚴經入法界品](#)

This file is part of the **WWW Database of Chinese Buddhist texts** maintained by Christian Wittern. All rights reserved. Usage permitted for non-commercial purpose, but please acknowledge my work. Any comments and suggestions are welcome. [Click here to send me a message.](#)

## The WWW Database of Chinese Buddhist Texts: A Bibliographic Database?

As time went by, the content of the database grew beyond what was conveniently handled in a single file and the maintenance became cumbersome. To overcome this, I reconsidered using a database, but I looked at different kinds of databases now. Since a catalog like the one I was trying to build could be seen as a kind of bibliographic database, I thought that databases that are specifically designed to work with bibliographic collections might be useful here. In difference to relational databases, based as they are on a tabular paradigms, these bibliographical database applications are designed to work with inherently hierarchical data.

As it happens a database of this type, Allegro-C, was developed in a small city not too far from where I lived,<sup>7</sup> and this database program was widely used in Sinological Seminars in Germany, so it was already proved that it can be used with Chinese data. This could be used to not only maintain the database, but at the same time also to query and browse the existing data. With this interface, it was for the first time also feasible to put a part of the work of editing the database into the hands of research assistants.

The transformation of the WWW Database of Chinese Buddhist texts into the new format was quite straightforward, since the existing TEI-SGML records also form a hierarchical, nested structure, which mapped very well to the hierarchical structure that could be created as an Allegro-C database.

In the years following this, Sanskrit and Pali equivalents of the titles had been added to the database, as well as more information on the dates of creation of the texts. While Allegro-C does come with its own solution for the Web publication of databases, it did not satisfy all the requirements, so an update of the published version had to be delayed.

### WWW Database of Chinese Buddhist Texts as a Topic Map?

The addition of new data, such as parallel titles did however force a new thinking about exactly what is the target of the information recorded in the database. If the records are about the texts in the Chinese canon, adding a Sanskrit or Pali title on the same level does not really seem appropriate. The title in Sanskrit, Pali or

---

<sup>7</sup> More information about this database program at [www.allegro-c.de](http://www.allegro-c.de). A short note on the WWW Database of Chinese Buddhist texts use of Allegro-C is at A99/Alcarta databases with Chinese data.

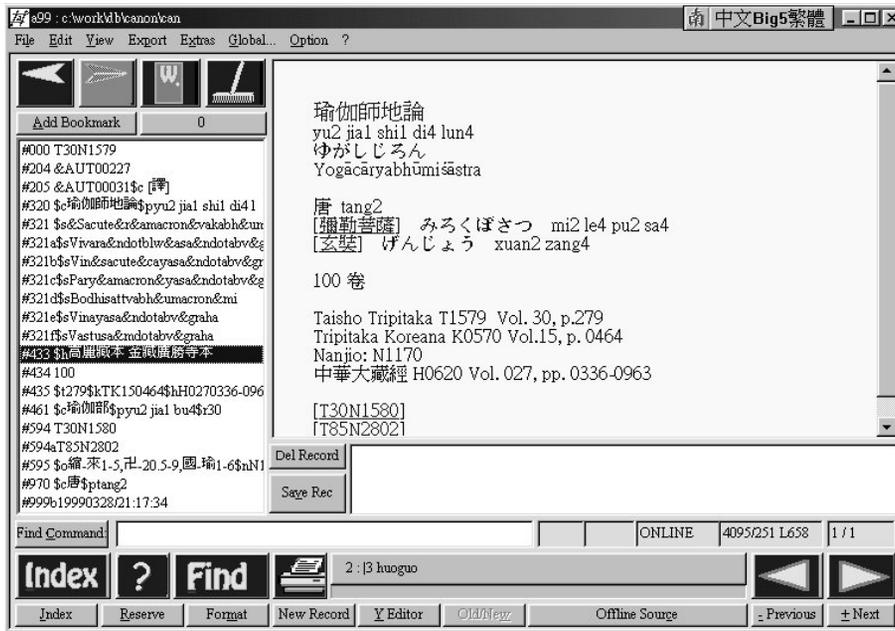


Figure 1.

even Tibetan would not be the title of the work itself, but rather refer to a text that might have been used as a source for the translation, which produced the Chinese text catalogued here. This is especially important, since not only parallel titles seemed desirable, but also an alignment of textual witnesses in different languages at the level that described the structural divisions of the text. Furthermore, where different Chinese translations exist, they could be seen as being derived from the same source. To reflect this, the following hierarchy of elements to be described was considered:

- At the highest level, the texts are seen on a very abstract level, which might correspond to a common source text, for example the *Avatamsaka sūtra*, but has no concrete instantiations in this schema, it is only an abstract, “idealized” placeholder.
- Derived from this are instantiations of a text in various languages, e.g. a translation into Chinese by Buddhabadra or Śikṣananda.
- These instantiations are mentioned in catalogues and included in collections.
- As canonical collections are distributed, individual items arrive in different collections.

There are therefore four levels to describe the content of the texts that are the target WWW Database of Chinese Buddhist texts, to complicate the fact, items

at different levels relate to each other in various ways, as parts or containers, but also as commentary, subcommentary, summary etc. An example of some of these features would look as follows:

- Avataṃsaka sūtra
  - translation by Buddhābhadra (ca. 418–420 or 398–)
    - \* contained in:
      - + T09N0278 → reprint by Hsinwenfeng → held by Kyoto University
      - + K0079 → reprint by Hsinwenfeng → held by Kyoto University
      - + H0086
    - \* listed by: N0087
  - translation by Śikṣānanda (ca. 695–699)
    - \* contained in:
      - + T10N0279
      - + K0080
      - + H0087: based on
        - Jin Canon [Guangsheng si]
        - the northern Ming Canon
        - the Jin Canon [Dabao si]
    - \* listed by: N0088
  - composed of:
    - \* (list of divisions . . . )
      - + translation by Prajñā (ca. 795–798)
        - contained in:
          - T10N293
          - K1262
        - H1522 based on:
          - K1262
          - Jin Canon
      - listed by: N0089

A format of this complexity can only with great difficulties be represented in a hierarchical database, so some other solution had to be found. After some consideration a completely different approach was attempted: The relations and information items that made up the WWW Database of Chinese Buddhist texts could all be expressed using the formal representation of Topic Maps,<sup>8</sup> a newly defined international standard that is derived from notation systems for first order logic.<sup>9</sup> This would

---

<sup>8</sup> Topic Maps are a systematic way to provide a descriptive navigation layer for any kind of resource. They have been adopted as an international standard in 2000, but the development of the underlying data model and other related parts is still continuing. For more information, see the homepage of SC34/WG3 and the article “The TAO of Topic Maps” by Steve Pepper.

<sup>9</sup> Or, to be more precise, one of the sources for this was the ‘Conceptual Graph’ notation that was

allow the representation of arbitrary information attached to any of the nodes in the database. A first attempt at re-defining the WWW Database of Chinese Buddhist texts using the Topic Map paradigms has been presented at an earlier occasion.<sup>10</sup> In addition to the ontology of texts and their representations outlined above, the authors, translators, compilers, commentators and other persons responsible for the content of the texts need also be represented in the Topic Map, as well as other information items, like the categories in which the texts have been placed, time and place of creation, revision or re-edition etc.

The result of this attempt as a whole could not be entirely called successful. While it allowed for a very fine-grained description of every aspect of the items contained in Buddhist textual collections, it proved difficult to maintain the data in this format. For a presentation in Bangkok 2002, a dedicated middleware layer for the processing of Topic Maps was written, with a relational database as a backend, but this proved to be too slow for practical work. Commercial software applications for Topic Maps also slowly emerged, but they suffered from similar deficiencies for the most part.<sup>11</sup> The editing of the data itself was also difficult, because there was no editor that could abstract from the low level of data representation that lies at the heart of a Topic Map.

## Divide et Conquere

Stepping back again, I looked at the data again and made some observations:

- While the bibliographic format chosen at the beginning can not express everything that is needed, the catalogues show clearly that there is a considerable amount of bibliographic data.
- This bibliographic part of the data needs to be enhanced and supplemented with other data.
- There is no requirement that the whole database needs to be in one single format. On the contrary, for different parts, different type of data, there might be natural fits of different types.
- The most obvious example of a successful design along these lines might be

---

developed by John F. Sowa from notation systems for Logic.

<sup>10</sup> This presentation was given at the International Conference of the International Association of Buddhist Studies in Bangkok, December 2002. The presentation notes are still available.

<sup>11</sup> The leading software application for Topic Maps, the OKS Suite from Ontopia timed out after one hour when trying to load a Topic Map of some size.

the web itself: it integrates vast arrays of completely unrelated data into one interface. On the other hand, it fails miserably in expressing the relationship between different information items.

The next, and to date latest attempt at defining a data structure for the WWW Database of Chinese Buddhist texts thus operates as follows:

- The “bibliographic records,” that is the entries for the text items (the lower levels in the table above) are expressed in a format that was designed for such data: the Metadata Object Description Schema (MODS).<sup>12</sup>
- Information about the persons and places mentioned in the records and elsewhere in the database is centered around authority file entries for these entities, recorded in the Metadata Authority Description Schema (MADS).<sup>13</sup>
- To connect the individual items in these different formats, a Topic Map will be used. The main difference to the previous design is that now the map will mainly be used to connect information items, not to contain them.
- A link to digital versions of a text will be routed through a METS (Metadata Encoding & Transmission Standard)<sup>14</sup> record. This will allow a unified access to multiple digital representations of the object.
- At some point, a link to the appropriate entries in a common shared ontology, for example the CIDOC-CRM<sup>15</sup> will be added to assist in cross-collection sharing of data.
- Additional notes, references and description will be added in free text using the XML schema provided by the Text Encoding Initiative (TEI).<sup>16</sup>

---

<sup>12</sup> This format was defined by the Library of Congress as a simplified, XML based version of the MARC standard.

<sup>13</sup> Not surprisingly, this format is also maintained by the Library of Congress, further description is available here.

<sup>14</sup> Again, this is one member of the family of standards defined by the Library of Congress, the description is available here.

<sup>15</sup> The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. The CIDOC CRM is the culmination of over 10 years work by the CIDOC Documentation Standards Working Group and CIDOC CRM SIG which are working groups of CIDOC. It is currently being elaborated by the International Standards Organization as Draft International Standard ISO/DIS 21127 and the CIDOC CRM Special Interest Group to become an ISO standard. More information is available at the CIDOC CRM Home page.

<sup>16</sup> The schemas and documentation are available from the TEI Consortiums homepage. For this project, the development version of P5 is used.

At this point, one question can not be avoided: Is this really necessary? This looks much too complicated for the purpose at hand.

The answer to this of course depends on how the task is defined. There might indeed be more modest ways to achieve retrieval of records by title, author and the like. The reason to construct such a complex array of descriptive data from different standards is twofold. On the one hand, a maximum interoperability and potential for datasharing is made possible by using specialized but widely adopted standards. On the other hand, the task at hand is not seen as simply to construct a list of bibliographical records, but rather to build a central element of an evolving resource for the study of Buddhism. This will require many references to the texts that serve as a source for the transmission of this tradition, thus it makes sense to start this effort with the construction of this database, which will be most useful in conjunction with electronic versions of these texts, especially if they are also in XML format. Some of the things that will thus become possible are demonstrated below.

## Using the Database

One of the important purposes of traditional catalogues of Buddhist scriptures was to separate orthodox texts (translations) from spurious works of doubtful provenience. The inclusion in a scriptural catalog thus meant endorsement and ensured continued transmission and safekeeping of a text.<sup>17</sup>

In contrast to this, the current scriptural database serves a completely different purpose. Rather than serving as a gatekeeper, who restricts access and controls entry, the database tries to provide as many entrance points and doorways into the complex written tradition of the Buddhist as possible.

For this purpose, the earliest versions already provided access to the more than 4000 titles listed in the WWW Database of Chinese Buddhist texts by canonical division, authors and periods, in addition to the access through the locations in the canonical collections.

The 32 canonical divisions simply record in which division a text has been included to the canon.<sup>18</sup> Now this categorization is not always very helpful, for the

---

<sup>17</sup> The earliest extant catalog is *Chu sanzang jiji* 出三藏記集 by Sengyou, based in part on the catalog *Congli zhongjing lu* 淙理眾經 by Daoan of 374; cf. Zürcher 1972:30 and 195.

<sup>18</sup> In the current version of the WWW Database of Chinese Buddhist texts only one attribution is given and this attribution is based on the *Taishō*, *Tripitaka Koreana*, *Zokuzōkyō*, and *Zhonghua Canon* in descending order of preference. However, since the attribution changed over time in

purpose of providing better access to the canon. In the division 16 經疏部, for example are commentaries and subcommentaries to a variety of different scriptures recorded. It would be more helpful for the user, to group these together with the scriptures they comment on, as has been done in the CBETA 'bulei' 部類 categorization.<sup>19</sup>

The access by "authors" provided lists of texts attributed to individuals, in any capacity, be it as author, translator, compiler, commentator or other function. The list given in the older version of WWW Database of Chinese Buddhist texts is fixed and sorted by number of texts attributed to a given author, but this was not very flexible. It would for example be interesting to also weight this list by the volume (amount of scrolls) of the texts translated, or by date of the work. But even the limited list already gave some quite interesting insights in the process of the creation of the Chinese Buddhist canon and those involved in it. So is for example the person whose name appears most frequently in an attribution Amoghavajra (Bukong Jingang 不空金剛; often abbreviated to Bukong 不空), the disciple of Vajrabodhi who worked during the reign of Tang Xuanzong and subsequent emperors and has 176 attributions, followed by those text with lost attributions at 153.<sup>20</sup>

Access by periods lists the number of texts translated or written by the period in which it was created. The overview page for this section lists the total amount of works and number of active authors.<sup>21</sup> Since the area and duration of the various dynasties varies greatly, it is again difficult to draw conclusions from this list. With the new format of the database, it is much easier to analyze these data and

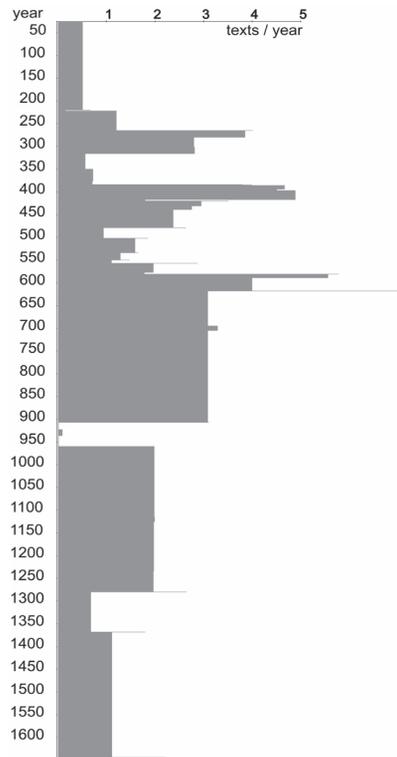


Figure 2

some cases, a more detailed description is necessary, which will also include the section where a text is listed in the scriptural catalogues.

<sup>19</sup> See Shi Huimin 2002.

<sup>20</sup> This list is still available at Authors (Translators, Commentators etc.) #1.

<sup>21</sup> This list is available at Historical periods and areas.

normalize (at least on the temporal axis), the result of such a normalization is given in Figure 2. In this case, since for a large number of texts the date of creation is unknown or imprecise, the dynasty under which the text was created is used as a very rough temporal indicator. The number of texts produced during one dynasty is then divided by the number of years this dynasty existed to arrive at an indicator for one single year; for dynasties that existed in parallel the indicators for the corresponding years are then summed up. It goes without saying that this provides only a very rough overview of the areas of productivity, which is, especially for longliving dynasties like the Tang or Song very much levelled to a meaningless average. The graph nevertheless does give some indication on which where the more productive phases in Chinese Buddhism.

### Combining the Database with the Electronic Version of the Scriptures

The most obvious enhancement of the WWW Database of Chinese Buddhist texts is its combination with the full text of the scriptures, as for example prepared by the Chinese Buddhist Electronic Text Association (CBETA).<sup>22</sup> While many ways of enhancement and mutual analysis of data is possible, I will limit myself to introduce just one example of the possibilities that are now at the disposal of scholars.<sup>23</sup>

*Table 1*

<b>Document</b>	<b>Before</b>	<b>Match</b>	<b>After</b>
T36N1742	佛身入一	佛身。一	佛身入一
T36N1742	身入一切	佛身。一	切刹入己
T46N1923	如是一切	佛身。一	切眾生身
T46N1923	如是一切	佛身。一	切眾生身
T46N1924	。一切諸	佛身。一	切無礙人
T47N1961	。即是一	佛身。一	心一智慧
T48N2016	教。唯一	佛身。一	實性佛。
T85N2883	地是一切	佛身。一	切佛道場
T18N0901	瓔珞莊嚴	佛身。七	者寶帳。

<sup>22</sup> See CBETA Chinese Electronic Tripitaka collection and CD-ROMs published by CBETA.

<sup>23</sup> It should be mentioned here that this “disposal” implies a rather heavy investment in the underlying concepts and methodologies, such as XML, XSLT and XML Query, since most of this is not yet exposed through easy-to-use interfaces that would hide this complexity. Especially the XML Query language does provide rich possibilities for exploring the richness of information contained in XML texts, so the investment of time and effort will be rewarded.

Table 1 (cont'd)

Document	Before	Match	After
T35N1735	之記。六	佛身。七	即種性。
T35N1733	王下現十	佛身。丈	六遍於十
T04N0211	虛空還現	佛身。三	十二相八
T09N0265		佛身。三	十二相種
T12N0322	當以成就	佛身。三	十二大士
T33N1707	菩薩得成	佛身。三	三恒河沙
T37N1748	二願願攝	佛身。三	次三願願
T47N1957	光明莊嚴	佛身。三	者。佛無
T19N0965	遍。現證	佛身。上	根者。得
T25N1509	方來欲量	佛身。上	過虛空無
T35N1735	絕。結歸	佛身。上	來體性寂
T38N1776	從之得生	佛身。上	來別竟。
T38N1781	患厭當樂	佛身。上	來第一。
T36N1736	。我剎見	佛身。下	取意引。
T37N1751	地樹座像	佛身。下	去諸境皆
T38N1776	自行教求	佛身。下	約利他教
T38N1776	別門而顯	佛身。下	六約對諸
T39N1796	界者即是	佛身。下	句云我者

In this example, a list of terms is compared by its usage in different periods. For this purpose, for every term, a KWIC (KeyWord In Context) list is constructed. These custom KWIC sets, of which a small excerpt is shown in Figure 3, are then further analyzed by retrieving additional information from the database, for example about the time of creation of the text. A new report is then generated, which uses the information gleaned from the WWW Database of Chinese Buddhist texts to group and further analyze the occurrences received from the text database.

Over the many centuries that witnessed activities in the translation of Buddhist scriptures, the terminology used to translate Buddhist key terms changed considerably. In the context of the current essay, I can not elaborate in the soteriological and philosophical background, but I would like to use the terminology surrounding the various buddhakāya schemes, as given by Sharf 2002:103.<sup>24</sup> The following table

<sup>24</sup> Sharf does refer only to the terms as they are used in the three extant versions of the *Laṅkāvatāra Sūtra* (by Guṇabhadra T16N0670, Bodhiruci T16N0671 and Śikṣananda T16N0672). I am completely ignoring his argument here and extending the search to the whole text base of the Taishō

shows the distribution of these terms over time, the occurrences in the text have been assigned the creation dates as recorded in the database, in order to present them in a single table, the occurrences have been summed up for periods of 100 years and only the median value is given here.<sup>25</sup>

Table 2.

<b>Term</b>	<b>Total</b>	<b>100</b>	<b>200</b>	<b>300</b>	<b>400</b>	<b>500</b>	<b>600</b>	<b>700</b>	<b>800</b>	<b>900</b>	<b>1000</b>	<b>1100</b>	<b>1200</b>	<b>1300</b>
佛身	4715	7.179	7.179	84.31	180.7	38.37	197.1	90.44	80.44	80.44	30.35	30.35	29.89	12.71
依佛	1416	2.051	2.051	17.64	89.19	23.76	112.0	66.57	56.57	56.57	21.46	21.46	21.46	7.029
化佛	1632	2.051	2.051	23.52	73.59	18.66	93.27	38.40	38.40	38.40	11.77	11.77	11.77	4.756
報佛	1003	2.564	2.564	11.76	65.05	13.54	101.8	29.92	29.92	29.92	10	10	10	3.409
報生佛	5	0	0	0	0	0	0	0.346	0.346	0.346	0.312	0.312	0.312	0
報相佛	1	0	0	0	0.675	0.675	0	0	0	0	0	0	0	0
報身	1475	0.512	0.512	1.960	38.98	7.624	109.4	48.75	38.75	38.75	13.33	13.33	13.33	5.892
應化佛	44	0	0	0	2.809	2.809	13.68	2.594	2.594	2.594	1.875	1.875	1.875	0
應身	1441	1.025	1.025	7.843	24.15	6.086	115.0	41.31	31.31	31.31	10.42	10.42	10.42	6.103
根本佛	21	0	0	0	4.381	1.351	5.405	2.076	2.076	2.076	0.625	0.625	0.625	0
根本如來	5	0	0	0	0.675	0.675	2.702	0	0	0	0	0	0	0
法佛	877	1.538	1.538	28.09	78.59	14.23	100.6	32.52	32.52	32.52	16.25	16.25	16.25	2.272
法佛報佛	10	0	0	0	1.961	1.961	10.81	0	0	0	0	0	0	0
法佛報身	0	0	0	0	0	0	0	0	0	0	0	0	0	0
法化佛	1	0	0	0	0.675	0.675	0	0	0	0	0	0	0	0
法性佛	26	0	0	0	3.705	0.675	5.405	4.152	4.152	4.152	0.625	0.625	0.625	0
法性所流佛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
法身	16588	4.615	4.615	74.50	189.2	37.00	205.6	112.5	102.5	102.5	31.51	31.51	31.05	15.82
眞如來	60	0.512	0.512	5.882	4.676	5.023	13.51	3.806	3.806	3.806	0.625	0.625	0.625	1.136
眞身	543	0.512	0.512	3.921	23.98	6.151	49.16	36.12	26.12	26.12	9.273	9.273	9.273	2.483
變化佛	26	0	0	1.960	6.233	0.172	2.875	3.978	3.978	3.978	0.937	0.937	0.937	0
變化如來	9	0	0	0	0.970	0	0	1.384	1.384	1.384	0.625	0.625	0.625	0

part of the CBETA electronic Tripiṭaka.

<sup>25</sup> Since this is only a rough summary, which does not answer to more specific questions, I am making the raw data that have been used in the analysis for this paper available for those who would like to make there one analysis. The datafiles can be downloaded from <http://www.kanji.zinbun.kyoto-u.ac.jp/~wittern/can/index.html>, the same place as the WWW Database of Chinese Buddhist texts itself.

## Conclusions

It has been shown in this essay how the WWW Database of Chinese Buddhist texts has evolved over the last years, not so much in terms of the addition of content or functionality, but rather in terms of how the underlying data are understood and in what context they will be used. There has been a rather frequent remodeling of the whole structure to experiment with a new understanding of how the scriptures and the existing texts are related. Although it is hoped that the current vision of the database will prove to be a sound foundation for further development, there can be no guarantee that this will be the case.

It might be in fact the very activity of creating and recreating different views, or models on the canonical scriptures, that will be the main reward from this project. The way this model needs to be constructed itself already makes implicit assumptions about the potential and possible usages and is thus limiting in a way. In order to adopt to new questions, the question then might not be, which is the right model, but rather, how can I construct a view that most easily allows further modifications and re-modeling. This could be seen in fact as one of the central activities in the evolving field of Computing Humanities, of which this project forms a part.<sup>26</sup>

A lot of work still has to be done with regard to a user interface to the database that allows to explore how it can be used in conjunction with the full-text database of Chinese Buddhist texts developed by CBETA and other evolving databases, for example the Resources for Research on Tang Civilization currently under development in Kyoto.

## References

- McCarty, Willard. *Humanities Computing*. Palgrave MacMillan: Houndmills, Basingstoke, Hampshire, 2005
- Sharf, Robert H. *Coming to Terms with Chinese Buddhism*. Kuroda Institute: Honolulu, 2002.
- Shi, Huimin 釋惠敏. “CBETA ban dianzi Fodian jicheng bulei mu (CBETA ban jing) jianjie” CBETA版電子佛典集成部類目(CBETA版經)簡介. *Fojiao tushuguan guanxun* 佛教圖書館館訊 32 (Dec. 2002): 218–239.
- Zürcher, Erik. *The Buddhist Conquest of China*. E.J. Brill: Leiden, 1972.

---

<sup>26</sup> More about modeling and its relation to Humanities Computing can be found in McCarty 2005, Chapter 1 and passim.

# 佛教研究的新入口

## ——佛教經目與電子全文

維習安 (Christian Wittern)

京都大學人文科學研究所副教授

### 提要

傳統中國佛教從一開始，爲了掌握浩瀚龐大的佛陀教導，佛經的編目和記錄便是很重要的活動。

筆者在此篇文章中將陳述線上漢文佛教文獻資料庫 (<http://www.kanji.zinbun.kyoto-u.ac.jp/~wittern/canwww>) 的歷史，它的目的同樣是爲了掌握經典，不同的只是使用了電子媒介。本文的主題將是此資料庫在十年持續發展中所遭遇的問題，格式、方法原則與外觀呈現的各種變化，資料庫目前的狀態與未來展望。此篇文章依下列問題陳述：

1. 漢傳經典由什麼組成？
2. 我們是如何處理「同一」經典的不同譯本？
3. 這些譯本與源於印度或中亞的「原典」有何關連？

這篇文章超越傳統學術論文的部份，而去嘗試結合經典目錄與全文電子版，進而增進我們對於經典以及當時時代背景與人物的了解。

關鍵詞：1. 佛教藏經 2. 目錄 3. 文獻資料庫 4. 模型 5. 主題圖