

# The Digital Dictionary of Buddhism [DDB] as a Functional Model for Web Collaboration

A. Charles Muller

Professor, University of Tokyo

Over twenty-two years have passed since the beginning of the lexicographical compilation that has resulted in what is presently named the *Digital Dictionary of Buddhism* (DDB), and over thirteen years have passed since its installation on the WWW. Originally uploaded with approximately 3,200 entries, this compilation of terms, text names, person names, school names, etc., contains, at the time of this writing, over 47,000 entries, based on the contributions of more than 57 individuals. The DDB is also subscribed to by twenty-five university libraries from top-rated institutions in North America, Europe, and Asia. Originally viewed by its creator primarily as a lexicographical tool for the translation of Buddhist canonical texts, the DDB is now fulfilling that role to a degree that is enhanced greatly by the concurrent maturation of canonical text digitization projects undertaken by the SAT Taishō Daizōkyō, Chinese Buddhist Electronic Text Association (CBETA), Research Institute for Tripiṭaka Koreana (RITK), and the digital *Hanguk bulgyo jeonseo* (HBJ). As the usage of these digital canons grows in scope and sophistication, translators around the world can benefit immensely by the integrated usage of digital canons and the DDB, both through its web implementation and the usage of localized tools. This paper discusses some of the main benefits of combined usage of digital text and digital lexicon.

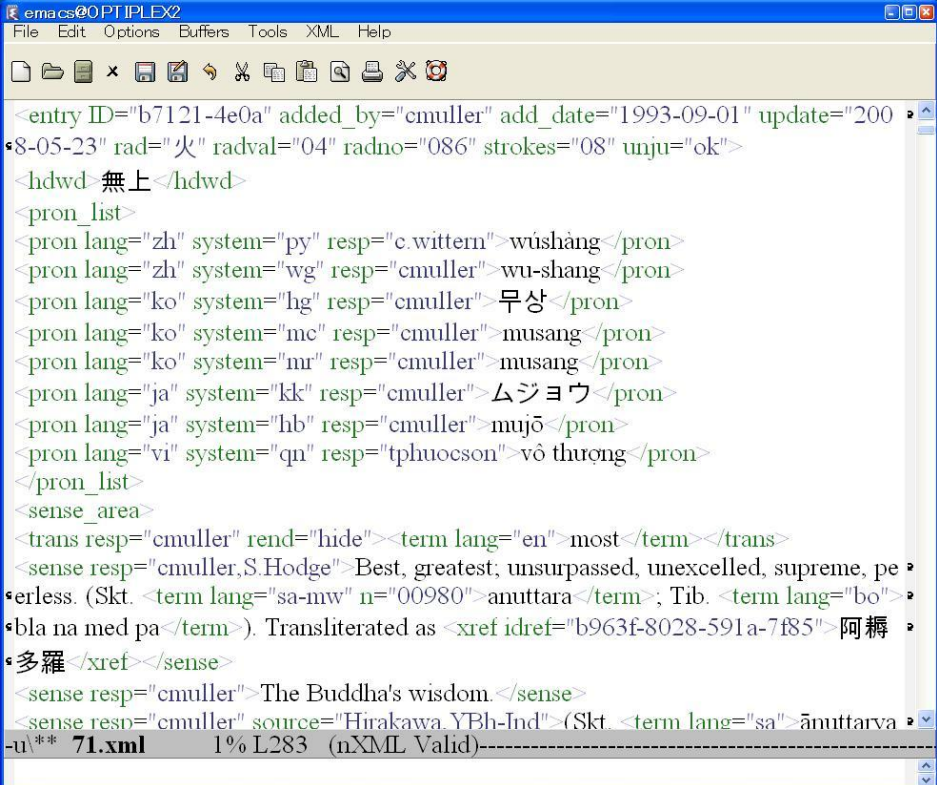
## I. Background

The compilation of the *Digital Dictionary of Buddhism* (DDB) and *CJKV-English Dictionary* (CJKV-E) began in 1986, at a time when none except the most forward-thinking of computer scientists had even dreamed of such a thing as the WWW. Therefore, what was envisioned at the outset was simply the eventual publication of a standard printed dictionary. In 1994, the Web made its appearance, bringing about the possibility that the publication of the compiled material as a web resource might provide the dual advantage of (1) allowing the broad availability far sooner than the case of waiting until the development of a proper print compilation (which could conceivably take decades), and (2) allowing the possibility of gathering collaborators to hasten the compilation, broaden the scope of the coverage, and improve the accuracy of the material. Responding to this potential, in the middle of 1995, the WordPerfect 6.0 word-processor files were converted to HTML, and placed on the web. It did not take long for the latter possibility to become a reality, as several scholars appeared, offering both technical help and data contributions,<sup>1</sup> and many professors of Buddhism in North America began to use the DDB as a reference work in their university courses. The work on the project continues actively up to the present, including terminology from scores of classical texts from Buddhism, Confucianism, and Daoism.

During the first five years on the web, the DDB/CJKV-E dictionaries were maintained on the web in a simple, hard-linked HTML format. A major turning point in the history of the project came in January 2001, when Dr. Michael Beddow, a scholar of German Studies who was extremely knowledgeable regarding the application of XML/XSLT technology with textual corpora, offered to program the DDB data such that XSLT and X-Linking functionality could be produced in the latest versions of the standard browsers. To this effect, he developed a search engine in PERL to call up dictionary entries based on user queries.<sup>2</sup>

Our decision to settle on XML as the storage-and-delivery format for the DDB was in also somewhat novel, in that most other comparable data sets (especially lexicographical projects) were (and still are) handled with some kind of popular database format. But we found that XML provided a far greater range of flexibility for our textual materials than a tabular database, and there were many ways in which the appending of metadata through XML attributes turned out to be very useful.

We also discovered the Text Encoding Initiative (TEI; <http://www.tei-c.org/index.xml>), which had already done extensive work in creating systematic tag sets for various forms of textual corpora—including lexicons. Thus we were able to create the tags and attributes we needed largely based on the TEI system. This in turn allowed us the possibility of utilizing style sheets and various other structures from the TEI system. A sample short entry in XML format is as follows:



```
emacs@OPTIPLEX2
File Edit Options Buffers Tools XML Help
[Icons]
<entry ID="b7121-4e0a" added_by="cmuller" add_date="1993-09-01" update="200
8-05-23" rad="火" radval="04" radno="086" strokes="08" unju="ok">
  <hdwd>無上</hdwd>
  <pron_list>
  <pron lang="zh" system="py" resp="c.wittern">wúshàng</pron>
  <pron lang="zh" system="wg" resp="cmuller">wu-shang</pron>
  <pron lang="ko" system="hg" resp="cmuller">무상</pron>
  <pron lang="ko" system="mc" resp="cmuller">musang</pron>
  <pron lang="ko" system="mr" resp="cmuller">musang</pron>
  <pron lang="ja" system="kk" resp="cmuller">ムジヨウ</pron>
  <pron lang="ja" system="hb" resp="cmuller">mujō</pron>
  <pron lang="vi" system="qn" resp="tphuocson">vô thượng</pron>
  </pron_list>
  <sense_area>
  <trans resp="cmuller" rend="hide"><term lang="en">most</term></trans>
  <sense resp="cmuller,S.Hodge">Best, greatest; unsurpassed, unexcelled, supreme, pe
erless. (Skt. <term lang="sa-mw" n="00980">amuttara</term>; Tib. <term lang="bo">
bla na med pa</term>). Transliterated as <xref idref="b963f-8028-591a-7f85">阿耨
多羅</xref></sense>
  <sense resp="cmuller">The Buddha's wisdom.</sense>
  <sense resp="cmuller" source="Hirakawa.YBh-Ind">(Skt. <term lang="sa">ānuttarva
-u*** 71.xml 1% L283 (nXML Valid)
```

In addition to the basic data contained in the DDB, over the years a variety of groups, institutions, and individual scholars dealing with East Asian Buddhism (including Muller, his collaborators, and paid assistants) have been developing a comprehensive, composite index drawn from the indexes of dozens of major East Asian Buddhist reference works, which now includes almost 300,000 entries (described in further detail below). With this valuable resource in mind, Michael built the search engine so that if a given item was not found in the DDB proper, it

could be searched for in this comprehensive index. If found, its location in relevant reference works could be provided, a great benefit to users of the dictionary.<sup>3</sup>

## II. Content Development

The first public presentation of the DDB at an academic conference was at the meeting of the Electronic Buddhist Text Initiative (EBTI)<sup>4</sup> held at the Fo Guang Shan temple in Taipei in 1996. At that time, the DDB contained approximately 3,200 entries. At the time of the present writing (May 2009), that number has reached to over 47,000 and is continuing to grow rapidly. This rapid growth is due to a number of factors, the most important of which is the steady growth in size and efficiency of related digital tools and resources. The availability of the above-mentioned comprehensive index, which allows for the rapid location of all entries, has been of critical importance.

Another major reason for the rapid growth of the DDB is the concurrent digitization of the Chinese Buddhist canons, a project first undertaken by the Research Institute of Tripiṭaka Koreana (RITK, which digitized the Korean Buddhist canon; <http://www.sutra.re.kr/>), and followed upon SAT *Taishō* Database (which digitized the Japanese *Taishō* canon; <http://21dzk.l.u-tokyo.ac.jp/SAT/index.html>) and by CBETA (which has digitized the Japanese *Taishō* as well as the *Zokuzōkyō*; <http://www.cbeta.org/index.htm>). The availability of the canonical source texts in digital format has allowed us to develop local applications that can quickly extract terms from these texts and match them with entries in these reference works to include new entries "on the fly." At the same time (as will be discussed below) users of these text databases can also have direct access to the DDB — if their developers choose to provide it. Finally, the overall effect of availability on the Web and the steadily growth of the acceptance of the DDB as standard reference tool has naturally brought about an increase in the number of contributors, of whom there are now more than sixty.<sup>5</sup>

Further mention should also be made regarding technical help. In addition to the above-explained central role played by Michael Beddow, Christian Wittern has been a continual guiding force regarding the technical trajectory of the project. In addition, the creator of the DDB's first XML Document Type Definition (DTD), Louis-Dominique Dubeau (presently a Ph.D. candidate at the University of Virginia) played an important role at a critical juncture, and has recently been working on an application of the DDB for OpenOffice.<sup>6</sup>

## III. Usage

When the DDB was originally placed on the Web in 1995, users accessed its data solely through hyperlinks attached to various indexes on the top page of the web site.<sup>7</sup> These indexes were broken down into terms, texts, persons, schools, temples, places, etc., which were in turn broken down linguistically — as appropriate — into English, Chinese, Korean, Japanese, Sanskrit, Pali, Tibetan, etc. These indexes are still included on the top page, serving a useful role for study and research, also serving as extensive glossaries. Since nowadays all these of materials are indexed by Google and other web search entities, the presence of the indexes provides access to the dictionary to people who are just performing general web searches.

**Digital Dictionary of Buddhism**  
電子佛教辭典  
Established July, 1995  
DDB Entry Total (2009/03/31): 47,490  
[Password Access Policies](#)

---

<p><b>(1) Dictionary Access: Search</b></p> <div style="border: 1px solid black; padding: 2px;"><input style="width: 80%;" type="text"/> <input type="button" value="Search"/></div> <p>[non-members can log in with "guest"]</p> <p><b>(2) Dictionary Access: Indexes</b></p> <p><a href="#">Radical (部首)</a> <a href="#">Total Strokes (總畫)</a> <a href="#">East Asian Romanized Readings</a> <a href="#">Terms</a> <a href="#">Texts</a> <a href="#">Persons</a> <a href="#">Temples</a> <a href="#">Schools</a> <a href="#">Places</a> <a href="#">Names (gods, buddhas, bodhisattvas, demons, etc.)</a></p> <p><b>(3) Front Matter</b></p> <ul style="list-style-type: none"><li>• <a href="#">Introduction</a></li></ul>	<p><b>(4) External Resources: Lexical</b></p> <ul style="list-style-type: none"><li>• <a href="#">CJKV-English Dictionary</a></li><li>• <a href="#">WWJDic Server</a> (Jim Breen)</li><li>• <a href="#">Cologne Digital Sanskrit Lexicon</a></li><li>• <a href="#">Monier-Williams</a> (local cache)</li><li>• <a href="#">The Online Tibetan to English Dictionary and Translation Tool</a></li><li>• <a href="#">Thesaurus Linguae Sericae</a> An Historical and Comparative Encyclopaedia of Chinese Conceptual Schemes</li><li>• <a href="#">Nitartha Tibetan-English Dictionary</a></li><li>• <a href="#">Pali-English Dictionary (Pali Text Society)</a></li><li>• <a href="#">A Dictionary of Chinese Buddhist Terms</a>, by Soothill and Hodous</li></ul> <p><b>(5) External Resources: Textual Databases and Collections</b></p> <ul style="list-style-type: none"><li>• <a href="#">SAT Taisho shinshu daizokyo</a> (Japan — linked with DDB)</li><li>• <a href="#">CBETA Buddhist Texts</a> (Taiwan)</li></ul>	<p><b>(6) External Resources: Searchable Article and Bibliographical Databases</b></p> <ul style="list-style-type: none"><li>• <a href="#">NTU Bibliography Retrieval</a></li><li>• <a href="#">Article Database of the Japan Association of Indian and Buddhist Studies</a></li><li>• <a href="#">Indica et Buddhica</a>: Materials for Indology and Buddhology (Richard Mahoney)</li></ul> <p>For OpenOffice users: <a href="#">OpenOffice Hanzil Toolbar</a> (including direct lookup for the DDB and CJKV-E dictionaries) by Louis-Dominique Dubaut</p> <p><a href="#">Subscribe to the DDB-User Newsletter</a></p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The main form of access is the search applet available from the top page, which can also be accessed from the pages generated from searches. This feature, created by Michael Beddow in 2001, has remained remarkably durable, still working fine after eight years to deliver data at an acceptable speed to users around the globe.

### A. Sticks and Carrots (飴と鞭)

Once the search function was established and we had developed the coverage to a significant degree, usage of the DDB increased rapidly. Yet despite our repeated pleas for user contributions, except for a very small number of unselfish and aware individuals who somehow naturally grasped the meaning of *collaboration*, we gradually came to realize that despite the large number of heavy users of the dictionary (readily seen in log data), virtually no one was willing to take the time to send us even a couple of terms from their own research work.

It also turned out that after the search function was working well, we began to be plagued with the problem of selfish individuals attempting to download the full data set. This created a technical headache, as these hackers would attempt to achieve their aims by writing scripts (or "robots" ), that made several requests per second on our server, thus slowing our system to a halt and preventing access by honest users. We dealt with this problem primarily by the setting of a quota limit, which would terminate a guest user's access at fifty within a twenty-four period. Contributors, on the other hand, could get an unlimited-use password.

Aside from this technical issue, the problem of the lack of contributions was extremely frustrating — especially given the awareness that many of our users were scholars or advanced students of East Asian Buddhism, or East Asian thought, history, etc., who were clearly quite capable contributing if they were merely willing to take a few minutes to compose some notes derived from their research. We then decided to apply the password system not only for hackers who wanted to take all of the data, but as a means to put pressure on heavy users of the resource, to force them into contributing in some way or another. We did this by creating two tiers of access privileges. The first level of access was that wherein any user could access the data a limited number of times, logging in with the user ID of *guest*, with no password. We started off setting the limit at fifty. Leaving it at this amount for about three months, generated virtually no reaction, in terms of either complaints or contributions. We then began to gradually drop the number down to forty, thirty, and then twenty searches in a day. At twenty, there was still barely a complaint made or contribution to be seen. When we tried the number of ten, however, everything changed. We were first bombarded with complaints, but holding the line, eventually these complaints began to turn into contributions. It is not an understatement to say that this was a watershed moment for the project, because we found that once people made their initial contribution, many of them continued to do so on a regular basis.

The basic policy to which we continue to adhere is that if someone wants to have full access for two years, they need to contribute the equivalent of one

single-spaced A4 page of their own materials, include one or multiple entries. There is flexibility in this policy, as there are a few steady users who in addition to offering new data, have been frequently proofreading and letting us know of errors and other shortcomings. We also accept contributions of a technical nature, and for those whose scholarly background is insufficient, but who have the requisite linguistic background, we offer source materials from East Asian reference works to be translated into English. For those who can convince us that they are absolutely not qualified to offer any kind of data contribution whatsoever, we also allow for the possibility of paying for a two-year subscription.

#### A. Institutional Acknowledgment

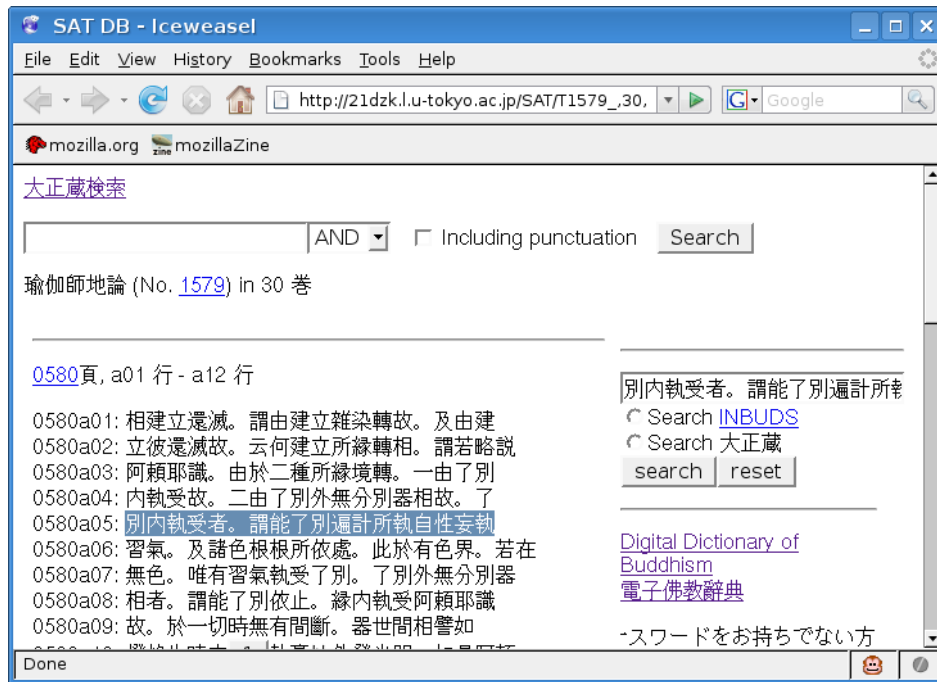
The continued growth in popularity of the DDB, especially as a reference work for graduate and undergraduate courses in Buddhist Studies in North America and Europe generated one more problem that needed to be solved in terms of access — that of how to provide for the use of the DDB in the kinds of situations where the instructor of a course wanted to use the dictionary for an undergrad or graduate course where there were constraints in the basic ability of the students to contribute, or the logistics of putting contributions together from the members of an entire class. To deal with these kinds of situations, without breaking our principle of making someone, somewhere, feel a certain sense of responsibility, we decided to begin to offer subscriptions to university library networks based on IP address. For a modest fee, university libraries may offer the DDB and CJKV-E dictionaries to their faculty and students. The creation of this policy brought about an unforeseen benefit, in that we could now provide a list of reputable institutions that deem the DDB to be an academic reference tool of high standards. At the time of this writing, we have subscriptions from twenty-five institutions, including many of the most prestigious universities and colleges from around the world.<sup>8</sup>

## IV. Applying the DDB to Specific Tasks

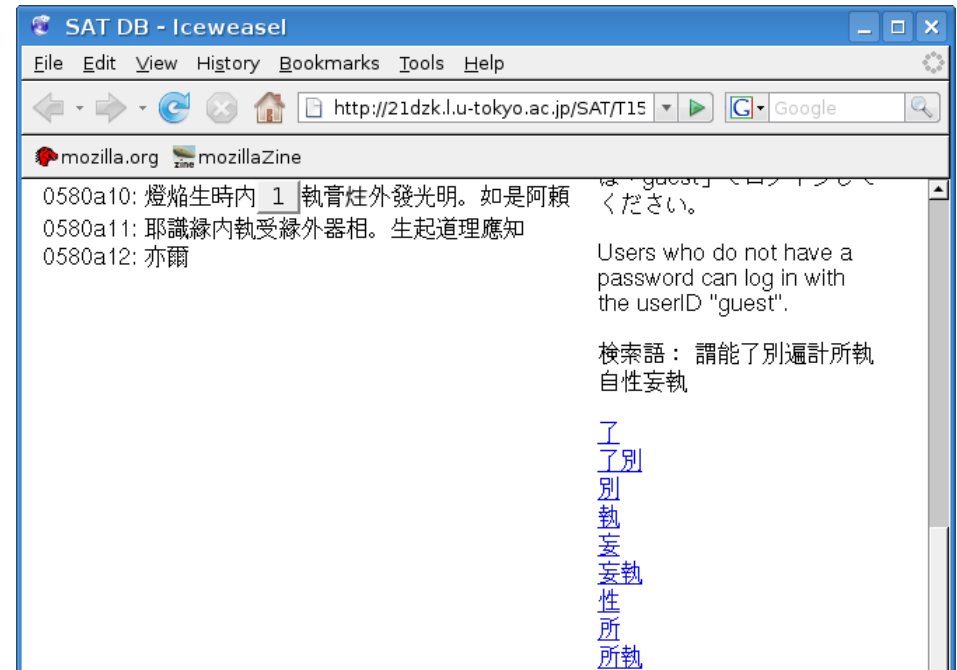
### A. Digital Canons Online

As noted earlier, the basic value of the DDB as a reference tool has been significantly enhanced by the change in the character of the very texts to which the DDB is intended to be applied — for understanding, interpretation, and translation. At the outset of the compilation of the DDB in 1986, the notion of the existence of a digital *Taishō Daizōkyō* was barely possible, but by the time the DDB first went on the Web in 1995, Urs App and Christian Wittern at Hanazono University had released their ZenBase CD ROM, including most of the important Chan canonical classical Chinese texts. Ven. Chongnim and his collaborators in Seoul were soon hard at work in the task of digitizing the *Tripitaka Koreana* [KT]. By 2000, the digitization of the KT was complete, and the CBETA group in Taipei and the SAT group in Tokyo were well on the way toward the completion of their respective digitized versions of the Taishō Canon. This was soon followed by the digitization of the *Hanguk bulgyo jeonseo* (Collected Works of Korean Buddhism) at Dongguk University (<http://ebti.dongguk.ac.kr/>). Today, in 2008, all of these canons are fully digitized and are available for usage via the web or locally, and are also being equipped with various applications for organizing, analyzing, and reading the data contained therein.<sup>9</sup>

These texts in digital format are a perfect match for a digital lexicon such as the DDB, as there is a wide range of ways that one may use computer technology in a way that both sides can take mutual great advantage. One of the most valuable accomplishments is the creation of methods of marking up terms in the subject text that are contained in the DDB "on-the-fly." This sort of function has recently been implemented online in the SAT *Daizōkyō* Database. Kiyonori NAGASAKI has developed an application wherein when a user/reader of the SAT database selects a portion of text with one's mouse...



...characters and compound words within that area of text that are contained in the DDB appear in the form of a vertical column on the right side. All the user needs to do is to click on one of these to consult the DDB:

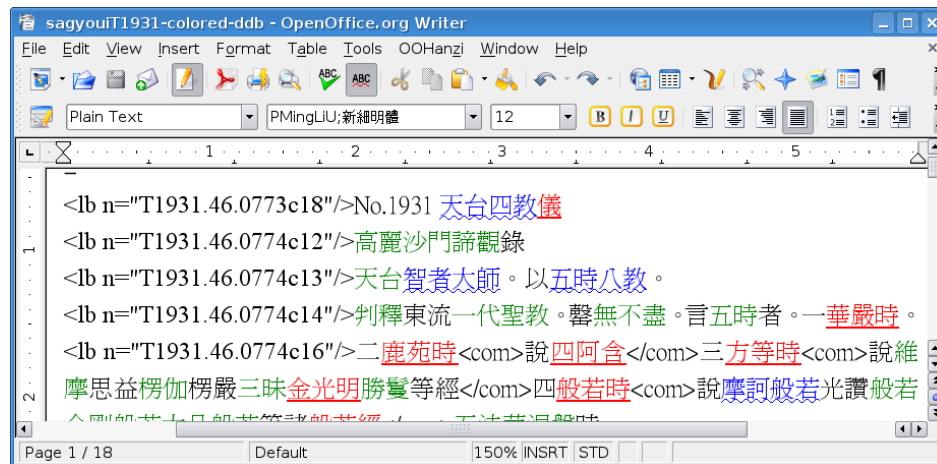


This development represents a huge step forward, one which, it would seem, any serious translator of classical Chinese Buddhist texts cannot possibly ignore. It is a tremendous boon to readers of any *Taishō* text, and equally auspicious for the continued development of the DDB, because it will certainly lead an increasing number of scholars to pay attention to the DDB, induce them to notify us of deficiencies, and hopefully stimulate more of them to contribute the fruits of their own research.

## B. Using the DDB Locally

This web-based approach is just one way to take advantage of the DDB in direct application to a text being studied. What is also of great help is an application

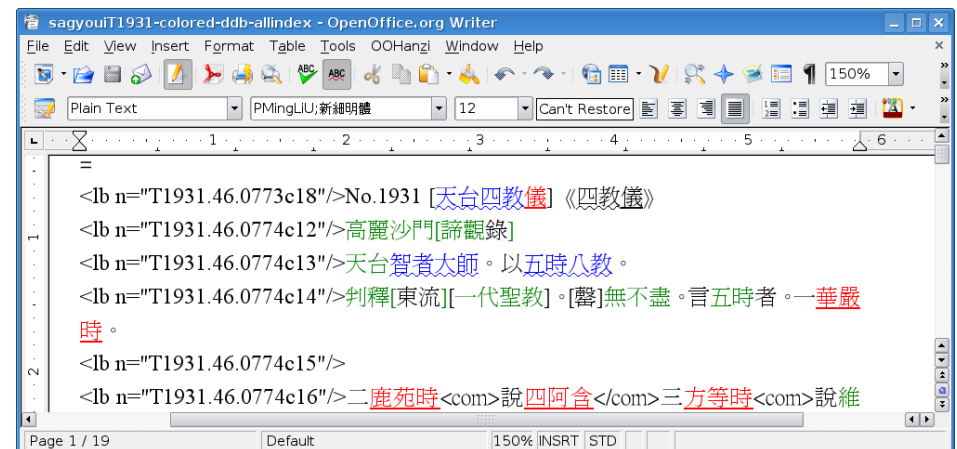
that allows researchers mark up a text on their local desktops without having to be connected to the web, and in a way that the terms contained the DDB can be retained for viewing. Such an application does exist, but unfortunately only on my own computer and a handful of others on whose machines it has been installed. This is a function that allows one to quickly mark (with color, brackets, etc.) text on-screen. This can be done in MS-Word or any other word processor that has a concordance feature, which will allow one to compare the text of the scripture with the current headword list from the DDB, and color the terms that are identified, like this:



Within a local system while working in MS-Word, if we select any of the colored terms with the mouse, that entry will be immediately retrieved from the DDB (or CJKV-E, as the case may be), with the option of editing. In this case, the data comes from the main source of the DDB on the local system. As a translator, this is an incredible advantage over the old ways of doing things.

### C. Attaching Markup from the Master Index File

The DDB is also intimately connected with an ongoing project called *Allindex*. This is a master index of a few dozen of the most popular East Asian lexicons and encyclopedias of Buddhism.<sup>10</sup> The *Allindex* file is used locally when working with the DDB to identify words in a given text that are not yet contained in the DDB, but which are contained in some reference work. So the next step after color-coding the DDB headwords, is to run another macro that marks all of the terms contained in the *Allindex* file with square brackets.<sup>11</sup> The result looks something like this:



For those of us who enjoy having a choice of platform and word-processor, being locked into Windows and Word like this is not really a very good thing. But this will eventually be done in a more open and efficient manner by a person with

real programming know-how. The availability of these kinds of tools serves to vastly increase the speed and accuracy of translation work, and is of course equally as helpful in the task of proofreading and editing the translations of others.

## Notes

1 The first scholar to offer help was Christian Wittern, at the time a graduate student affiliated with Hanazono University. Soon after, a number of scholars offered significant content contributions, including Gene Reeves, Jamie Hubbard, and Iain Sinclair.

2 At that time, building a search engine that could deal with mixed Western/CJK text in UTF-8 encoding was not a straightforward matter, so Michael's search engine was a novel creation — and it is still serving its purpose quite well today, eight years later. Fairly soon after the completion of this framework, Michael and I were asked to submit an article to the online *Journal of Digital Information*. That article, entitled "Moving into XML Functionality: The Combined Digital Dictionaries of Buddhism and East Asian Literary Terms," can be read at <http://journals.tdl.org/jodi/article/view/jodi-65/82> . (*Journal of Digital Information: Special Issue on Chinese Collections in the Digital Library*, Volume 3, issue 2, October 2002).

3 I have focused here on developments in the DDB, but please note that all of the same technological enhancements have been applied to the CJKV-E, except for the search through a comprehensive index, which, at present, has not yet been developed.

4 The EBTI is an open, expanding liaison group, comprised primarily of representatives of academic institutions and Buddhist clerical organizations from around the world, all of whom hold the common interest in meeting the new challenges, and taking advantage of the new opportunities presented with the advent of the electronic age into the area of humanistic studies. For details of the founding and ongoing activities of this group, please see <http://buddhism-dict.net/ebti/> .

5 The names of all contributors are listed in the middle column at <http://www.buddhism-dict.net/credits/credits-ddb.html>, in the approximate order of size or significance of their contributions.

6 See, for example <https://launchpad.net/oohanzi> .

7 <http://buddhism-dict.net/ddb.>

8 These institutions are listed at [http://www.buddhism-dict.net/ddb/subscribing\\_libraries.html](http://www.buddhism-dict.net/ddb/subscribing_libraries.html).

9 In prior articles dealing with the DDB and related topics, I had provided the URLs for all of these projects, but experience has proven that to be a somewhat wasteful exercise, as it is rarely the case that these remain stable for more than a couple of years. Additionally, it is nowadays relatively easy to find

these resources using Google.

10 This project was initiated by Urs App and Christian Wittern at Hanazono in the early 1990's under the name of ZenDicts. After the completion of that project, the ZenDicts file was further supplemented by myself (with the aid of JSPS grants) as well as collaborators at CBETA, RITK, and Christian himself. For details, see <http://www.buddhism-dict.net/ddb/allindex-intro.html>.

11 Both this process and the above coloring process can be done in a matter of seconds with the concordance function in MS-Word's indexing-concordance application, so it only takes a couple of minutes.