# The Digital Dictionary of Buddhism [DDB]

: A Model for the Sustainable Development of a Collaborative,

### Field-wide Web Reference Service

A. Charles Muller\*

목 차

- 1. History
- 2. Content Development
- 3. Usage
- 4. Applying the DDB to Online Digital Canons
- 5. "DDB/CJKV-E 2.0:" A Recent Major Upgrade

## 1. History

The compilation of the Digital Dictionary of Buddhism (DDB) and CJKV-English Dictionary (CJKV-E) began in 1986, a month or so into my first Buddhist texts readings class in

<sup>\*</sup> Center for Evolving Humanities, University of Tokyo

graduate school at the University of Virginia, upon my realization of the dearth of adequate lexicographical and other reference works in English language for the textual scholar of East Asian Buddhism — as well as for the student of the broader field of East Asian philosophy and religion. I decided, at this time, to save every term I looked up, and have continued that practice down to the present, through the course of studying scores of classical texts.

In 1986 none except the most forward-thinking of computer scientists had dreamed of such a thing as the WWWeb, so at the outset I was simply envisioning the eventual publication of the usual printed work. In 1994, the Web made its appearance, and it soon occurred to me that publication of my compiled material as a web resource might provide the dual advantage of (1) allowing me to make these materials available far sooner than if I waited until I had fully developed a proper print compilation (which could conceivably take decades), and (2) it might enable me to garner collaborators to hasten the compilation, broaden the scope of the coverage, and improve the accuracy of the material. So in the middle of 1995, I converted my WordPerfect word-processor files to HTML, and placed the dictionary on the web. It did not take long for my hopes to receive their first confirmation, as Christian Wittern (at the time a PhD student working at Hanazono University) discovered the DDB and applied a basic SGML structure to the data, which is the ancestor of the XML markup system used today. In addition to Christian's aid with the technical foundations of the work, other scholars slowly began to appear who offered data content contributions, and many of my colleagues began to use the DDB as a reference work in their university courses.

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 I had the great fortune to encounter Michael Beddow through the Web. Michael, a scholar of German Studies with a long career in humanities computing, who was extremely knowledgeable regarding the application of XML/XSLT technology with textual corpora, offered to program the DDB data such that XSLT and

<sup>1)</sup> For the record, the very first scholar to offer his data as DDB content was Gene Reeves, who contributed a glossary for his then in-progress translation of the Lotus Sūtra. He was soon followed by Jamie Hubba rd, who contributed a glossary of terms derived from his research on the Three Stages Sect.

X-Linking functionality could be produced in the latest versions of the standard browsers, and wrote a search engine in PERL to call up dictionary entries based on user queries. At that time, building a search engine that could deal with mixed Western/CJK text in UTF-8 encoding was a not at all a straightforward matter, so Michael's search engine was a bit of a novel creation — serving its purpose for a full ten years, up to his recent renovation of the technical framework of the project.<sup>2)</sup>

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 myself and my collaborators and 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.

## 2. Content Development

My first public presentation of the DDB at an academic conference was at the meeting of the Electronic Buddhist Text Initiative (EBTI)<sup>3)</sup> held in at the Fo Guang Shan temple in

<sup>2)</sup> Fairly soon after the completion of this framework, Michael and I were asked to submit an article to the e 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 <a href="http://journal.stdl.org/jodi/article/view/jodi-65/82">http://journal.stdl.org/jodi/article/view/jodi-65/82</a>. (Journal of Digital Information: Special Issue on Chinese Collections in the Digital Library, Volume 3, issue 2, October 2002).

<sup>3)</sup> The EBTI is an open, expanding liaison group, comprised primarily of representatives of academic instituti ons 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

Taipei in 1996. At that time, the DDB contained approximately 3,200 entries. At the time of the present writing (November 2010), that number has jumped to over 53,000 and is continuing to grow rapidly. This rapid growth is due to a number of factors, the most important of which is no doubt 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, along with the fact that reference works such as the *Fo Guang Dictionary, Ding Fubao, and Iwanami bukkyō jiten* are available in (legal) digital format, while a number of other dictionaries have been digitized and circulated privately.

The third major reason for the ability of the DDB to grow rapidly is that of the digitization of the Chinese Buddhist canons, a project first undertaken by the Research Institute of Tripitaka Koreana (RITK, which digitized the Korean Buddhist canon), and followed upon SAT Taishō Database (which digitized the Japanese Taishō canon) and by CBETA (which has digitized the Japanese Taishō was well as the Zokuzōkyō). 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 users of these text databases can 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 seventy. While all contributions are vitally important, there are a number which have been of special size or significance, which I would like to acknowledge here:

· In terms of sheer number of individual entries, the largest individual contribution has come from Seishi Karashima of Soka University, who contributed his glossaries from the Kumārajīva and Dharmaraksa translations of the Lotus Sūtra (T 262 and 263), accounting

the electronic age into the area of humanistic studies. For details of the founding and ongoing activities of this group, please see <a href="http://buddhism-dict.net/ebti/">http://buddhism-dict.net/ebti/</a>.

for more than 3,000 new entries.

· Stephen Hodge (U.K.) has contributed almost 1,000 new entries from his ongoing translation/glossary work on the Yogâcārabhūmi-śāstra

The above two data sets are large in terms of number of individual entries, but are, in terms of structure, relatively short, glossary type materials that do not contain fully developed explanations. In addition, we have had major contributions that are fewer in number of overall entries, but are comprised mostly of full-length explanatory sections, some as long as several pages. Among these are:

- The glossary of Tiantai terms provided by Paul Swanson (Nanzan University). This includes approximately 840 entries specific to Tiantai discourse, many of which go into considerable depth. Prof. Swanson has also included lexicographical materials from French sources.
- Approximately 2,000 terms added by Ockbae Chun (Dongguk University), primarily terms
  of relevance to Korean Buddhism. Many of the entries on persons and temples are also
  extensive. Mr. Chun has also done a great deal of valuable proofing of Korean readings
  and other facets of the work.
- The number of actual new entries provided by Dan Lusthaus is around 200, but many of these are extensive and detailed, providing a perspective on the material based on extensive research and deep insight. Dr. Lusthaus has also done extensive re-working of many pre-existent materials. The fact that the DDB has strong coverage in the area of Buddhist Logic (emanating mostly from his essay on sādhana 能立) and seminal Yogâcāra concepts is much due to his efforts.
- · Iain Sinclair's data on texts from Esoteric Buddhism, along with some terminology from that area (approximately 350 entries).

This is to mention only a few of the most prominent contributors. There is a long list of other scholars who have contributed entries, supporting lexicographical data sources, or who

have continued to volunteer extensive time in proofreading DDB materials and sending in their corrections. Their names are listed in the middle column at <a href="http://www.buddhism-dict.net/credits/credits-ddb.html">http://www.buddhism-dict.net/credits/credits-ddb.html</a>, in the approximate order of size or significance of their contributions.

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 OpenOffic e.4)

In terms of content development, we have also gained much from the digitization and assimilation of materials that we have been able to do through grants received from Japan Society for the Promotion of Science (JSPS). One grant in particular allowed us to digitize the *Dictionary of Chinese Buddhist Terms* by Soothill and Hodous. While much of this material was dated and needed reworking, the majority of the basic definitions were useful enough for us to include at least provisionally, allowing us to provide a basic coverage of more than 14,000 entries. At first this material was simply added "as is," but we have been steadily working through it, utilizing what we can in the most effective way. The same grant also enabled us to digitize Lewis Lancaster's landmark work, *The Korean Buddhist Canon: A Descriptive Catalogue*. Using the data from this compilation whenever we create a new entry on a text from the Chinese canon, we are able to quickly gain all the basic information of dating, provenance, translation, variant works, and so forth. To this we are able to add content information for the given text from other sources. And of course, we can at the same time include corrections based on interim research.

4) See <a href="https://launchpad.net/oohanzi">https://launchpad.net/oohanzi</a>.

### 3. 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. 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 certain purpose for study and research, such as when one is looking for a term, person, place, etc. for which she or he does not know the proper spelling, or has forgotten it. These indexes also serve as extensive glossaries.

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.<sup>5)</sup>

### 3.1. Sticks and Carrots

Once Michael had set up the search function, 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 his or her own research work.

<sup>5)</sup> It should be noted that the continued function of this search application has necessitated frequent attention from Michael, as it is often the case that when our former web hosting service upgraded FreeBSD support libraries (such as for PERL and so forth), some portion of the search program ended up being broken, and thus needed some rewriting. We have recently moved to a new host that uses straight Linux, and this has solved these problems.

It also turned out that after Michael got the search function working well, we began to be plagued with the problem of selfish individuals attempting to download the full data set. While I was of course bothered by such an attitude being taken toward a compilation that I had labored at for a couple of decades, the greater problem was actually a technical one—as these hackers would attempt to achieve their aims by writing scripts (or "robots"), that made several requests per second on our server, thus clogging our system and preventing access by honest users. Michael dealt with this problem with a couple of different strategies, one of which included the setting of a quota limit, which would terminate a guest user's access at fifty for a twenty-four period. Registered 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 of contributing material from their own studies. The idea then came 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. Thus, we decided to experiment with creating two tiers of access privileges. The first level of access was one 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, we received neither complaints nor 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 nary a complaint made nor 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. This was a watershed moment for the project, because we found that once people contributed one time, 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. This page can include one entry or ten — it doesn't matter. There is some 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.<sup>6)</sup> 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 at the rate of US \$60.

### 3.2. 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 the modest fee of \$600 for two years, 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 which had deemed the DDB to be an academic reference tool of high standards. At the time of this writing, we have subscriptions from thirty institutions, including many of the most prestigious universities and colleges from around the world.<sup>7)</sup>

<sup>6)</sup> A few of the more steady contributors of this category include Gene Reeves, Robert Kritzer, Dan Lusthau

s, Hudaya Kandhajaya, Charles Jones, Jimmy Yu, Karen Mack, Ockbae Chun, and John McRae.

## 4. Applying the DDB to Online Digital Canons

As noted earlier, the popularity, indeed, 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 hard at work in the task of digitizing the Tripitaka Koreana [KT]. By 2000, the digitization of the KT was complete, and the CBETA team in Taipei and the SAT team in Tokyo were well on the way toward their respective digitizations of the Taishō Canon. Today, in 2010, all of these canons are 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. In addition to the basic set of Taishō texts, CBETA has also digitized and made available the Zokuzōkyō. A team at Dongguk University led by Ven. Bogwang has provided us with a digitized version of the Collected Works of Korean Buddhism (Hanguk bulgyo jeonseo), and the digitization of other East Asian Buddhist collections is in progress both inside of the above-mentioned groups and elsewhere.

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 for both sides to take great advantage of each other. This possibility has recently been elegantly met in the recent re-release of the SAT Daizōkyō Database (<a href="http://21dzk.l.u-tokyo.ac.jp/SAT">http://21dzk.l.u-tokyo.ac.jp/SAT</a>). Kiyonori Nagasaki, a Madhyamaka specialist who is also a first-rate database/web programmer, 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

<sup>7)</sup> These institutions are listed at <a href="http://www.buddhism-dict.net/ddb/subscribing\_libraries.html">http://www.buddhism-dict.net/ddb/subscribing\_libraries.html</a>>.

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.

At this point in time, 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.

## 5. "DDB/CJKV-E 2.0:" A Recent Major Upgrade

After almost ten years of operation since Michael Beddow's initial creation of the programming structure for the online CJKV-E/DDB dictionaries, we are delighted to announce a major upgrade of these web services.

The most basic components of this upgrade are (1) a move to a dedicated server which will be able to deliver more power to search functions and greater stability to Unicode-related programming, and (2) an entire rewriting of the underlying search and indexing routines, resulting in a noticeable increase in speed and variety of search results, and links to both internal and external resources. Some major specific additions and enhancements include:

### 5.1. Basic Search

- 1. A middle-level of search results, showing a list of head words that contain the search term. Previously, searches for a term would produce only the headword itself (when it existed), along with a long, scattered list of entry body matches.
- 2. The list of body entry matches, which was previously delivered without any particular

- ordering, is now sorted according to traditional ascending radical + stroke count (basically equivalent to Unicode hex number).
- 3. The list of matched body entries now includes a snippet of context, to give the user some hint of the usefulness of each listed match.
- 4. Head word searches via Pinyin, Hangeul, Korean romanization, Katakana, and Japanese romanization. Previously, searches for headwords via their various renderings in East Asian and romanized syllabaries would only yield matches as body entries. Now, dedicated search indexes for Pinyin, etc. will yield head word matches in a very fast search.
- 5. Searches with or without diacritics are equally and transparently supported. Searches employing those romanization systems that use diacritics may also be made with or without diacritical marks (though in the nature of things the latter may produce some false positives). This also applies to searches for Sanskrit and Pāli terms in entry bodies.

### 5.2. Entry Results

- 1. Previously, hyperlinks to terms within displayed entries sometimes lacked actual targets, or led to the comprehensive external index in a roundabout manner. Now, if a term currently has no target in the dictionary concerned or (in the case of the DDB) in the external index, it will be shown without a hyperlink.
- 2. If the link goes to the comprehensive external index rather than the DDB itself, the user will be taken directly to that information, with no other message or page in between.
- 3. If the headword of a DDB entry is also present in the CJKV-E, a hyperlink to that entry will automatically be added to the DDB entry when it is displayed. The converse applies to CJKV-E entries: if the DDB has an entry for the same headword, a link to it will be added to the CJKV-E entry on the display.
- 4. A link for a direct search to the SAT Taishō Database will automatically be generated for DDB entries (we are also able and willing to generate links directly into other

web-based canonical collections if the administrators of those collections are willing to provide us with the requisite code for such links).

### 5.3. Behind-the-scenes

There are other enhancements which, while not visible to users, will greatly improve the function of both dictionaries. Most importantly:

- 1. The two main indexes (on headwords and fulltext) previously used have been completely re-implemented to give faster and more flexible matching. In addition, a number of specialized supplementary indexes have been added which are automatically invoked alongside or instead of the main indexes as and when appropriate.
- 2. Index updating has been made significantly faster and extensively automated. This means that all the indexes can be regenerated as frequently as desired. So from now on, corrections to existing entries, as well as newly-contributed entries, will be browsable and searchable in their entirety very shortly after editorial acceptance (assuming of course, that the Human in-charge is not indisposed for some reason or other!).
- 3. Great care has been taken to ensure that hyperlinks on external sites to DDB and CJKV-E entries which employ the syntax of the previous implementation of the Dictionaries continue to function exactly as before. No existing external links made in accordance with the methods previously specified for creating such links will be broken as a result of the new infrastructure.

### 5.4. CJKV-E

 In the process of preparing this upgrade, a great amount of work has been put into improving the structure and content of the CJKV-E dictionary, which has stayed pretty much on the back burner for the past decade or so. Greater attention will henceforth be given to the development of this resource. 2. In fact, I am presently working with a small grant that will have the effect of drastically increasing the coverage of the CJKV-E over the next few years.

### 5.5. Real Collaboration

Due credit must be given to those scholars who have provided staunch and enduring support for the DDB over the past decade. Most importantly to Michael Beddow, who has, without any monetary remuneration whatsoever, provided state-of-the art programming of these dictionaries (along with web security and all other related functions), buttressed by a matched level of understanding of lexicographical and linguistic principles that has provided us with so much of the structure and precision that these online references currently exhibit. Many of the technical enhancements are based on Michael's work on the Anglo-Norman Dictionary (<a href="http://www.anglo-norman.net/">http://www.anglo-norman.net/</a>), funded by the Arts and Humanities Research Council of the United Kingdom, whose indirect but significant support is gratefully acknowledged.

There is also a core group of approximately 20 scholars, many of them recognized as leading figures in their own areas of expertise, who have continued to generously contribute large amounts of material from their own research notes and glossaries. They have also spent much time in scouring previously-existent entries, amending, appending, and entirely rewriting, such that the DDB and CJKV-E are in a steady state of growth in size and accuracy (the names of these scholars can be browsed at <a href="http://www.buddhism-dict.net/credits/credits-ddb.html">http://www.buddhism-dict.net/credits/credits-ddb.html</a>). I would also like to thank those scholars who have convinced their libraries of the value of an institutional subscription. The resulting funds, albeit modest, have been invaluable to help pay for infrastructure, web hosting, and the employment of part-time assistants to do input and editing.

I believe we can say that there are few, if any, other examples in the academic humanities field where a body of scholars, bonded by overlapping interests but spread across the globe, have contributed to a central resource on such a scale, upholding rigorous standards of composition, accreditation, and citation, and providing an eminently practical and useful

example of how we can collaborate to build resources that are far more substantial than mere anonymous aggregations.

Digital Dictionary of Buddhism: http://buddhism-dict.net/ddb

CJKV-E Dictionary: http://buddhism-dict.net/dealt

### **Abstract**

## The Digital Dictionary of Buddhism [DDB]

: A Model for the Sustainable Development of a Collaborative, Field-wide Web Reference Service

A. Charles Muller
Center for Evolving Humanities
University of Tokyo, Japan

The Digital Dictionary of Buddhism [DDB] (<a href="http://buddhism-dict.net/ddb">http://buddhism-dict.net/ddb</a>), now on the Web for more than 15 years, has become a primary reference work for the field of Buddhist Studies. Containing over 53,000 entries, it is subscribed to by more than 30 university libraries(<a href="http://www.buddhism-dict.net/ddb/subscribing\_libraries.html">http://www.buddhism-dict.net/ddb/subscribing\_libraries.html</a>), and supported by the contributions of over 70 specialists, many of these recognized leaders in the field. It can perhaps be described as example of the type of web resource that has reached a degree of status and sustainability such that it has been able to grow and thrive as a collaboratively-developed online reference—despite having little funding or the support of a major organization or team of programmers—in the age where such resources are so readily washed away by the combination of Wikipedia and Google. Thus, the field of Buddhist Studies has its own reliable, scholarly-edited, fully documented and responsible resource that has developed a center of gravity sufficient for it to continue to grow as the resource that specialists turn to first without hesitation, and to which they may contribute knowing that they will be clearly accredited, and that what they write will not be deleted or changed in the following moment by, for example, a junior high school student.

Recently, the technical advisor to the DDB, Michael Beddow, has completed a full overhaul of the supporting structure of the DDB and CJKV-E dictionaries, which will include a broad range of enhanced functions, both internal to the dictionary and in terms of interoperation

47

The Digital Dictionary of Buddhism [DDB](A. Charles Muller)

with other lexicons and web corpora. This presentation will start off with a demonstration of the most advanced functions of the DDB, to be followed by a brief overview of its technical framework (P5-influenced XML, delivered through XSL and Perl). We will then outline the key factors of the management of the DDB that we believe have most directly contributed to its success.

논문투고일: 2010.11.30. 심사완료일: 2010.12.16. 게재확정일: 2010.12.18.