

The Digital Dictionary of Buddhism [DDB]: Present Status and Future Developments

A. Charles MULLER

The University of Tokyo

I. 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 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 I 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 even dreamed of such a thing as the WWW, 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 6.0 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 promptly downloaded it. He 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,¹ and many of my colleagues began to use the DDB as a reference work in their university courses.

¹ 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 Hubbard, who contributed a glossary of terms derived from his research on the Three Stages Sect.

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 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 — and it is still serving its purpose quite well today, eight years later.²

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.³

II. Content Development

My first public presentation of the DDB at an academic conference was at the meeting of the Electronic Buddhist Text Initiative (EBTI)⁴ held at the Fo Guang

² 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/>> .

Shan temple in Taipei in 1996. At that time, the DDB contained approximately 3,200 entries. At the time of the present writing (May 2008), that number has jumped to over 45,000 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 (too bad the publishers of these dictionaries didn't take advantage when the time was right).

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 Tripiṭaka 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 as 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 (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. 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 Dharmarakṣa translations of the *Lotus Sūtra* (T 262 and 263), accounting for almost 7,000 new entries.
- Stephen Hodge (U.K.) has contributed almost 2,500 new entries from his ongoing translation/glossary work on the *Yogācārabhūmi-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 1,200 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 100, 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 Yogācāra 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 <http://www.buddhism-dict.net/credits/credits-ddb.html>, 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 OpenOffice.⁵

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.

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.⁶ 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, 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 in themselves as extensive glossaries. Furthermore, since nowadays all these of materials are indexed by Google and other search entities, the presence of the indexes provides access to the dictionary to people who are just performing general web searches.

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

⁵ See, for example <<https://launchpad.net/oohanzi>> .

⁶ <<http://buddhism-dict.net/ddb>>

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.⁷

A. 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.

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 slowing our system to a halt 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 doing so. The idea then came to me 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 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

⁷ 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 web hosting service upgrades its support libraries (such as for PERL and so forth), some portion of the search program ends up being broken, and thus needs some rewriting.

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 mostly 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 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.⁸ 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 \$110.

B. 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 \$500 for two years, university libraries may offer the DDB and

⁸ A few of the more steady contributors of this category include Gene Reeves, Robert Kritzer, Dan Lusthaus, Hudaya Kandhajaya, Charles Jones, Jimmy Yu, Karen Mack, Ockbae Chun, and John McRae.

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 twenty institutions, including many of the most prestigious universities and colleges from around the world.⁹

IV. Applying the DDB for Specific Tasks

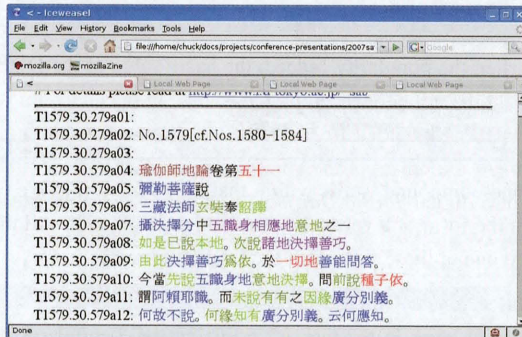
A. Digital Canons Online

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 group in Taipei and the SAT group in Tokyo were well on the way toward their respective digitizations of the Taishō Canon. Today, in 2008, 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

⁹ These institutions, which are listed at <http://www.buddhism-dict.net/ddb/subscribing_libraries.html>, presently include: Australian National University, Columbia University, Duke University, Harvard University, International College for Postgraduate Buddhist Studies, McMaster University, Princeton University, Smith College, Stanford University, University of California, Berkeley, University of California, Los Angeles, University of California, Santa Barbara, University of Chicago, University of Pennsylvania, University of Heidelberg, University of Southern California, University of the West, University of Wisconsin-Madison, and Yale University. For the record, the first institution to take up a subscription was Columbia University.

collections is in progress both inside of the above-mentioned groups and elsewhere.¹⁰

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. One of the things we began to do early on was to generate color-coded versions of popular texts, wherein determined-length strings of characters of terms that were contained in the DDB would be marked with a distinguishing color, and linked to their location in the DDB, as below (figure 1):

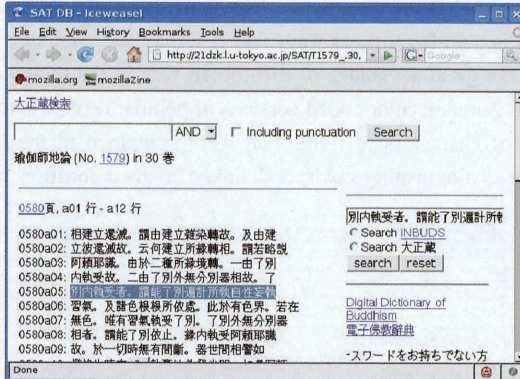


While this sort of application was appreciated by a number of scholars, it was subject to the limitations inherent in any hard-linked document, since these color-coded versions would need to be periodically updated to be useful, and the continuous updating of the huge volume of texts contained in the Taishō canon is not really a viable option. What is needed, is a way to flag terms contained in the DDB “on-the-fly,” — i.e. according to the status of the DDB at that very moment. Also, one does not really need to mark up an entire text (especially, for example, such a large text as the *Yogācārabhūmi*, shown above graphic); one only needs to have lexicographical information on the data contained in section of text one is working with at the moment.

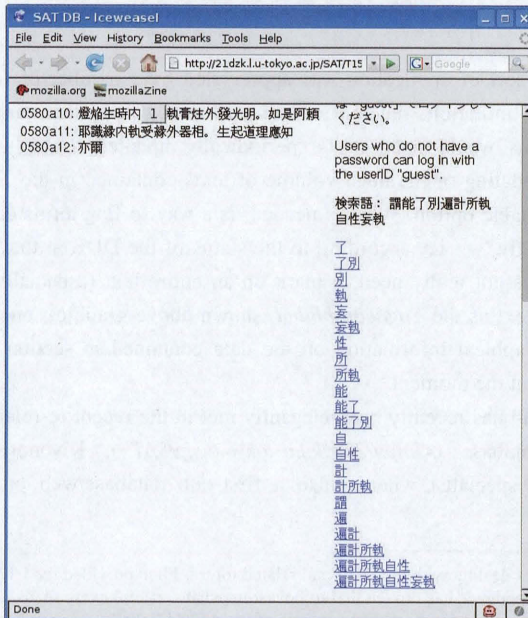
This demand has recently been elegantly met in the recent re-release of the SAT Daizokyo Database (<<http://21dzk.l.u-tokyo.ac.jp/SAT>>). Kiyonori Nagasaki, a Madhyamaka specialist who is also a first-rate database/web programmer, has

¹⁰ 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.

developed an application wherein when a user/reader of the SAT database selects a portion of text with one's mouse (figure 2):



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 (figure 3).



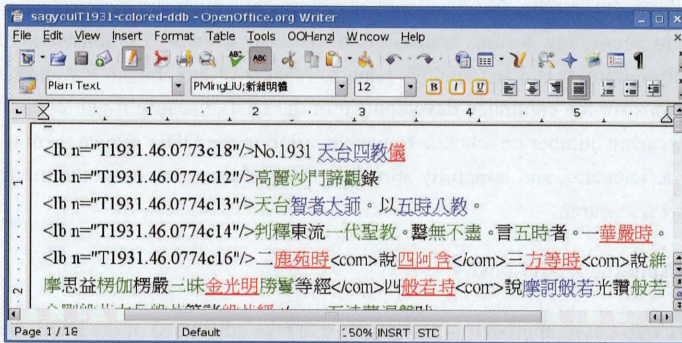
At this point in time, this development represents a huge step forward, one which, it would seem, any serious translator of classical Chinese Buddhist texts cannot possibly ignore. It is a tremendous boon to readers of any Taishō text, and equally auspicious for the continued development of the DDB, 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.

B. Using the DDB Locally

This web-based approach is just one way to take advantage of the DDB, and there is also a demonstrated level of usefulness in color-coding as shown above. Thus, in the future, it would be wonderful if we could develop a readily portable and installable application that allows users to do the same sort of thing on their own desktop without having to be connected to the web, or necessarily using a predetermined digital version of the Taishō, or other canon.

Such an application actually does exist, but unfortunately only on my own desktop and a handful of others on whose machines I've taken the time to install it. What I am talking about here is an array of MS-Word VBA macros that allow me to quickly mark (with color, brackets, etc.) text on-screen, as well as look up, and add terms on the fly. The problem with this system is that as marvelously as it works on my own system, it usually takes a few hours to properly set up on someone else's computer, even if they are using the same version of Windows and Office as I am. There are endless tiny differences in configuration that gradually work their way into an individual's PC that need to be addressed. Nonetheless, the basic structure of my local DDB system stands as a model for an application that could someday be developed by a person with the right programming skills, although it would certainly be more desirable to create such a system in an Open Source environment such as that of OpenOffice.

The system basically works like this. When I am getting ready to translate or study a text, I take a version of a text from the SAT, CBETA, or HBJ digital corpora and run a Word macro on it that compares the text of the scripture with the current headword list from the DDB, and I color the terms that are identified in the same way shown above in the example of the *Yogācārabhūmi*, like this:

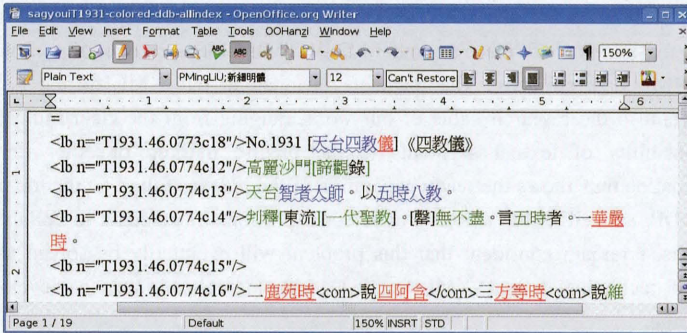


Within my local system while working in MS-Word, if I select any of the colored terms with my mouse, that entry will be immediately retrieved from the DDB (or CJKV-E, as the case may be), with the option of editing. In this case, the data comes from the main source of the DDB on my local system. As a translator, this is an incredible advantage over the old ways of doing things, and most importantly, perhaps, not for the terms that I *don't* know (which I would certainly have to look up, even in the old days), but for the terms that I *think* I know, but regarding which I am not 100% certain. But there is one more important step here.

As briefly noted above, the DDB is also intimately connected with an ongoing project which I have labeled with the name *Allindex*. This is a master index of a few dozen of the most popular East Asian lexicons and encyclopedias of Buddhism. The project was initiated by Urs App and Christian Wittern at Hanazono 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.¹¹ I use the Allindex file locally when I am 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 when I am working with a canonical text, the next step I take 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.¹² The result looks something like this:

¹¹ For more details, see <http://www.buddhism-dict.net/ddb/allindex-intro.html> .

¹² 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.



At the time I first began to translate this particular text, all the terms contained inside the square brackets were not yet contained in the DDB, but were available in some lexicon. Working inside of Word, if I select one of these terms, I will immediately receive information as to what lexicon contains the term, along with the page number. This means that I can also quickly add the term to the DDB through another array of Adding and Editing macros that I have developed.

Once again, being embedded in Windows and Word like this is not really a very good thing, but if it can be done this way by someone with computing skills as limited as mine, something like this can, and no doubt will be done in a more open and efficient manner by a person with real programming know-how. In any case, the availability of these kinds of tools has vastly increased the speed and accuracy of my own translations, and is of course equally as helpful in the task of proofreading and editing the translations of others, a task that is coming to occupy a steadily greater portion of my time.

V. The Big Wish: Real-time User Input and Editing

In thinking of possible useful developments and applications of the DDB and CJKV-E dictionaries, there are many, many things that can be done, but the most important single development that I would like to achieve at the earliest time possible is that of the capacity for real-time editing of entries, something like we see in Wikipedia. Of course, in the case of the DDB, editing privileges would be limited to accredited and trustworthy specialist scholars of the highest caliber. But I have no doubt that many of the regular scholarly users of the DDB who currently do not contribute would be considerably more motivated to correct an error or shortcoming that they see on the screen in front of them if they were able to do so.

The reason that we have not yet been able to provide this capacity is that it is not really feasible to try to place a standard flat-HTML CMS system on top of our TEI-like XML data structure. The metadata included in our XML format is vitally important to the research value of our work, ranging from the clear indication of responsibility of textual segments to the precise markup of other important information that allows the ready generation of the indexes found on the top page of the DDB, as well as the delivery of selected portions of our data to collaborating projects. I remain confident that this problem will eventually be solved as web-related technologies and XML related technologies continue to develop and converge.