

The transmission of the Buddha's teachings in the digital age

Sumachaya Harnsukworapanich and Phatchareporn Suphipat
Dhammachai Institute

Abstract

The objective of this paper is to report on and delineate the creation of an online version of the Dhammachai Tipiṭaka Edition, named D-Tipitaka, which has been in development since 2012 by the Dhammachai Tipiṭaka Project (DTP), Thailand. D-Tipitaka is a web page giving access to a series of Pāli texts from the Dhammachai Tipiṭaka Edition(De). It also provides an easy access to manuscript images and other relevant evidence by clicking key words. From over 10,000 digitized Pāli Tipiṭaka manuscripts from Sinhalese, Burmese, Khom, and Tham traditions, the D-Tipitaka displays digital images of 21 sets, totaling approximately 5,000 folios, of selected manuscripts utilized in the creation of the first volume in the series. Their readings were transcribed into Roman script to produce a digital text database which has been emended by researchers. With the help of a state-of-the-art text alignment and editing program, the text has been examined, and critically edited using the academic methods of modern philology. Finally, the edited text and their respective manuscript images are shown on the D-Tipitaka online version. Thus, this developed tool will serve as an alternative and convenient platform to consult Buddhist teachings along with the primary source.

1 Introduction

Buddhist tradition holds that after the Buddha's time, devoted disciples preserved and transmitted the Dhamma originally by means of oral recitation. This made it possible for the authentic and authoritative teachings of the Buddha to be accurately handed down word for word in its original state. After the Buddha attained Nirvana, this form of oral transmission continued to be used in the first three Buddhist Councils. Until approximately 400 B.E., the method of writing was introduced in the Fourth Buddhist Council where canonical text divided into three parts known as Tipiṭaka was written down on palm leaves for the first time in Sri Lanka during the reign of King Vaṭṭa Gāmaṇī Abhaya. Since then, the written transmission has been maintained by many Theravada countries in which palm-leaf manuscripts were inscribed in increasingly different local scripts. In present, the extant palm-leaf manuscripts can be divided into at least four major traditions: Sinhalese, Burmese, Khom, and Tham scripts. Despite their immense historical and cultural value, palm-leaf Pāli manuscripts are susceptible, over time, to many factors, such as climate, insects, ignorance of people, and adverse storage. Taking this matter into consideration, Wat Phra Dhammakaya launched the Dhammachai Tipiṭaka Project (DTP) in 2010 with the aims to preserve the existing palm-leaf manuscripts of various traditions, and to create digitalized and computerized databases of a new critical edition of the Pāli canon (D-Tipitaka) in order to support Buddhist studies. The work is classified into 4 steps as follows: collecting and digitizing the palm-leaf manuscripts, creating databases of digitized images and readings, text analyzing and editing, and finally publishing the critical edition of the Pāli canon. To effectively achieve the Project's goal, the Digital Humanities Center (DHC) of the DTP has developed several computer programs to support the work in every step. This paper will explain the process and the programs created in the production of the first section of the Dīghanikāya, the Sīlakkhandhavagga (DN1).

2 Collecting and digitizing Pāli Tipiṭaka palm-leaf manuscripts

The initial step of the work is to preserve large amounts of palm-leaf manuscripts from various countries in digital format. These color digital images safeguard the primary sources against deterioration over times and permit display or reproduction with image fidelity. Since 2010, the project has digitized over 10,000 bundles from the Sinhalese, Burmese, Khom, and Tham Pāli traditions. The criteria for selection for digitizing are: monolingual Pāli canonical texts, physical condition, completeness of content, age over 100 years old, historical value, and regional diversity. Based on these criteria, the collection of digital images are primarily given to Pāli Canon of Theravada Buddhism and secondarily to other canonical Buddhist literatures. According to DTP's manuscript data the average age of digitized manuscripts is approximately 260 years old, and the oldest dated manuscript is Sinhalese script manuscript of Saṃyutta Nikāya created in the 7th century. Besides yielding preservation benefit, an extensive collection of palm-leaf manuscripts greatly helps in the creation of the critical edition of the Pāli canon as the more digitized manuscripts we have in hand, the clearer lineage within and across manuscript traditions we can establish. The preservation process includes manuscript selection, manuscript registration, colophon reading and age determination, cleaning, page order sequencing, and digitization. To support digitization work, the DHC has developed 2 online programs: "Manuscript Journey" (MSJ) and "Application of Palm-leaf Manuscript Segmentation" (APMS). MSJ has been specifically designed to create metadata describing features of palm-leaf manuscripts. In the registration process, instead of filling in a paper-based information card, MSJ enables staff members to systematically record metadata schemas including bibliographic information (e.g. title, date of inscription, name of donor, name of scribe, etc.), physical characteristics (e.g. script, language, number of pages, size, condition, etc.), and administrative information (e.g. original registration code, location, date of digitization, legibility, etc.). These electronic information cards are useful for long-term management and enable searching and sorting for the rapid retrieval of information. The second program of APMS is developed under the Lanna OCR Project, a cooperation between the DTP and the Department of Computer Science, Chiang Mai University (CMU). It is a palm-leaf segmentation tool, which automatically cuts a single five-folio digital image file into five separated one-folio digital image files. It then renames and sequences the files according to manuscript page numbers and inserts descriptive labels on images. So far, over 10,000 manuscripts have been segmented into an estimated 2,000,000 one-folio image files. The information contain in the MSJ and APMS programs is essential for the on-going development of an online palm-leaf manuscript index database and a digital library. These programs also provide easy access to publicly available collections from any location thereby facilitating the study of Buddhist heritage without traveling to manuscript repositories.

3 Creating databases of digitized images and readings

The work in this step is to create databases of selected manuscript images collected in the previous steps and their readings. On average, in the production of a book in the series, at least 5 sets of the best representatives of palm-leaf manuscripts from each tradition are selected. For the creation of Dīghanikāya Sīlakkhandhavagga (DN1), Suttantapiṭaka, a total 20 sets of palm-leaf manuscripts from four main traditions plus one set of a rare palm-leaf manuscript from the Mon tradition have been used. Almost 5,000 one-folio image files from these 21 manuscripts were assigned to local staff members in Thailand, Sri Lanka, Myanmar, and Cambodia. Each local office has the responsibility to transliterate the Tham, Sinhalese, Burmese, and Khom readings into Roman. To support the distribution of the very large image files to different locations, and the centralization of input digital text across locations to the main server in Thailand, the DHC has developed an Online Data Entry of Manuscripts (ODEM) program [Figure 1]. It makes palm-leaf manuscript images available online page by page for staff members to read and then convert to Roman script. The program uses as a base text the Burmese Pāli text (Be) of the Chaṭṭha Saṅgāyana Tipiṭaka. It serves as a preliminary text for staff members to read, check,

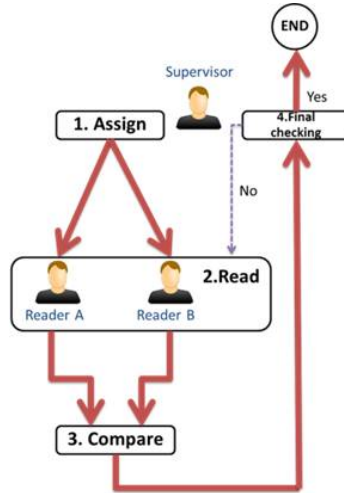


Figure 1: Data Flow of ODEM

alter, and edit in accordance with the text as appeared on the palm-leaf manuscript images by the principle of “Key as You See”. In other words, staff members, without taking Pāli grammar into consideration, simply read the Pāli manuscripts and input their reading into the database by altering Burmese Pāli text (Be) of the Chaṭṭha Saṅgāyana Tipiṭaka. To assure accuracy, ODEM allows paired-work in which 2 staff members, called readers, can simultaneously work on the same image files and compare their input data until the result indicates “zero difference” meaning that the transliterated work is exactly the same. Furthermore, for the ultimate accuracy, ODEM allows a third person with Pāli and IT code knowledge, called the final checker, to ascertain that it is 100 % correct and technical symbols have been thoroughly applied.

4 Text analyzing and editing

The work in this step involves textual aligning and computer-based editing. To facilitate its critical editing by Buddhist scholars, the textual database is transferred into a software system developed by DHC named Indic Text Analysis Program (ITAP) and Editing Page (EP). Applying the idea from CollateX, one of the most well-known programs for collating textual sources, the DHC has designed ITAP with Matching and Editing Page (EP) features: Matching helps automatically collate and split transcribed textual data obtained from ODEM at the word level. After that the segmented words will be aligned and all variant readings from every palm-leaf manuscript will be dynamically arranged in a synoptic table along with links to relevant manuscript images. However, due to the complexity of canonical Pāli texts, this automatic variant alignment done by the program reaches only 80 % accurate; human involvement is required to manually adjust the text into a synoptic form. Next, this well-aligned text is transferred to the Editing Page (EP), in which the text is rearranged in a conveniently editable format; it is here that the Pāli text can be critically edited. After the editing process, the selected readings or, in some cases, emendations are kept into EP database together with their annotation, if any. In the final process, EP generates the entire edited text into format ready for the publication of the critical edition of Pāli canon in printed edition and online version. The printed edition will contain the critically established texts and selected critical apparatus showing only variant readings of historical and grammatical value. The online version will contain three selective displays (Text View, Book View, and Synoptic View) of the critically established text. Moreover, it can link to manuscript digital images together with other printed references and a list of every single word appearing in both the critically established texts and critical apparatus.

5 Publishing critical edition of Pāli canon

5.1 Introduction

After researchers have finished editing the text, the final text will be sent to create text layout using Adobe InDesign, a publishing and typesetting software. The program will produce the printed and digital editions in PDF and EPUB formats. The EPUB format is utilized to display the online version of Dhammachai Tipiṭaka called D-Tipitaka because it supports many e-readers and is compatible with software and applications available in most devices, such as smartphones, tablets, and computers. Moreover, instead of creating a Windows-Based Application requiring program installment, the Web-Based Application format will be able to be conveniently accessed via the internet anywhere and anytime without installing any program.

5.2 Algorithm

The process of data preparation of the D-Tipitaka starts from importing the edited text from researchers from the Editing Page (EP) into Microsoft Word (MS Word) for a chief editor to make final amendments. The PDF file alone is used for the publication of the Dhammachai Edition (De) printed edition, while both PDF and EPUB files are imported into the database of D-Tipitaka through Linking Process. It uses EPUB to generate XML and HTML tags with diverse attributes, e.g. idword, idgroup, and span. The encoded data preserves character format (e.g. upper case, italicized, superscript, or bold text), document format (page layout), and paragraph format (e.g. indentation, line spacing, or alignment) of printed edition. However, the formatting tags alone are insufficient as they indicate position PX not text line number. Therefore, PDF is necessary to locate line number and to encapsulate all formats. Through the collaborative performance of EPUB and PDF, the data of text line number and text position are combined. The illustration below [Figure 2] shows the overall process of data preparation.

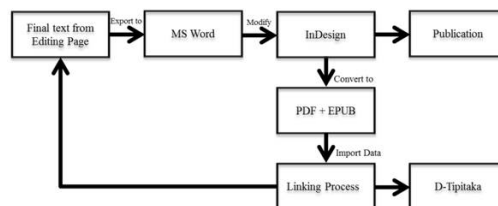


Figure 2: The Process of Data Preparation

The illustration number [Figure 3] reveals the database flow. The ITAP and EP databases operate using Microsoft SQL Server (MSSQL) while Linking and D-Tipitaka databases using PostgreSQL (PGSQL). The reason for using duo databases instead of single database is the enhancement of processing capabilities. MSSQL has been used since the beginning of the project and later PGSQL was utilized for its claim to be the world's most advanced open-source database, and for its safety of server connection and capacity of complicated analytical processes. Using two databases enable users to achieve abrupt analytical response which expedites overall workflow of D-Tipitaka.

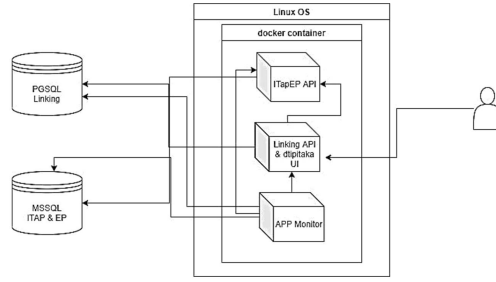


Figure 3: D-Tipitaka Database Flow

5.3 Main Functions of D-Tipitaka

D-Tipitaka consists of 3 main functions: display, manuscript and printed edition link, and search.

5.3.1 Display

- Text View Function

The Text View displays [Figure 4] the content of a selected Sutta of the Dīghanikāya Sīlakkhandhavagga (DN1) divided in units. They are systematically identified by unit numbers and sub-heading names. There are 2 types of words in Text View: non-problematic words and problematic words. Both types are clickable to access a table showing the manuscript and printed edition abbreviations. Each abbreviation links to a manuscript image or an image from the printed editions [See Appendix A]. However, only the underlined problematic words are linked to critical apparatus; by hovering the mouse pointer over any underlined words, a Tool Tip pops up. A click on it leads to the critical apparatus showing all variants and abbreviations, which link to original manuscript images or printed editions. Further details regarding manuscript images and printed edition links are discussed in topic no. 5.3.2.

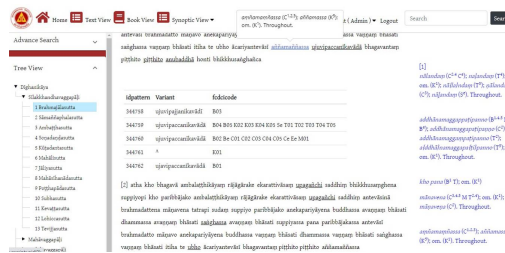


Figure 4: Text View of D-Tipitaka

- Synoptic View Function

D-Tipitaka also provides Synoptic View function [Figure 5] for easy comparison across manuscript traditions and printed editions. Users firstly select a Sutta and its paragraph number. The program will automatically align all variant readings of that selected paragraph in a table showing 10 columns at a time. The first row of the table is a header row containing the edited text of the Dhammachai Edition (De) followed by rows of variant readings from the different manuscript traditions and printed editions. Every word on the table is clickable and linked to images of the manuscript and printed editions. For easy comparison, variant readings deviating from the header can be highlighted in different colors if users click on “diff” menu.

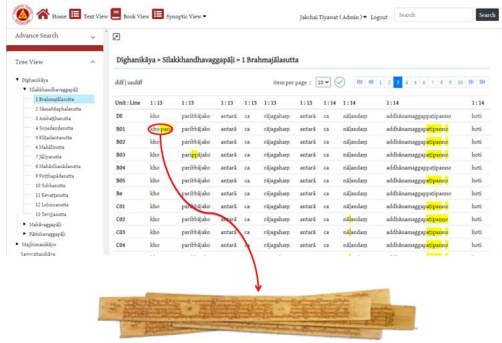


Figure 5: Synoptic View of D-Tipitaka

- **Book View Function**

The last display of D-Tipitaka is Book View function [Figure 6]. It offers users the convenient experience of reading the electronic version of the Dhammachai Edition page by page plus user friendly features, such as a page turning effect, bookmarks, etc. The main components of Book View are paragraphs and critical apparatus.

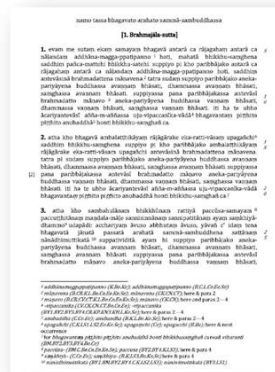


Figure 6: Book View of D-Tipitaka

Paragraphs contain non-problematic and underlined problematic words. Clicking on any word in the paragraph locates its page and paragraph number in the status bar on top of the page and leads to a table of variants which links to images of the manuscripts and printed editions. Similar to Text View, only the underlined problematic words are linked to the critical apparatus, which is revealed by hovering a mouse pointer over any underlined words, causing Tool Tip to pop up leading to the variants of the critical apparatus. In addition, the paragraph part includes unit numbers linking with Chaṭṭhasaṅgīti and PTS (Pali Text Society) printed editions. According to the Dhammachai Edition’s superficial study 94 % of the text consists of non-problematic words, while only 6 % of text is problematic. However, less than 6 % of problematic words have been critically analyzed in the Dhammachai Edition.

5.3.2 Manuscript and printed edition link

Another main function of the D-Tipitaka is the linkage to images of selected manuscripts and printed editions [Figure 7]. The linkage is divided into 2 types: single-step access and double-step access. The single-step access applies to the linkage to both manuscript images and printed editions by clicking on any required word on the Synoptic Table or in the critical apparatus of

the Book View. It also applies to direct linkage to the Chaṭṭhasaṅgīti and PTS Editions by clicking on a paragraph number of the Chaṭṭhasaṅgīti Edition and a page number of the PTS Edition in Book View. The double-step access means users need to perform 2 steps to access images. In Text View and Book View, clicking on a word will bring users to a variant table where they have to choose an abbreviation to access required manuscript images or printed editions. In short, single-step access shows images right after users click on a word while double-step access requires users to select words from text followed by abbreviation from the variant table.

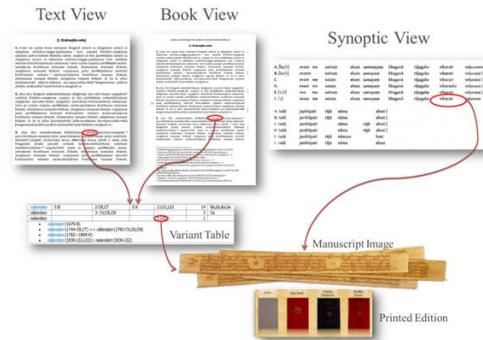


Figure 7: Manuscript images and printed editions link of Text View, Book View, and Synoptic View in D-Tipitaka

The retrieval of text data and image data is shown in the picture below [Figure 8].

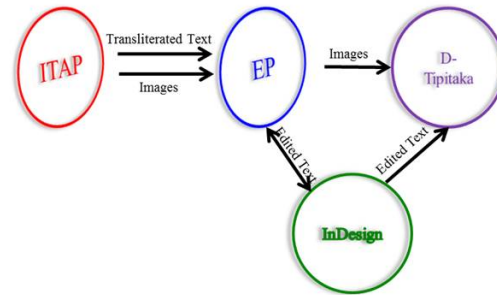


Figure 8: The retrieval of text data and image data

- **Image Data**

The data of palm-leaf manuscript images and printed editions stored in the Directory are retrieved and displayed on D-Tipitaka through a communication process among 3 programs. Firstly, ITAP reports picture name and address of required images to EP, which then relays all information to the D-Tipitaka program. In other words, EP acts as a medium of communication between ITAP and D-Tipitaka.

- **Text Data**

Text Data are classified into transliterated text and edited text.

- The transliterated text data created by readers, who are responsible for producing the Romanized Pāli text from manuscripts written in local scripts (as mentioned earlier). The transliterated data from 4 Pāli traditions is stored in ITAP and sent to EP program for researchers to edit.
- Edited text data refers to data of text edited in EP programs onwards.

* Transliterated text data is used to generate the tables used in text alignment suitable for editing. Researchers then examine the text and make critical decisions.

This edited text is then exported from EP Program to MS Word (.DOC) for a chief editor to finalize.

- * Edited text data in InDesign: The chief editor will deliberately go through edited work once again. After being emended, this final edited text along with design features, e.g. stylish typography and rich graphics, images and tables, etc., will be sent to update EP database and to D-Tipitaka program for publication.

The commands used for text and data retrieval are as follows:

- The command to get names of manuscript images and printed editions
`this.subscription.add(this.http.get < EditionTextItap[] >
(this.APIEndpoint + '/ws/api/ieps/editiontext/findbyfcrd/' +
this.fcbaseid, headers : this.httpHeaders)`
- The command to get directory of manuscript images and printed editions
`this.subscription.add(this.http.get < Picture > (this.APIEndpoint +
'/ws/api/ieps/editiontext/findfcpicture?picturename =' + fcedition + '/' +
filename, headers : this.httpHeaders)`
- The command to demonstrate manuscript images and printed editions
`this.tipitakaImage = this.getBypassSecurityTrustResourceUrl(encodeURI('data :
image/jpeg;base64,' + this.picture.binaryfile));`

5.3.3 Search

The D-Tipitaka provides users with 3 searching methods:

1. Search by using search box [Figure 9]
2. Search by clicking on variant readings in De (to compare with relevant witnesses including manuscript images and printed editions)
3. Search by using PTS page number and Chaṭṭhasaṅgīti paragraph number

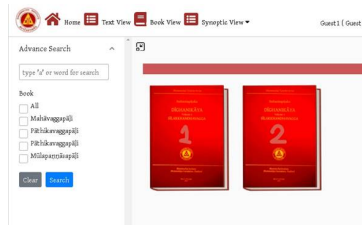


Figure 9: Search Function of D-Tipitaka

- **Search by using search box:**
The search box in D-Tipitaka provides users with 2 options: “partial matches searching” and “exact matches searching”.
 - **Partial matches searching** is the default of the system. It lets users enter part of a query word, and the program will automatically show related results. For example, if the desired word contains “Puri”, once a user types “P”, the related results started with “P” will be automatically shown, and when any syllable or vowel is added after “P”, the related results will be shown continuously. Until a user types “Puri”, the searching will show the desired word along with others related results such as Purisa, Purise, Purisasa, etc.
 - **Exact matches searching** is an alternative feature that helps users to find only a query word by selecting the EXACT check box. For example, if the target word “Pura” is entered into the search box and the exact check box is ticked, on top of the navigation panel will show the total number of occurrences in D-Tipitaka and in the panel displays all occurrences of the query word along with a couple of surrounding sentences. Each

occurrence is highlighted, and its Sutta heading, unit number, and page are identified. Users can jump directly to any location by clicking on the tab.

- **Search by clicking on variant readings in De:**

The Book View feature provides a book shelf for users to select any volume to read. It shows all pages exactly the same as those of the printed version divided into the main text and the critical apparatus. The variant readings of the main text can be accessed by double clicking on a word, and then the variant reading table will appear on a new page as shown in [Figure 10]. The first column is a variant reading part arranged as follows. The critical word of the main text appears on top of the list followed by different manuscript readings in a systematic order based on deviation from the critical reading. After each reading, the manuscripts which contain it are arranged in alphabetical order by abbreviations of different traditions (B: Burmese, C: Ceylon, K: Khom, M: Mon, T:Tham). This is a default order of the program.

Variant ↓	B	C	K	H	T	PTS ↓	Eds	Comm
☞ addhānamaggapattipanno	B02 B03	C01 C03 C04 C05	K02 K03 K04 K05		T01 T03 T04	13	Ce Ee Se	<input type="checkbox"/>
☐ addhānamaggapattipanno	B01 B04 B05			M01		4	Be	<input type="checkbox"/>
☐ addhānamaggapattipanno					T02	1		<input type="checkbox"/>
☐ ^			K01			1		<input type="checkbox"/>
☐ a (ddhā)namaggapa (ti)panno					T05	1		<input type="checkbox"/>
☐ addhānamaggapattipanno		C02				1		<input type="checkbox"/>
☐								<input type="checkbox"/>

Figure 10: Variant Table

The columns of manuscript evidence are followed by a column showing the total number of manuscript evidence from all traditions that can be sorted in ascending or descending order. The last column indicates printed editions containing the word, shown in alphabetical order (Be: Burmese Edition, Ce: Ceylon Edition, Ee: European Edition, Se: Syāmaratṭhassa Edition) in the “Eds” column. When users click on any manuscript abbreviation, manuscript image will be popped up.

- **Search by using PTS page number and Chaṭṭhasaṅgīti paragraph number**

In book view, the D-Tipitaka supports the comparison with two well-known printed editions, the PTS or European Edition (Ee) and the Chaṭṭhasaṅgīti or Burmese Edition (Be). The consultation of any De paragraph can be done by clicking on the PTS page number or Chaṭṭhasaṅgīti paragraph number on the outer margin. The PTS page number is on the furthest margin without brackets while the Chaṭṭhasaṅgīti paragraph number is in the square brackets. Clicking these brings up a new window showing pages of the printed edition and its transliterated text.

5.4 Conclusion / Limitation

D-Tipitaka is an on-going program development divided into 6 main parts including Login and Registration feature, Text View, Synoptic View, Book View, manuscript and printed editions link, and search. The development of every part except “Book View” has been accomplished over 80 %; Book View is only 30 % developed. Thus, the overall accomplishment marks up to approximately 60 %; completion is expected by mid 2022. However, as the project is dealing with newly designed programs and a massive database, unexpected problematic are bound to be encountered.

5.4.1 Login and Registration Feature

The Login and Registration feature is 100 % complete and allows user access in 3 levels: standard, registration, and restricted levels. Standard Level allows a user to use the search box and read text in 3 displays (Text View, Synoptic View, and Book view). Registration is not required. Registration Level allows users to search and link text with all evidence excluding palm-leaf manuscript images, and to utilize additional options for reading e.g. give feedback on a word,

insert bookmark, highlight text, etc. Restricted Level permits certain groups of users unlimited access to all evidence including palm-leaf manuscript images. Registration and permission are required. The process of login and registration of all user levels is outlined in the picture below [Figure 11].

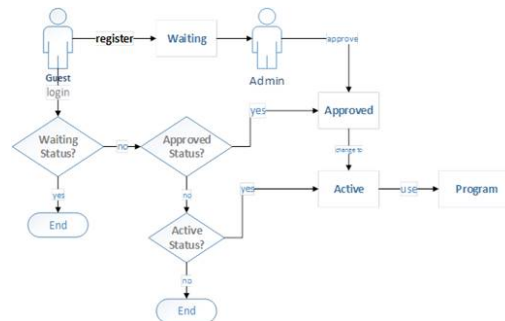


Figure 11: D-Tipitaka login and registration of all user levels process

5.4.2 Synoptic View

The development of Synoptic View is roughly 80 % completed. The cause of 20 % of incomplete work primarily lies on the discrepancy between computer and electronic devices display capacity and the needs of users and secondarily the constant adjustment to serve users' needs. For example, the Synoptic Table was initially designed to show 7 columns per page display as requested by users. After trial, users found them too fragmented to continuously read on computer or electronic device screen and requested programmers to increase columns from 7 to 15 which seemed impossible in terms of program design. The reconciliation between programmers and users is 10 columns per pages. The equipment's capacity limitation and constant adjustments consumed time on recoding and layout rearranging.

5.4.3 Text View

The Text View is almost finished, but has been delayed because of the difficulty in finding the best method to present the critical apparatus along with the canonical text. Unlike the critical apparatus of a printed book appearing at the bottom of the page, the critical apparatus of e-book version should appear in view of text, when scrolling up or down. To fulfill users' needs, programmers spent time on gathering examples from other electronic reading materials to get some ideas and to design several possible critical apparatus displays, each of which possesses a unique code writing and layout design. Users also spent time on testing and finding pros and cons of real-life use. Once the most suitable critical apparatus display was identified, programmers continued adjusting and designing until the critical apparatus display came to satisfied completion.

5.4.4 Book View

Among 3 displays, Book View gains least progress with roughly 30 % of completion due to internal and external factors. The internal factor is the unfamiliarity of users within the project in using a newly-designed program when editing. The whole process from editing until preparing publication artwork had been designed to take place in Editing Page (EP). But in practice, users felt more comfortable working with MS Word (.DOC) than in EP. Amendments in MS Word outside of the program have to be imported back to EP database to make EP and D-Tipitaka equal. This causes an external factor of difficulty in finding an accurate and consistent method to maintain the amended version in MS Word (.DOC) format when being processed in the EP program. To deal with this external factor, programmers tried multiple methods to find the stable format, such as .XML tag from Microsoft Office, .PDF, .PDF plus .IDML, and .PDF plus .EPUB. Eventually, after 3 years of attempts, they came to the conclusion of using the program

InDesign which can export .PDF and .EPUB. The use of these file types together help maintain accuracy and consistency of content, logical structure and formatting information when sent to EP to update the database and to D-Tipitaka for publication. Furthermore, another external factor relating to the database arose. The first database of MSSQL that had been used since the beginning of the project could not work at its maximum speed when operated with the D-Tipitaka program, due to massive text analysis and image links. To expedite speed, a new database of PGSQL was added. The implementation of a new database caused programmers to redesign databases and programs to make co-operating of MSSQL and PGSQL compatible. It can be concluded that the unfamiliarity that users will face with a newly developed programs will take several attempts to find qualified/specific file types, as well as engaging with a new database are the key factors that hinder the progress of the Book View development.

5.4.5 Manuscript and printed edition link

The work on manuscript and printed edition link is 80 % complete; the 20 % of unfinished work is the user interface. So far, we have succeeded in retrieving and presenting images relevant to text, but the interaction between images and users is limited. To make it user-friendly, programmers will build an interface enabling users to have more interaction with the palm-leaf manuscripts and printed edition images, such features as zooming in and out, and pan or scroll from one part of the image to another. It might be possible to incorporate a whole page of text alongside any particular image, but it might be too intricate. Thus, in the meantime we are focusing on improving user interface and leave text incorporation for the next version.

5.4.6 Search by using search box

As explained earlier, the D-Tipitaka provides partial matches searching and exact matches searching. At the moment, the completion of work in this part is more than halfway. Now, users can type part of word or exact word in the search box which seems to work perfectly for searching. However, we plan to provide a short cut of searching by adding “partial search” and “exact search” buttons. This part has not been commenced yet, not because of the complexity of work but time constraint and work prioritization, that requires programmers to focus on crucial work in Book View display.

In conclusion, since 2010 Digital Humanities Center (DHC) of DTP has developed several computer programs, ODEM, ITAP, EP, and D-Tipitaka to support work in every step from palm-leaf manuscript preservation until the publication of the Dhammachai Edition. The database of the previous program has been designed to be used as the database of the consecutive program. At present, the first 3 programs work effectively with minor adjustment according to the users’ additional requests or program revisions as effected by web-browser updates. Now, we are in the process of developing our latest program of D-Tipitaka, which has been found to be challenging. Emphasizing on creating a user-friendly interface, dealing with immeasurable palm-leaf manuscript and printed edition image retrieval, and finding a consistent and effective method for accurate electronic output are time consuming. Although we are confronted with several unexpected internal and external factors, which obstruct the achievement of the D-Tipitaka webpage in a timely fashion, we are determined to complete the D-Tipitaka webpage promptly to make it as an alternative and convenient platform for the transmission of the Buddha’s teachings in the digital age.

A Abbreviation

ABBR	DCI Code	Location	AD	No.of folios
B01	MM_04_01_006_00	Universities'Central Library, Yangon	1679	435
B02	MM_03_01_001_01	Bakara Temple, Amarapura	1768	237
B03	MM_04_01_001_00	Universities'Central Library, Yangon	1774	240
B04	MM_04_01_005_00	Universities'Central Library, Yangon	1792	301
B05	MM_04_01_008_01	Universities'Central Library, Yangon	1806	228
C01	SR_99_01_003_01	Ridīvihāraya, Kurunegela, Sri Lanka	1744	140
C02	SR_05_01_001_01	Sri Vijayasundarārama Vihāraya, Kadihāre, Kurunegela, Sri Lanka	1783	214
C03	SR_11_01_001_01	Koṭikāgoda Rajamahavihāraya, Walgama, Mātara	1832	193
C04	SR_02_01_001_12	Sri Narendrasingha Rajamaha Viharaya, Kunḍasāle	1855	172
C05	SR_01_01_007_01	Sri Daladā Maligāva Oriental Library, Kandy	Unknown	235
K01	TH_05_01_003_00	National Library, Bangkok	1777	535
K02	TH_05_01_002_00	National Library, Bangkok	(1782-1809)	551
K03	TH_04_01_001_00	Wat Pho, Bangkok	(1824-1851)	591
K04	TH_05_01_006_00	National Library, Bangkok	(1851-1868)	561
K05	TH_03_01_001_00	Wat Hongrattananan, Bangkok	Unknown	601
T01	LA_01_01_007_00	Digital Library of Lao Manuscripts	1598	526
T02	TH_10_01_063_00	Wat Mahadhatu, Yasothon	1822	652
T03	TH_01_01_002_00	Wat Sung Men, Phrae	1836	573
T04	TH_01_01_003_00	Wat Sung Men, Phrae	1836	717
T05	TH_10_01_007_00	Wat Mahadhatu, Yasothon	Unknown	501
M01	TH_05_01_026_00	National Library, Bangkok	Unknown	571

References

- Papangkorn Inkeaw, Jakramate Bootkrajang, Phasit Charoenkwan, Sanparith Marukatat, Shinn-Ying Ho, Jeerayut Chaijaruwanich: Rule-Based Page Segmentation for Palm Leaf Manuscript on Color Image *International Conference on Asian Digital Libraries (ICADL)*, pp.127-136, 2016
- David Doermann, Karl Tombre: Handbook of Document Image Processing and Recognition *Springer*, Heidelberg, 2014
- Suchada Srisetthaworakul: Developing Database of the Pāli Canon from the Selected Palm-leaf Manuscripts: Method of Reading and Transliterating the Dīghanikāya in Khom and Tham Scripts *Journal of Indian and Buddhist Studies*, Vol. 65, No. 3, 2017
- Ronald Haentjens Dekker, Dirk van Hulle, Gregor Middell, Vincent Neyt and Joris van Zundert: Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project *Digital Scholarship in the Humanities*, pp.1-19, 2014
- Elisa Nury: Visualizing Collation Results Elisa Nury *The Journal of the European Society for Textual Scholarship*, pp.75-94, 2019
- Ciprian-Octavian Truica, Alexandru Boicea and Florin Radulescu: Asynchronous Replication in Microsoft SQL Server, PostgreSQL, and MySQL *International Conference on Cyber Science and Engineering*, 2013
- <http://lannakadee.cmu.ac.th/LannaOCR/>
- <http://www.tipitaka.org/chattha>
- <https://collatex.net/>