

A Model of the Unified Tripitaka: Various Versions of the Saddharmapundarika-sutra Processed by XML

Jae Sung Kim
The Research Institute of Tripitaka Koreana, Korea

Project Planners : Ven. Dae Seok, Ven. Joong Hyun, Dr. Hur, In-sub (RITK)

Person-in Charge : Jaesung Kim

Data Processing (Proofreading, Correcting, Tagging): Mi-sook Kim (Dongguk University), Jong Min Kim (Korea KnowledgeWare Inc.) Minsu Kang, Jaesung Kim, Eun-ju Sin, Yoon-sim Lee, Young-Keun Jeon, Sun-Hwa Joo, Hye-Sung Han (RITK)

XML Solution (Editor, Designer, Browser): Jong Min Kim (Korea KnowledgeWare Inc.)

1. The Digitalization of the Saddharmapundarika-sutra: A Synopsis

For the past ten years the digitization of the world's Buddhist texts has been vigorously advancing, and it has come to pass that we have many commonly shared databases. In several nations of the world, at the heart of this advancement is EBTI. This advancement includes the crafting of tools to assist in projects related to Pali texts, Chinese texts, Tibetan texts and Sanskrit texts, and the development of digital electronic dictionaries.

If indeed, it is said that the 20th century was the era when basic Buddhist texts were digitized, then we can consider the start of the 21st century a time when we attempt to satisfy the demands of not only Buddhist scholars, but also the public who want to approach and understand Buddhism.

For purposes of specialized Buddhist research, our primary task is

the inputting of data, accurately and completely. We will continue to gather qualitative data at all times, while amassing academic research data for input such as subsequent comparisons, analysis, and translations. We must open this data to the general public so that non-specialists are able to read Buddhist texts translated into modern languages.

In this digitized Saddharmapundarika-sutra, for Buddhist specialists we have prepared the Tripitaka Koreana (Unicode and punctuated text), the Taisho version, Sanskrit text, and images of Tripitaka Koreana print blocks. For non-specialists who desire to read the Saddharmapundarika-sutra in a modern language, we have produced translations into Korean, and an English language translation of the original Sanskrit text.

The digitized Saddharmapundarika-sutra, with the accompanying five kind of texts and the images of the Tripitaka Koreana print blocks linked through XML (eXtensible Markup Language), constitutes a primary feature of the program. Therefore, a very worthwhile capability of this program is to be able to work with images of the Tripitaka Koreana, and the encoded text of the Tripitaka Koreana. The program can fill the role of providing a literary strictness in conveying the data through original images, for those who are reading the Tripitaka Koreana or the Taisho. We hope, this digitized Saddharmapundarika-sutra will serve as a model for the development of an advanced, digitized unified Tripitaka. Because at present the digitized Lotus Sutra is being analyzed by the initial phases of DTD, we must do more work on the construction of a specialized DTD for analysis of Buddhist texts. Therefore, I think this task needs to be carried by Buddhist specialists and XML specialists must work on together.

2. Selected Text

The structure of all texts is seen as appendices Picture 1,2,3.

2.1 Korean Translation of Saddharmapundarika-sutra (Korean Koryo: kk):

The Korean translation of the Saddharmapundarika-sutra, was published in 1993 by scholars of the People's Cultural Research Center, Socialist Science Research Bureau in Pyongyang, North Korea. The Lotus Sutra was selected as a Korean (Hangul) version in one volume from the 17 volumes collection entitled "The 80,000 Tripitaka," published by the North Korean Socialist Science Research Bureau. Our research center took a printed text of the Lotus Sutra from the North Korean Socialist Science Research Bureau, compared and edited its Hangul spelling and usages of terminologies in the field of Buddhism, and produced a modified version of the North Korean translation of the Lotus Sutra text. Additionally, cooperative work between the Dongguk Translation Center, the Electronic Buddhist Text Institute will result in the production a Hangul version of the tripitaka, and the Tripitaka Koreana, which will serve as a model for a unified electronic text for tripitaka, that is, unified Chinese-Korean tripitaka.

2.2 Koreana Tripitaka Unicode Text (Koryo: k)

Regarding the characteristics of Koreana Tripitaka Unicode text, the Research Institute of Tripitaka Koreana has supervision over punctuation and compilation that results in finalized text. Currently, we are advancing the compilation of the text with precision. The punctuated Tripitaka Koreana is being worked on, and we expect this will be very helpful for those who are reading the Chinese Buddhist texts.

2.3 Tripitaka Koreana Imaged Text (Koryo image: ki) (see appendices Picture 4)

The Dongguk Translation Center published a photocopied version of the Tripitaka Koreana. I used this to contrast and compare the coded text file. This work has been extremely helpful to philological research. Tripitaka Koreana, Volume 9, pages 725a-800b, Dongguk University, Seoul, 1994.

2.4 Taisho Tripitaka (Taisho Tripitaka: t)

The Taisho Tripitaka version as used in the digitized version of the Saddharmapundarika-sutra has been proofread by CBETA, Taiwan. The comparison between the Taisho Tripitaka and the Tripitaka Koreana is also helpful for philological research.

2.5 Sanskrit Edition (Sanskrit Edition: s)

We selected the Kern-Nanjio version of Sanskrit Saddharmapundarika-sutra published in 1912. The digitized data was provided by Professor Jong-chol Lee. Proofreading remains to be done. Nonetheless, the inclusion of the original Sanskrit version into the Unified Tripitaka is a very important step.

2.6 English Translation of the Sanskrit Text (English Translation: se)

Kern rendered this translation using a manuscript which was part of a collection in the Cambridge University Library. The text was finally published as the 21st volume of “Sacred Books of the East.”
<http://www.sacred-texts.com/bud/lotus/index/htm>

3. Structure of the Electronic Texts

3.1 Structure of the Chinese version of the Lotus Sutra (same as Korean translation version, Koreana and Taisho Tripitakas) 7 volumes, 28 chapters, 114 Tripitaka Koreana wooden blocks consisting of a total of 227 columns. The structural analysis of the basic Saddharmapundarika-sutra was predicated on its volumes, chapters and wooden blocks. The focus of the comparison has been on the imaged text of the Tripitaka Koreana.

3.2 Structure of the Sanskrit Text and the English Translation Text: The Sanskrit text of the Saddharmapundarika-sutra is organized into 27 chapters. In the Chinese version of the Lotus Sutra, Chapter 12 (Devadatta-parivarta) is independent, but in the Sanskrit version, what

is called the Devadatta-parivarta is part of Chapter 11 (Stupa-samdarsana-parivarta).

3.3 Photocopied Wooden Block Images: Photocopied wooden block images consist of a total of 227 portions. These files are linked by separated external files, which allows for the reading of some texts.

3.4 Basic Classification by Document Type The tagging of selected texts from the Saddharmapundarika-sutra is essentially done as follows: There are minor differences in the basic structures between the Chinese version and Sanskrit text.

First, in the case of the Chinese version, the components are broken down by volume, of which there 7, tagged s <chapter>; chapter, of which there are 28, tagged as <section1>; and, the number of wooden block images, of which there are 227, tagged as <section2>.

In the case of the Sanskrit text, the highest classification is chapter, of which there are 27, tagged as <section1>. Wooden block images are the next classification, consisting of 227, tagged as <section2>.

There are paragraphs beneath <section1> and <section2>. Below that, organized in rows, are general text, quotes, sentences, and verses.

<Appendices 1>

Structure of the Chinese Text

<chapter> 7 volumes

<section1> 28 chapters

<section2> 227 wooden block images

<para> paragraph

<text> descriptive sentence

<quote> quoted sentence

<verse> verse

Structure of Sanskrit Text

<chapter> 27 chapters
 <section1> 227 wooden block images
 <para> paragraph
 <text> descriptive sentence
 <quote> quoted sentence
 <verse> verse

DTD (Document Type Definition)

```
<!ELEMENT CHAPTER (CHAPTER-TITLE,(PARA|SECTION1)*)>
<!ATTLIST CHAPTER id ID #IMPLIED>
<!ELEMENT CHAPTER-TITLE (%txt-ref)*>

<!ELEMENT SECTION1 (SECTION1(TITLE,(PARA|SECTION2)*,DIALOG*))>
<!ATTLIST SECTION1 id ID #IMPLIED>

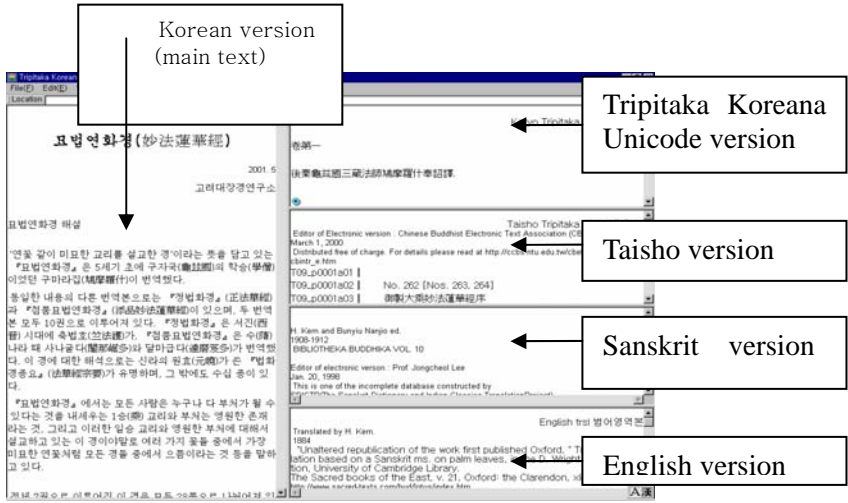
<!ELEMENT PARA (TEXT|VERSE|QUOTE|LIST|TABLE|FIGURE|MEDIA)*>
<!ELEMENT TEXT (%txt-ref)*>
<!ELEMENT QUOTE (%txt-ref)*>
<!ELEMENT VERSE (%txt-ref)*>
```

English translation by William J. Strnad III

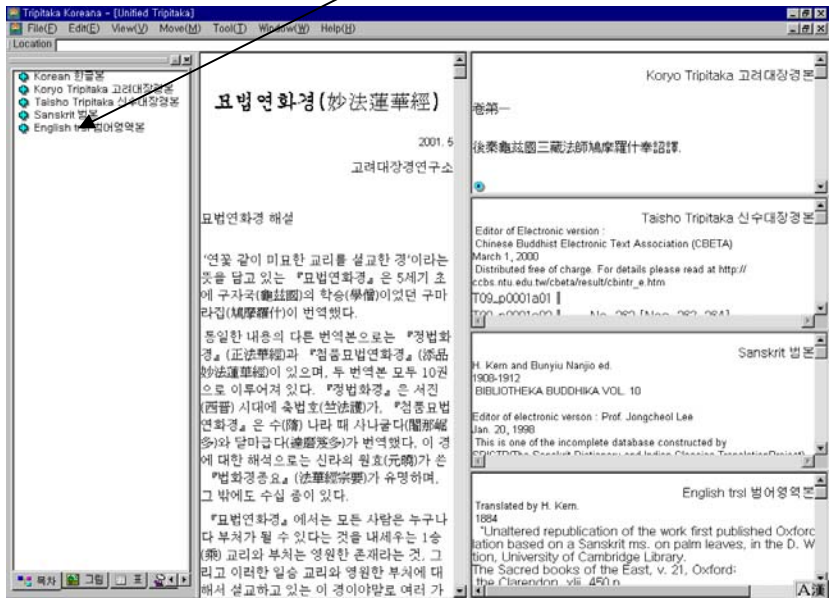
<Appendices 2>

(picture 1)

[TMp@pyrus](#) XML Browser – 5 frames form (We can select various form of frames)

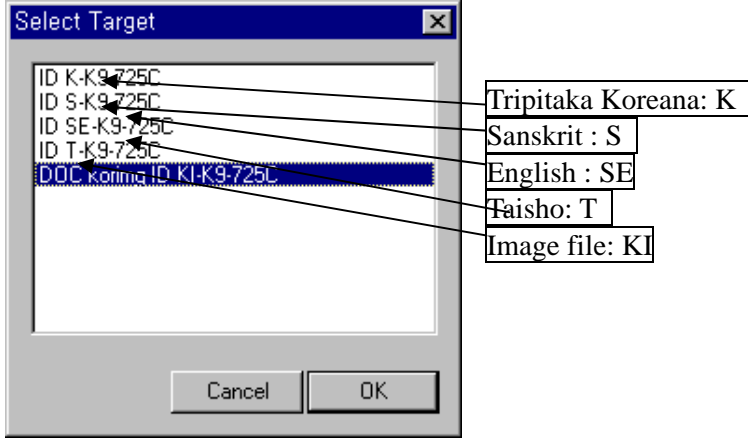


(picture 2) 5 frames form with Table of contents(5 text files)



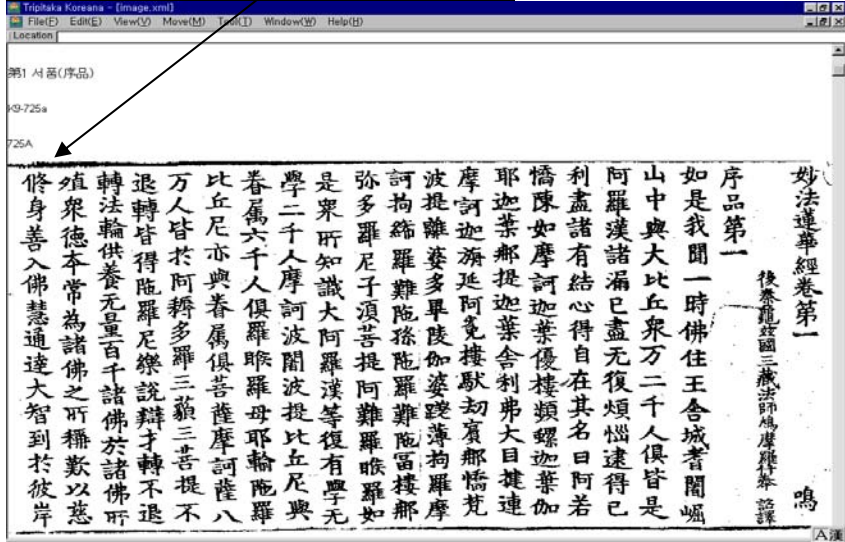
(picture 3)

<ref> link of 5 texts and 1 image file



(picture 4) view photocopy image file

K9-725a (Tripitaka Koreana Vol. 5 p.725a)



회 보

목 차

- 연구소 설립목적
- 연구소 조직임원
- 연구소 중점 사업
- 연구소 연혁

연구소 설립목적

동국대학교 전자불전연구소 (EBTI)는 1997년 12월 1일 연구소장 한태식(보광) 교수를 비롯한 불교학과, 선학과, 국어국문학과, 컴퓨터공학과 등 전자불전 관련 분야의 여러 교수들을 중심으로 한국불교전적 전산화, 불교 전자도서관 구축, 불교 사이버 박물관 구축 등 전자불전 관련 제분야의 연구 및 사업의 종합적이고 체계적인 추진을 목적으로 설립되었습니다.

연구소 조직임원

전자불전연구소 연구부 :

불전전산화 연구부
불전전자도서관 연구부
불전정보매체 연구부

전자불전연구소 연구위원회 :

연구소장 한태식(보광) 교수 (선학과)
간사 이용규 교수 (컴퓨터·멀티미디어공학과)
위원 홍영식 교수 (컴퓨터·멀티미디어공학과)
위원 이금식 교수 (컴퓨터·멀티미디어공학과)
위원 임기중 교수 (국어국문학과)
위원 박부권 교수 (교육학과)

불교전산화연구부 연구위원 :

목정배 교수 (불교학과)
권기중 교수 (불교학과)
김용표 교수 (불교학과)
정승석 교수 (인도철학과)
김호성 교수 (인도철학과)
이법산 교수 (선학과)
한용수 교수 (중어중문학과)
장태무 교수 (컴퓨터·멀티미디어공학과)
이강우 교수 (정보통신공학과)
이행구 교수 (불교학과)
이계영 교수 (컴퓨터학과)

연구소 중점 사업

한국불교전적 전산화

-한글대장경 전산화

-한국불교전서 전산화

세계 불교 전자도서관 구축

한국 불교 사이버 박물관 구축

멀티미디어 사이버 법당 건립

사이버 불교대학

연구소 연혁

- 1997년 3월 불교학, 선학, 컴퓨터공학과 교수들을 주축으로 전자불전연구소 설립 추진
- 1997년 12월 전자불전연구소 설립 및 소장에 한보광 스님 취임
- 1998년 7월 한국불교전적 전산화 시범사업 수행
- 1999년 1월 세계전자불전협의회 대만총회에서 한국지부 설치 및 지부장에 전자불전연구소장 한보광 스님 선출
- 1999년 2월 동국대학교 불교자료 종합전산화 추진위원회 구성 및 전자불전연구소를 전산화 사업 실행 기구로 선정
- 1999년 7월 한국불교전적 전산화 시범사업 시연회
- 1999년 7월 한국불교전서 전산화 사업 (1차) 수행
- 1999년 10월 제1회 전자불전 국내 학술대회 개최
'21세기의 불전 전산화 방향' (10.6 덕암세미나실)
- 1999년 12월 전자불전 창간호 발간
- 2000년 1월 국제전자불전학회 총회 및 학술회의 참가
(1. 11-17 미국 캘리포니아대학 주립 버클리 대학)
- 2000년 7월 한국불교전서 전산화 1차 사업 시연회
- 2000년 7월 한국불교전서 전산화 사업 (2차) 수행
- 2000년 10월 제 2 회 전자불전 국내 학술대회 개최
'불전 전산화의 제문제' (10.26 덕암세미나실)
- 2000년 12월 전자불전 2집 발간
- 2001년 1월 한글대장경 전산화 사업 (1차) 수행
- 2001년 5월 동국대학교 개교 95주년 기념 국제전자불전학회 학술대회 개최 '불전 전산화에서의 국제협력'
(5. 25-26 다향관 세미나실)

- 2001년 5월 국제전자불전학회 정기총회에서 전자불전연구소장
한보광 스님을 학회 회장으로 선출
- 2001년 7월 한국불교전서 전산화 2차 사업 시연회
- 2001년 7월 한국불교전서 전산화 사업 (3차) 수행
- 2001년 9월 동국 사이버 박물관 구축 사업 수행
- 2001년 12월 전자불전 3집 발간
- 2002년 1월 한글대장경 전산화 2차 사업 수행
- 2002년 1월 동국 사이버 박물관 구축 사업 시연회
- 2002년 2월 한글대장경 전산화 1차 사업 시연회

전자불전 편집위원

연 구 소 장 : 한보광 (동국대학교 선학과 교수)
간 사 : 이용규 (동국대학교 컴퓨터·멀티미디어공학과 교수)
편집위원장 : 한보광
편 집 위 원 : 홍영식 (동국대학교 컴퓨터·멀티미디어공학과 교수)
이금석 (동국대학교 컴퓨터·멀티미디어공학과 교수)
이용규
연 구 원 : 이성렬 (동국대학교 전자불전연구원)
박영희 (동국대학교 전자불전연구원)
조환기 (동국대학교 전자불전연구원)
조 교 : 김신우 (동국대학교 컴퓨터공학과 박사과정)

전자불전 (第3輯)

인 쇄 : 2001년 12월 25일

발 행 : 2001년 12월 30일

발 행 인 : 한보광

편 집 인 : 전자불전연구소

인 쇄 : 광진문화사

발 행 처 : 여래장(如來藏)

발행처주소 : 경기도 성남시 수정구
상적동 346번지

등록번호 : 제 1-20호

등 록 일 : 1999. 4. 26

서울특별시 중구 필동

3가 26번지 동국대학교

정보문화관 Q동 107호

전화 : (02)2260-3504

<http://ebti.dongguk.ac.kr>

ISSN 1229-6171

JEON JA BUL JEON

Journal of Electronic Buddhist Texts

Volume 3

December 2001

2001 EBTI International Conference
Commemorating the 95th Anniversary of Dongguk University

International Collaboration for the Digitization of Buddhist Texts

TABLE OF CONTENTS

PHOTOGRAPHS	4
WELCOME MESSAGE	9
COMMITTEE	14
GENERAL INFORMATION	15
PROCEEDINGS	16
Keynote Speech	19
Session 1: Textual Analysis	23
Session 2: Canonical Collections	69

Session 3: Online Reference Works
..... 131

Session 4: Characters and Encoding
..... 239

Published by
ELECTRONIC BUDDHIST TEXT INSTITUTE
office :Electronic Buddhist Text Institute
3-26 Phil-Dong Jung-gu, Seoul, Korea
Tel : (02) 2260-3504
<http://ebti.dongguk.ac.kr>