

CBETA-RP -- an Integrated Digital Research Platform of Chinese Buddhism Texts

Dharma Drum Institute of Liberal Arts Department of Buddhist Studies Jen-Jou (Joey) HUNG

International Symposium on Grids and Clouds (ISGC) 2016 ECAI Workshop 2016/03/17 @ Academia Sinica, Taipei, Taiwan

Rapid Growth of Digital Resources

In the last decade, we have executed more than 20 digital archive projects related to Buddhist studies.



CBETA 2014: near 190,000,000 Characters



Digital Archive of Chinese Buddhist Temple Gazetteer: 240 volumes of books, more than 14,000,000 Characters

Buddhist Studies Authority Databases: basic information of 23,000 persons, 58,000 places, Calendar information of 2000+ years





Rapid Growth of Digital Resources (II)



Glossaries for Buddhist Studies: 11 dictionaries, more than 100,000 entries

Catalogues of Chinese Buddhist canons: 21 Catalogues , 33,370 entries



and more....

- Ideal: scholars will save time on acquiring relevant materials, and become able to do larger scale research.
- Reality: the potential of these materials has been hindered by a lack of integration.



Building the Integrated Digital Research Platform of Chinese Buddhism Texts

- Instead of keeping creating new resources of important materials, we should also consider: how to integrate the resources that we have created so far.
- From 2013, we propose to build an online platform for bringing digital resources of Buddhist studies together.

CBETA Research Platform: an Integrated Digital Research Platform of Chinese Buddhism Texts

http://cbeta-rp.dila.edu.tw

new research platform
(analyze, integrate and share data)



CBETA Research Platform

- Begin at 2013, now is the third year.
- Integrating as more as possible digital Buddhist resources, to build a better research environment
- No only for "reading" texts, but would also have data editing, organizing, analysis functions.
- To transform the traditional "Read-only" digital reader to "read-write-analysis" research platform



Three Function Modules of CBETA-RP



Integrated Text Reader

Integrating scripture texts, catalogues, dictionary, phrase statistics and bibliography database in one place.



Deep Search

Not only the full-text search, but guiding user to compare the search results from different conditions or targets.



Digital Textual Analysis toolset

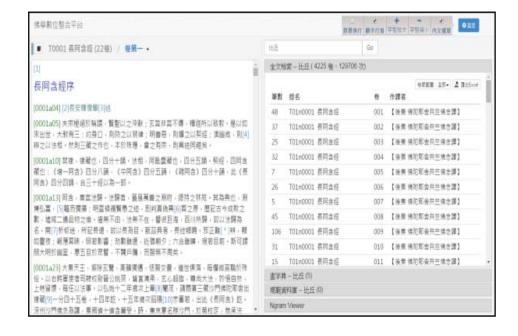
tools for performing textual analysis on scripture texts

Integrated Text Reader



Current Status of the Integrated Text Reader

- Rich functional web application.
- □ With rich data set from CBETA : 大正藏(100%)、卍 續藏(100%)、嘉興藏(c.a 80%-90%).
- Implement core functions of CBReader (the desktop reader of CBETA corpus).





Various Way of Selecting Texts— By Category

10

經文選擇

依據部類

依據冊別

經目查詢

Goto查詢

依據作譯者查詢

依據時間查詢

依據部類

輸入搜尋過濾字串

- 출 01 阿含部類 T01-02,25,33 etc.
- **출** 02 本緣部類 T03-04, X20-21 etc.
- **☎** 03 般若部類 T05-08,25,33,40,85, X24-26,46,74 etc.
- **■** 04 法華部類 T09a,26a,33-34,40,46,85, X27-35,46,55-57,74 etc.
- **■** 05 華嚴部類 T09b-10,26a,35-36,45,85, X03-05,07-09,45,58,74 etc.
- **查** 07 涅槃部類 T12b,26a,37-38,40b,85, X36-37,53 etc.
- **☎** 08 大集部類 T13,26a, X21,74 etc.
- **■** 09 經集部類 T14-17,19,21,26a,38-39,85, X01,09-10,17-21,37,39,7...
- **■** 10 密教部類 T18-21,39,46, X01-02,10-16,23,59,74 etc.

Various Way of Selecting Texts—

By Book Volumes



Various Way of Selecting Texts— By Searching



Various Way of Selecting Texts— By Location Reference



Select Text by Time

朝代「曹魏	溾(6) 220 CE	~ 265 CE」,共 6 筆。(返回上一頁)	×
		Export To E	xcel
部	經號	經名	作譯者
寶積部	T12n0328	佛說須賴經	曹魏 白延譯
寶積部	T12n0360	佛說無量壽經	曹魏 康僧鎧譯
律部	T22n1432	曇無德律部雜羯磨	曹魏 康僧鎧譯
律部	T22n1433	羯磨	曹魏 雲諦譯
印度撰述	X01n0011	佛說無量壽佛名號利益大事因緣經	曹魏 康僧鎧譯
印度撰述	X02n0205	佛說阿彌陀佛根本祕密神咒經	曹魏 菩提流支譯

Select Text by Author



Too Many Author Names:

- 1911 different author names
- 440 different first characters of author names

The Main Interface of the Platform

to Interface

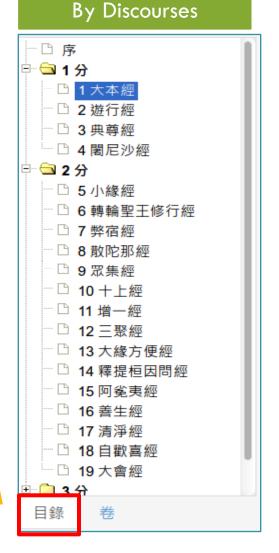


Two Different TOC views

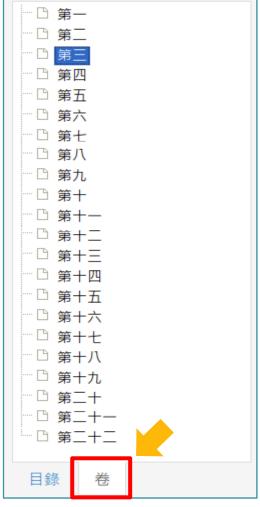
to Interface

Switch from different TOC

- By Discourses
- By Fascicles



By Fascicles



Various Text Arrangements

Arrange Text as in Taisho

原書換行 顯示行首 字型放大 字型縮小 內文選單

Display Page/Line Reference

```
T0001 長阿含經 (22卷) /
T01n0001 p0001a02 : [1]長阿含經序
T01n0001 p0001a03:
T01n0001 p0001a04: [2]長安釋僧肇[3]述
T01n0001 p0001a05: 夫宗極絕於稱謂, 賢聖以之沖默; 玄旨非言
T01n0001_p0001a06:不傳,釋迦所以致教。是以如來出世,大教
T01n0001_p0001a07: 有三: 約身口,則防之以禁律;明善惡,則導之
T01n0001 p0001a08:以契經;演幽微,則[4]辨之以法相。然則三藏
T01n0001 p0001a09: 之作也,本於殊應,會之有宗,則異途同趣矣。
T01n0001 p0001a10: 禁律,律藏也,四分十誦。法相,阿毗雲藏也,四
T01n0001 p0001a11:分五誦。契經,四阿含藏也:《增一阿含》四分八
T01n0001 p0001a12: 誦,《中阿含》四分五誦,《雜阿含》四分十誦,此《長
T01n0001 p0001a13:阿含》四分四誦,合三十經以為一部。阿含,秦
T01n0001 p0001a14:言法歸。法歸者,蓋是萬善之淵府,總持之林
T01n0001 p0001a15: 苑。其為典也,淵博弘富,[5]韞而彌廣;明宣禍
T01n0001 p0001a16: 福賢愚之迹, 剖判真偽異[6]齊之原, 歷記古
T01n0001_p0001a17: 今成敗之數, 墟域二儀品物之倫。道無不由,
T01n0001 p0001a18: 法無不在,譬彼巨海,百川所歸,故以法歸為
T01n0001 p0001a19: 名。開[7]析修途,所記長遠,故以長為目。翫
T01n0001 p0001a20:茲典者,長迷頓曉。邪正難[*]辨,顯如晝夜;報
T01n0001 p0001a21: 應冥昧,照若影響;劫數雖遼,近猶朝夕;六合
T01n0001_p0001a22:雖曠,現若目前。斯可謂朗大明於幽室,惠五
T01n0001_p0001a23:目於眾瞽,不閱戶牖,而智無不周矣。大秦天
T01n0001_p0001a24: 王,滌除玄覽,高韻獨邁,恬智交養,道世俱
```

Various Text Arrangements



Access Catalogue Information

CBETA 線上閱讀 🗘 選項 T0001 長阿含經 🚺 / 整合功能搜尋 搜尋 □ 序 經典資訊 🗎 🗀 1分 information about the author/translator, 1 2分 經名: T0001 長阿含經 created time and place 🕀 🗀 3 分 譯者:佛陀耶舍、竺佛念 🗎 🛅 4分 翻譯時間: From:0412, To:0413, 後秦(弘始一四一一五 A.D. 412-413) 翻譯概況: Translation by Buddhayasas and (Chu) Fo-nien: 15th year of Hung Shih (弘始), [2]長安釋僧肇[3]述 Later Ch'in dynasty (後秦) (A.D. 413) in Ch'ang An (長安). 夫宗極絕於稱謂,賢聖以之沖默;玄旨非言不傳,釋迦所以致教。是以如來出世,大教有 三:約身口,則防之以禁律;明善惡,則導之以契經;演幽微,則[4]辨之以法相。然則三藏之作 也,本於殊應,會之有宗,則異途同趣矣。 禁律,律藏也,四分十誦。法相,阿毗曇藏也,四分五誦。契經,四阿含藏也:《增一阿 含》四分八誦,《中阿含》四分五誦,《雜阿含》四分十誦,此《長阿含》四分四誦,合三十經 以為一部。



One Click Search Function

🔍 資料查詢

1 引用複製

□ 問題回報

■ T0001 長阿含經 (22卷) / 卷第一 -

如是住。云何,諸賢!如來為善別法性,知如是事,為諸天來語,乃知此事?」

[0001b22] 爾時,世尊在閑靜處,天耳清淨,聞諸比丘作如是議,即從座起,詣花林堂,就座而坐。

[0001b24] 爾時,世尊知而故問,謂:「諸<mark>比丘</mark>!汝等集此,

何所語議?」時,諸比丘具以事答。

[0001b25] 爾時,世尊告諸比丘:「善哉!善信,出家修道,諸所應行,凡有二業:一曰賢聖默然。汝等所論,正應如是。如來神通,展去無數劫事,以能善解法性故知,亦以諸天來頌曰:

「比丘集法堂, 講說賢聖論;

如來處靜室, 天耳盡聞知。

佛日光普照, 分別法界義:

亦知過去事, 三佛般泥洹。

名號、姓、種族,受生分亦知;

1. Full Text Search

2. Dictionary Look up

3. SearchPerson/Place Name BuddhistStudies Authority Database

4. Statistic Information about the selected term

5. [DBLM] Bibliography
Database

Search different sources at once!!

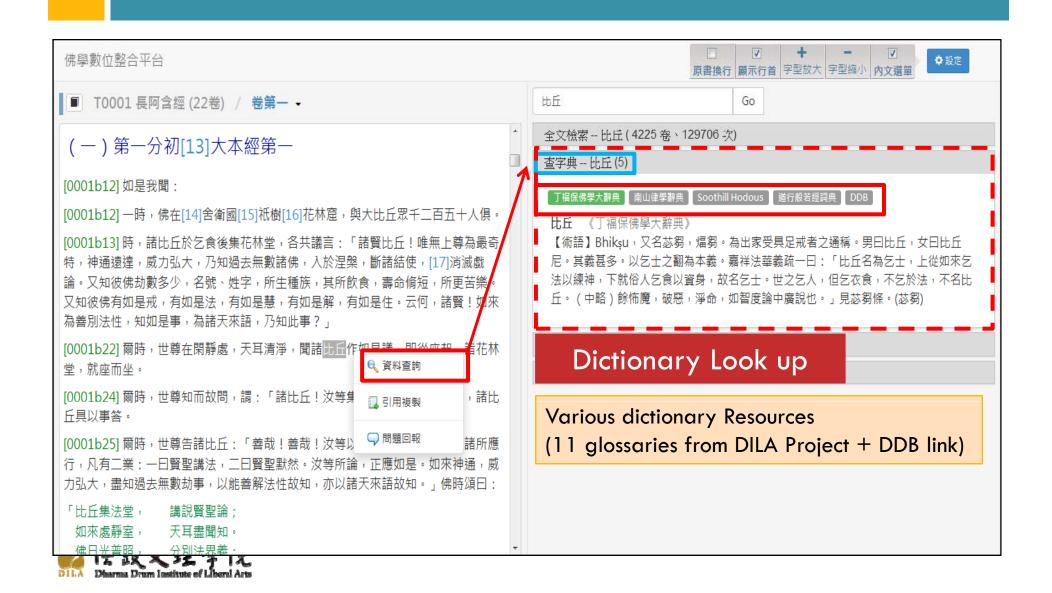


Select and Search – Full Text Search

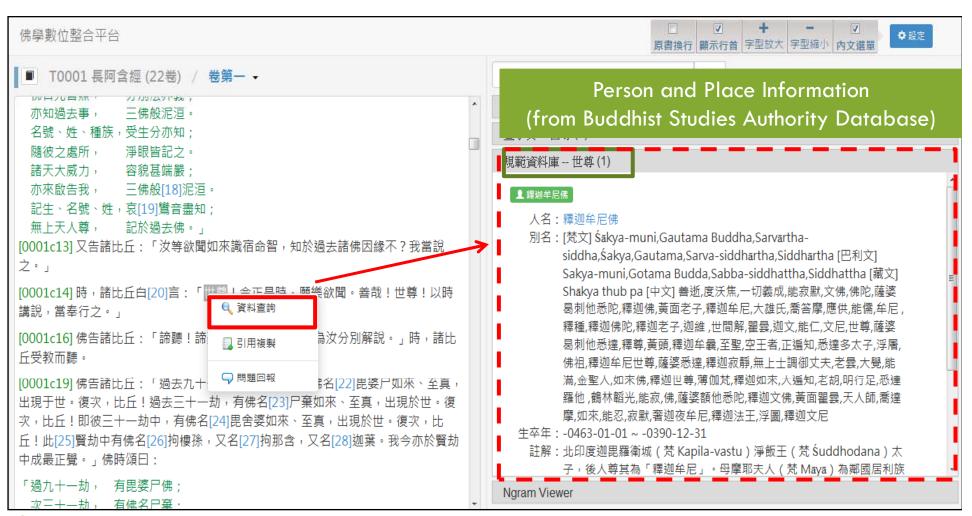




Select and Search – Dictionary Lookup



Select and Search – Person and Place Information





Select and Search -- Phrase Statistics



DBLIVI -- Digital Library & Museum of Buddhist Studies (from NTU)





CBETA DATA Service



CBETA Data Service

- The design of this CBETA research platform strictly flow the principle of separating data and interface.
- The whole data set is stored and managed by another program and a set of CBETA Data API is provided.
- Everyone can access CBETA's data from this API



CBETA Data API

- http://cbdata.dila.edu.tw
- including: catalogue information, full text of scriptures, toc of scriptures, full-text search services.
- Response to standard http request
- The result is in JSON format

Example of TOC information

Example of Full Text Search

Example of getting Scripture Text



Future Plan



Generate Multi-editions Data

- The current CBETA database only consists of one edition data (CBETA edition)
- The Taisho edition contains critical apparatus from 20+ different editions.
- Able to use these differences to rebuild the full texts of different editions.

Display the Difference of Editions

Need a new design of reading interface for showing difference from the text of different editions.

◉大正○宋○元○明

No. 109 [No. 99(379), No. 110]

佛說轉法輪經

後漢後漢安息三藏安世高譯

聞如是:一時,佛在波羅繚國鹿野樹下坐,時

有千比丘諸天神,皆大會側塞空中,於是有

picture source: Yap Yeh · Ksanaforge

Project

Deep Search

 Comparing the usage of phrases in different texts is the basic form of quantitative textual analysis

Phrase	Matches in MĀ	Matches in EĀ
我聞如是一時佛遊	216 times	0
聞如是一時佛在	0	441 times

These information is actually from search results



Deep Search (II)

- The traditional search function is aiming for locating the places of "search terms"
 - usually only allow us to perform search for "one search condition" in "one target collection"
- Deep Search: trying to help users to explore the different meaning of search results, ex:
- comparing the result of
 - different search conditions in the same text collection
 - same search condition in different text collections



Thank You for Listening.

Q & A!!

http://cbeta-rp.dila.edu.tw

