Using XML for Dunhuang Manuscript Database: The Dilun Manuscript Project

Kosei ISHII, Komazawa Junior College, Japan

Abstract

This presentation reports our attempt to use XML (Extensible Markup Language) in developing a database for manuscripts of the Dilun or "Stage Treatise" school. Since only a few of the works of the Dilun masters have survived, their doctrinal systems and their influence on other schools have not yet been clarified. Mr. Takashi Aoki and I have been discovering many handwritten manuscripts of several lineages within the Dilun school in the Dunhuang materials. We started to tackle a project in which those manuscripts are transcribed into electronic texts with various pieces of information marked up with XML. Through the flexibility of XML, we believe it is suiting to our needs in this particular project by offering the sought after support necessary for our database.

Introduction

In this presentation we report our attempt to use XML (Extensible Markup Language) in developing a database for manuscripts written by the Dilun masters (   ) or scholarly priests of the "Stage Treatise" school (   ). The study on the doctrinal systems of the Dilun masters have not progressed much. This is due primarily to the fact that except Hui-yuan (   ) (AD523-592) of Jingying temple (   ), there remain only a few fragments of manuscripts written by the Dilun masters. In addition to this, unlike the Chan/Zen scripts which includes vivid dialogue or sharp aphorisms, most of the Dilun manuscripts are not considered interesting because they only explain the terminology. These are some of the reasons why the study of the Dilun manuscripts have not advanced.

However, the doctrinal systems of the Dilun masters influenced the Chan school and others such as the Shelun school (   ), the Tiantai school (   ), and the Huayan school (   ). In addition, it is well-known that the Awaking of Faith in the Mahayana (the Dasheng qixin lun   ), which greatly influenced Buddhism in China, Korea and Japan, has similar elements of the Dilun masters' doctrines. Although it has not been confirmed whether the Awaking of faith was translated from Sanskrit into Chinese or an apocrypha written in China. Even so, many of the scholars who do see it as an apocrypha assert that this treatise was written by a Dilun master. This only stresses the importance in advancing the study of the Dilun masters by all means.

Fortunately, there exist many hand-written manuscripts of Dilun masters in the materials discovered at Dunhuang. From these findings, Mr. Takashi Aoki (　　　) and I have been trying to identify them. We have also started tackling a project to be titled "Dilun Manuscript Project" in which a database composed of transcribed texts and related information are released on the Web. For this, we decided to adopt XML to add various pieces of information to our transcribed texts and then to link them to texts of the Dilun school and other schools and to a database of related papers we are currently working on. Although XML is still in the middle of being implemented, we expect it will offer the support necessary to our project with its flexibility through functions such as "extended link" and "namespace".

Dilun manuscripts found at Dunhuang

It was Dr. Keiki Yabuki (　　　　) who discovered for the first time that manuscripts of the Dilun school are included in the Dunhuang materials. He provided to the Taisho Tripitaka pictures taken from fragments of some Dilun Manuscripts such as a commentary of Hui-guang (　　) and that of Fa-shang (　　) , his disciple. The volume 85 of the Taisho Tripitaka contains transcriptions of them. Second came the Dunhuang manuscripts research group led by late Professor Akira Fujieda (　　　) at Kyoto University. They studied in detail many manuscripts copied in the period of the Northern Dynasties, and in their study of them, they discovered Dilun manuscripts embedded within them. Professors Enjun Koizumi (　　　) and Noritoshi Aramaki (　　　　) of the group conduct excellent researches on some Dilun manuscripts such as S.613V.

There are some more scholars who are interested in the doctrinal systems of the Dilun masters and studying them. In particular, Mr. Aoki and I established a method to identify Dilun manuscripts and to assess the age of its compilation. Based on our method, I discovered many of manuscripts written by the Dilun masters of various lineages in the Dunhuang manuscripts, the Kanazawa Bunko (　　　　) Collection, and the texts collected in the Taisho Tripitaka. Mr. Aoki also discovered from the Dunhuang manuscripts many of Dilun school's manuscripts written by lineages other than Hui-yuan. Some of those Dilun masters insisted that the Mahasamnipata-sutra (the Daji jing　　　) is the most complete sutra, some asserted the Avatamsaka (the Huayan jing　　　) to be the most profound sutra, and some believed another sutra was the most perfect one. These viewpoints prove that there were several doctrinal systems completely different from that of Hui-yuan who emphasized that all of the Mahayana sutras were equally important. Parts of these assertions were cited in the Huayan manuscripts of Silla (　　), indicating that the priests of Silla were under the influence of the Dilun masters.

Since most of the discovered Dilun manuscripts were in fragments, identifying which manuscript to which school it belonged to has been trying - some showing only partial lines, with the exception of a few works. It was researchers of the Chan school who studied Buddhist manuscripts of the Dunhuang manuscripts most earnestly and comprehensively. However, they threw away these fragments of the Dilun manuscripts saying that "this is not a Chan manuscript" or that "this is not a Chan manuscript either". Even erudite compilers who published inventories of the Dunhuang Materials often could not identify them. For example, S.613V, typical a Dilun school reference of technical terminology, is only noted as "Commentary on a Buddhist scripture" in the Newest Inventory of Dunhuang Manuscript ( ) edited by Dr. Yong-yu Huang ( ). Owing to their unintentional efforts of putting aside unidentifiable fragments; accumulating into a mound of trash - only to be found later - a treasure mountain for the researchers interested in the Dilun masters!

Off course, transcription and revision must be done to survey the doctrinal systems of the Dilun masters. Further more it is necessary to make our transcribed texts accessible to researchers in various fields in order to clarify the influence of the Dilun masters. From this I consulted with Mr. Aoki and from there we started tackling this project since last summer. Due to the size of our task; in actuality it has just started, with an extremely small working group, only a part of our database is expected to become accessible on the Web from May.

Our adoption of XML

To create our database, we decided to adopt XML. XML is designed to make it easy to use SGML on the World Wide Web. Although XML is still being implemented, even at this stage of development it enables us to use various functions that exceed those of HTML. As for example, the hypertext linking abilities of XML are much greater than those of HTML.

Although Professor Urs App's "Computerized Collation of a Dunhuang Text" and Dr. Christian Wittern's "Minimal Markup and More - Some Requirements of Public Texts" are based on SGML/TEI, their visionary attempts at marking up the Buddhist manuscripts have provided very useful examples. Along with their efforts, we are marking up our texts using trial and error subtraction.

It would have been very useful if we were able to mark up the manuscripts as complete as in full SGML or TEI/TEI-Lite. Since the major objective aimed by Mr. Aoki and I is to clarify the kind of influence Dilun masters had on other schools, we are planning a markup of indispensable minimum for that purpose. The markup is to enable access from our transcribed texts to other electronic texts of relevant works of the Dilun school and other schools and to

related papers.

Related texts

With regard to the related texts of the Dilun school and other schools, we will mainly use the electronic texts of the Taisho Tripitaka on the Web released by SAT (Samganikikrtam Taisho-tripitakam: The Machine-readable Taisho Tripitaka) and CBETA (Chinese Buddhist Electronic Texts Association). As one of the directors of SAT, I am very happy to be able to tell you that these two projects are cooperating together. As for SAT, Mr. Shigeki Moro (　　), a colleague of SAT, is going to report it after my presentation. With regard to CBETA, Ven. Hui-Min (　　　) is expected to release their announcement on 20th.

We will also use, as resource, the Zenbase published by IRIZ, the Tendai Electronic Buddhist Scriptures by the Tendai Research Foundation (TRF), and other text databases released freely on the Web. As for the images of the Dilun manuscripts, the International Dunhuang Project Database offers a great deal to our database. We appreciate their efforts very much and would like to cooperate with all of the people who worked on those projects.

Once our transcribed text has been marked up with XML, we approve of free use by scholars all over the world, letting each of them work on their own necessary markup. We believe this to avoid useless repetition in work, and would like to proceed this task in cooperation with scholars in various fields who are interested in the Dilun manuscripts.

Code set and missing characters

As for code set, we are adopting either UTF-8 or UTF-16, which have been recommended by XML Working Group of the W3C for handling multi-language tasks. Regarding to the handling method of missing characters, we have followed the way which is used in SAT and INBUDS (Indian and Buddhist Studies Treatise Database) published by the Japan Association of Indian and Buddhist Studies. It is based on the KanjiBase method developed by Dr. Wittern. From there, SAT and INBUDS are developing a XML way of handling missing characters and we are to adopt it. A more detailed description will be explained in "On the Missing-Characters (GAIJI) of the Taisho Tripitaka Text Database Published by SAT" presented by Mr. Moro, in charge of technical area of SAT and INBUDS.

Variant characters

Many variant characters, not concluded in big dictionaries, have been found in the Dunhuang

manuscripts. This, until recently, has been a distress to scholars. However, the Mojikyo Institute opened the Mojikyo Font Center since last December. The server of the Mojikyo Font Center makes more than 80,000 Chinese characters of TTF and GIF available on the Web without charge. Those not included in the Mojikyo Fonts can be included. Assigned with Mojikyo's numbers, fonts can be created and become ready for distribution if researchers report missing characters along with the source. Mr. Tokio Furuya (　　　) and Mr. Tsuneo Yatagai (　　　) of the Institute will be presenting "The activities of the Mojikyo Font Center" in regard to their work.

With the Mojikyo numbers and fonts, we are able to discriminate variants in the Dunhuang manuscripts, and use them for searching, on the monitor and in print-outs. Once variants are marked up with their standard forms and Mojikyo numbers by our XML way, it becomes possible to carry out a search not only by variants but also by standard characters.

In addition, if bibliographical information of the manuscripts or its structure is marked up by XML, a more complicated search will become possible. For example, we will become aware of the fact that: this variant only appears in the colophons of the sutras copied for the the merit of the deceased before the Sui era and it only appears at the part of the description of paradise after death etc.. Such an attempt is not only useful when assessing a lineage of the manuscript but is also useful when assessing if a manuscript is forged. Of course, having such information, this might very well lead to more contrived and academically plausible forgery in the future.

Conclusion

As mentioned before, we have just started our project, and it is far from complete.　However, it holds new possibilities to databases on the Web. Our database is not just an input project but an attempt at using and linking the data of many projects on the Web such as SAT, CBETA, IDP, IRIZ, TRF, MFC etc. Or rather, we should say that these new possibilities were brought about by efforts of people through their working on such projects. We pray from our hearts that our project will fruit in the advancement of study of Dilun manuscripts, and at the same time, become a supportive means for creating a closer association between the people involved.