

The Technical Solution in the Computerization of Tripitaka Koreana: 4 Byte Code System, Word Processing and Corrective Programs

Chongrim Sunnim, Korea

One of the main problems that we have had during computerization of the Tripitaka Korean is to deal with the transformed Chinese characters. Handling of the transformed Chinese characters contains technological problems as well as academic ones at the same time.

Considering one of technological problems, approximately 60,000 transformed characters, that have founded so far, can not be expressed in the existing 2 Byte code system. For that reason, we could not but give up the technological guidances on the existing computerization and have tried to seek for new computerization system based on 4 Byte Code system (Tripitaka Koreana Code System). Design of 4 Byte code system became the first and the most significant part in the new computerization system. Success of computerization by our institute depends on the efficiency of the new code system in design.

The main purpose of design for the Tripitaka Koreana code system is to make Chinese characters shown in the Tripitaka Koreana be processed in computer including storing, inspecting and outputting by endowing codes for the whole Chinese characters. The most important factor in the process is an efficiency. The principles in designing new Tripitaka Koreana Code system by our institute are as follows:

a) The efficient and logical structure in disposition and input method

The new code system should realize easy and efficient disposition by the sound, core character and stroke count through the classified disposition on the basis of specific standard. Logical structure such as sound, core character, stroke count, and the number of frequency, etc. can be used as a standard of disposition.

b) Maintenance of consistency

Consistency on the standard which has already defined should be continued between the character codes contained in the code system. This is one of the considerable factors to support efficient disposition, inspection, output and letter editing.

c) Uniqueness of character code

A character should not be expressed as more than two different codes. Lack of uniqueness makes processing of disposition and inspection very difficult, as well as breaches a basic spirit of code designing.

d) Extensibility

One of the most difficult problems to be solved in Tripitaka code system is an extensibility. Range of Chinese characters shown in the Tripitaka Koreana is very vast and it frequently needs to add new characters which have not known yet. That's because we can not but consider the extensibility of the character code system. Accordingly, the Tripitaka Koreana code system should have a function for new character to be easily and rapidly registered. Not only alteration or deletion for the already registered character codes, but containing of every character shown in the Tripitaka Koreana should be available.

e) Comparability

Computerization of Buddhism scriptures in progress by each country has a limitation of localization since it has been operated by their own particular code system. This fact has worked as an obstruction to share international data and to keep comparability. So, for the international share and comparability of data, the new code system should have a logical structure to convert it into other code systems perfectly and easily.

We have produced a new word processing program for input and corrective work in future since the existing word process can not be used anymore since we adopt new code system. The word processing program is based on Tripitaka Koreana Code, so that it will be functioned as a tester of the Tripitaka Koreana Code. The New code system and word processing program are very important result of our study.

We have treated the corrective work manually so far, but we are trying to develop new method owing to its limitation in the accuracy. It is a 'corrective program'. The program is designed to work with 'Koryo Word' in the same process for the correction. performance at the EBTI next year.