

Publishing Pre-modern Chinese Classification Schemes as Linked Data: an experiment report

Haiqing Lin

C.V. Starr East Asian Library
University of California, Berkeley
Phone: (510)643-0892
Email: hlin@library.berkeley.edu

This experiment is an implementation of a SKOS based linked data service for publishing Chinese pre-modern classification schemes online. It intends to demonstrate the usage of SKOS in representing non-Western historical classification systems. This demonstration will be closely related to the themes of this workshop (2).

A pre-modern Chinese classification scheme is usually a component of an individual bibliography work to index pre-modern Chinese books. The pre-modern Chinese books (published before 1912), are valuable time capsules which reflect the contemporary Chinese knowledge and wisdom at the time. However traditional Chinese classification systems are fundamentally different from modern classification systems, which are rooted in a Western perspective of knowledge.

The history of pre-modern Chinese classification schemes can be traced back to as early as the late Western Han Dynasty (206 BC – 24 AD). Two imperial bibliographers: Liu Xiang (77–6 BCE) and his son Liu Xin (50 BC – 23 CE) established the Seven Epitomes, which is widely regarded as the first systematic classification scheme in China. In the following centuries, hundreds of classification schemes were developed. These schemes all had their own purposes, scopes and special features. However, they usually shared a common hierarchical structure which consists of two or three levels: the top level being division, called Bu (部), the second level being class, called Lei (類). Some schemes also have a third level called Shu (屬).

In the effort to serve and support the emerging digitization projects for pre-modern Chinese books, we found that it is challenging to represent these classification schemes using modern data models and vocabularies. The theoretical intricacy of modeling pre-modern Chinese classification schemes is the result of the differences in the epistemological foundations between Chinese traditional classification schemes and modern data models for knowledge organization systems. The latter is commonly based on the analytic approach in organizing knowledge which stems from classical Greek philosophy.

The structures of traditional Chinese classification schemes usually lack correlations and are divergent and inconsistent from an analytic philosophical perspective. The elements of the pre-modern Chinese classification schemes are sometime organized subjectively by bibliographers rather than by their objective interrelationships. In order to represent Chinese traditional classification schemes in SKOS, it is vital to develop a concept map to demonstrate the logical relationships among the elements of classification schemes as well as a framework to bridge the gap between traditional classification schemes and SKOS data models.

This demonstration reports an experiment in representing Chinese pre-modern classification schemes using SKOS data model and publishing these schemes as linked data. The purpose of this experiment is to examine an architecture which transforms existing Chinese pre-modern classification schemes to a linked data service. The core of the architecture is a data model which is a universal representation of Chinese pre-modern classification schemes structures. The data model is derived from the SKOS data model and consists of a universal concept scheme and following postulates:

1. A pre-modern Chinese classification scheme is a set of entities
2. Each entity is a concept
3. Each entity must have properties or attributes
4. Each entity must have a label
5. Each entity must have a label with Chinese language attribute
6. Each entity must have at least one broad-narrow relationship with other entities
7. Each scheme must have at least one top-concept
8. A Bu is an entity but also an Ordered Collection of Lei
9. A Lei is an entity but also an Ordered Collection of Shu
10. A Shu is an entity.
11. Each scheme is a part of a bibliography work

The universal concept scheme which contains 25 entities serves as a common concept map to represent the logical relationships among the elements of Chinese traditional classification schemes. It is also used to achieve the interoperability among the Chinese pre-modern classification schemes through a pre-established SPARQL query templet.

The demonstration system is composed of three layers, including a semantic scheme layer, a data storage layer and a data access layer.

- The function of the semantic scheme layer is to manage data representation schemes, such as SKOS, DC and other vocabularies
- The data storage layer is implemented using ARC2 RDF Store, a simple RDF/SPARQL solution for the LAMP system. This layer performs CRUD (create, read, update, and delete) operations on the RDF data store
- The data access layer provides two interfaces which make pre-modern Chinese classification schemes available via the web. The HTML based interface allows human

users to access the schemes through browsers. The system also provides RESTful API which makes the classification schemes accessible programmatically.

Data model:

