Collecting the Name of a Historical Person from Related Historical Material

Taizo Yamada
t_yamada@hi.u-tokyo.ac.jp
The University of Tokyo, Japan

Satoshi Inoue
inoue@hi.u-tokyo.ac.jp
The University of Tokyo, Japan

This proposal describes a method of collecting and managing the name of a person from historical materials related to Japanese history from the 8th century until the 19th century.

There are several tasks related to personal name such as collecting, identifying, representing. For historians and researchers, the recording and management of a personal name is an integral (and unavoidable) task. In Japan, historical materials are held by both a public institute and a private house in which they are found. Because there are no catalogs listing all of the material, the research staff at our institute (the Historiographical Institute of the University of Tokyo) spend several weeks each year investigating and examining historical records kept in Japan and abroad. The staff has over 100 years' combined experience conducting these examinations, and their journeys have taken them all over Japan as well as to many different parts of the globe. Ho (2015) and Bol et al (2015) introduced a method of personal name management in the context of Chinese historical materials (like a "地方志 (difang-zhi)") that used the China Biographical Database (CBCB) as a biographical dictionary. In Japanese history, there is no such exhaustive encyclopedia or dictionary for a name of a historical person.

A personal name appears in a variety of historical materials such as an old diary, an old document, a classical book, a family pedigree, a document related to appointment, and so on. The materials contain diverse name representations, such as a real name or an original name, a nickname, an epithet, a role name which indicates a person, and a Kao (which is a stylized signature or a mark), among others.

We constructed a repository which can store a collection containing both the names and the historical materials in which they appear. There are several different representations of a name in each material type. In order to represent and preserve these variations, the repository can store a variety of name data regardless of differences in data schema. In order to ensure efficient data-searching, we prepared common metadata that consist of an identifier, a personal name, a correspondence of the name, a related resource, a database where the name has been stored if such a database has already been constructed and the date on which the name appeared in the resource.

The name data can be expressed as RDF data and searched with SPARQL in the repository which provides an RDF store and a SPARQL endpoint. Furthermore, the repository can integrate the name data regardless of differences among personal name metadata schema, and provide the result of the integration. In addition, we will show an application for representing the outcomes of search results.

Bibliography

Ho, H. I. (2015). MARKUS – A Fundamental Semi-automatic Markup Platform for Classical Chinese. Proceedings of the 2015 International Conference on Digital Humanities.

Bol, P., Liu, Ch.-L., and Wang, H. (2015). Mining and Discovering Biographical Information in Difangzhi with a Language-Model-based Approach. Proceedings of the 2015 International Conference on Digital Humanities.