

## **Development of Digital Gazetteer using Semantic Web Technologies - A Case Study for “Knowledge of Areas”-**

Shoichiro HARA

Center for Integrated Area Study, Kyoto University

"Knowledge of Areas" is a new interdisciplinary paradigm between area studies and information science, which is intended to develop big repositories to store various resources of area studies, realize quantitative methods for area studies (area informatics) and expand application fields of area studies. An information technological aspect of "Knowledge of Areas" is to develop methods flexibly to link heterogeneous repositories and intelligently to retrieve data.

Repositories of area studies include resources of various research domains such as environment, ecology, agriculture, medical science, history etc. As vocabularies and data structures adopted in these repositories are different from domains, linking and retrieving repositories to discover new facts and/or hypotheses are very difficult. We have developed "Resource Sharing Systems (RSSs)" to link heterogeneous repositories and "REST-like-APIs (APIs)" flexibly to retrieve repositories. RSSs use a kind of synonym dictionary to link data elements/fields which have the same meaning but described by different vocabularies in each database. As this dictionary tightly depends on the system and is not easy to edit vocabularies, RSSs are not flexible to add new repositories. APIs are just keywords-retrieval-functions, that is, retrievals such as "find cities/towns in the county to which Berkeley belongs (retrieve hierarchical data)" and "find all ancestors of Prof. Lew Lancaster (transitive inference)." RSSs and APIs are excellent systems but not flexible and intelligent for "Knowledge of Areas"

In "Knowledge of Areas," we are trying to introduce Semantic Web technologies to develop intelligent information infrastructures, which can describe meaning/structure of data, link data flexibly and realize intelligent data retrieval. These are essentials to discover new knowledge from repositories on the Web. As its first step, we have reconstructed the digital gazetteer of Japanese Historical Place Names. In this reconstruction, we designed data model (structure of place names), defined vocabularies to describe place names, defined schemas (RDF, RDFs and OWL), defined URI, and then developed RDF store and retrieval interfaces (using SPARQL).

In the session, we will explain RSSs, APIs and their strength and weakness, then current states of the reconstructed digital gazetteer, and finally, discuss the possibilities of Semantic Web technologies to intelligent information infrastructures for "Knowledge of Areas."