Classification and Representation of Scenes in Field Note by Spatiotemporal Characteristics Using Linked Data

Taizo YAMADA
Historiographical Institute, The University of Tokyo
E-mail: t_yamada@hi.u-tokyo.ac.jp

In the paper we introduce to a method to quantitatively represent and analyze a spatiotemporal characteristic of a scene in a field note which a research of area studies described.

First we constructed a scene of a field note which was written concerning an appearance of Sumatera Island in Indonesia from October 1984 to January 1985, and extracted a spatiotemporal characteristic from the text of the field note. In the extraction we performed to extract words concerning a place name, date and a term which can characterize a scene and was extracted using Mecab as a morphological analyzer and IPAdic as a dictionary. Furthermore, we characterized a sentence in the field note with a latent topic which is hidden in the field note and can be detected by LDA (Latent Dirichlet Allocation) which is one of a topic model, in LDA each text can be represented as a mixture of various latent topics and each topic can be represented as a mixture of various terms.

Next we prototyped a search system of the field note. Because the system supports semantic web technology, data of a scene in the system can be represented by RDF (Resource Description Framework). The system enables to link a similar scene based on a characteristic of a scene automatically.

Using Linked Data, the system can link the extracted term and a place name to a resource which is located other systems such as a DBpedia, geonames.org, and other LOD sites.