

# **Text Analysis Method Using Latent Topics for Field Notes in Area Studies**

Taizo Yamada<sup>1</sup>

1 Historiographical Institute, The University of Tokyo

Recently Area Studies has made remarkable progress because researchers in Area Studies can search and analyze large volumes of data and search easily and quickly by using information technology such as web technology, data analysis, data engineering, and so on. In order to promote the analysis, the researchers have published databases related on Area Studies, such as catalogues, images, statistical data, spatial data and temporal data. For more the progress of the study, we believe that text analysis is one of the essential elements. For example, a text such as a field note has a description of sights, scenes and customs but latent topics or subjects which can be key elements characterizing the area.

In this paper we mention a text analysis method for field notes in Area Studies. We prepare a field note database in which the data unit is a description of a sight or a scene. In order to detect latent topics, we use latent Dirichlet allocation (LDA). LDA is one of a topic model, in LDA each text can be viewed as a mixture of various (latent) topics and each topic can be viewed as a mixture of various words. We show a prototype of a search system which can categorize texts in the database using LDA and present texts in same topic. We expect that the user can obtain a connection or a relation between the texts which cannot directly observed and is not expressed such as catalogues and statistical data yet.