

## Abstract of Main Thesis

### Title of Thesis

# Information Access Techniques for Digitized Historical Materials Across Languages and Time Periods

ばとじゃるがる びるげさいはん  
BATJARGAL BILIGSAIKHAN

### Abstract on the Content of the Applicant's Thesis

Searching and accessing digitized historical materials in multiple languages, covering periods ranging from ancient to modern times, along with unifying, merging and displaying the retrieved results dynamically in real-time, is a challenging task due to the historical evolution of languages and the diversity of textual data representation and interfaces.

We provide information access techniques for digitized historical materials across languages and time periods.

The first contribution of this thesis is the creation of a traditional Mongolian script digital library, which solves two major problems in making ancient Mongolian documents available to the public in digital formats: 1) encoding and displaying the traditional Mongolian script, and 2) retrieving ancient documents.

We propose a novel method for retrieving ancient Mongolian documents using queries in modern Mongolian language, which takes into account differences in the writing systems and grammars of ancient and modern Mongolian. The experiments show 96.72% precision and 77.85% of recall respectively for words in base forms. Usability studies conducted with human subjects yield a high level of satisfaction with the proposed system.

The second contribution is the development of methods for simultaneously accessing multiple databases of digitized historical materials in multiple languages on the Internet.

A federated search system has been developed for simultaneously accessing diverse databases, and solving the problems with unifying the returned results by using an effective automatic mapping method of metadata elements that has been proposed.

This system had been applied to online bilingual databases of Japanese Ukiyo-e woodblock prints for retrieving multiple databases in parallel and providing real-time integration and combination of bilingual results. The experiments show the proposed approach unifies metadata elements of various types of databases with precision between 85.7% and 100%.