

Translingual Digital Libraries: Beyond Interoperability

Ronald L. Larsen

Maryland Applied Information Technology Initiative (MAITI)
University of Maryland, College Park, Maryland 20742-2411 USA

rlarsen@deans.umd.edu

Position Statement

Semantic interoperability remains a research goal for large-scale digital libraries. But as digital library technology becomes more important to international scholarship, issues of semantic interoperability across language barriers take on greater significance. The technologies of machine translation, search, and text processing must transparently interoperate in order for an individual to effectively specify a query for materials in another language, and subsequently to extract desired information from retrieved documents, to summarize documents occurring in unfamiliar languages, and to translate relevant materials into the user's native language.

A translingual digital library's (TDL's) goal is to make language materials accessible and usable across language barriers. This includes the ability to accept the expression of an information need in the user's native language, use this query to search among materials in the TDL, retrieve relevant materials, translate their content into the user's language, extract relevant information (e.g., names of people, places, organizations, and related entities), identify events of interest, and correlate the content of documents in different languages. The objective is to enable the user to quickly and accurately develop a comprehensive understanding of information resources relevant to a particular topic, regardless of the language in which they occur.

Monolingual information retrieval is a relatively mature technology. This progress is reflected in a highly successful commercial market. Information retrieval has been an active field of research for decades. But despite the successes, it remains the case today that on formal evaluations measured by precision and recall, the best retrieval systems perform only modestly well. Less progress has been made on more recent fields of exploration, such as topic detection, information extraction, and summarization.

Moving into the multilingual arena, issues of mapping a query into the target language and interpreting the results have dominated the research. Performance has typically been around 50% of the precision and recall results returned on monolingual collections. Many challenges remain to be addressed before we can claim to have developed a translingual digital library. Those that are particularly relevant to Korea and the US will be identified in this workshop.

Collaboration between Korea and the US on a TDL holds great promise for advancing the state of the art in translingual digital libraries, as well as providing a valuable online scholarly resource to the citizens of each country and the world. This workshop provides the first step in building the coalition that can achieve this vision.

Biographical Information

Ronald L. Larsen is director of the Maryland Applied Information Technology Initiative (MAITI), a consortium of eight universities in Maryland working on Information Technology workforce issues. He is also an affiliate associate professor of computer science at the University of Maryland. Prior to his current position, he was assistant director of the Information Technology Office (ITO) at the Defense Advanced Research Projects Agency (DARPA), where he developed and managed national research programs in Information Management and Translingual Information Detection, Extraction, and Summarization (TIDES). Dr. Larsen has eight years experience in the University of Maryland Libraries developing and managing networked library information management systems. His current research interests include the modeling and understanding of the role of temporal information in the operation and control of large-scale information systems.