Seminar on
Geographic Information Retrieval:
Algorithms and Approaches

Prof. Ray R. Larson
University of California, Berkeley
School of Information

13:30-15:00, Monday, August 3rd, 2009
Room No. 1208 & 1210, National Institute of Informatics, Tokyo, Japan

The goal of Geographic Information Retrieval (GIR) is to retrieve relevant information resources in response to queries with geographic constraints. GIR implies that the indexing and retrieval of objects in a digital collection takes into account some form of georeferencing, and may use various forms of geographical proximity, containment, or other spatial relations in estimating or predicting geographic relevance. Systems that provide searches using GIR methods, including geographic digital libraries, and location-aware web search engines, are based on a collection of georeferenced information resources and methods to spatially search these resources with geographic location as part of their search specifications.

Information resources in digital library collections can be considered georeferenced if they are spatially indexed by one or more regions or points on the surface of the Earth, where the specific locations of these regions are encoded using spatial coordinates directly (geometrically), or indirectly by toponyms (place names). In this lecture we will examine the effectiveness of Geographic Information Retrieval (GIR) methods in IR systems.

We will show how various types of information may benefit from explicit geographic search, and where text-based place name search may be sufficient. We will also show how implicit geographic search (or geographic browsing) can be used to dynamically generate geographic searches in geographic interfaces like Google Earth. We will describe the algorithms used for Geographic search and how these may be combined with topical text searches. In addition we will show results from the GeoCLEF IR evaluation for text-based geographic search.

Prof. Ray R. Larson
University of California, Berkeley
School of Information

BIOGRAPHICAL SKETCH

Professor Ray R. Larson is a faculty member of the University of California, Berkeley, in the School of Information. Prof. Larson specializes in the design and performance evaluation of information access systems, and the evaluation of user interaction with those systems. His current research focuses on Geographic Information Retrieval, Cross-Language Information Retrieval and Structured (XML) retrieval using probabilistic methods. His background includes work as a programmer/analyst with the University of California Division of Library Automation (DLA) where he was involved in the design, development, and performance evaluation of the UC public access online union catalog (MELVYL). His past and current research has concentrated on the design and evaluation of information retrieval systems, with an emphasis on digital libraries. Prof. Larson was the principal investigator for the "CHESHIRE Demonstration and Evaluation Project" sponsored by the US Dept. of Education, that developed a next-generation online catalog and full-text retrieval system. He was a co-principal investigator for the "Searching Unfamiliar Metadata Vocabularies" project sponsored by DARPA. Prof. Larson was also the principal investigator of the "Cross-Domain Resource Discovery: Integrated Discovery and Use of Textual, Numeric and Spatial Data" project sponsored by NSF as part of the International Digital Libraries program.

Prof. Larson was Co-PI along with Michael Buckland and Fredric C. Gey for the IMLS National Leadership grants "Seamless Searching of Numeric and Textual Resources", "Going Places in the Catalog: Improved Geographic Access" and most recently, "Support for the Learner: What, Where, When and Who". Prof. Larson is the PI on the recently awarded IMLS National Leadership grant "Bringing Lives to Light: Biography in Context." Prof. Larson also teaches courses on the design and evaluation of information systems, including IS202 "Information Organization and Retrieval", IS257 "Database Management", IS240 "Principles of Information Retrieval", and IS245 "Organization of Information in Collections".