

NTCIR from the Beginning: A Personal Research Journey

Fredric C Gey
University of California, Berkeley

Abstract

My participation in the first NTCIR workshop starting in 1997 led to a fifteen year research journey. It was a journey through Kanji, Katakana, Hiragana, segmentation, bigrams, phonetic recognition, decompounding and parallel corpora alignment for lexicon development. In languages the travels went through Japanese, Chinese, Korean, European languages, Russian, Arabic, and Hindi, as well as mathematics. It produced three SIGIR workshops (2002, 2006, 2009) on cross-language search and multilingual information access as well as a special issue of Information Processing and Management on cross language information retrieval. In later years the research direction changed toward geographic and geo-temporal information retrieval evaluation. This talk will cover highlights of my personal research journey and pay tribute to colleagues and students with whom I have been fortunate enough to collaborate.

Biography

After receiving a Master's degree in Mathematics from UC Berkeley in 1964, Fredric Gey worked for 3 1/2 years at Bell Laboratories (later ATT Labs Research). In 1967 he returned to Berkeley and worked for 21 years as a staff scientist at Lawrence Berkeley Laboratory in the Computer Science Research Department. In 1989 he returned to the Berkeley campus as Data Archivist and Librarian for Social Science and Health Statistics for UC Berkeley, while simultaneously pursuing a PhD in Information Science, which was conferred in 1993. His dissertation "Probabilistic Dependence and Logistic Inference in Information Retrieval" developed the first of several logistic regression ranking models which have stood the test of time. In 1996 he received a USA National Science Foundation Grant to develop logistic regression search models. In 1998 he turned the research direction of this grant toward cross-language information retrieval, leading to participation in the first NTCIR workshop in 1997-1999, and in 2000 to CLEF, the Cross Language Evaluation Forum for European languages. In addition to multilingual information access, he has also done research and development in nuclear forensics, geographic information retrieval, digital humanities and social science information systems. He was the General Chair of ACM-SIGIR 1999, the 22nd International Conference on Research and Development in Information Retrieval.