TextScope: Enhance Human Perception via Text Mining

Dr. ChengXiang Zhai, Professor, University of Illinois at Urbana-Champaign, USA

Abstract:

Recent years have seen a dramatic growth of natural language text data (e.g., web pages, news articles, scientific literature, emails, enterprise documents, blog articles, forum posts, product reviews, and tweets). Text data contain all kinds of knowledge about the world and human opinions and preferences, thus offering great opportunities for analyzing and mining vast amounts of text data ("big text data") to support user tasks and optimize decision making in all application domains. However, computers cannot yet accurately understand unrestricted natural language; as such, involving humans in the loop of interactive text mining is essential. In this talk, I will present the vision of TextScope, an interactive software tool to enable users to perform intelligent information retrieval and text analysis in a unified task-support framework. Just as a microscope allows us to see things in the micro world, and a telescope allows us to see things far away, the envisioned TextScope would allow us to see useful hidden knowledge buried in large amounts of text data that would otherwise be unknown to us. As examples of techniques that can be used to build a TextScope, I will present some general statistical text mining algorithms that we have recently developed for joint analysis of text and non-text data to discover interesting patterns and knowledge. I will conclude the talk with a discussion of the major challenges in developing a TextScope and some important directions for future research in text data mining.

Biography:

ChengXiang Zhai is a Professor of Computer Science and a Willett Faculty Scholar at the University of Illinois at Urbana-Champaign (UIUC), where he is also affiliated with School of Information Sciences, Carl R. Woese Institute for Genomic Biology, and Department of Statistics. He received a Ph.D. in Computer Science from Nanjing University in 1990, and a Ph.D. in Language and Information Technologies from Carnegie Mellon University in 2002. He worked at Clairvoyance Corp. as a Research Scientist and a Senior Research Scientist from 1997 to 2000. His research interests are in the general area of intelligent information systems, including specifically intelligent information retrieval, data mining, natural language processing, machine learning, and their applications. He has published over 200 papers in these areas and a textbook on text data management and analysis. He is the America Editor of Springerís Information Retrieval Book Series and an Associate Editor of BMC Medical Informatics and Decision Making, and previously served as an Associate Editor of ACM Transactions on Information Systems, Associate Editor of Elsevierís Information Processing and Management, Program Co-Chair of NAACL HLT 2007, ACM SIGIR 2009, and WWW 2015. He is an ACM Distinguished Scientist, and received a number of awards, such as ACM SIGIR Test of Time Paper Award (three times), the Presidential Early Career Award for Scientists and Engineers (PECASE), Alfred P. Sloan Research Fellowship, IBM Faculty Award, HP Innovation Research Award, UIUC Rose Award for Teaching Excellence, and UIUC Campus Award for Excellence in Graduate Student Mentoring.