# Research on Chinese discourse rhetorical structure representation scheme and corpus annotation

Guodong Zhou
Soochow University, China
gdzhou@suda.edu.cn

**Abstract:** It is well-known that interpretation of a text requires understanding of its rhetorical relation hierarchy since discourse units rarely exist in isolation. Such discourse structure is fundamental to document-level applications, such as text understanding, summarization, knowledge extraction and question-answering. In comparison with English, there are only a few studies on Chinese discourse analysis, due to the lack of appropriate theories to Chinese discourse structure representation and large-scale well-accepted corpora. In this talk, I will present a novel discourse structure representation scheme for Chinese, called Connective-driven Dependency Tree (CDT), and describe our adventure in corpus annotation of the Chinese Discourse Treebank (CDTB) of 500 documents, using a top-down strategy to keep consistent with Chinese native's cognitive habit.

**BIO:** Zhou Guodong received the Ph.D. degree in computer science from the National University of Singapore in 1999. He joined the Institute for Infocomm Research, Singapore, in 1999, and had been an associate scientist, scientist and associate lead scientist at the institute until August 2006. Currently, he is a distinguished professor at the School of Computer Science and Technology, Soochow University, Suzhou, China. His research interests include natural language processing, information extraction and machine learning. Currently, he is an associate editor of ACM Transaction on Asian Language Information Processing(2010.07-2016.06), an editorial member of Journal of Software (Chinese)(2012.01-2014.12) and a vice chair of Technical Committees on Chinese Information/China Computer Federation(2010.12-2016.12), Computational Linguistics/Chinese Information Processing Society of China and Natural Language Understanding/Artificial Intelligence Society of China. Besides, he had been a member of the Editorial Board of Computational Linguistics (2010.01-2012.12).



*Proceedings of the Third CIPS-SIGHAN Joint Conference on Chinese Language Processing*, page 1,
Wuhan, China, 20-21 October 2014