

**NUS Graduate School for Integrative Sciences and Engineering
Research Project Write-up**

Title of Project : Chinese Language Discourse Parsing

Name of Supervisor : KAN Min-Yen

Contact Details: kanmy@comp.nus.edu.sg

Short Description

The state-of-the-art in Natural Language Processing is largely confined to the understanding of single sentences. However, the topic of intersentential discourse processing is becoming more pressing and in recent years, more attention has been placed on this nascent area of research for the English language.

The Chinese language differs greatly in how it structures ideas and sentences and arguments from English, and Chinese NLP also is in need of discourse analysis programs, especially ones aimed at much of current Chinese language production by the masses: microblogs, SMS and email.

The student taking on this project must be a quick study of Chinese language, linguistics and statistical processing. We hope to collaborate with leaders in Chinese discourse processing (Brandeis U.) to realize the first Chinese discourse processing system and make subsequent refinements in the course of this PhD topic.

References:

- [1] Ziheng Lin, Hwee Tou Ng and Min-Yen Kan (2010). A PDTB-Styled End-to-End Discourse Parser. Technical Report TRB8/10, School of Computing, National University of Singapore, August. (pdf, demo/code)

- [2] Nianwen Xue. 2005. Annotating the Discourse Connectives in the Chinese Treebank. In Proceedings of the ACL Workshop on Frontiers in Corpus Annotation, Ann Arbor, Michigan.

- [3] Yuping Zhou and Nianwen Xue. 2012. PDTB-style Discourse Annotation of Chinese Text. ACL 2012.