

Title: One Million Sense-Tagged Instances for Word Sense Disambiguation and Induction

Abstract:

Supervised word sense disambiguation (WSD) systems have achieved the best performance when evaluated on standard benchmark datasets. However, the lack of large amounts of sense-tagged data poses a major hurdle to scaling up supervised WSD systems to disambiguate all words of English. In this talk, I will present a semi-automatic approach to extract and annotate a large sense-tagged corpus. This one-million-word sense-tagged corpus has been publicly released since 2015 and has been used by other researchers working on automated WSD. When trained on this one-million-word sense-tagged corpus, the open source IMS (It Makes Sense) WSD system created in my research group achieves good performance on standard WSD tasks and another word sense induction task.

Bio:

Professor Hwee Tou NG is Provost's Chair Professor of Computer Science at the National University of Singapore (NUS) and a Senior Faculty Member at the NUS Graduate School for Integrative Sciences and Engineering. He received a PhD in Computer Science from the University of Texas at Austin, USA. His research focuses on natural language processing and information retrieval. He is a Fellow of the Association for Computational Linguistics (ACL).

His papers received the Best Paper Award at EMNLP 2011 and SIGIR 1997. He is the book review editor of Computational Linguistics, an action editor of the Transactions of the Association for Computational Linguistics (TACL), an editorial board member of Natural Language Engineering, and a steering committee member of ACL SIGNLL.