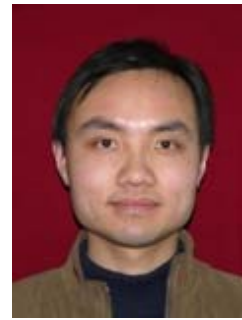


CV



Personal

Name: Li Bin
Gender: Male
Year of Birth: 1981
Address: School of Chinese Language and Literature,
Nanjing Normal University,
No.122 Ninghai Road, Nanjing, P.R. China
Postal code: 210097
Email: David.Been.Lee@gmail.com
Mobile: +86+13813878144

Education

PhD student in Computational Linguistics, Nanjing Normal University, China. 2006~2009.
MA in Computational Linguistics, Nanjing Normal University, China. 2003~2006.
BA in Chinese Language and Literature, Nanjing Normal University, China. 1999~2003.

Professional Experience

- **Max-length overlapping string detection in very large corpus(Jan ~ June, 2005).**
In the research, a program is designed and implemented to detect the max length overlapping strings in 4GB news corpus, using C++; then a team with 15 graduate students was organized to tag and analyze 10,000,000 tokens randomly selected from the corpus. As a result, a database for disambiguation of overlapping strings is created and used in memory-based word segmentation system. A paper is written to summarize the research result.
- **Human Computer Interaction System for unfamiliar text segmentation with no lexicon(2005~2006).** The system can extract unknown words using an improved Mutual Information based algorithm, and with little correction work done by the annotator. A text with segmentation F-score of over 0.8 can be easily obtained. This is mainly my MA dissertation work. The system is now applied to segment Ancient Chinese texts (3BC ~ 10AD) which have few lexicon resources to use.
- **Sentiment computation on English and Chinese topics(June, 2006 ~ June, 2007).** This is a student project funded by Nanjing Normal University. Its aim is to compute people's opinion on different topics from the web, such as ipod, DELL, Sadam, etc. Clustering search engines (Vivisimo, BBMao) are queried to get the sub-topics and normal engines (Google, Baidu) are queried to compute the sentiments of the topic and sub-topics. The system is designed and written in C++ and 2 papers were written on it.

- **Simile Recognition in Chinese(March ~ August, 2008)**. The paper is funded by the National Science Fund “Key Technologies of Chinese Metaphor Computation”. The research employs MaxEnt and CRFs models to tag the tenor, vehicle and similarity elements in simile sentences with the marker “像 (like)” in Chinese news texts with the mean F-score over 0.8.
- **Chinese Collocation and Selectional Preference(2007~)**. This is my PhD dissertation still under construction. Collocation and selectional preference of verbs are traditionally treated in semantic class or hierarchy approach, such as WordNet based studies. I just take it from another perspective, the semantic features. The real collocations of verbs are extracted from 2 treebanks, and the feature based lexicon HowNet is applied to analysis the semantic restrictions to nouns from verbs. The results are very attractive comparing to former studies. The current research is focused on the automatic recognition of collocations based on semantic features.

Awards

Rewarded paper, 2nd Student Workshop on Computational Linguistics. Beijing. 2004.8.
Recommended Paper, 8th Joint Seminar on Computational Linguistics. Nanjing. 2005.8
Excellent Graduate Student, By Nanjing Normal University. Nanjing. 2005.11
Third Prize in 2nd National Post-Graduate Mathematical Contest in Modeling. 2005

Research Interests

Computational Linguistics: Chinese Word Segmentation, Tagging and Chunking, Syntax-Semantic Interface, Semantic Orientation, Collocation, Metaphor Computing

Linguistics: Syntax, Lexical Semantics, Metaphor

Papers

2008

Li Bin, Yu Lili , Shi Min, Qu Weiguang. Computation of Chinese Simile with “像(like)”. *Journal of Chinese Information Processing*. 2008(6).

Li Bin, Chen Xiaohe. Direction Tendency Analysis on Chinese Derogatory and Commendatory Words. In *Proceedings of SWCL2008*. 2008.7.

Xiaohe Chen, **Bin Li**, Junzhi Lu, Hongdong Nian, Xuri Tang Nanjing Normal University Segmenter for the Fourth SIGHAN Bakeoff. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing*. 2008.

2007

Li Bin, Chen Xiaohe. A HCI Word Segmentation Method Adapting to Chinese Unknown Texts. In *Journal of Chinese Information Processing*. 2007(3).

Li Bin. Error Tagging Method inter-language Corpus Construction. *Journal of College of Chinese Language and Culture of Jinan University*. 2007(3).

Li Bin, Lu junzhi, Zhang Chengzhi, Chen Xiaohe. Semantic Orientation of Topics Based on Clustering Engines. In *Frontiers of Content Computing: Research and Application*.

2007.5.

2006

Li Bin, Chen Xiaohe. A HCI Word Segmentation Method Adapting to Chinese Unknown Texts. In *Proceedings of SWCL 2006*. 2006.8.

Li Bin, Chen Xiaohe, Fang Fang, Xu Yanhua. Corpus Based Investigation on High Frequent Maximal Overlapping Ambiguity String in Chinese Word Segmentation. *Journal of Chinese Information Processing*. 2006(1).

2005

Li Bin, Chen Xiaohe, Fang Fang, Xu Yanhua. The Research of High Frequent Maximal Overlapping Ambiguity String. In *Natural Language Understanding and Large-scale Content Computing*. Tsinghua University Press. 2005.7 (Recommended Paper)

Zhang Chengzhi, **Li Bin**. Hybrid Strategic Similarity Measures of Query. In *Natural Language Understanding and Large-scale Content Computing*. Tsinghua University Press. 2005.7

Xu Yanhua, Chen Xiaohe, **Li Bin**, Chen Zhong. “V+V” Structure Ambiguity Study in Contemporary Chinese for Automatic Parsing. In *Natural Language Understanding and Large-scale Content Computing*. Tsinghua University Press. 2005.7

Li Bin, Fang Fang. Auto Conversion from Arabian Number String to Chinese Number String. *Journal of College of Chinese Language and Culture of Jinan University*. 2005(2).

2004

Li Bin. Single Chinese Character Country Name Recognition. In *Proceedings of SWCL 2004*. 2004.8. (Rewarded Paper)

Academic Main Courses (MA & PhD)

C Programming & Chinese Language Processing

VC++ & Introduction to Chinese Language Processing

Corpus Linguistics

Introduction to English Papers on Computational Linguistics

Data Structure (C++)

Modern Logic

Artificial Intelligence

Machine Translation

Foundations of Statistical Natural Language Processing

Phonetics

Semantics

Chinese Semantics

Introduction to Chomsky's Theory

Modern Chinese Linguistics

Modern Linguistics

Applied Linguistics

History of the Grammar of Chinese Language

History of China's Linguistics

Teaching Chinese as Second Language