

# 计算语言学教育部重点实验室

## 2011 成果清单

### 一、论文

---

杨敏,常宝宝, 基于北京大学中文网库的语义角色分类, 中文信息学报, 2011 年第 3 期.

---

韩冬煦, 常宝宝, 基于边界熵和卡方统计量的多领域适应性中文分词方法, 全国第十一届计算语言学联合学术会议文集, 清华大学出版社, 2011.

---

白易, 王雷, 常宝宝, 汉语成语与典故知识库的自动映像策略与实现, 第四届华语文教学国际研讨会暨工作坊论文集, 2011 年 3 月.

---

董秀芳.语法化研究: 争议与发展并存, 《中国社会科学报》2011 年 8 月 2 日第 15 版。

---

李素建, 文本内容自动处理的相关研究, 《术语标准化与信息技术》, 2011 年 3 月第 1 期, pp43-48,

---

Decong Li, Sujian Li: Hypergraph-based inductive learning for generating implicit key phrases. WWW (Companion Volume) 2011: 77-78

---

---

Ouyang You, Wenjie Li, Sujian Li, Qin Lu: Applying regression models to query-focused multi-document summarization. *Inf. Process. Manage.* 47(2): 227-237 (2011)

---

Sujian Li, Wei Wang, Wenjie Li, Exploring strategies for developing link analysis based question-oriented multi-document summarization models, ICMLC. July 11, 2011, Guilin. (accepted)

---

王荀, 李素建, 宋涛, 姜伯平, 服务于内容侧面发现的框架识别, 全国第十一届计算语言学会议(CNCCL-2011), 2011.8. (accepted)

---

Chen Wang, Sujian Li, CoRankBayes: Bayesian Learning to Rank under the Co-training Framework and Its Application in Keyphrase Extraction, CIKM '11, October 24 - 28 2011, Glasgow, United Kingdom.

---

Sujian Li, Tao Song, Xun Wang, TAC 2011 guided summarization of ICL, TAC 2011 workshop, 2011, Gaithersburg, Maryland USA.

---

Sujian Li, Wenjie Li. Extracting Person Descriptions from Chinese Newswire through Ontological Semantic Calculation. *International Journal of Knowledge and Language Processing*, vol. 2, no.4, pp.1-27, 2011.

---

---

Wenjie Li, Zhifang Sui, A Hybrid Pattern-Based Method to Extract Concept Instances and Concept Attributes, Journal of Computational Information Systems.2011

---

Xingxing Zhang,Zhifang Sui, Semantic Class Learning with Deep Coordinate Structures in Web Pages, Journal of Computational Information Systems.2011

---

俞士汶、穗志方、朱学锋，综合型语言知识库及其前景，《中文信息学报》，第25卷第6期，2011年

---

王鑫、孙薇薇、穗志方，基于浅层句法分析的中文语义角色标注研究，《中文信息学报》，2011年第1期

---

凌晨添，张璐. 基于词可见度模型的图片相关文本信息提取技术. 第七届全国信息检索学术会议，2011年10月，山东大学.

---

凌晨添. 进化神经网络在信用卡欺诈检测中的应用. 微电子与计算机，2011年10月.

---

Weiwei Sun. 2011. A stacked sub-word model for joint Chinese word segmentation and part-of-speech tagging. ACL.

---

Weiwei Sun and Jia Xu. 2011. Enhancing Chinese word segmentation using unlabeled data. EMNLP.

---

Xiaojun Wan. Bilingual Co-training for Sentiment Classification of Chinese Product Reviews. Computational Linguistics, 37(3): 587-616.

---

---

Xiaojun Wan. Using Bilingual Information for Cross-Language Document Summarization. In ACL2011, pages 1546-1555.

(Long Paper)

---

Xiaojiang Huang, Xiaojun Wan, Jianguo Xiao. Comparative News Summarization Using Linear Programming. In ACL2011, pages 648-653. (Short Paper)

---

Xiaojun Wan, Houping Jia, Shanshan Huang, Jianguo Xiao. Summarizing the Differences in Multilingual News. In SIGIR2011, pages 735-744. (Full Paper, Acceptance Rate =19.8%)

---

Rui Yan, Xiaojun Wan, Jahna Otterbacher, Liang Kong, Xiaoming Li, Yan Zhang. Evolutionary Timeline Summarization: a Balanced Optimization Framework via Iterative Substitution. In SIGIR2011, pages 745-754. (Full Paper, Acceptance Rate =19.8%)

---

Rui Yan, Liang Kong, Congrui Huang, Xiaojun Wan, Xiaoming Li, Yan Zhang. Timeline Generation through Evolutionary Trans-Temporal Summarization. In EMNLP2011, pages 433-443. (Oral Paper, Acceptance Rate =15%)

---

Xiaojun Wan. Collaborative Data Cleaning for Sentiment Classification with Noisy Training Corpus. In PAKDD2011, pages 326-337. (Short Oral Presentation)

---

---

Xiaojun Wan, Liang Zong, Xiaojiang Huang, Tengfei Ma, Houping Jia, Yuqian Wu, Jianguo Xiao. Named Entity Recognition in Chinese News Comments on the Web. In IJCNLP2011, pages 856-864. (Full Oral Paper, dataset)

---

Huiying Li, Yue Hu, Zeyuan Li, Xiaojun Wan, Jianguo Xiao. PKUTM participation at TAC 2011 Summarization Track. In TAC2011.

---

王厚峰, 汉语缩略语自动处理研究现状, 中文信息学报, 2011, Vol.25 (5)

---

焦妍,王厚峰, 基于机器学习方法与搜索引擎验证的缩略语预测, 第十一届全国计算语言学学术会议论文集《中国计算语言学前沿研究进展》, 清华大学出版社, 2011年8月

---

陈晨 王厚峰, 基于社会网络的跨文本同名消歧, 中文信息学报, 2011, Vol.25 (5)

---

何正炎 王厚峰, 商品品牌名称挖掘, 全国计算语言学联合学术会议论文集《中国计算语言学前沿研究进展》, 清华大学出版社, 2011年8月

---

徐幸 王厚峰 中文文本蕴涵的推理模型, 全国计算语言学联合学术会议论文集《中国计算语言学前沿研究进展》, 清华大学出版社, 2011年8月

---

---

Yang Song; Houfeng Wang; Jing Jiang. *Link Type Based Pre-Cluster Pair Model for Coreference Resolution*. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, 2011.

---

HE Pei, Colin G. JOHNSON & WANG Houfeng, Modeling grammatical evolution by automaton, *SCIENCE CHINA(Information Sciences)*, 2011 Vol. 54 (12)

---

吴云芳, 石静, 金澎. 2011. 基于图的同义词集自动获取方法. 《计算机研究与发展》 Vol. 48(4), Pp 610-616.

---

Miaomiao Wen, Yunfang Wu. 2011. Mining the sentiment expectation of nouns using bootstrapping method. *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*, Pp 1423-1427.

---

魏雪、袁毓林。《汉语词类测试平台的设计与实现》，《语言教学与研究》2011年第2期，第83-90页。：

---

袁毓林、王明华。《“Neg+Wh+VP”和“Wh+Neg+VP”意义同异之辨——兼谈全称否定的排他性保留和特称容忍的逻辑机制》，《中国语文》2011年，第3期，第195-204页。

---

张秀松、袁毓林、陈振宇、周强。《汉语亲属关系的语义表达和句法实现》，《语言学论丛》第42辑，商务印书馆，2011年，第211-235页。

---

---

张秀松、袁毓林、陈振宇、周强。《基于情境网络的汉语亲属名词的语义描写体系》，《世界汉语教学》，2011年，第3期，第318-333页。

---

袁毓林。《“差点儿”和“差不多”的意义同异之辨》，《语言教学与研究》2011年第6期，第66-75页。中国人民大学复印报刊资料《语言文字学》2012年第3期，第22-30页。全文转载。

---

Lei Zou, Jinghui Mo, Lei Chen, M. Tamer Özsu, Dongyan Zhao, “gStore: Answering SPARQL Queries Via Subgraph Matching”, International Conference on Very Large Databases (VLDB), Seattle, USA, pp. 482-493, Aug. 2011

---

Lei Zou, Jinghui Mo, Lei Chen, M. Tamer Özsu, Dongyan Zhao. gStore: Answering SPARQL Queries Via Subgraph Matching. International Conference on Very Large Data Bases, VLDB, 2011. Lei Zou, Lei Chen. Pareto-Based Dominant Graph: An Efficient Indexing Structure to Answer Top-k Queries, IEEE Transactions on Knowledge and Data Engineering (TKDE), TKDE, 2011.

---

王雷，常宝宝，Harkness, J., 科技文献中多重词缀词的一种自动翻译方法，第十一届汉语词汇语义学研讨会（CLSW2010）论文集，COLIPS Publishing Society, 2011年3月

---

王雷 编著、《中国成语 1000（汉英对照）》、978-7-301-19376-1、北京大学出版社、2011年9月

---

## 二、专著

《词汇化：汉语双音词的衍生和发展（修订本）》，商务印书馆，2011年11月。30万字。2013年5月出版重印本。董秀芳。

《语言学纲要学习指导书（修订版）》，北京大学出版社，2011年1月。董秀芳（第四章和第九章）。

## 三、科研项目

类型	批准部门	编号	负责人	时间	经费
自然科学基金	基金委	61003009 基于图数据库理论的海量 RDF 数据存储和查询方法研究	邹磊	2011.1- 2013.12	19 万
自然科学基金青年项目	基金委	61003009	邹磊	2011.1- 2013.12	19



教育部博士点新教师基金	教育部	20100001120029	邹磊	2011.1-2013.12	3.6
科技支撑计划子课题	科技部			2011.1-2015.01	
国家科技支撑计划子课题项目	科技部	2011BAH10B04-03	李素建	2011.7-2014.3	15 万
国家自然科学基金（重大研究计划培育）	国家自然科学基金委	91024009	王厚峰	2011-2013	35 万元
NSFC 2010	国家自然科学基金委	61075076	穗志方	2011年1月—2013年12月	37 万元

教育部人文社会科学 重点研究基地重大研 究项目	教育部	10JJD740008	袁毓林	2011年1月— 2013年12月	20万
国家语言文字工作委 员会“十二五”规划项 目	教育部	YB125-47	袁毓林	2011年9月— 2014年8月	5万元