

语言工程与计算实验室

工作简报

(2014年第2期 总第2期)

2014年12月31日

新闻要目

1. 协同创新中心协同实验室老师召开读后续写自主学习系统设计讨论会
2. 赵东岩教授做题为“自然语言处理技术与大数据分析”学术报告
3. 实验室老师参加第十三届全国计算语言学学术会议
4. 实验室学术委员宗成庆教授谈国家级项目申报要点
5. 赵军教授围绕信息抽取和问答系统做精彩报告
6. 实验室老师参加“面向网络文本处理的统计学习方法” CCF 学科前沿讲习班
7. 实验室老师参加第三届自然语言处理与中文计算学术会议
8. 实验室喜获广东省高校国家级重点培育平台立项
9. 香港城市大学 Alex Fang 教授来校做报告
10. 北京语言大学宋柔教授应邀来我校讲学

要闻一：协同创新中心协同实验室老师召开读后续写自主学习系统设计讨论会

按照阳爱民副校长的部署，2014年10月10日上午10:00-12:00在北校七教留学生院会议室举行了读后续写自主学习系统设计讨论会，会议由蒋盛益主持，王初明、王凤兰、王启、郝天永、李霞、柯晓华、姜琳、范香娟、林江豪、许琪等参加了研讨会。

首先郝天永老师对前期调研工作做了总结和汇报，给出了目前国内已有相关学习系统的相关概况，并给出了读后续写自主学习系统的架构设计初步方案。王初明老师对所给出的初步方案给与了肯定，并指出系统需要明确所面向的对象是公共英语还是专业英语、是大学生还是小学生使用等，系统应该具备互动性强、多样化等特点。

会上老师们畅所欲言，并对如何将读后续写在软件系统中的实现细节进行了深入的讨论。通过来自语言学领域和信息技术领域的不同学科方向老师的深入讨论，会议初步确定在10月底先让语言学领域专家制作一个能够体现读后续写精髓的案例模板，以及这个模板所对应系统应具备的功能和方法脚本，在此基础上，双方再次碰头争取下次能够拿出一个可以展示的DEMO出来，为系统的最终实施打下基础。

要闻二：赵东岩教授做题为“自然语言处理技术与大数据分析”学术报告

2014年12月10日上午10:10—12:00，赵东岩教授在南校区C308进行了“自然语言处理技术与大数据分析”讲座。

大数据复杂、多样，其有价值的信息需要深度挖掘和语义理解，因此，利用自然语言处理和知识工程方法来分析大数据是十分重要的。根据研究与实践经验，在大数据分析任务中，分析方法、研发框架与数据积累均十分重要。赵东岩教授报告中举例说明大数据分析中典型的自然语言处理方法，利用基于语义的数据挖掘和信息抽取来分析异质大数据（如文本、图像、多媒体）的语义概念、实体知识及其语义关联，并构建大规模多媒体语义知识库，提升大数据分析性能。



赵东岩教授在演讲中

附：赵东岩，北京大学计算机科学技术研究所研究员，博士生导师。主要研究方向为文本挖掘、语义信息处理、数字出版技术。近年来承担国家级项目13项、省部级科研项目8项，主持的项目获得国家科技进步二等奖和北京市科学技术一等奖。目前的研究工作聚焦于互联网异质文本数据的语义理解、中文语义知识库构建以及大规模多媒体语义知识库的语义搜索与智能服务，发表了包括ACL、AAAI、WWW、EMNLP、VLDB以及TKDE、VLDB Journal等国际一流会议与期刊在内学术论文60余篇。个人获第十届中国青年科技奖和北京市第七届“科技之光”技术创新特别奖等荣誉。计算机学会CCF高级会员，CCF中文信息技术专委会秘书长，CCF大数据专家委员会委员、网络与数据通信专委会委员。

要闻三：实验室老师参加第十三届全国计算语言学学术会议

2014年10月18日—19日第十三届全国计算语言学学术会议（CCL 2014）在武汉华中师范大学举行。作为国内最大的自然语言处理专家学者的社团组织—中国中文信息学会（CIPS）的旗舰会议，全国计算语言学会议从1991年开始每两年举办一次，从2013年开始每年举办一次。CCL着重于中国境内各类语言的计算处理，为传播计算语言学最新的学术和技术成果提供了广泛的交流平台。

实验室老师蒋盛益教授、李霞副教授、郝天永副教授、丘心颖副教授等参加了会议，其中蒋盛益教授还宣读和播放了下一届暨第十四届全国计算语言学学术会议（CCL 2015）即将在我校举行的报告和视频。另外，实验室学术委员孙茂松教授、宗成庆教授、周国栋教授、刘挺教授也参加了这次会议，会后，实验室教师与实验室学术委员进行了合影。



实验室教师与学术委员在会议上合影

要闻四：实验室学术委员宗成庆教授谈国家级项目申报要点

2014年12月11日下午2:30在信息学院会议室，实验室学术委员宗成庆教授对国家自然科学基金项目如何进行选题、如何进行申报书的撰写、需要注意的事项等与参会的教师进行了深入的指导。参会的教师多数为当年需要申请国家课题的青年教师，还包括学院的其它对申报国家课题感兴趣的教师。

通过宗教授的讲解，以及与宗教授的交互，大家对申报国家课题有了更多的了解，并增加了更多的信心。个别教师还将实现准备好的申报书的选题、内容等与宗教授进行了探讨和交互，其它参会教师也积极探讨和提建议。

宗教授不仅从专家的角度探讨了国家课题申报的选题、撰写方法、注意事项，还从如何将计算机与语言学结合角度，深入探讨了作为广外的教师如何选题进行申报，并对未来广外的科研增长点提出了积极的建议。

要闻五：赵军教授围绕信息抽取和问答系统做精彩报告

2014年12月4日下午2:30-4:30, 中国科学院自动化研究所模式识别国家重点实验室研究员赵军教授在南校E203举行“信息抽取和问答系统”讲座。问答系统被认为是下一代搜索引擎的重要形态, 而信息抽取是支撑问答系统等互联网应用的关键技术之一。赵军教授围绕信息抽取和问答系统两个研究方向, 系统地介绍其中的基本概念、主要方法、最新研究进展、需要解决的问题和发展趋势, 使听者能够对信息抽取和问答系统研究领域的重点问题和主要方法有较为系统的了解。



赵军教授在演讲中

附：赵军，中国科学院自动化研究所模式识别国家重点实验室研究员、博士生导师。1998年于清华大学计算机科学与技术系获得博士学位，1998年-2002年在香港科技大学计算机科学系进行博士后访问，2002年至今在中科院自动化所模式识别国家重点实验室工作。目前负责国家重点基础研究发展计划973课题、国家高科技863计划课题、国家自然科学基金等多项课题的研究。兼任ACM Transactions on Asian Language Processing副主编，中国中文信息学会计算语言学专业委员会副主任委员，《中文信息学报》编委，973项目“面向三元空间的互联网中文信息处理理论与方法”专家组成员。在IEEE Transactions on Knowledge and Data Engineering (TKDE), Journal of Machine Learning Research等国际顶级期刊和SIGIR、ACL、AAAI、IJCAI、COLING、CIKM等国际顶级会议上发表论文40余篇，获得COLING-2014最佳论文奖(1/705)，ACM KDD CUP-2011亚军(2/1297)等。研究方向为自然语言处理和信息检索，目前主要从事网络信息抽取和问答系统等方面的研究与开发工作。

要闻六：实验室老师参加第三届自然语言处理与中文计算学术会议

自然语言处理与中文计算会议（NLPCC）是由中国计算机学会主办的 CCF 中文信息技术专业委员会年度学术会议。NLPCC 专注于自然语言处理及中文计算领域的研究和应用创新，致力于推动该领域学术界和工业界的交流，成为覆盖全国、具有国际影响力的学术交流平台。

NLPCC 2014 第三届全国自然语言处理及中文计算会议于 2014 年 12 月 5-9 日在深圳麒麟山庄举行，会议以“自然语言深度理解和知识整合”为主题，聚焦于 Web Mining & Big Data、Multi-Lingual Access、Machine Learning for NLP、Knowledge Acquisition in Social Networks、Search & Ads、Question Answering，以及 Fundamentals and Applications on Language Computing 等领域。会议组织了一系列学术活动，包括特邀报告、论文报告、ADL/Tutorials、Workshops、Panel、技术评测、工业展示、校园开放日等多项活动。其中会议邀请了 Bing Liu 教授（UIC）、Hwee Tou Ng 教授（NUS）和 Eduard Hovy 教授（CMU）等为大会做特邀报告。实验室蒋盛益等 8 位老师参加了本次会议，同时在本次会议中，实验室李霞老师被选为中国计算机学会中文信息处理专委会委员。下图为实验室老师与实验室学术委员宗成庆教授、周国栋教授在本次会议上的合影。



实验室老师与实验室学术委员在 NLPCC2014 上合影

要闻七：实验室老师参加“面向网络文本处理的统计学习方法” CCF 学科前沿讲习班

互联网与移动网络的飞速发展网络文本处理技术的发展带来了前所未有的机遇和挑战。网络文本处理涉及信息检索、信息抽取、观点挖掘、语义搜索等任务，统计学习是其中的重要方法。中国计算机学会《学科前沿讲习班》第 52 期《面向网络文本处理的统计学习方法》2014 年 12 月 5 日-7 日在深圳举行，讲习班邀请了学术界和工业界的著名专家学者对信息检索、信息抽取、观点挖掘和语义搜索中统计学习的基础理论、方法和应用以及其他热点问题进行了系统的讲解。讲习班为青年学者和学生提供了学习和交流机会，让青年学者和学生快速了解本领域的基本概念、研究内容、方法和发展趋势。实验室老师参加了本次培训。



参加培训班的学员合影



实验室郝天永老师提出问题

要闻八：实验室喜获广东省高校国家级重点培育平台立项

据悉，我实验室申请的广东省高校国家级重点实验室培育平台近期获得教育厅项目立项。该培育项目是于2014年6月30日经校长办公会讨论决定推荐并申报的。

2014年7月12日上午9:00，学校针对“广东省教育厅关于组织广东高校国家级重点培育平台、重大培育项目与成果申报遴选工作”在北校区行政楼第三会议室进行专家指导答辩会，其中广东高校国家级重点培育平台-语言工程与计算实验室进行了答辩，蒋盛益教授在答辩会上介绍了申报材料，李霞、郝天永、丘心颖、张新猛等陪同参会，并根据专家提出的建议，对最终申报材料进行了修改、补充和完善，2014年7月14日正式提交申报材料。

实验室进一步厘清了主要发展方向为：1) 计算机辅助第二语言学习研究；2) 语言测试与评估技术；3) 面向安全战略的网络大数据语言应用研究。详细研究内容包括：计算机辅助第二语言学习模式研究、基于计算机网络技术的二语习得与语言新现象研究、基于语料库的外国学生汉语语块习得研究、满足模型校准等测试理论的测试题库建设、大规模考试中计算机辅助自动阅卷及自动评分研究、面向网络信息安全领域的信息抽取、法律信息抽取、社会情绪演化模型研究、基于网络语言理解的舆情研判与情感倾向性研究等。

要闻九：香港城市大学 Alex Fang 教授来校做报告

2014年11月12日上午，来自香港城市大学的方称宇教授在南校区图书馆报告厅，从语料库语言学家的角度，向大家介绍了基于文本的信息处理的历史发展，同时全面展现了自己最新的文本聚类实验成果。本次讲座由思科信息学院、语言工程与计算广东省社会科学重点实验室承办，思科信息学院副院长蒋盛益主持讲座。

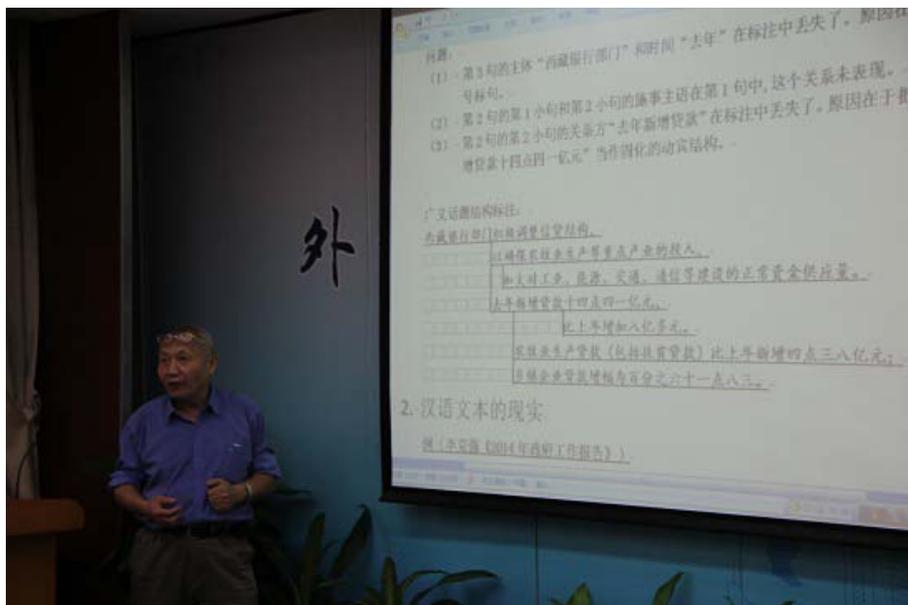


方称宇正在作讲座

附：方称宇，先后任职于广州外国语学院英语系、英国伦敦大学学院的英语系和计算机系，并任英国伦敦大学学院英语调查中心副主任。他现任香港城市大学翻译及语言学系副教授，讲授语料库语言学、认知语言学、文体学和机器翻译等课程，同时兼任北京航空航天大学外国语学院的客座教授。他于2008年创办了对话系统实验室，并兼任实验室主任，指导在计算语言学、语料库语言学和自然语言处理等领域的研究，课题包括中英文语料库建设、自动句法分析、对话行为分析、文本特征提取、术语提取、文本信息处理等。

要闻十：北京语言大学宋柔教授应邀来我校讲学

2014年11月27日下午，由我校科研处主办，语言工程与计算实验室、中国语言文化学院和外语研究与语言服务协同创新中心联合承办的首期云山科研大讲坛在北校区第六教学楼外语研究与语言服务协同创新中心讲学厅举行。讲坛邀请了北京语言大学二级教授、博士生导师宋柔解读“什么是汉语的句子”。学校科研处有关负责人出席了此次讲坛。



宋柔教授演示“广义话题结构流水模型”

另外，宋柔教授还与实验室教师、外语研究与语言服务协同创新中心教师、词典学研究中心教师等举办了一场有关如何开展计算语言学研究的座谈会。实验室教师蒋盛益、李霞、郝天永、柯晓华、丘心颖等老师参加了讨论会。

报送：广东外语外贸大学科研处

抄送：学校各科研单位

2014 年 12 月 31 日印
