

语言工程与计算实验室

工作简报

(2015 年第 1 期 总第 3 期)

2015 年 7 月 30 日

新闻要目

1. 王挺教授访问和指导实验室工作
2. 实验室开展语音语言信息化成果展
3. 实验室老师指导学生获得广东大学生课外学术科技作品竞赛一等奖
4. 宗成庆教授开展机器翻译的专题系列讲座
5. 实验室举办研究人员聘任仪式
6. 实验室召开建设研讨会
7. 北京大学王厚峰教授应邀来我校讲学
8. 实验室老师参加 ACL2015 国际计算语言学顶级会议
9. 实验室专职研究员云山青年学者刘伍颖老师做专题讲座

要闻一：王挺教授访问和指导实验室工作

4月24日下午4时，国防科技大学王挺教授来实验室指导工作，并分别和实验室及信息学院的教师在实验室会议室A201进行了座谈。王挺教授分别从实验室未来研究方向的凝练、实验室在对学校的学科建设贡献和社会服务影响力等方面，给出了他的看法和指导意义。座谈会上，参加座谈的教师和王挺教授也分别进行了讨论和交流，取得了很好的效果。



王挺教授与实验室及信息学院教师交流会场



王挺教授与实验室老师合影

本次王挺教授来访，还给我校师生带来了一场关于网络信息的话题与观点分析的精彩讲座。



王挺教授谈网络信息的话题与观点分析

王挺在讲座中提到，互联网信息的快速发展产生了海量的信息，也为信息的处理提出了新的要求。“我们不仅要知道大家在关心哪些事，更要知道大家是怎么看待这些事的。”他认为，从目前来看，人们对于信息的需求已经不局限于周围日常生活层面，经济活动、社会管理、科学研究都离不开信息。当信息像雪片般飞来，就需要信息呈现得更条理化和信息处理分析。随着信息技术的发展，信息出现了质量不高、规模庞大、话题演变迅速的特点，他以 H7N9 疫情为例，将 N7N9 视为一个大话题，而往下细分则可以分出多个子话题：饮食、治疗、预防……而随着时间的推移，这些话题又在内容上进行演化。从这话题的特性出发，王教授继续介绍了话题分析的主要内容，包括报道切分、话题关联识别、新事件发现、话题发现、话题追踪、分层话题发现等。王教授提出通过 Wikipedia 增强文档的语义表示，用技术手段过滤文本噪音，用图距离进行相似度计算等来解决问题。

除了话题和话题分析，王挺在讲座中还讲述了观点与观点分析方法。他提到，语言的功能包括描述、社交和表达情绪。后面两个功能都是人们在传递主观信息，表达自己的观点，为了实现观点分析目标，王教授提出了构建情感词典、抽取观点要素、抽取特定信息、分析感情倾向性等方法。

要闻二：实验室开展语音语言信息化成果展

由学校科研处和语言工程与计算实验室共同举办的“语音语言信息化处理及其应用”科技成果展于 2015 年 5 月 16 日在南校区实验楼开展。

本次成果展是 2015 年广东省科技进步活动月、广州市科技活动周和广外建校五十周年系列学术活动的一个组成部分，活动大力宣传实施创新驱动发展战略和创新创业发展的理念，普及科学知识、弘扬科学精神、提高全民科学素养。活动的主要内容是向广州市民展示语音、语言信息化处理的手段、方法、技术，以及语言信息化处理在实际生活中的应用。成果覆盖了语言评测、网络舆情分析和信息搜索三大领域，包括软件、手机 APP 和在线网站，展示者通过系统演示、视频播放、以及现场讲解等多样化的形式向前来参观的我校师生和广州市民进行了展示。



参观者阅读成果概要

在此次成果展中，有“在线英语发音质量评测系统”、“英汉口语互译智能评分系统”、“英文作文批改助手”和“英汉翻译测评系统”等四款系统可应用于解决英语学习中的口语发音、写作评分及反馈、翻译评分及反馈等问题，这些应用系统是我校师生将计算机科学和外语教学与测试相结合开展交叉学科研究产出的重要成果。在现场参观的一位我校英语教师表示，如果此类英语自动评分系统能够应用到实际教学中去，可以大大缩短教师的课后批改作业时间，帮助老师投入更多的时间和精力到创造性的工作中去，另外，系统自动生成的即时教学反馈对学生的自主学习也有很大帮助。



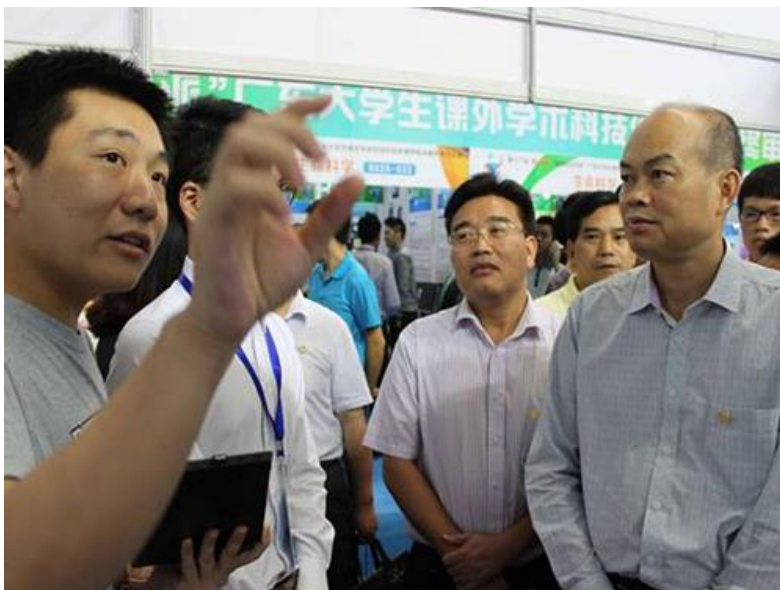
系统作者向参观者讲解和交流

附：广东省“科技进步活动月”介绍

根据 2015 年全国科技活动周的部署和省委、省政府的工作安排，今年的全省“科技进步活动月”（以下简称“活动月”）时间将安排在 5 月中旬至 6 月中旬进行。“活动月”将围绕“创新创业科技惠民”这一主题，开展提升广东自主创新能力、营造创新创业环境、科技惠及民生等系列活动。主要目的是宣传创新驱动经济社会发展、创新创业成果服务改善民生，进一步提高公众科技意识和科学素养。在内容和形式上，重点展示一批大众创业万众创新的科技创新成果，开展科技服务企业、青少年、农村的活动，组织高校、科研院所等科技资源向社会开放，加强低碳节能、健康生活、食品安全、空气质量等社会热点问题的科普宣传，在全省掀起科学技术普及的新高潮。（来源：广东外语外贸大学）

要闻三：实验室老师指导学生获得广东大学生课外学术科技作品竞赛一等奖

2015年5月23日，为期3天的第十三届“挑战杯”课外学术科技作品竞赛于华南师范大学大学城校区拉开帷幕，我校共有11个作品成功入围终审决赛，其中科技发明制作类2个作品均是由实验室老师李心广指导，“基于移动互联网的在线英语口语发音质量评价系统”和“具有语义分析功能的英汉口语互译智能评分系统的研究”实验室李心广老师指导的决赛作品荣获了一等奖1名，二等奖1名。实验室蒋盛益老师和丘心颖老师指导的决赛作品均获得三等奖。



选手介绍参赛作品

附:第十三届“挑战杯”广东大学生课外学术科技作品竞赛我院获奖名单:

奖项	作品名称	指导老师	作者
一等奖	基于移动互联网的在线英语口语发音质量评价系统	李心广	陈君宇、张胜斌、朱小凡、李升恒、王泽铿、许港帆
二等奖	具有语义分析功能的英汉口语互译智能评分系统的研究	李心广	陈广豪、何智明、李婷婷、陈泽群、马晓纯、王晓杰、陈伟峰、郭婉华
三等奖	面向热点公共事件的社会情绪分析	蒋盛益	吴佳林、林先铭、黄卫坚、蔡茂丽、邓远挺、谢铭熙、骆家焕、甘穗福
三等奖	面向网络空间安全领域的情报获取与分析系统	丘心颖 蒋盛益	林小铂、黄宇轩、陈舒扬、杨逸峰、陈兴武、郭佳明、王晓锐、陆罗炜

要闻四：宗成庆教授开展机器翻译的专题系列讲座

2015年6月15-16日，我实验室云山讲座学者宗成庆教授在南校区院系办公楼201会议室面向广外学生和教师做了有关机器翻译的专题系列讲座。



宗成庆教授讲座现场



宗成庆教授与听讲座的教师合影

15日下午3:30-5:00，宗教授首先给我们带来了系列讲座第一场：机器翻译概论。在第一场讲座中，宗教授分别从引言、机器翻译的产生与发展、机器翻译的困难、机器翻译研究现状、机器翻译基本方法以及语音翻译等几个方面，对机器翻译的历史和现状、以及研究方法等进行了详细的阐述。

16日上午9:30-11:00，宗教授继续广外师生带来了机器翻译系列讲座的第二场：统计机器

翻译。在这一场讲座中，宗教授从统计翻译原理、翻译模型、解码算法以及系统实现等几个方面向大家详细讲解了当前统计机器翻译所涉及的概念、原理和优秀方法。

宗教授的系列讲座不同于一般的概述性讲座，他将机器翻译的相关专题分解、详细的给大家讲解，让我们深刻了解机器翻译的理论，并且了解到如何将语言学理论、语言特征的获取、语言分析结果结合，应用于机器翻译及其它相关领域。

此次驻校过程中，宗教授还分别参加了信息学院的研究方向论证会，对信息学院的 3 个团队建设给出了建设性的建议。

要闻五：实验室举办研究人员聘任仪式

2015年7月3日，语言工程与计算实验室在南校区举行研究人员聘任仪式。16位校内教师和科研人员受聘为语言工程与计算实验室研究人员。

我校副校长阳爱民，科研处处长陈彦辉，云山青年学者刘伍颖以及实验室受聘科研人员等出席活动。活动由语言工程与计算实验室主任李心广主持。



阳爱民校长为实验室研究人员颁发聘书

附：现场受聘研究人员名单(排名不分先后)：

东语学院 陈多友

教育学院 董金伟

英文学院 冯光武

商英学院 郭桂杭

思科信息学院 郝天永

思科信息学院 蒋盛益

教育学院 吕剑涛

翻译研究中心 莫爱屏

信息学院 丘心颖

文科基地 任伟

中文学院 魏慧斌

英文学院 温宾利

思科信息学院 谢柏林

思科信息学院 张晶

词典学研究中心 章宜华

思科信息学院 周咏梅

要闻六：实验室组织研究人员召开建设研讨会

2015年7月3日，实验室组织研究人员召开实验室建设研讨会。实验室研究人员积极地为实验室的建设出谋划策，围绕“如何更好地发展建设与管理实验室？”和“如何更好地发挥实验室的协同合作？”等议题展开讨论。

词典学研究中心主任章宜华表示要明确语言学科和计算机学科的分工合作。首先，语言学研究人员要根据对外语的研究现状和研究进展提出运用计算的方式解决外语问题的需求计划。其次，计算机学科的研究人员要根据需求计划，提出切实可行的需求技术方案。最终，通过语言学科和计算机学科的分工合作，产出合理有效的计算语言学成果。

翻译学研究中心主任莫爱屏对章宜华教授的建议进行了补充。他指出所谓成果即商业性产品。怎样产出成果是实验室应重点考虑的问题，首先要挖掘出实验室发展的重点方向，然后再根据学校的相关政策、配套措施等有利支持条件全面营销，将成果推向市场，产生社会效益，真正做到产学研一体化。

此外，其他研究人员也积极建言献策。东语学院院长陈多友提到实验室的发展应该充分发挥广外特色的多语种资源。教育学院院长董金伟提出实验室应向全国招标以发展优秀项目等建议。



研究人员积极为实验室建设出谋划策

要闻七：北京大学王厚峰教授应邀来我校讲学

2015年7月8日下午，北京大学信息科学技术学院教授、北京大学计算语言学研究所所长王厚峰受邀做客我校，为师生带来了一场题为“语篇分析与指称”的学术讲座。



王厚峰讲学中

为了让听讲者更好地理解语篇分析，王厚峰着重介绍了语篇的衔接性现象，包括 Reference(照应)、Substitution(替换)、Ellipsis(省略)、Conjunction(连接)和 Lexical cohesion(词汇衔接)五种衔接关系，并用生动的例子讲解语篇中的衔接性现象和出现衔接的三个原因：一是保持话题的延展，二是丰富表达形式，三是追求语言的经济。最后他介绍了进行语篇分析所使用的工具、方法，以及相关的实验结果。



王厚峰教授与教师座谈

讲座开始前，王厚峰与语言工程与计算实验室和思科信息学院研究人员在南校区第一行政会议室进行了学术交流会。交流会由思科信息学院院长蒋盛益主持。蒋盛益首先向王厚峰介绍了思科信息学院和语言工程与计算实验室的研究工作开展情况。王厚峰向与会研究人员介绍了北京大学在自然语言处理领域和计算语言学研究所的发展情况，作为国家自然科学基金委的评审专家，王厚峰讲解了当前国家自然科学基金课题的申报与管理情况。全体与会人员就计算语言学研究中遇到的问题进行了深入交流。

附：王厚峰，北京大学信息科学技术学院教授、北京大学计算语言学研究所所长。主要从事自然语言处理相关的研究，包括语言知识资源建设、语篇分析技术、文本观点挖掘、实体识别与歧义消解、智能问答等。王教授是国家“十二五”863项目“大规模中文语义信息处理技术与系统”的首席专家、国家社科基金重大项目“面向网络文本的多视角语义分析方法、语言知识库及平台建设研究”的首席专家，主持多项国家自然科学基金项目。

要闻八：实验室老师参加 ACL 会议

2015 年 7 月 27 日-29 日，2015 年度自然语言学协会年度会议 (ACL2015) 在北京举行。ACL 会议 (Annual Meeting of the Association for Computational Linguistics) 是自然语言处理与计算语言学领域最高级别的学术会议，由计算语言学协会主办，每年一届。涉及对话 (Dialogue)、篇章 (Discourse)、评测 (Eval)、信息抽取 (IE)、信息检索 (IR)、语言生成 (Language Gen)、语言资源 (Language Res)、机器翻译 (MT)、多模态 (Multimodal)、音韵学/形态学 (Phon/Morph)、自动问答 (QA)、语义 (Semantics)、情感 (Sentiment)、语音 (Speech)、统计机器学习 (Stat ML)、文摘 (Summarizations)、句法 (Syntax) 等多个方面。



国际计算语言学协会 (The Association for Computational Linguistics) 是世界上影响力最大、最具活力的国际学术组织，其会员遍布世界各地。ACL 旗下的 ACL 年会是国际上最有影响的计算语言学及自然语言处理学术会议，而 IJCNLP 则是亚洲自然语言处理联合会 (Asian Federation of Natural Language Processing) 的旗舰会议。实验室老师李霞副教授参加了会议。

要闻九：实验室专职研究员云山青年学者刘伍颖老师做专题讲座

2015年7月3日，实验室专职研究员云山青年学者刘伍颖老师做题为“面向大数据的多语言信息处理”的专题讲座。讲座共讲述三部分内容：

首先，刘伍颖老师分析大数据时代网络信息的特性，从体量巨大、高速实时、格式繁多、语种混杂、跨网传播等5个方面给出数据例证，并从宏观态势和微观内容两个层面提出网络信息处理面临的新挑战。其次分析美国多语言信息处理研究项目，包括TIDES、GALE、MADCAT、DEFT、XDATA在内的10个多语言信息处理研究项目，展现美国在多语言信息处理上的战略规划和实施步骤，从语言和计算相结合的视角阐释语言科学与工程的重大意义。最后，刘老师从语言资源建设、自然语言处理基础构件研发、多语言信息处理系统构建等3个方面全景展示刘伍颖研究团队取得的研究成果，并现场演示相关软件系统。



刘伍颖老师分析大数据下的多语言信息处理

附：刘伍颖，男，汉族，1980年出生，江西九江人。学士、硕士、博士毕业于“天河”超级计算机的诞生地——国防科学技术大学计算机学院。硕士阶段主要从事超级计算技术——分布式计算环境研究工作，硕士论文题目是《异构多域环境的全局资源调度技术研究是实现》。博士阶段主要从事信息检索和信息过滤研究工作，博士论文题目是《面向垃圾信息过滤的主动多域学习文本分类方法研究》。当前主要研究领域包括自然语言处理、信息检索、计算语言学、机器翻译、机器学习、数据挖掘、大数据、分布式计算等。刘伍颖已在国内外顶级学术会议和期刊上发表高水平论文40余篇，其中被SCI检索4篇，被CPCI-S检索4篇，被EI核心检索15篇，出版学术著作2部，主持和参加科研项目20余项，获得智能摘要软件等计算机软件著作权3项。刘伍颖已主持开发具备完全知识产权的软件系统10余个、大型语言资源5个，其中“银河摘星”自动文摘软件、高效能幂律分类器、“星汉传思”机器翻译软件、智能语义搜索引擎、“哲译通”英汉双向术语词典等系统和同文术语矩阵、星汉词库、星汉句库等资源已在搜索引擎、电子商务、机器翻译、网络舆情等应用领域广泛部署，实用效果理想，经济效益显著。

报送：广东外语外贸大学科研处

抄送：学校各科研单位

2015年7月30日印
