

中文信息处理技术评测与语言 资源建设专题总结

北京语言大学 杨尔弘

刘 群：中文语言资源建设与评测的进展与展望

陶建华：中文语言资源联盟进展

何婷婷：国家语言资源监测语料库介绍

俞士汶：北京大学综合语言知识库

刘奕群：海量规模中文网络信息检索评测语料库的设计与实现

赵 军：信息检索和情感分析评测情况介绍

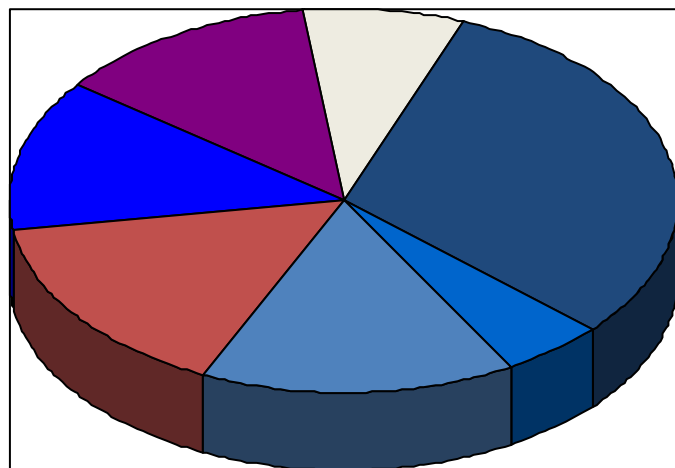
周 强：中文信息学会句法分析评测介绍

邹嘉彦：香港城大泛华语共时同题语料库介绍

易绵竹：基于本体语义学的语言资源观

自由发言：萧国政

资源分布（按数量）



- 评测语料
- 翻译语料
- 词典、人名、地名库
- 分词、句法标注语料
- 语音识别与合成
- 自然口语语料库
- 其它

语音类、语言类、评测类、其他
(脸部表情动画等)

刘 群：中文语言资源建设与评测的进展与展望

陶建华：中文语言资源联盟进展

何婷婷：国家语言资源监测语料库介绍

俞士汶：北京大学综合语言知识库

刘奕群：海量规模中文网络信息检索评测语料库的设计与实现

赵 军：信息检索和情感分析评测情况介绍

周 强：中文信息学会句法分析评测介绍

邹嘉彦：香港城大泛华语共时同题语料库介绍

易绵竹：基于本体语义学的语言资源观

自由发言：萧国政

语言资源建设及其应用

刘 群： 中文语言资源建设与评测的展望

任务驱动
自动构建

陶建华： 中文语言资源联盟进展

何婷婷： 国家语言资源监测语料库介绍

俞士汶： 北京大学综合语言知识库

刘奕群： 海量规模中文网络信息检索评测语料库的设计与实现

赵 军： 信息检索和情感分析评测情况介绍

周 强： 中文信息学会句法分析评测介绍

邹嘉彦： 香港城大泛华语共时同题语料库介绍

易绵竹： 基于本体语义学的语言资源观

自由发言： 萧国政

刘 群： 中文语言资源建设与评测的进展与展望

陶建华： 中文语言资源联盟进展

何婷婷： 国家语言资源监测语料库介绍

俞士汶： 北京大学综合语言知识库

刘奕群： 海量规模中文网络信息检索评测语料库的设计与实现

赵 军： 信息检索和情感分析评测情况介绍

周 强： 中文信息学会句法分析评测介绍

邹嘉彦： 香港城大泛华语共时同题语料库介绍

易绵竹： 基于本体语义学的语言资源观

自由发言： 萧国政

刘 群: 中文语言资源建设与评测的进展与展望

陶建华: 中文语言资源联盟进展

何婷婷: 国家语言资源监测语料库介绍

俞士汶: 北京大学综合语言知识库

刘奕群: 海量规模中文网络信息检索评测语料库的设计与实现

赵 军: 信息检索和情感分析评测情况介绍

周 强: 中文信息学会句法分析评测介绍

邹嘉彦: 香港城大泛华语共时同题语料库介绍

易绵竹: 基于本体语义学的语言资源观

自由发言: 萧国政

成果

- 语言资源的建设已经取得一些成果
- 评测推动了资源的建设和技术的发展
 - 与国际评测接轨，推动了国内研究的进展

问题

- 资源建设：
 - 具有汉语特色的工作还不多，以语言学理论为指导的大规模语料库建设仍然比较少见，影响也不大
 - 语言资源建设与算法模型研究仍然处于割裂状态，语言学家和计算机科学家的交流太难、太少
 - 系统性
 - 面向应用、或需求拉动：整合、建立资源

问题:

- 评测方面

- 还在学术层面

- 技术研讨仍然不够深入，还很少有从国内评测中诞生的新思想和新方法

- 还未从适合国情的应用中提炼评测

- 需求拉动、应用背景、学术意义并重

一些想法

- 资源的建设和应用跟踪
- 资源建设规划
- 评测中提炼
- 评测
 - 当数据
- 精雕
 - 精
 - 技术细节

专家、学者在参与政府、国家的科学战略规划，以学会的身份来号召、建议、组织相应的研究立项

名称实体
IC: 实体、实体的关系、事件
TDT: 话题

数

谢谢!