

gitbook.cn

GitChat - 《自然语言处理研究报告》的文章

63 ~ 79 分

3.1 国外实验室及人才介绍

AMiner 基于发表于国际期刊会议的学术论文，对自然语言处理领域全球 top1000 的学者进行计算分析，绘制了该领域学者全球分布地图。

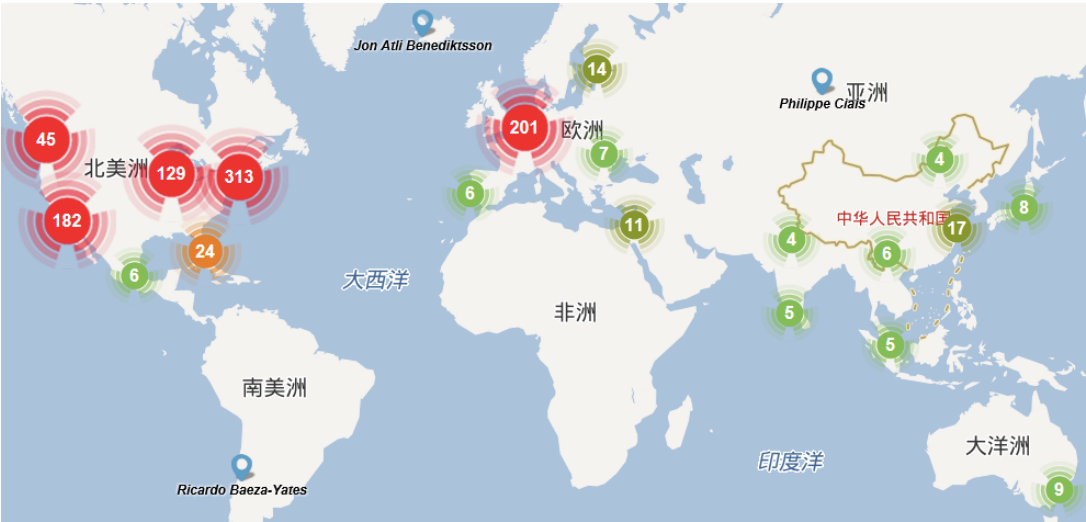


图 5 自然语言处理人才全球分布图

根据上图，我们可以得出以下结论——从国家来看，美国是自然语言处理研究学者聚集最多的国家，英国、德国、加拿大和意大利紧随其后；从地区来看，美国东部是自然语言处理人才的集中地，

而西欧、美国西部等其他先进地区也吸引了大量自然语言处理的研究者。

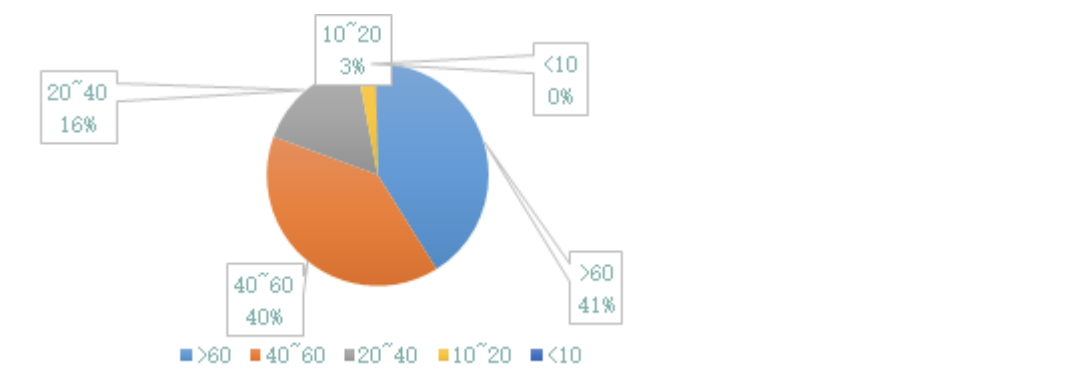


图 6 自然语言处理学者 h-index 统计

全球自然语言处理学者的 h-index 平均数为59， h-index 指数大于60的学者最多占41%， h-index 指数在40到60之间的学者次之， 占比40%。仅从 h-index 指数来说， 全球自然语言处理学者的学术水平还是较高的。

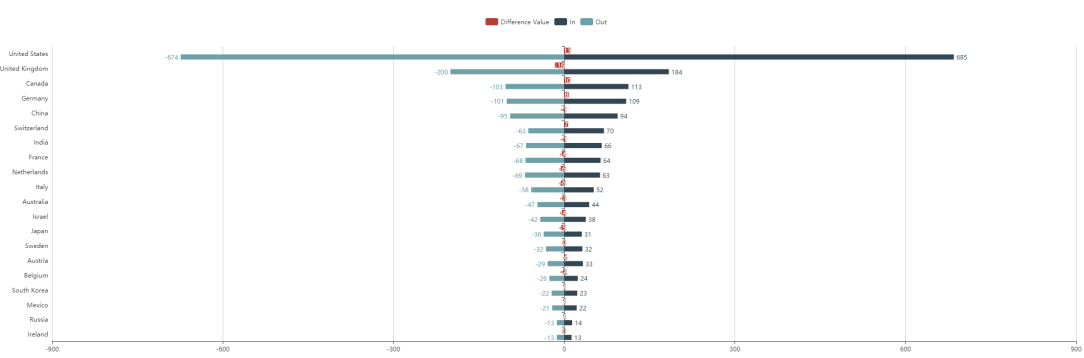


图 7 自然语言处理各国人才顺逆差

AMiner 选取自然语言处理领域影响力排名前1000的专家学者， 对其迁徙路径做了分析。由上图可以看出， 各国自然语言处理领域人才的流失和引进是相对比较均衡的， 其中美国是自然语言处理领域人才流动大国， 人才输入和输出幅度都大幅度领先，

且从数据来看人才流入略大于流出。英国、德国、加拿大和中国等国落后于美国，其中英国和加拿大有轻微的人才流失现象。

以下对国外知名自然语言处理实验室进行介绍：

• Natural Language Processing research at Columbia University

哥伦比亚大学自然语言处理研究室在计算机科学系、计算学习系统中心和生物医学信息系的支持下进行的，将语言洞察力与严谨前沿的机器学习方法和其他计算方法结合起来进行研究。在语言资源创造如语料库、词典等，阿拉伯语 NLP，语言和社交网络，机器翻译，信息提取，数据挖掘，词汇语义、词义消除歧义等方面有着比较深入的研究。

现在该实验室计算机方面的主要负责人为 Michael Collins。



Michael Collins

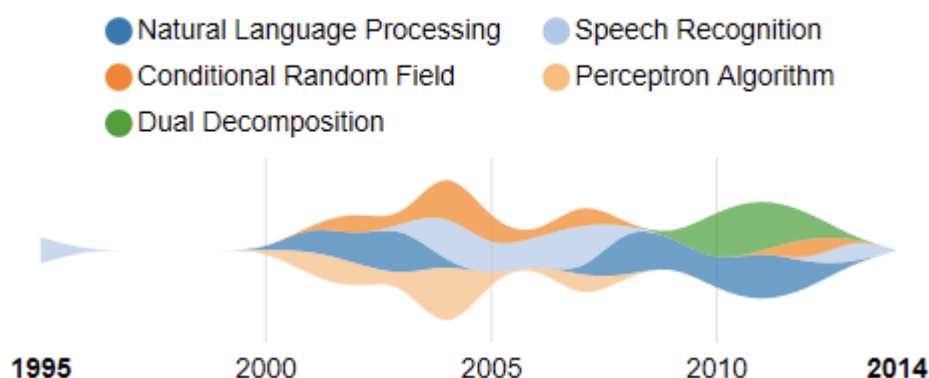
H 55 A 0.17 S 0 c 22976 P 97

Professor

Computer Science, Columbia University

Natural Language Processing Speech Recognition Conditional Random Field Perceptron Algorithm
Dual Decomposition Language Model Exact Solution Lagrangian Relaxation

Research Interests



Michael Collins, 哥伦比亚大学计算机科学系教授, 谷歌 NYC 研究科学家。1998年在宾夕法尼亚大学获得计算机科学博士学位。主要研究兴趣是自然语言处理和机器翻译。多次在国际顶级会议上发表文章, 例如在 EMNLP 2010, CoNLL 2008, UAL 2055等会议上都获得最佳论文奖, 同时还是 ACL 的研究员, 获 NSF 生涯奖。

- **The Stanford Natural Language Processing Group**

斯坦福大学自然语言处理小组包括了语言学和计算机科学系的成员, 是斯坦福人工智能实验室的一部分。主要研究计算机处理和理解人类语言的算法, 工作范围从计算语言学的基本研究到语言处理的关键应用技术均有涉猎, 涵盖句子理解、自动问答、机器翻译、语法解析和标签、情绪分析和模型的文本和视觉场景等。该小组的一个显著特征是将复杂和深入的语言建模和数据分析与 NLP 的创新概率、机器学习和深度学习方法有效地结合在一起。

- **The Berkeley NLP Group**

伯克利大学自然语言处理小组分属于加州大学伯克利分校计算机科学部。主要从事以下几方面的研究工作, 语言分析、机器翻译、计算机语言学、基于语义的方法、无监督学习等, 多次在顶级国际会议 (ACL、EMNLP、AAAI、IJCAI、COLING 等) 上

发表多篇论文，下表是2018年最新被选用的论文。

Constituency Parsing with a Self-Attentive Encoder

作者：Nikita Kitaev、Dan Klein

收录会议：ACL 2018

Policy Gradient as a Proxy for Dynamic Oracles in Constituency Parsing

作者：Daniel Fried、Dan Klein

收录会议：ACL 2018

Learning with Latent Language

作者：Jacob Dan Klein、Sergey Levine

收录会议：NAACL 2018

Unified Pragmatic Models for Generating and Following Instructions

作者：Daniel Fried、Jacob Andreas and Dan Klein

收录会议：NAACL 2018

What's Going On in Neural Constituency Parsers? An Analysis

作者：David Gaddy, Mitchell Stern and Dan Klein

收录会议：NAACL 2018

现在该小组主要负责人为 Dan Klein。

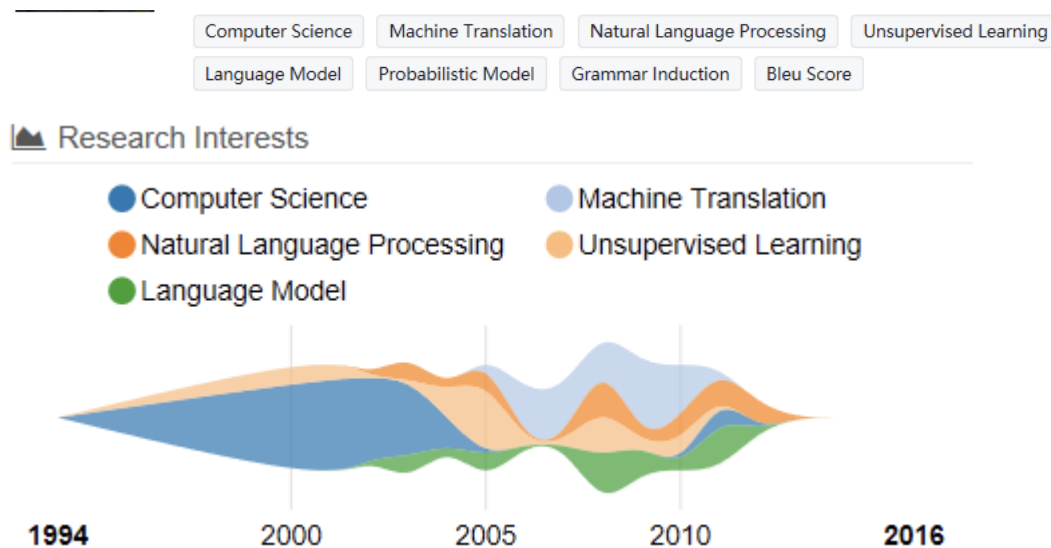


Dan Klein

H 56 A 24.93 S 0 c 17326 P 161

Associate Professor

UC Berkeley Berkeley Engineering



Dan Klein, 伯克利大学自然语言处理小组负责人。2004年在斯坦福大学取得计算机科学的博士学位。主要研究重点是自然语言信息的自组织, 兴趣领域包括无监督的语言学习、机器翻译、NLP 的高效算法、信息提取、语言丰富的语言模型、NLP 的符号和统计方法的集成以及历史语言学等。多次在国际顶级会议上发表论文并获奖, 如在2012年 EMNLP 上获得 Distinguished Paper "Training Factored PCFGs with Expectation Propagation"。

- **Natural Language Processing Group at University of Notre Dame**

圣母大学自然语言处理小组主要关注机器翻译领域, 并有多多个项目的研究, 如由 DARPA LORELEI 和 Google 赞助的无监督多语言学习模型和算法研究; 由亚马逊学术研究奖和谷歌教师研究奖赞助的研究, 主要研究课题为基于神经网络的机器翻译模型, 以及使用神经网络进行翻译和语言建模的算法

等。多次在国际顶级期刊和会议上发表论文，最新的论文有：

Leveraging translations for speech transcription in low-resource settings

作者：Antonis Anastasopoulos, Davis Chiang

收录会议：INTERSPEECH. 2018

Part-of-speech tagging on an endangered language: a parallel Griko-Italian resource

作者：Antonios Anastasopoulos, Justin

DeBenedetto, and David Chiang.

收录会议：COLING 2018

Composing finite state transducers on GPUs

作者：Arturo Argueta and David Chiang.

收录会议：ACL 2018

Algorithms and training for weighted multiset automata and regular expressions

作者：Justin DeBenedetto and David Chiang.

收录会议：CIAA 2018

Synchronous hyperedge replacement graph grammars

作者：Corey Pennycuff, Satyaki Sikdar, Catalina

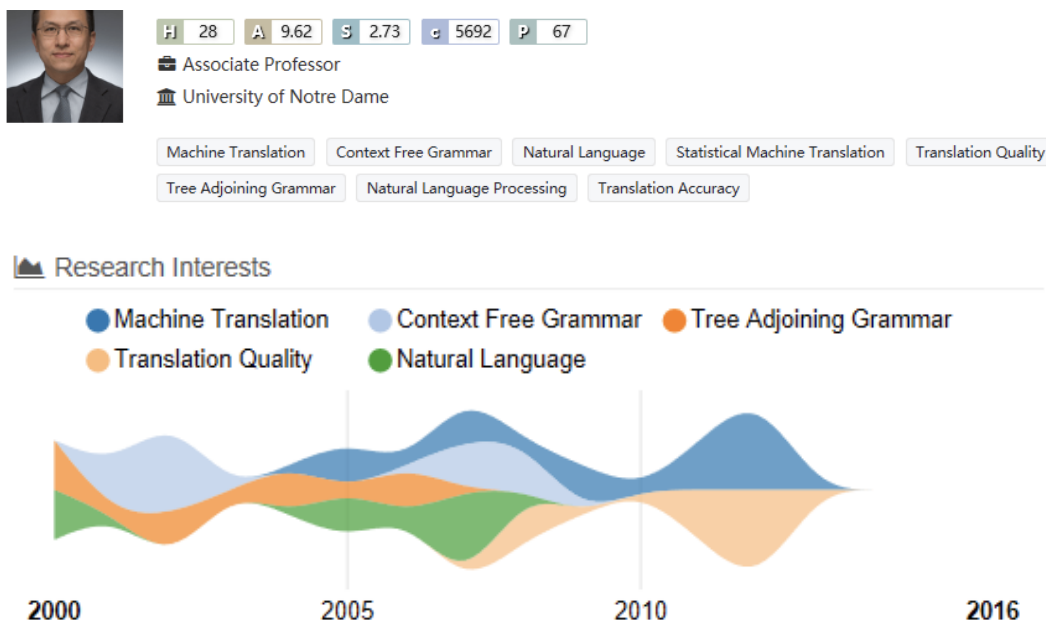
Vajiac, David Chiang, and Tim Weninger.

收录会议ICGT 2018

目前该小组主要负责人是 David Chiang。



David Chiang (蔣偉)



David Chiang, 美国圣母大学教授, 在宾夕法尼亚大学计算机与信息科学获得博士学位。主要研究领域是自然语言处理, 同时在语言翻译、句法分析等方面也有研究。David Chiang 在2005年提出的基于短语的翻译模型, 对机器翻译来说是一个巨大的进步, 他把机器翻译从平面结构建模引向了层次结构建模。

- **The Harvard natural-language processing group**

哈佛自然语言处理小组主要通过机器学习的方法处理人类语言, 主要兴趣集中在数列生成的数学模型, 以人类语言为基础的人工智能挑战以及用统计工具对语言结构进行探索等方面。该小组的研究出版物和开源项目集中在文本总结、神经机器翻译、反复神经网络的可视化、收缩神经网络的算法、文档中实体跟踪的模型、多模态文本生成、语法错误

修正和文本生成的新方法等方面。

Seq2Seq-Vis: A Visual Debugging Tool for Sequence-to-Sequence Models

作者：Hendrik Strobelt, Sebastian Gehrmann, Michael Behrisch, Adam Perer, Hanspeter Pfister, Alexander M. Rush

Semi-Amortized Variational Autoencoders

作者：Yoon Kim, Sam Wiseman, Andrew C. Miller, David Sontag, Alexander M. Rush

收录会议：ICML 2018

Compressing Deep Neural Networks with Probabilistic Data Structures

作者：Arturo Argueta and David Chiang.

收录会议：ACL 2018

Algorithms and training for weighted multiset automata and regular expressions

作者：Brandon Reagen, Udit Gupta, Robert Adolf, Michael M. Mitzenmacher, Alexander M. Rush, Gu-Yeon Wei, David Brooks

收录会议：SysML 2018

Stuart .Shieber 是该小组的主要负责人。



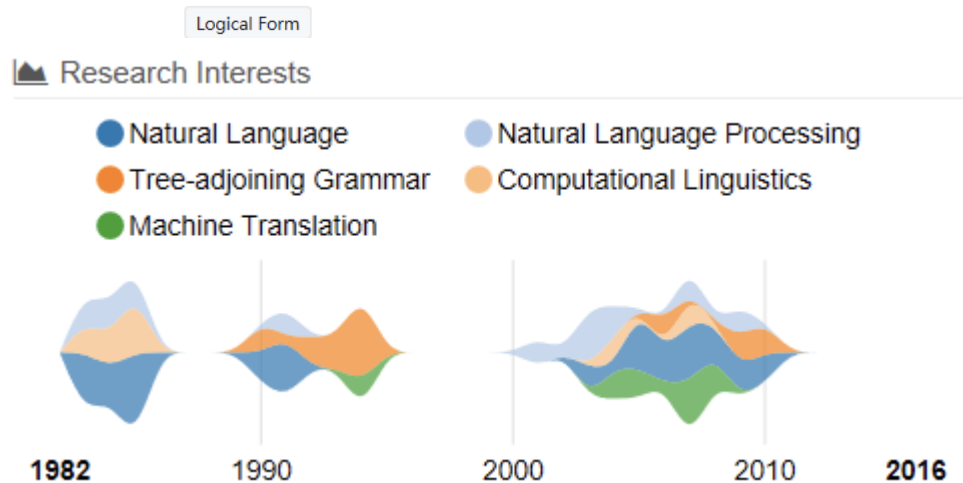
Stuart M. Shieber

H 48 A 20.94 S 19.68 c 11682 P 190

Professor Director

Maxwell-Dworkin Laboratory Center for Research on Computation and Society

Natural Language Natural Language Processing Tree-adjoining Grammar
Computational Linguistics Machine Translation Context Free Grammar Heuristic Search



Stuart Shieber, 美国计算机协会 (Association for Computing Machinery) fellow 和美国人工智能协会 (American Association for Artificial Intelligence) fellow。他综合语言学、理论计算机科学、计算机系统以及人工智能等领域的知识, 研究计算机语言学, 从计算机科学的角度研究自然语言, 在该领域的研究以科学和工程目标, 以基础形式和数学工具为基础。具体研究领域包括计算语言学、数学语言学、基于语法的形式、自然语言生成、计算语义、机器翻译以及人机交互等。

- **The Language Technologies Institute (LTI) at Carnegie Mellon University**

卡内基梅隆大学语言技术研究所主要研究内容包括自然语言处理、计算语言学、信息提取、信息检索、文本挖掘分析、知识表示、机器学习、机器翻译、多通道计算和交互、语音处理、语音界面和对话处理等。

3.2 国内实验室及人才介绍

AMiner 基于论文数据抓取了367位自然语言处理领域的华人专家，其中包括了 NUS, HKUS, THU, PKU, FDU 等知名高校的实验室以及百度、科大讯飞、微软等公司及研究院的专家和学者。下面基于自然语言处理华人库中的数据对其进行分析。



图 8 AMiner 自然语言处理华人库专家全球分布



图 9 AMiner 自然语言处理华人库专家国内分布

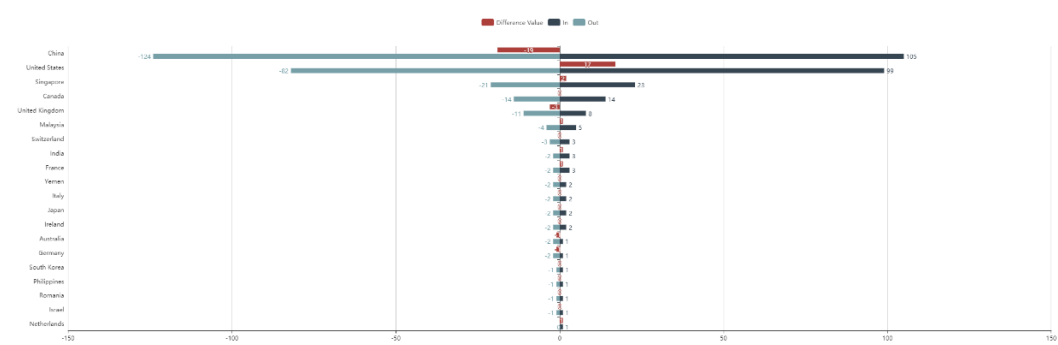


图 10 AMiner 自然语言处理华人库专家地区统计

自然语言处理领域中华人专家在中国最多，美国次之。从地区来看，中国大陆是自然语言处理华人人才的最主要聚集地，尤其是北京、哈尔滨及东南沿海地区等具有自然语言处理学术基础的地区。美国东部和西部等其他地区排在其后。由图11可以看出，华人专家在中国流出量大于流入量，美国则正好相反，这也说明就自然领域而言，中国对人才的吸引力要小于美国。

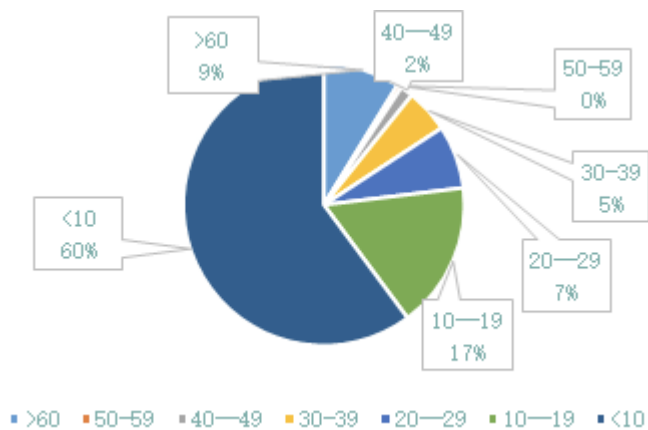


图 11 AMiner 自然语言处理华人库专家流动图





图 12 AMiner 自然语言处理华人库专家 h-index 统计

AMiner 自然语言处理华人库中专家 h-index 指数的平均数为14，这一数值是远远低于自然语言处理全球 top1000 学者 h-index 指数平均数的。而且，在华人库中，h-index 指数<10的专家人数最多，占比60%；10-19次之，占比17%；>60的专家占比仅占9%。这也说明，自然语言处理华人专家整体水平低于自然语言处理领域全球 top1000 的学者，尤其是在 h-index 指数>60的学者方面有所欠缺。

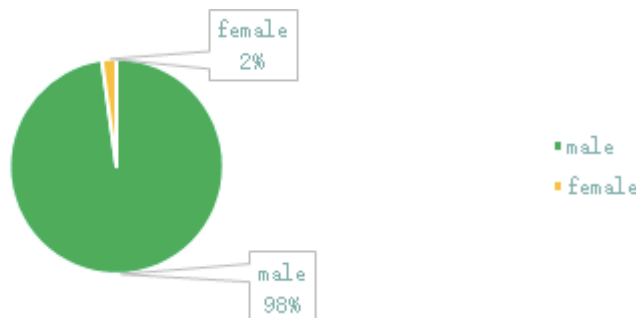


图 13 自然语言处理华人库男女比

AMiner 自然语言处理华人库367位专家中，男性专家占98%，女性专家仅占2%，二者比例约为49:1。

国内从事自然语言处理的学者人数众多，AMiner 基于发表于国际期刊会议的学术论文，对自然语言处理领域内专家进行深入挖掘，并按照相关度和影响力等对专家进行排序和分类，排序和分类规则主要参考专家的 h-index、paper、citation、专家所获得

的荣誉、任职机构排名、专家 Activity、Sociability 以及 Diversity 等，得到结果如下：

有来自清华大学的孙茂松、朱小燕、马少平、李涓子等；来自北京大学的王厚峰、李素建、穗志方、万小军等；哈工大赵铁军、刘挺等；中国科学院刘群、宗成庆、赵军、孙乐等；复旦大学黄萱菁等人；苏州大学张民、周国栋等；东北大学朱靖波等。

以下是 AMiner 挖掘出国内自然语言处理重要实验室及其主要负责人简介。

• 清华大学自然语言处理与社会人文计算实验室

清华大学计算机系自然语言处理课题组在20世纪70年代末，就在黄昌宁教授的带领下从事这方面的研究工作，是国内开展相关研究最早、深具影响力的科研单位，同时也是中国中文信息学会计算语言学专业委员会的挂靠单位。现任学科带头人孙茂松教授任该专业委员会的主任（同时任中国中文信息学会副理事长），其余教师还有刘洋、刘知远等人。目前该课题组对以中文为核心的自然语言处理中的若干前沿课题，进行系统、深入的研究，研究领域的涵盖面正逐步从计算语言学的核心问题扩展到社会计算和人文计算。

该课题组多篇论文被 ACL 2018、IJCAI-ECAI 2018、WWW 2018 录用，内容涉及问答系统、信

息检索、机器翻译、诗歌生成、查询推荐等多个领域。具体见下表：

Denoising Distantly Supervised Open-Domain Question Answering

作者：林衍凯、计昊哲、刘知远、孙茂松

收录会议：ACL 2018

Incorporating Chinese Characters of Words for Lexical Sememe Prediction

作者：金晖明*、朱昊*、刘知远、谢若冰、孙茂松、林芬、林乐宇（*同等贡献，本篇文章与腾讯微信合作）

收录会议：ACL 2018

Entity-Duet Neural Ranking: Understanding the Role of Knowledge Graph Semantics in Neural Information Retrieval

作者：刘正皓、熊辰炎、孙茂松、刘知远（本篇文章与 CMU 熊辰炎合作）

收录会议：ACL 2018

Towards Robust Neural Machine Translation

作者：程勇、涂兆鹏、孟凡东、翟俊杰、刘洋（本篇文章与腾讯 AI 实验室合作）

收录会议：ACL 2018

Chinese Poetry Generation with a Working Memory Model

作者：矣晓沅、孙茂松、李若愚、杨宗瀚

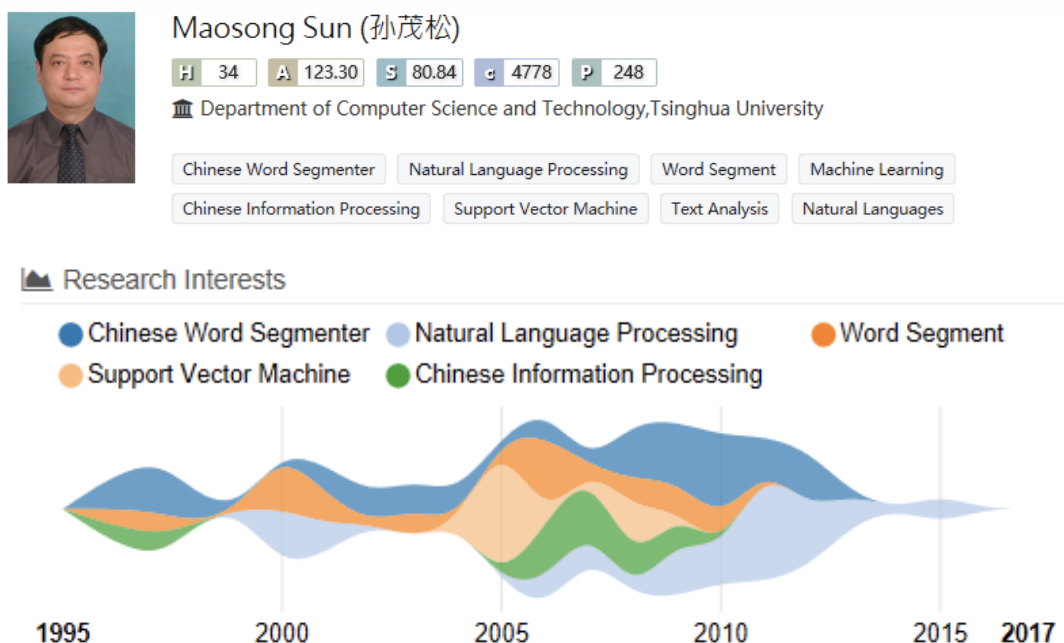
收录会议：IJCAI 2018

Query Suggestion with Feedback Memory Network

作者：武彬、熊辰炎、孙茂松、刘知远（本篇文章与 CMU 熊辰炎合作）

收录会议：WWW 2018

该实验室主要负责人为孙茂松。



孙茂松，清华大学计算机科学与技术系教授。

2007-2010年任该系系主任，主要研究领域为自然语言处理、互联网智能、机器学习、社会计算和计算教育学。国家重点基础研究发展计划（973计划）项目首席科学家，国家社会科学基金重大项目首席专家。

在国际刊物、国际会议、国内核心刊物上发表论文160余篇，主持完成文本信息处理领域 ISO 国际标

准2项。2007年获全国语言文字先进工作者，2016年获全国优秀科技工作者以及首都市民学习之星。多次担任相关领域国际会议和全国性学术会议大会主席或程序委员会主席。

• 清华大学智能技术与系统国家重点实验室

智能技术与系统国家重点实验室依托在清华大学，1987年7月开始筹建。1990年2月通过国家验收，并正式对外开放运行。从1990年至2003年这十三年间，实验室顺利通过国家自然科学基金委受科技部委托组织的全部三次专家组评估，并被评估为A（优秀实验室）。1994年10月在庆祝国家重点实验室建设十周年表彰大会上，智能技术与系统国家重点实验室获集体“金牛奖”。1997年被科技部列为试点实验室。2004年庆祝国家重点实验室建设二十周年表彰大会上，本实验室再次荣获集体“金牛奖”。从2004年开始，实验室参与筹建清华信息科学与技术国家实验室。实验室学术委员会由17名国内外著名专家组成。实验室学术委员会名誉主任为中科院院士张钹教授，主任为应明生教授、副主任为邓志东教授。

现任实验室主要负责人为朱小燕。



Xiaoyan Zhu (朱小燕)

H 36 A 136.20 S 31.09 d 5859 P 323

Professor

Department of Computer Science and Technology, Tsinghua University

Speech Recognition

Privacy

Data Privacy

Mobile Computing

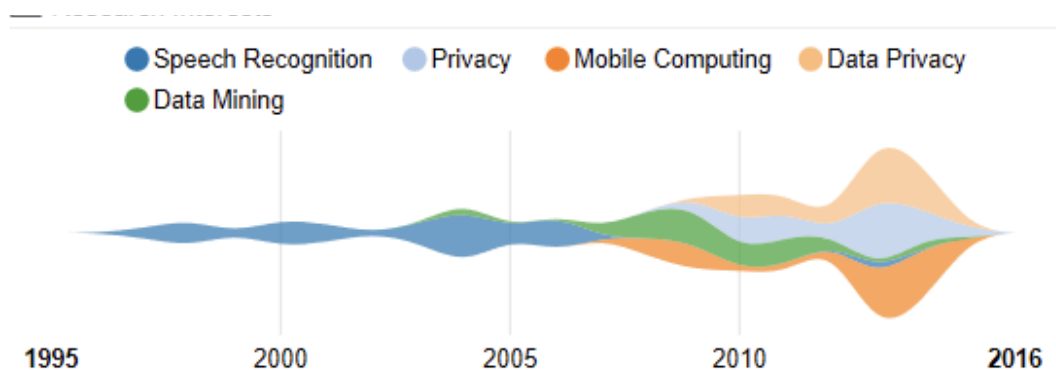
Data Mining

Protein-protein Interaction

Ad Hoc Network

Information Retrieval

Research Interests



朱小燕，清华计算机系教授，博士生导师，智能技术与系统国家重点实验室主要负责人。

主要研究领域为智能信息处理，其中包括：模式识别、神经网络、机器学习、自然语言处理、信息提取和智能问答系统等。近年研究工作主要集中于生物领域文本信息处理和新一代智能信息获取的研究。作为项目负责人先后承担国家863，973项目，自然科学基金项目、国际合作项目多项。1997年获国家教委科技进步二等奖，2003年获北京市科技进步二等奖。获得国家发明专利3项。在各种国际刊物和会议上发表论文近100篇。其中包括国际刊物 Genome Biology, Bioinformatics、BMC Bioinformatics、Medical informatics、IEEE Transactions on SMC、IEE Electronics Letters、Neural Parallel & Science Computations、Document Analysis and Recognition，以及国际会议 SIG KDD、ACL、COLING、CIKM 等。

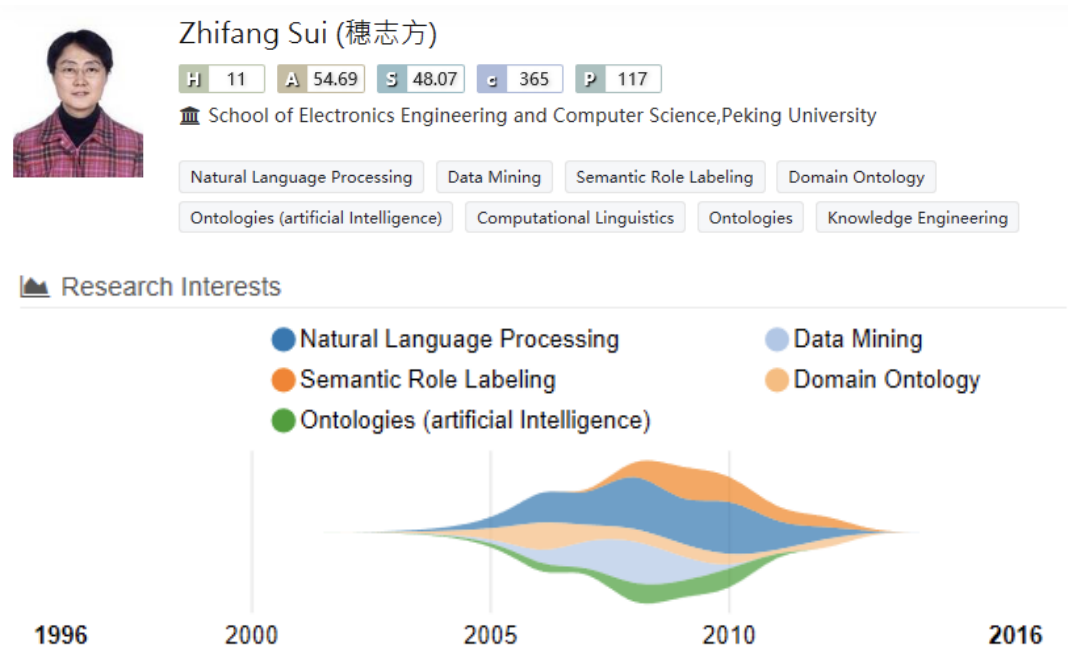
- **北京大学计算语言学教育部重点实验室**

计算语言学教育部重点实验室依托北京大学建设。

实验室研究人员由北京大学信息科学技术学院计算语言学研究所、中文系、软件与微电子学院语言信息工程系、计算机技术研究所、心理系和外语学院的相关研究人员构成。

主要研究方向包括：中文计算的基础理论与模型；大规模多层次语言知识库构建的方法；国家语言资源整理与语音数据库建设；海量文本内容分析与动态监控；多语言信息处理和机器翻译。

该实验室主要负责人为穗志方。



穗志方，北京大学信息科学技术学院计算语言学实验室主任，教授、博士生导师。2011年度国家科技进步二等奖“综合型语言知识库”项目第二完成人。长期从事自然语言处理方面的研究。

在计算语言学国际顶级会议 ACL 2000、COLING 2008、CONLL 2008、ACL 2009、EMNLP 2009、AIRS 2008 上发表多篇学术论文。作为课题负责人

主持的科研项目有：国家自然科学基金项目“汉语动词子语类框架自动获取技术研究”、“基于结构化学学习的语义角色标注研究”、“基于 Web 的概念实例及其属性值提取方法研究”，国家社科基金项目“面向文本内容提取的生成性组件库研究及建设”等。

• 北京大学语言计算与互联网挖掘研究组

语言计算与互联网挖掘研究室从属于北京大学计算机科学技术研究所，成立于2008年7月，负责人为万小军老师。研究室以自然语言处理技术、数据挖掘技术与机器学习技术为基础，对互联网上多源异质的文本大数据进行智能分析与深度挖掘，为互联网搜索、舆情与情报分析、写稿与对话机器人等系统提供关键技术支撑，并从事计算机科学与人文社会科学的交叉科学研究。

研究室当前研究内容包括：1) 语义理解：研制全新的语义分析系统实现对人类语言（尤其是汉语）的深层语义理解；2) 机器写作：综合利用自动文摘与自然语言生成等技术让机器写出高质量各类稿件；3) 情感计算：针对多语言互联网文本实现高精度情感、立场与幽默分析；4) 其他：包括特定情境下的人机对话技术等。

该研究室主要负责人为万小军。

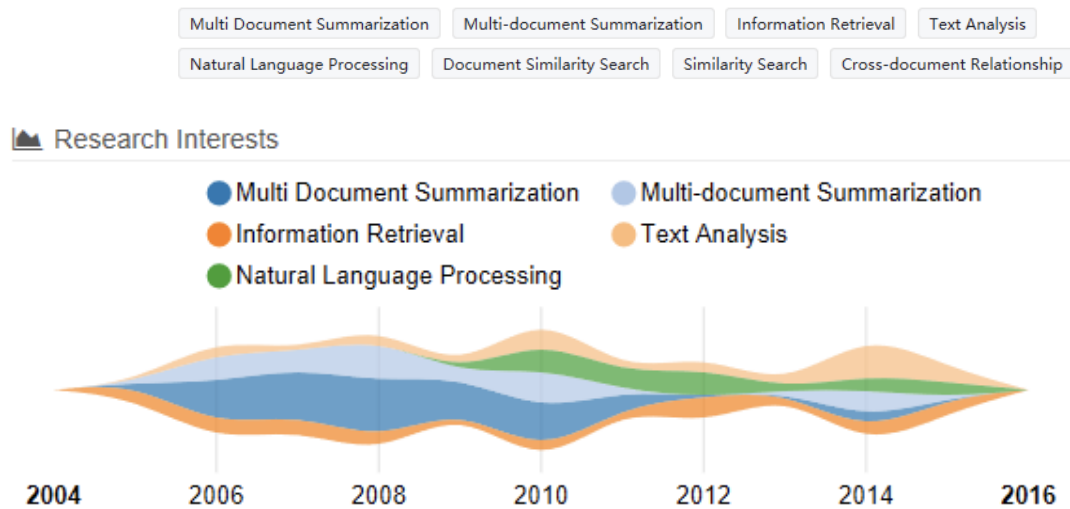


Xiaojun Wan (万小军)

H 27 A 102.43 S 45.33 c 4105 P 202

Professor

Peking University, Institute of Computer Science and Technology



万小军，北京大学计算机科学技术研究所教授，博士生导师，语言计算与互联网挖掘实验室负责人。研究方向为自然语言处理与文本挖掘，兴趣领域包括自动文摘与文本生成、情感分析与观点挖掘、语义计算与信息推荐等，在国际重要学术会议与期刊上发表高水平学术论文上百篇。

担任计算语言学顶级国际期刊 Computational Linguistics 编委，TACL 常务评审委员（Standing Reviewing Committee），多次担任自然语言处理领域重要国际会议领域主席或 SPC（包括 ACL、NAACL、IJCAI、IJCNLP 等），以及相关领域多个国际顶级学术会议（ACL、SIGIR、CIKM、EMNLP、NAACL、WWW、AAAI 等）程序委员会委员。研制了自动文摘开源平台 PKUSUMSUM，与今日头条合作推出 AI 写稿机器人小明（Xiaomingbot），与南方都市报合作推出写稿机器人小南等应用系统。

- 复旦自然语言处理研究组

复旦大学自然语言与信息检索实验室，致力于社会媒体海量多媒体信息处理的前沿技术研究。主要研究方向包括：自然语言处理、非规范化文本分析、语义计算、信息抽取、倾向性分析、文本挖掘等方面。实验室开发了 NLP 工具包

FudanNLP，FudanNLP 提供了一系列新技术，包括中文分词、词性标注、依赖解析、时间表达式识别和规范化等。

实验室先后承担和参与了国家科技重大专项、国家 973 计划、863 计划、国家自然科学基金课题、上海市科技攻关计划等，并与国内外多所重点大学、公司保持着良好的合作关系。研究成果持续发表在国际权威期刊和一流国际会议（TPAMI、TKDE、ICML、ACL、AAAI、IJCAI、SIGIR、CIKM、EMNLP、COLING 等）。

该实验室主要负责人为黄萱菁。



黄萱菁，复旦大学计算机科学技术学院教授、博士生导师。在 SIGIR、ACL、ICML、IJCAI、AAAI、NIPS、CIKM、ISWC、EMNLP、WSDM 和 COLING 等多个国际学术会议上发表论文数十篇。

曾任2014年 CIKM 会议竞赛主席，2015年 WSDM 会议组织者，2015年全国社会媒体处理大会程序委员会主席，2016年全国计算语言学会议程序委员会副主席，2017年自然语言处理与中文计算国际会议程序委员会主席。

多次在人工智能、自然语言处理和信息检索的国际学术会议 IJCAI、ACL、SIGIR、WWW、EMNLP、COLING、CIKM、WSDM 担任程序委员会委员和资深委员。兼任中国中文信息学会常务理事，社会媒体专委会副主任，中国计算机学会中文信息处理专委会委员，中国人工智能学会自然语言理解专委会委员，ACM 和 ACL 会员，《中文信息学报》编委，国家自然科学基金、教育部高校博士点基金和863计划同行评议专家。

• 哈工大社会计算与信息检索研究中心

哈工大社会计算与信息检索研究中心（HIT-SCIR）成立于2000年9月，隶属于计算机科学与技术学院。研究中心成员有主任刘挺教授，副主任秦兵教授，教师包括张宇、车万翔、陈毅恒、张伟男等。

研究方向包括语言分析、信息抽取、情感分析、问

答系统、社交媒体处理和用户画像6个方面。已完成或正在承担的国家973课题、国家自然科学基金重点项目、国家863重点项目、国际合作、企业合作等课题60余项。在这些项目的支持下打造出“语言技术平台 LTP”，提供给百度、腾讯、华为、金山等企业使用，获2010年钱伟长中文信息处理科学技术一等奖。

研究中心近年来发表论文100余篇，其中在 ACL、SIGIR、IJCAI、EMNLP 等顶级国际学术会议上发表20余篇论文，参加国内外技术评测，并在国际 CoNLL’2009七国语言句法语义分析评测总成绩第一名。研究中心通过与企业合作，已将多项技术嵌入企业产品中，为社会服务。双语例句检索等一批技术嵌入金山词霸产品中，并因此获得2012年黑龙江省技术发明二等奖。

该实验室主要负责人是刘挺。



刘挺，哈尔滨工业大学教授，国家“万人计划”科技创新领军人才。多次担任国家863重点项目总体组专家、基金委会评专家。中国计算机学会理事，中国中文信息学会常务理事/社交媒体处理专委会（SMP）主任，曾任国际顶级会议 ACL、EMNLP 领域主席。

主要研究方向为人工智能、自然语言处理和社会计算，是国家973课题、国家自然科学基金重点项目负责人。2012-2017年在自然语言处理领域顶级会议发表的论文数量列世界第8位（据剑桥大学统计），主持研制“语言技术平台 LTP”、“大词林”等科研成果被业界广泛使用。曾获国家科技进步二等奖、省科技进步一等奖、钱伟长中文信息处理科学技术一等奖等。

• 苏州大学人类语言技术研究所

苏州大学人类语言技术研究所成立于2013年9月，隶属于苏州大学计算机科学与技术学院，同时也是自然语言处理课题组的一部分，主要从事的研究方向包括自然语言分析、知识图谱构建、机器翻译等，所长张民教授。研究所成立以来在顶级国际会议（ACL、EMNLP、AAAI、IJCAI、COLING）上发表多篇论文。下表是2018年最新被录用的论文。

*Improved English to Russian Translation by
Neural Suffix Prediction*

作者：Kai Song, Yue Zhang, Min Zhang, Weihua Luo

收录会议：AAAI 2018

Adversarial Learning for Chinese NER from Crowd Annotations

作者：YaoSheng Yang, Meishan Zhang, Wenliang Chen*, Haofen Wang, Wei Zhang, Min Zhang

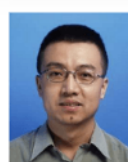
收录会议：AAAI 2018

SEE: Syntax-aware Entity Embedding for Neural Relation Extraction

作者：Zhengqiu He, Wenliang Chen*, Zhenghua Li, Meishan Zhang, Wei Zhang, Min Zhanga

收录会议：AAAI 2018

该实验室主要负责人为张民。



Min Zhang (张民)

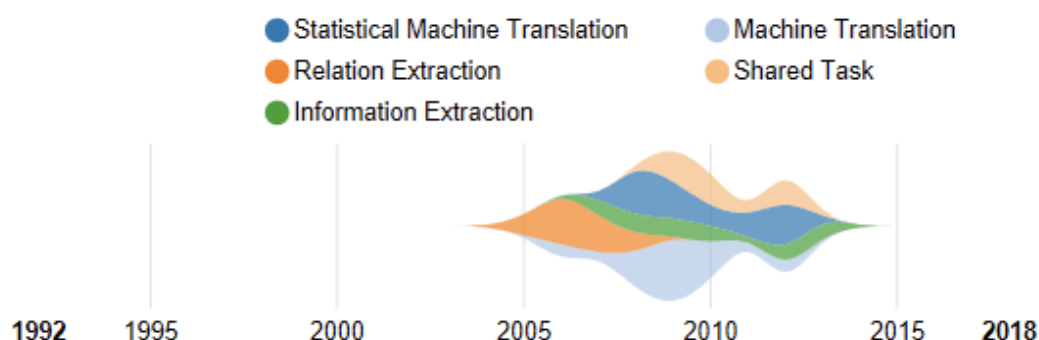
H 31 A 28.05 S 12.68 C 3942 P 137

Professor

苏州大学计算机科学与技术学院

Statistical Machine Translation Machine Translation Relation Extraction Information Extraction Shared Task
Machine Transliteration Common Benchmarking Platform Language Model

Research Interests



张民，苏州大学计算机科学与技术学院副院长。

2003年12月，他加入新加坡信息通信研究所并于2007年在研究所建立了统计机器翻译团队。2012年加入苏州大学，并于2013年在该大学成立智能计算研究所。

目前的研究兴趣包括机器翻译、自然语言处理、信息提取、社交网络计算、互联网智能、智能计算和机器学习。近年来在国际顶级学报和顶级会议发表学术论文150余篇，Springer 出版英文专著两部，主编 Springer 和 IEEE CPS 出版英文书籍十本。他一直积极地为研究界做贡献，组织多会议并在许多会议和讲座中进行演讲。

• 苏州大学自然语言处理实验室

苏州大学自然语言处理实验室2006年成立，隶属于江苏省计算机信息处理技术重点实验室，致力于基础研究，推动计算机处理和理解人类语言的水平。工作范围从计算语言学的基础研究到人类语言技术的关键应用，涵盖了词汇分析、句法分析、浅层语义分析、深层语用理解、信息提取、情绪分析、统计机器翻译等领域。

该实验室主要负责人为周国栋。



Guodong Zhou (周国栋)

H 39 **A** 60.64 **S** 22.91 **c** 6740 **P** 411

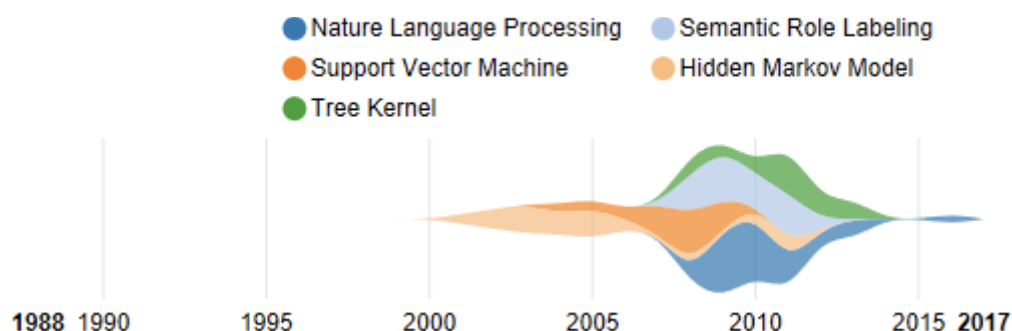
Professor

Natural Language Processing Lab, Soochow University

Nature Language Processing Semantic Role Labeling Support Vector Machine Hidden Markov Model

Coreference Resolution Tree Kernel Chinese Information Processing Sentiment Classification

 Research Interests



周国栋，苏州大学计算机科学与技术学院教授，苏州大学 NLP 实验室负责人。主要研究兴趣是自然语言处理、中文计算、信息抽取等。自1999年起，一直是 ACM、ACL、IEEE computer society 的会员，负责了多项国家863项目、国家重点研究项目等。

• 中科院计算所自然语言处理研究组

自然语言处理研究组隶属于中国科学院计算技术研究所智能信息处理重点实验室。研究组教师有刘群、冯洋等人。研究组主要从事自然语言处理和机器翻译相关的研究工作，研究方向包括机器翻译、人机对话、多语言词法分析、句法分析和网络信息挖掘等。研究组已完成和正在承担的国家自然科学基金、863计划、科技支撑计划、国际合作等课题40余项，在自然语言处理和机器翻译领域取得了多项创新性研究成果。研究组自2004年重点开展统计机器翻译方面的研究并取得重大突破，并于2015年起转向神经机器翻译并取得很大进展。

在自然语言处理的顶级国际刊物 CL、AI 和顶级国际学术会议 ACL、IJCAI、AAAI、EMNLP、

COLING 上发表高水平论文70余篇，取得发明专利10余项，在国际机器翻译评测中多次名列前茅。研究组已经成功将自主开发的统计机器翻译和神经机器翻译技术推广到汉语、维吾尔语、藏语、蒙古语、英语、韩语、泰语、日语、阿拉伯语等多种语言。部分语种的翻译系统已经在相关领域得到了实际应用，获得用户的好评。

实验室在2017年发表论文见下表。

ME-MD: An Effective Framework for Neural Machine Translation with Multiple Encoders and Decoders

作者：Jinchao Zhang, Qun Liu, Jie Zhou

收录会议：IJCAI 2017

Deep Neural Machine Translation With Linear Associative Unit

作者：Mingxuan Wang, Zhengdong Lu, Jie Zhou, Qun Liu

收录会议：ACI 2017

Incorporating Word Reordering Knowledge into Attention-based Neural Machine Translation

作者：Jinchao Zhang, Mingxuan Wang, Qun Liu, Jie Zhou

收录会议：ACI 2017

Memory-Augmented Neural Machine Translation

作者：Yang Feng, Shiyue Zhang, Andi Zhang,
Dong Wang, Andrew Abel

收录会议：EMNLP 2017

*Further Investigation into Reference Bias in
Monolingual Evaluation of Machine Translation*

作者：Qingsong Ma, Yvette Graham, Timothy
Baldwin, Qun Liu

收录会议：EMNLP 2017

*Blend: a Novel Combined MT Metric Based on
Direct Assessment — — CASICT-DCU submission
to WMT17 Metrics Task*

作者：Qingsong Ma, Yvette Graham, Shugen
Wang, Qun Liu

收录会议：WMT 2017

*CASICT-DCU Neural Machine Translation
Systems for WMT17*

作者：Jinchao Zhang, Peerachet Porkaew, Jiawei
Hu, Qiuye Zhao, Qun Liu

收录会议：WMT 2017

该实验室主要负责人为刘群。



Qun Liu (刘群)

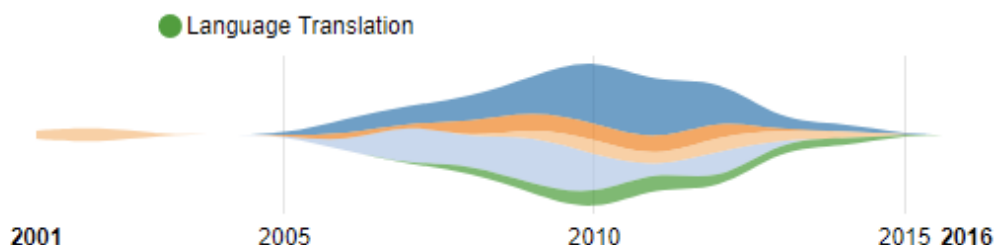
H 33 A 35.49 S 21.13 c 5596 P 194

Institute of Computing Technology, Chinese Academy of Sciences

Statistic Machine Translation Machine Translation Language Model Natural Language Processing
Language Translation Word Segmentation Artificial Intelligence Translation Performance

Research Interests

● Statistical Machine Translation ● Machine Translation
● Language Model ● Natural Language Processing



刘群，中国科学院自然语言处理研究组组长，都柏林大学自然语言处理组组长、项目负责人。主要研究方向是中文自然语言处理，具体包括汉语词法分析、汉语句法分析、语义处理、统计语言模型、辞典和语料库、机器翻译、信息提取、中文信息处理和智能交互中的大规模资源建设、中文信息处理以及智能交互中的评测技术等。曾负责863重点项目“机器翻译新方法的研究”和“面向跨语言搜索的机器翻译关键技术研究”等。

• 中科院中文信息处理研究组

中科院模式识别国家重点实验室自然语言处理组主要成员有宗成庆、赵军、周玉、刘康、张家俊、汪昆、陆征等。该小组主要从事自然语言处理基础、机器翻译、信息抽取和问答系统等相关研究工作，力图在自然语言处理的理论模型和应用系统开发方面做出创新成果。目前研究组的主要方向包括：自然语言处理基础技术（汉语词语切分、句法分析、语义分析和篇章分析等）、多语言机器翻译、信息抽取（实体识别、实体关系抽取、观点挖掘等）和智能问答系统（基于知识库的问答系统、知识推理、社区问答等）。

近年来，研究组注重于自然语言处理基础理论和应用基础的相关研究，承担了一系列包括国家自然科学基金项目、973计划课题、863计划项目和支撑计划项目等在内的基础研究和应用基础研究类项目，以及一批企业应用合作项目。在自然语言处理及相关领域顶级国际期刊（CL、TASLP、TKDE、JMLR、TACL、Information Sciences、Intelligent Systems 等）和学术会议（AAAI、IJCAI、ACL、SIGIR、WWW 等）上发表了一系列论文。2009年获得第23届亚太语言、信息与计算国际会议（PACLIC）最佳论文奖，2012年获得第一届自然语言处理与中文计算会议（NLPCC）最佳论文奖，2014年获得第25届国际计算语言学大会（COLING）最佳论文奖。获得了10余项国家发明专利。

该小组主要负责人为宗成庆。

- 宗成庆



宗成庆，模式识别国家重点实验室研究员、博士生导师。主要从事自然语言处理、机器翻译和文本数据挖掘等相关领域的研究。主持国家自然科学基金项目、863计划项目和重点研发计划重点专项等10余项，发表论文150余篇，出版专著和译著各一部。

2013年当选国际计算语言学委员会（ICCL）委员。目前担任亚洲自然语言处理学会（AFNLP）候任主席、中国中文信息学会副理事长、学术期刊 ACM TALLIP 副主编（Associate Editor）、《自动化学报》副主编、IEEE Intelligent Systems 编委、Machine Translation 编委和 JCST 编委。2013年获国务院颁发的政府特殊津贴，2014年获“钱伟长中文信息处理科学技术奖”一等奖，2015年获国家科技进步奖二等奖，2017年获北京市优秀教师荣誉称号。

• 东北大学自然语言处理实验室

东北大学自然语言处理实验室于1980年由姚天顺教授成立，现在由朱靖波教授负责。主要研究领域包括自然语言处理、文本挖掘、机器翻译等。实验室研究团队在国内外测评中取得不少好成绩，并先后承担和完成了二十多个项目，包括国家自然科学基金资助项目、国家863高技术智能计算机系统主题专家组的资助项目和国家教委博士点基金项目。尤其是其研制的学知名统计机器翻译开源系统

NiuTrans, 已经实现产业化并且在多项测评中取得的好成绩。

以下是该实验室最新论文列表。

Neural Relation Classification with Text Descriptions

作者：Feiliang Ren, Di Zhou, Zhihui Liu, Yongcheng Li, Rongsheng Zhao, Xiaobo Liang, Yongkang Liu.

收录会议：COLING 2018

A Simple and Effective Approach to Coverage-Aware Neural Machine Translation

作者：Yanyang Li, Tong Xiao, Yinqiao Li, Qiang Wang, Changming Xu, Jingbo Zhu

收录会议：ACL 2018

SEE: Syntax-aware Entity Embedding for Neural Relation Extraction

作者：Zhengqiu He, Wenliang Chen*, Zhenghua Li, Meishan Zhang, Wei Zhang, Min Zhanga

收录会议：AAAI 2018

该实验室主要负责人为朱靖波。



Jingbo Zhu (朱靖波)

H 17 A 7.26 S 0.02 C 1549 P 111

Professor

Natural Language Processing Lab, Northeastern University

Natural Language Processing

Machine Translation

Text Classification

Machine Learning

Active Learning

Feature Selection

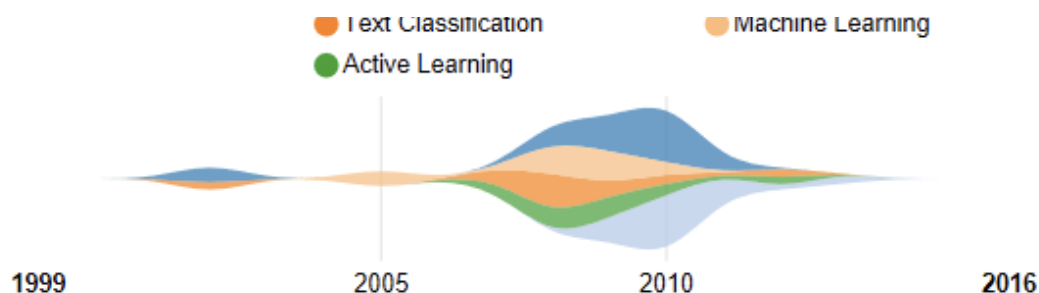
Text Analysis

Learning Artificial Intelligence

Research Interests

Natural Language Processing

Machine Translation



朱靖波，东北大学计算机学院教授、计算机软件研究所副所长、自然语言处理实验室主任、小牛翻译创造者、沈阳雅译公司创始人。至今作为第一作者或通讯作者在国内外期刊杂志和学术会议上共发表200多篇学术论文，包括一篇国际杂志 *Artificial Intelligence* 论文、一篇 *Journal of Artificial Intelligence Research* 论文，八篇 *IEEE/ACM Transactions* 国际杂志论文、十六篇国际自然语言处理三大顶级学术会议（ACL/EMNLP/COLING）论文，并作为副主编出版一本专著——《自然语言理解-一种让机器懂得人类语言的研究》。

作为课题负责人主持过五项国家自然科学基金面上项目、一项国家863计划课题、一项国家教育部科学技术重点项目和一项国家教育部博士点基金课题等。主持研制的学术界知名统计机器翻译开源系统 NiuTrans 和中文句法语义分析系统 NiuParser 等。

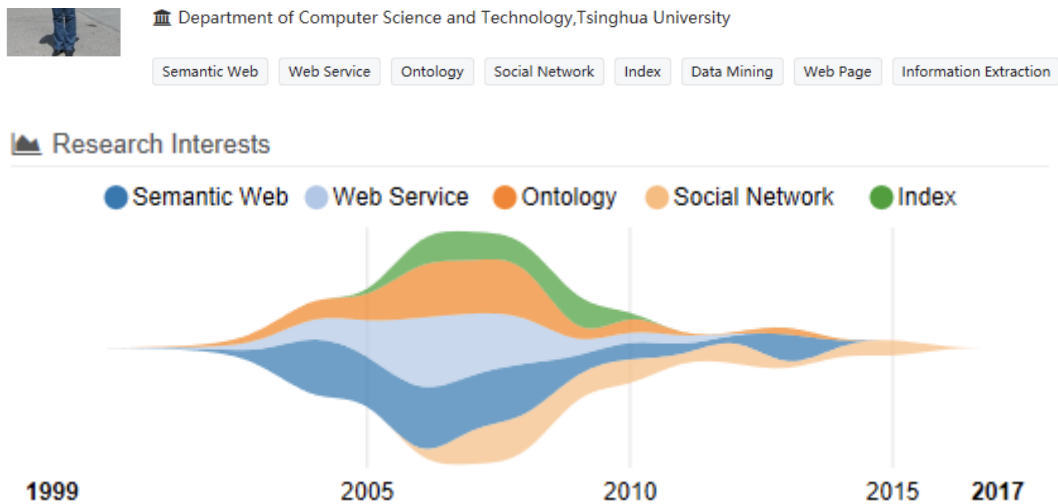
除以上实验室的各位负责人之外，还有以下学者在自然语言处理领域进行着深入的研究。

李涓子



Juanzi Li (李涓子)

H 41 A 86.70 S 34.51 C 6591 P 288
Professor



李涓子，清华大学教授，博士生导师。中国中文信息学会语言与知识计算专委会主任、中国计算机学会术语委员会执行委员。

研究兴趣是语义 Web，新闻挖掘与跨语言知识图谱构建。多篇论文发表在重要国际会议（WWW、IJCAI、SIGIR、SIGKDD）和学术期刊（TKDE、TKDD）。主持多项国家级、部委级和国际合作项目研究,包括国家自然科学基金项目重点，欧盟第七合作框架、新华社等项目。获得2013年人工智能学会科技进步一等奖，2013年电子学会自然科学二等奖。

赵铁军





赵铁军，哈尔滨工业大学计算机科学与技术学院语言技术研究中心教授，哈工大语言语音教育部-微软重点实验室副主任。

他的研究领域包括自然语言理解、机器翻译、基于内容的网页信息处理以及应用人工智能等。近年来主持完成国家自然科学基金、国家863项目、国防预研、省部委、国际合作项目等20余项。2009年获得国防科技进步奖。近4年来获得软件著作权7项，在国内外刊物和会议上发表论文150余篇。

赵军



赵军，中科院研究员，博士生导师。1998年在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做博士后、访问学者。2002年5月至今在中科院自动化所模式识

别国家重点实验室工作。

主持国家自然科学基金重点项目、973计划等国家级项目。研究方向为信息提取和问答系统等。在 IEEE TKDE、JMLR 等顶级国际期刊和 ACL、SIGIR、EMNLP、COLING 等顶级国际会议上发表论文六十余篇，获 COLING-2014 最佳论文奖，获 KDD-CUP2011 亚军 (2/1297)。研发了汉语文本分析、信息抽取和知识工程、百科问答等软件工具和平台，在中国大百科全书出版社、华为公司、讯飞公司等得到应用。

王厚峰



Houfeng Wang (王厚峰)

H 14 A 24.98 S 0 G 725 P 80

Professor

北京大学信息科学技术学院

Natural Language Processing

Maximum Entropy Model

Machine Learning

Sentiment Analysis

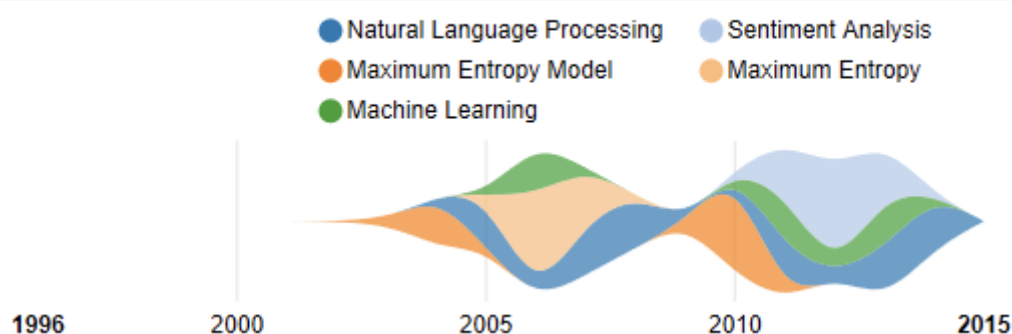
Computational Linguistics

Feature Selection

Semi Supervised Learning

Chinese Abbreviation

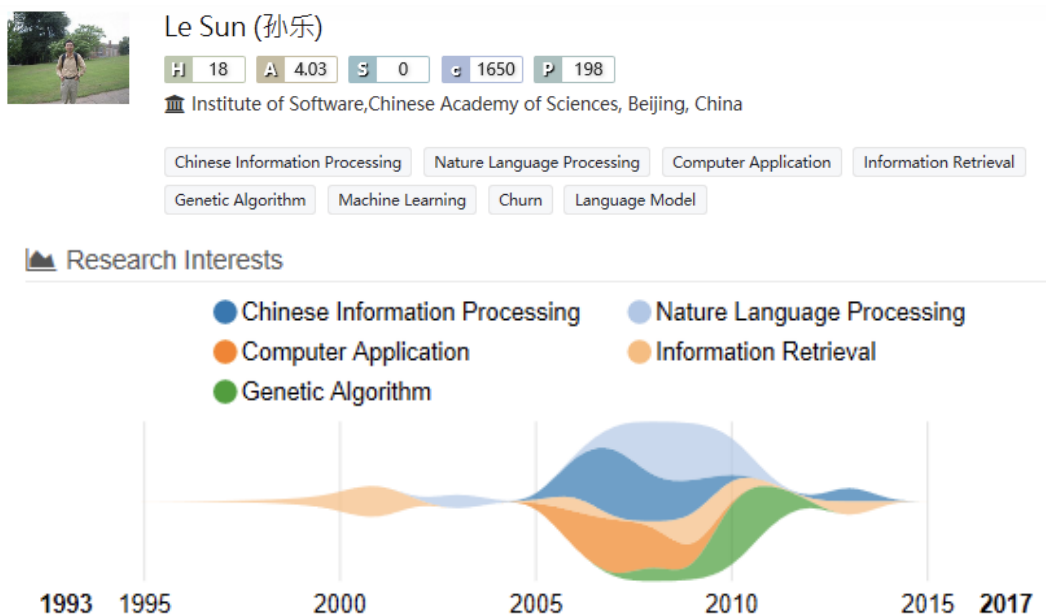
Research Interests



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。主要研究兴趣包括情感分析、问答与会话、自然语言语言语篇分析等，曾作为首席专家主持过国家863项目、国家社科基金

重大项目，负责国家自然科学基金重大研究计划等。在 ACL、EMNLP、COLING、AAAI、IJCAI、ICML 等会议以及 Computational Linguistics 等期刊发表论文70余篇。

孙乐



孙乐，中国科学院软件研究所，研究员，博士生导师。中国中文信息学会副理事长兼秘书长。《中文信息学报》副主编。2003至2005年，先后在英国 Birmingham 大学、加拿大 Montreal 大学做访问学者，从事语料库和信息检索研究。

目前主要研究兴趣：基于知识的语言理解、信息抽取、问答系统、信息检索等。在国内外主要刊物和会议上共发表论文80多篇。曾任2008和2009国际测评 NTCIR MOAT 中文简体任务的组织者、国际计算语言学大会（COLING 2010）组织委员会联席主席、机器翻译峰会（MT Summit 2011）组织委员会

联席主席、中文语言评测国际会议（CLP2010、2012、2014）大会主席、国际计算语言学年会（ACL 2015）组织委员会联席主席等。