· 中国自然语言处理年度盛典 ·

CCL 2018

第十七届中国计算语言学大会

第六届基于自然标注大数据的自然语言处理国际学术研讨会

学生研讨会

一线师生联袂教你做科研

报告人：刘知远

报告题目

文献综述与研究选题

报告摘要

与过去的学术报告不同，此报告不介绍具体的研究动态，而是向同学们介绍一下，在自然语言处理领域，如何更好地进行文献调研，了解某个研究领域或课题的最新进展与全貌；如何更好地进行研究选题，为做出高水平创新成果开好头。希望通过这个报告能够帮助同学更好地开展自然语言处理创新研究工作。

报告人简介

清华大学计算机系副教授、博士生导师。2006 年、2011 年于清华大学计算机系获得学士、博士学位。2011 年开始在清华大学计算机系担任博士后、助理研究员、助理教授、副教授，曾任新加坡国立大学高级研究员。主要研究方向为表示学习、知识图谱和社会计算。2011 年获得清华大学博士学位，已在 ACL、IJCAI、AAAI 等人工智能领域的著名国际期刊和会议发表相关论文 60 余篇，Google Scholar 统计引用超过 3000 次。承担多项国家自然科学基金。曾获清华大学优秀博士学位论文、中国人工智能学会优秀博士学位论文、清华大学优秀博士后、中文信息学会青年创新奖，入选中国科协青年人才托举工程、CCF-Intel 青年学者提升计划。担任中文信息学会青年工作委员会执委、副主任，中文信息学会社会媒体处理专委会委员、秘书，SCI 期刊 Frontiers of Computer Science 青年编委，ACL、COLING、IJCNLP 领域主席。

TsinghuaNLP

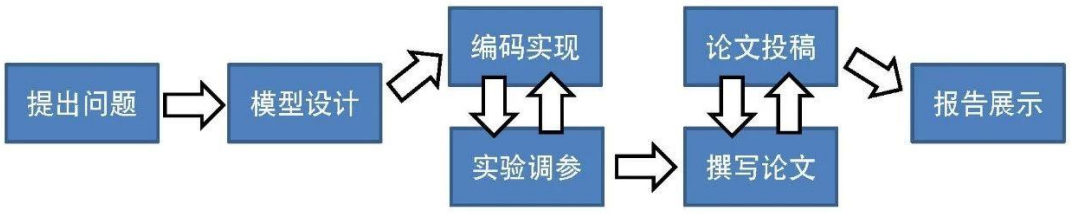
CCL 2018 学生研讨会

文献综述与研究选题

清华大学自然语言处理实验室

刘知远

学术研究是一项系统工程

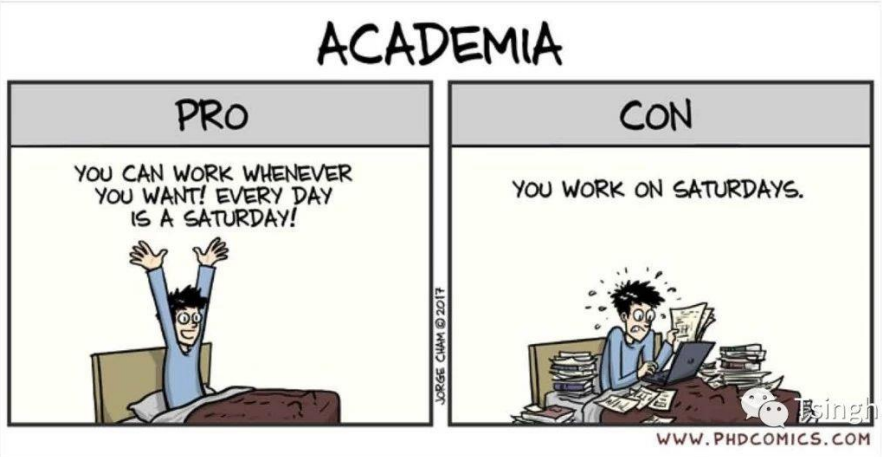


文献阅读　沟通能力　坚持不懈

学术研究需要天时地利人和

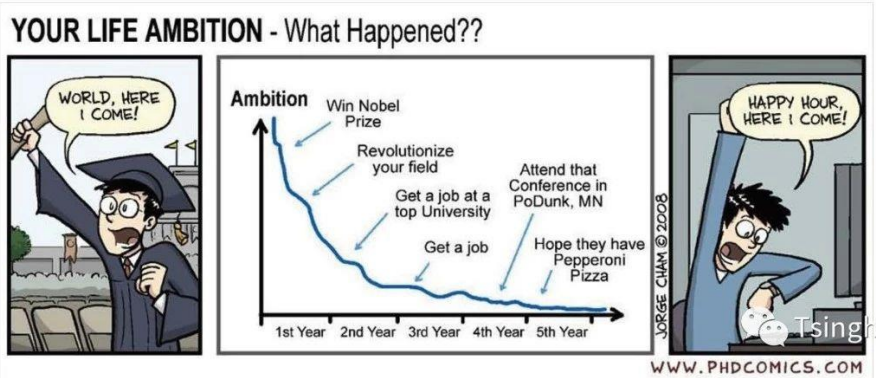
成功的研究 =

重要问题 + 新颖方法 + 努力、积累、坚持

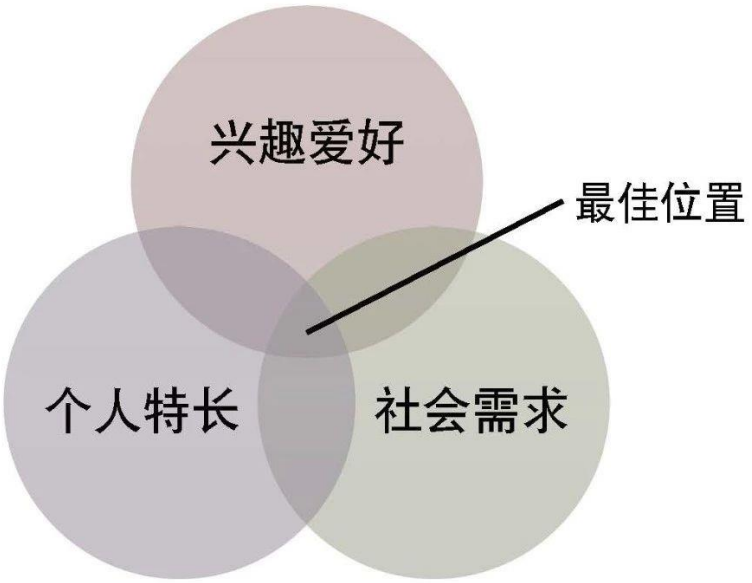


学术研究不同时期有不同追求

- 第一层：锻炼解决开放问题的能力
- 第二层：成为相关领域的知名专家
- 第三层：做出引领领域方向的工作



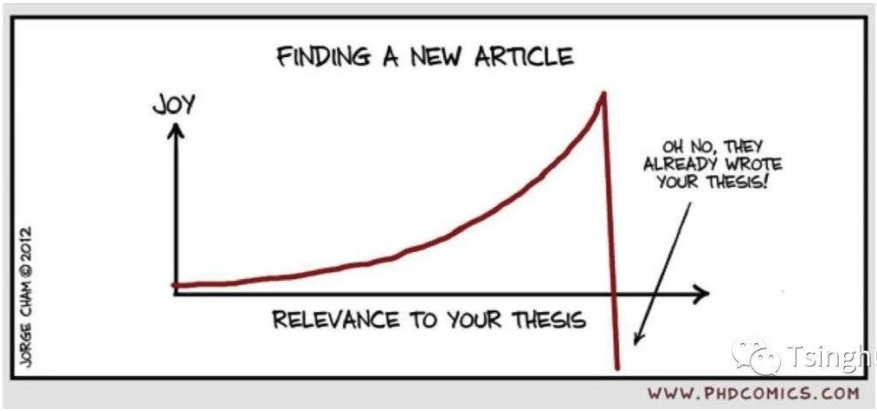
研究方向的选择



如何查阅文献

如何查找论文（给定关键词）

维基百科（搜索引擎） → 中文综述（CNKI） → 英文论文（Google Scholar）

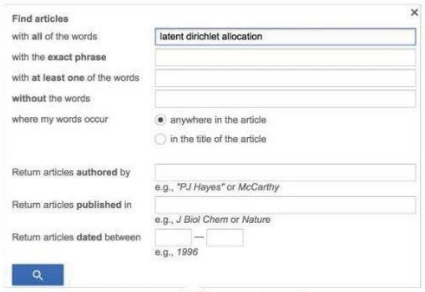


善用Google Scholar

- 查阅学者学术信息、引用情况，也提供引用格式文件

Latent dirichlet allocation
DM Blei, AY Ng, MI Jordan - Journal of machine Learning research, 2003 - jmlr.org
Abstract We describe latent Dirichlet allocation (LDA), a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying ...
Cited by 15978 Related articles All 124 versions Import into BibTeX Cite Save Fewer

- 学会使用相关搜索命令
  - Author: “DM Blei”
  - AllInTitle:“Latent dirichlet allocation”
  - …



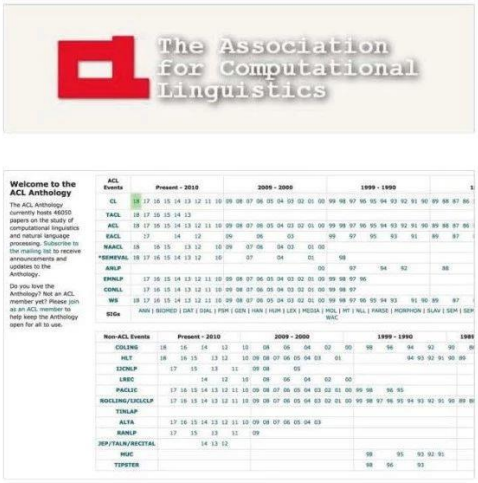
如何判断论文是否值得阅读

- 作者是否大牛学者？作者机构是否顶尖？
- 是否发表在顶级期刊/会议上？
- 论文社会关注度如何？是否获得最佳论文？引用情况如何？

学术资源-ACM

- 美国计算机学会
- 全球最大的计算机学术组织
- ACM DI拥有大量高水平论文
  - 信息检索
  - 数据挖掘
  - …

学术资源-ACL



- 国际计算机学会
- 全球最大的自然语言处理学术组织
- ACL Anthology囊括几乎全部的NLP重要论文（全部免费）
  - ACL
  - NAACL
  - EMNLP
  - COLING
  - …

学术资源-CCL

http://www.cips-cl.org/anthology

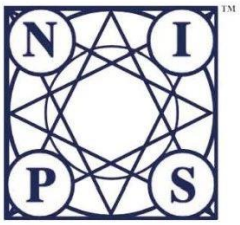
标题	作者	会议	下载量
基于深度学习的微博情感分析	梁军，柴玉梅，原慧斌，昝红英，刘铭	CCL 2014	2747
基于表示学习的中文分词算法探索	来斯惟，徐立恒，陈玉博，刘康，赵军	CCL 2013	1846
基于深度学习加强的混合推荐方法	丁冠原，张敏，谭云志，刘奕群，马少平	CCL 2016	1699
基于卷积神经网络的微博情感倾向性分析	刘龙飞，杨亮，张绍武，林鸿飞	CCL 2015	1690
结合卷积神经网络和词语情感序列特征的中文情感分析	陈钊，徐睿峰，桂林，陆勤	CCL 2015	1677
基于极性转移和LSTM递归网络的情感分析	梁军，柴玉梅，原慧斌，高明磊，昝红英	CCL 2015	1456
基于评论挖掘的药物副作用发现机制	程亮喜，赵明珍，林鸿飞	CCL 2014	1425
基于句法语义特征的中文实体关系抽取	郭喜跃，何婷婷，胡小华，陈前军	CCL 2014	1337
基于规则的越南语命名实体识别研究	闫丹辉，毕玉德	CCL 2014	1306
知识图谱中实体相似度计算研究	李阳	CCL 2016	1211

论文下载量排行榜

统计时间：2016年3月至2018年9月

学术资源-ICML/NIPS

- 机器学习领域的两大顶级会议



- 深度学习时代的新兴学术会议

学术资源-Arxiv

- 预印本文库
- 未发表的论文，良莠不齐
- 建议关注顶级组织的相关论文

subscribe Zhiyuan Liu
1 message

Zhiyuan Liu <liuzy@tsinghua.edu.cn>
To: cs@arxiv.org

add CL
add LG
add NE

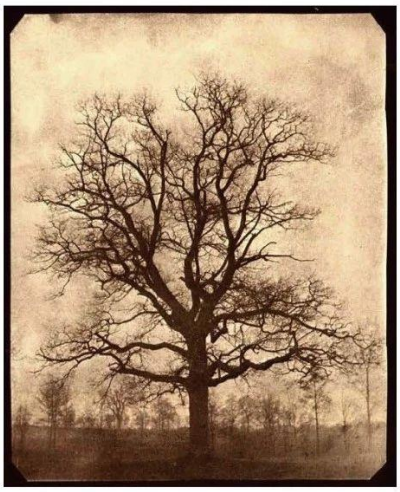
阅读论文顺序

- 题目（1）
- 摘要（2）
- 正文
  - 导论（3）、相关工作、自己工作（5）、实验结果（4）、结论
- 致谢
- 参考文献（6）
- 附录

如何研究选题

提出问题比解决问题更重要

- 一流学者提出问题
- 二流学者解决问题
- 三流学者打补丁



本页来自清华大学刘洋老师《撰写学术论文的若干技巧》。

为什么找问题更重要、更难？

- 提出问题者往往能影响整个领域的发展方向
- 解决问题往往是个技术活，能够后天培养（理论素养、编程能力、写作能力等），而提出问题则需要：
  - 站得更高
  - 看得更远
  - 嗅觉更好
  - 当机立断
  - 不畏风险

本页来自清华大学刘洋老师《撰写学术论文的若干技巧》。

如何找问题？

Think differently

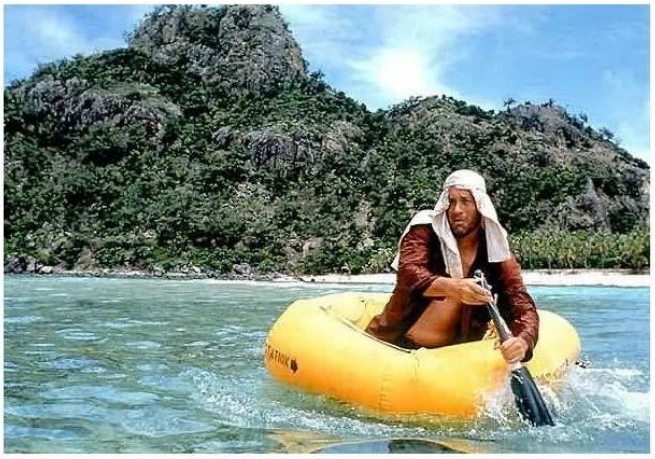
满腹经纶者固然可敬，擅长推陈出新者更值得推崇。

本页来自清华大学刘洋老师《撰写学术论文的若干技巧》。

哪里热闹去哪里

本页来自清华大学刘洋老师《撰写学术论文的若干技巧》。

哪里人少去哪里



*"It is not worth an intelligent man's time to be in the majority. **By definition**, there are already enough people to do that."*

--- G. H. Hardy (1877-1947)

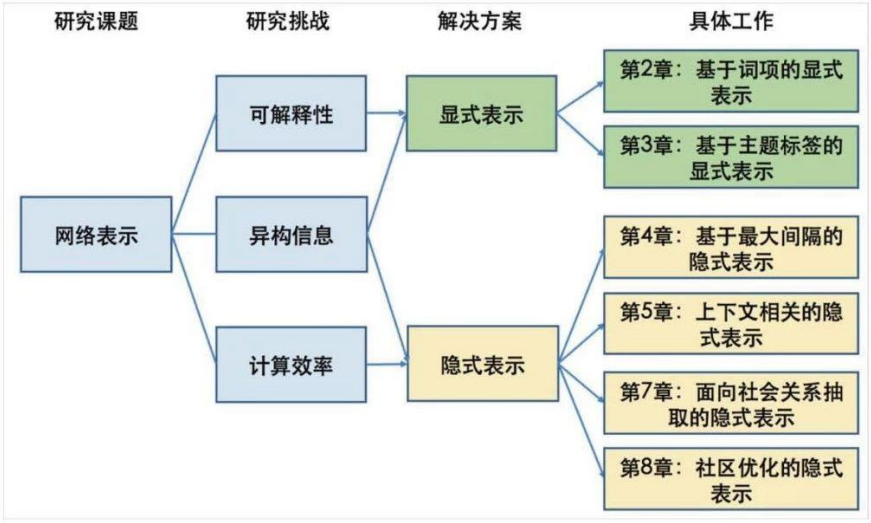
本页来自清华大学刘洋老师《撰写学术论文的若干技巧》。

如何找到好问题

- 博览群书，对整个领域有全貌式把握
- 熟知学术界动态，知道当前最热门问题是什么
- 明察秋毫，富有远见，结合个人兴趣选择一个数年后会变成热门的领域，并全力以赴去做

本页来自清华大学刘洋老师《撰写学术论文的若干技巧》。

对博士生的选题建议



涂存超（2018）：面向社会计算的网络表示学习

全国NLPers，联合起来！

http://nlp.csai.tsinghua.edu.cn/~lzy/

liuzy@tsinghua.edu.cn