

学术搜索入门

送给实验室刚入学的实验室硕士和博士们！本篇文章总结了学术搜索的几种常用方法。希望你们少走弯路，早日“入门”(高年级博士请直接跳到第4章)。

符号备注：

1. *: 博士需要，硕士不需要
2. **TIPS**: 小技巧
3. **WARNING**: 注意事项，小心中招

1. 学术搜索工具

现有的大部分学术数据库都要收费（国内学报除外）！而且很贵(5美金一篇)！幸好学校图书馆已经为我们准备了丰富的学术数据库(比大多数学校都多，而且全)。相当于帮我们论文的下载费用都先行支付了。那么，不好好利用就是太可惜了。

[东大学术数据库列表](#)

注意：大部分东大购买的数据库都是IP认证(其他是镜像)；所以不在学校时，或者使用电信线路访问时需要用到[VPN](#)。同时，由于这类数据库对于批量下载很抵触，希望大家不要用工具批量下载，否则会导致学校（和你）被加入黑名单。

1.1 简便和实用（检索入口）

[Google学术](#)

最强大的学术工具，没有之一。配合Google强大的搜索能力，搜索体验非常好，并支持多种格式的参考文献导出。**特别注意，Google学术可以跟踪作者和论文(如有论文被引用或者作者有新论文就会提醒)**。如果用东大IP登录，大部分文献可以直接下载。

[DBLP](#)

对会议检索很快，会将作者的相关工作整理归类。也会将会议的论文整理。

[微软学术](#)

对会议的分类和排序比较好，可以分领域搜索。微软学术本身对于会议和论文会有详细的分析(排名，引用状况等)，很适合详细分析某个会议或者一篇论文的发展趋势(引用趋势)。

1.2 专业型（版本有保障，会议扫墙专用）

两大计算机协会，外加Web of Science：

[ACM](#)

[IEEE](#)

[Web of Science](#)

WARNING：各大网站导出的参考文件格式不同(特别是IEEE和ACM)！记得在论文中统一

1.3 中文论文检索（检索学位论文专用）

两大中文检索数据库：

[万方](#)

[维普](#)

1.4 无法找到论文，或者无法下载（请Google）

如果前面几种方法搜索到了论文，但是无法下载（论文库问题，后者还未正式发表）。那么Google就是最好的方法。直接Google这篇论文，有些大牛会把论文挂在自己主页上。而且，现在Google支持论文内部检索（直接定位到论文中的段落）。需要注意的是：Google到的论文版本可能不是最终发布版本，部分细节可能会改变。

TIPS：实在不行，厚着脸皮问作者要论文（客气一点，作者还是会答应的）。

2. 鉴别论文好坏

论文很多，很杂。一篇优秀的论文需要很多心血和时间去不断的修改和凝练，而一篇垃圾论文可能只需要复制和粘贴。两个级别的论文差距很大，当然也会给我们带来不同的影响。从我的经验来看，阅读论文时要以优秀论文为主，避免被垃圾论文误导。所以，最好在检索时候就鉴别论文质量。

2.1 检索工具排序

上面提及到的各类检索工具都会对论文进行排序。排序方式有很多种

1. 被引用次数（很重要，大部分检索工具默认）

原理很简单：如果其他作者认可这篇论文中的工作，就会主动去引用这篇论文；因此，被引用次数越多，说明这篇论文的越重要。反之，如果没人引用，除非是新方向或者最新的论文，基本都可以判定为比较差。

2. 会议或者期刊的影响力

顶级期刊和会议会将论文交给领域内顶尖的大牛来审论文；这些大牛也很重视，会花大量时间和精力去审论文，以保证会议和期刊的质量。

2.2 会议和期刊分类列表

1. 领域列表（由导师或者师兄提供）
2. [CCF列表](#)（顶级会议和SCI期刊）
3. 学院认可的学报：

[计算机学报](#)

[软件学报](#)

[计算机研究与发展](#)

[通信学报](#)

[中国科学](#)

~~[东大学报](#)~~

3. 拓展知识面（找到更多论文）

搜一篇明确的论文需要不到一秒；但是通过搜索扩充自己对领域的认识，需要慢慢积累！对于一个新的领域，你至少需要了解：**关键词、关键技术、领域划分、重要的论文列表以及重量级大牛**。随后，再选择自己感兴趣的方向，进一步细分和研究。

了解一个领域的方式一般有以下几种：

3.0 综述和优秀学位论文（看运气，新领域一般没有）

如果这个领域有几篇很有影响力的综述（survey）或者几篇优秀的学位论文，那么以下三种方法所需数据都可以被大大缩短。因为通过他们，你就有了以下东西：

1. **关键词**
2. **关键技术**
3. **论文列表**（而且是分好类的）
4. **领域划分**
5. **大牛名单**（好的综述都是大牛写的，综述里重复出现的作者要么是他的朋友，要么就是大牛！）

TIPS: 一篇好的综述非常有用，可以省去很多时间和精力。不过有survey也说明该领域已经比较成熟。

3.1 关键词方式

关键词(keyword)是搜索的核心，也是找到论文的核心。好的关键词库能够串起一大片好论文，甚至一个领域，也能帮助我们快速识别某篇论文是否属于我们领域。不过，**寻找关键词不能心急，需要不断阅读不断积累**。等关键词积累到一定程度，你就会发现对领域的认识到了一个新的高度。

以下是我寻找关键词的流程：

1. 从中文关键词入手（如果已经知道英文关键词最好！）寻找英文关键词和论文
2. 用已知的关键词(英文)，搜索并阅读相关论文，**扩充关键词库**
3. 分类和细化关键词，删除一些干扰的结果
4. 可以尝试和作者交流，**一次好的交流可以获得很多最新的关键词**
5. 分析最近的热门论文，更新和扩充关键词

例如：最开始我只知道数据匿名和数据隐私，通过搜索和阅读文献，我发现数据匿名和数据隐私可以等同于data privacy, data anonymization, privacy preserving data publishing(PPDP)等；而数据的匿名操作在不同论文中的说法也不一致，有用anonymize, coarse, distort, clean, sanitize等等。那么这些关键词就可以扩充到我的关键词库中，方便我鉴别某篇论文是否属于我的研究领域。后期发现，其实用coarse, clean的人很少，只是在论文中避免重复使用anonymize时才会出现，所以可以弱化这两个关键词。

WARNING: 上面所述的关键词特指英文关键词，中文关键词由于翻译和语境问题，更加多样化，统一性太差。

TIPS: 关键词积累很漫长，需要阅读大量文献；但是随着时间推移，这些关键词会帮助我们更好的理解这个领域

3.2 参考文献方式

准备开始找：

1. 搜到一篇相关论文（论文越好参考文献越有价值）
2. 从该篇论文的参考文献（reference）入手
3. 不断增加论文，分类论文

WARNING: 很简单，很有效的方法！但是，需要有明确的目标。整个过程需要不断的增加论文，并分类论文。不然手头的论文只会越来越杂(论文多到不敢想象，但是真正对你有用的就那么几篇)。做好过滤和分类，可以大大提高入门速度。

TIPS: 相关工作（related work）一般会提供很好的文献分类，参考这个章节可以帮助我们划分手头的论文。

3.3 大牛效应

发现某些人在领域论文中出镜率很高？恭喜你，开始邂逅大牛了。大牛一般很牛X，理解得深入，实验做得好，论文写得好，不喜欢拿公式和理论恶心人。这样的前辈，对我们的研究会很有帮助。

找大牛一般有以下方法：

自动上门(实验室请过来交流):

1. 找和你方向相关的大牛，去听报告
2. 听完报告和大牛交流，留邮箱后面可以继续联系

主动找：

1. 找到一篇或者多篇好文章
2. 挨个查作者最近的论文，论文都很好，那么就收藏（论文+大牛）

被动发现：

1. 从已有的论文列表中获取大牛候补名单
2. 去这些大牛主页看看他的论文发表情况(可以了解他的研究进展状况和领域趋势)

WARNING: 跟着大牛的论文列表去入门肯定不会错。但是，小心别被大牛带坑里去（大牛一般都有团队，而我们往往只有一个人。特别是大牛说“这个很简单！”时）。

3.4 会议扫墙

最暴力的方法，工作量巨大，费眼睛。但是每年扫一次可以保持对领域的清晰认识。

1. 获取优秀会议和期刊列表（见2.2）
2. 选取和自己相关的会议和期刊
3. 从今年开始往前(反向也可以)，将这些会议期刊上相关的论文都下载下来

TIPS: 扫墙也是有技巧的。先浏览**abstract**和**introduction**，觉得相关或者感兴趣再进一步浏览。千万不要逐个通篇阅读，费力费脑，吃力不讨好。

4. 如何阅读论文

4.1 论文各部分的重要性

1. 摘要 abstract：论文最重要的部分，一般由导师直接把关
2. 引言 introduction：重要性仅次于摘要，内容更加易懂
3. 相关工作 related work：对领域内工作的分类和总结

4. 其他章节：论文细节，对于理解本论文非常重要。
5. 总结 conclusion: 全文总结和展望
6. 参考文献 reference: 论文引用的文献列表
7. 附录 Appendix: 部分无法放到正文中的内容，如过分细节的证明和算法

4.2 快速阅读（泛读）

如果时间有限，直接读abstract是最高效的方法。如果时间还很充裕，那么就把introduction也读一遍。读完这两部分，就应该能基本理解论文的核心思想。

WARNING：读得速度越快，对论文的理解越差，甚至会出现误解。因此，想要好好理解一篇论文，精读是非常必要的。

4.3 精读

通过泛读（快速）阅读筛选完论文之后，需要精读部分的重要论文。我个人建议将精读分为几个阶段进行：

1. 理解论文基本原理：motivation，所用的理论？
2. 理解论文详细内容：深入理解论文细节，包括定义、假设和相关公式等
3. ~~能够重现实验~~（未必需要）：能否重现整个实验，并得到相同结果？
4. 组织讨论：与其他人进行讨论，可能有没有疏忽的地方？或者理解不到位的地方？
5. 设计更好的方案：能不能更进一步，设计更好的方法？

每个阶段可以对应一次或者多次阅读，通过多次阅读来加深对论文的理解。不要指望一口气能够理解整篇论文。很多原因会造成理解不够深入甚至理解错误，例如现阶段的知识面不够，或者对于某些概念理解错误等。所以需要多次阅读（如果有必要，多次讨论）来避免这类问题。

注意：讨论是非常必要的（最好由阅读论文的人主讲），当其他人从他的角度提出疑问时，很可能会给意想不到的惊喜。

*5. 论文提醒和会议追踪(让论文来找你)

*5.1 论文提醒

通过添加论文被引用通知和作者论文通知，我们可以掌握领域内最新的研究进展。

已知可以添加被引用通知的搜索工具(均需要注册账号)：

1. [Google学术](#)
2. [微软学术](#)

3. [Web of Science](#)

Google学术添加提醒(强烈推荐):

1. 搜索结果提醒。搜索"WIFI", 点击左下角的"Create Alert", 添加后, 以后和wifi相关的论文有变动都会收到邮件。
2. 引用提醒。点击第一篇论文的引用列表(cited by), 点击"Create Alert"可以建立引用通知, 新论文引用该论文, 就会有邮件通知。
3. 作者提醒。点击作者链接, 再点击"Create Alert", 可以创建作者提醒。如果该作者维护自己的Google学术页面, 那么他有新文章时就会通知。

Web of science添加通知:

*添加"引用提醒。搜索论文, 点击进入, 点击"Create Citation Alert"添加。*收录期刊论文比Google快, 还是非常有必要的。

微软学术:

国内打开太慢了, 真心不想用。

TIPS: 有时候, 论文还没发表, 就已经能够通过引用关系找到该论文。建议师弟师妹们好好熟悉下。

*5.2 会议追踪(别再错过deadline)

核心会议的时间和地点基本固定, 但是投稿时间和deadline区别很大 (abstract没投就不让投正文)。为了避免晚一天的悲剧, 国内外各有一个工具, 都非常好。注册账号后, 只要搜索会议, 在会议页面点击"Track"或者"Add to my list"即可, 到特点时间点(例如, abstract deadline)就会提前发邮件通知你。

1. [WikiCFP](#) 老牌追踪工具
2. [Conference Partner](#) 后起之秀, [源于某个好心的博士前辈](#)

WikiCFP可以查看详细的timeline, 非常直观; 但是每次会议过期需要手动添加下一年会议, 比较麻烦(一般本年会议开完后, 下一年的会议时间才能基本确定, 等到WikiCFP收录, 估计要1-2个月)。Conference Partner最新改版后变得不是很好用, 但是可以显示会议地点, 直接在地图上显示。

6. 案例和练习

6.1 搜索: Data Privacy

1. 用Google学术搜索data privacy。发现"Data privacy through optimal k-

anonymization"被引用779次（大牛型论文），这样的话，关键词和引用列表就都有了，大牛也有了。

2. 点击相关论文，发现很多相关论文都被引用。选取相关论文中，引用次数最多的论文入手。
3. 点击Cited by 797, 了解该论文被引用情况，前几位的论文需要关注下。
4. 通过积累的关键词和论文，逐步熟悉这个领域。

6.2 搜索：室内定位(手机和无线定位)

1. 用Google学术搜索indoor location。发现很多都是RFID的论文。RFID是物联网方向的，偏硬件，不适合。但是第三篇论文"Robust indoor location estimation of stationary and mobile users"是INFOCOM2004(安全方向A类)的论文，而且是手机和无线。那么以这篇论文为后续的线索。
2. 搜索"Robust indoor location estimation of stationary and mobile users"相关论文。果然大量的wireless和WIFI的论文就出现了。找对方向了！
3. 通过积累的关键词和论文，逐步熟悉这个领域。

*6.3 添加一篇论文引用通知，追踪一个会议

1. 搜索website fingerprint，尝试添加一个引用提醒
2. 选择一篇论文，添加引用提醒
3. 在WikiCFP上面追踪下一届S&P会议

结束语

随手写了这么多，希望对各位师弟师妹的有帮助。希望你们站在我们肩膀上，比我们做得更好。

龚奇源 gongqiyuan@seu.edu.cn 2013-7-23

版本更新：

1. 2013-7-23 第一版完成
2. 2014-2-18 更新VPN地址，增加一些备注，修改部分语句，增加案例的详细说明
3. 2014-5-4 增加论文追踪，扩展论文引用追踪
4. 2014-5-5 修改部分语句，修改格式
5. 2014-5-6 根据凌振的建议，强化关键词小节
6. 2015-9-3 根据张骏雪的建议，增加DBLP
7. 2015-9-17 增加如何阅读论文

