

一个博士生的留学感受

刘立坤
清华大学

关键词：论文写作 问题驱动研究 与工业界交流

我是清华大学计算机系博士生，得到中国留学基金委“高水平”项目的资助，于2009~2010年在美国卡耐基·梅隆大学进行了为期一年的博士生联合培养，现已回国。留学期间的酸甜苦辣仍然历历在目，感受颇深。这些感受和经历也许是读者已经遇到或者将要遇到的。希望我的点滴感受对在读的研究生尤其是博士生，能有一点启发和帮助。

论文写什么？

科研论文是研究人员表达科研成果的主要方式之一。然而，头疼的是往往已经做了一些有成绩的科研工作，却写不出像样的论文来；或者说感到自己的工作很有价值，但总结不出高水平的论文——这也许是大多数科研人员遇到的问题，也是必须解决的问题。就我个人而言，主要有两个方面的体会：

首先是看问题的角度 从个人经历和与周围的人接触中，我发现“科研论文的一个误区来自于对创新的理解”。相当一部分科研人员往往不自觉地将科研创新和追求“与众不同、标新立异”等同起来，即为了创新而创新，追求解决方案的差异和改进，忽略了对问题本身的深入探讨。留学期间国外导师Ganger教授的一句话，让我有一种醍醐灌顶的感觉，颠覆了我以前的那种为创新而创新的科研动机。“好论文最重要的是让读者有所收获，即告诉读者那些本应该是他们知道但还不知道的知识。”所以，写好论文的一个重要认识是站在读者的角度考虑问题，动笔之前先要想一想，读者读了你的论文后有什么收获，你要表达的take home message是

什么。

问题的深度很重要 学术论文不应该追求大而全，而应该小而精，切忌浅尝辄止。通常，用于发表的学术论文短则6、7页，长也不过十几页，很难将一个大的问题完全地阐述清楚。因而学术论文本身应该是解决有一定深度和代表性的小问题。要想解决太大的问题，一般论文的深度往往是不够的。这样的观点，在我所接触的很多国外的著名学者中都有过相同或者相似的表达，也都身体力行地实践着。虽然他们的研究领域都很广泛，但具体的研究方向和研究选题都有所侧重，都很精、很细、也很系统。小而精的优势在于可以避免科研组织内部的相互竞争，进而促进组织内部的科研合作和共同创新。

开放性的科研

科研作为一种创新活动，其新颖性（novelty）是具有决定意义的。科研成果的发表往往是“只有第一、没有第二。”一旦某科研成果已经被他人公开发表，即便你是独立完成的，该成果依然失去了创新性，也不能再视为是你的成果。于是，学术成果在未发表之前有一定的保密要求，特别是不应该让具有竞争关系的人知道。但这并不意味着研究工作本身是封闭的，科研工作合理的范围内（比如实验室，合作人员或单位的范围内等）应该是完全开放的。科研灵感的重要来源之一是与别人交流，要在不断地辩论中才能碰撞出思维的火花。交流也是消除思维盲区、完善和提高科研成果的重要途径。初到并行数据实验室（Parallel Data Lab, PDL）的一个

明显的感觉就是会议室多，能够容纳百人左右的实验室里竟然有6个大大小小的会议室，利用率非常高，多数时候需要提前预约才能获得使用。每个会议室除了现代化的投影仪、视频设备之外，最显眼的是铺满整个墙面的白板和不同颜色的白板笔。此外，在学生工作的隔间（cube）之中也配有白板。上面通常写满了各种各样的公式和符号。

问题驱动而非答案驱动的研究

科研本身是为了解决问题。高水平的科研基础是找“问题”，找到了好的问题通常意味着科研已经成功了一半。但不幸的是，我们在科研中往往会自觉或不自觉地犯一些“答案驱动的研究”的错误，特别是在为了创新而创新的情况下。国内科研过程中还存在一些做了相关的研究然后再去找问题的情况。问题驱动的研究，其核心是寻找真正的问题。值得注意的是虽然有些问题看起来是个问题但其实也许不是问题。寻找问题的过程中要学会洞察问题的本质，尽量能用一句非常简单的话概括出该问题为什么是个问题。我在和导师的探讨中曾经提出过采用“列存储来组织文件系统元数据”的方法，导师给予的第一反应是“你想要解决什么问题”。显然，我在这里犯了答案驱动的错误。整个思考过程中关注解决方案的实现及其带来的好处，却忽略了所针对的具体问题。

重视跟工业界的接触和交流

计算机是一个工程性比较强的学科，真正的问题和尖端科研往往存在于工业界中。作为存储领域世界瞩目的顶级实验室，PDL给我最明显的感觉是重视同工业界的接触，强调科研问题从实际中来，到实际中去。主要表现在：（1）经常会有知名IT公司技术主管来作报告。由于工作在技术的第一线，距离实际的用户需求比较近，他们往往能提出许多真实问题，对问题的评价也是一针见血，避免了单纯的学术研究可能会脱离实践的尴尬。（2）每年有一个“PDL Retreat”，专门请工业界的人过来跟PDL

的人在一起交流。在Retreat上，PDL人会将自己的研究成果展示出来，听取工业界人士的意见和评价。PDL对Retreat上的报告要求特别高。在Retreat上的每个报告都要事前演练一至两遍，请实验室的人先提出改进意见。这个研究出发点使得PDL大部分科研问题都有实际的工业背景，也使得PDL的科研成果很容易应用到工业中去，从而对存储产业的科研和技术进步产生重要影响。

重视积累

科研本身是一个积累的过程，对于学校实验室这种“铁打的营盘，流水的兵”的科研模式尤其如此。然而，国内有相当一部分学校实验室往往注重科研课题的交差，缺乏对于实体知识积累的重视，甚至缺乏科研方向的传承。常常是一味地跟着科研项目来确定科研方向，“有什么项目搞什么科研”的现状比较普遍。当然这在一定程度上是由于大环境所致，与科研体制有关，不是简单说改就能改的。彻底的改变往往需要我们几代人的共同努力。但实验室级别的传承和积累是可以从现在做起的。PDL在这方面做得就比较好，实验室有自己的科研积累工具和平台，其中除了大家熟知的CVS源码仓库和wiki内部交流平台之外，给我印象最深的是一个自主研发的参考文献管理工具refmandb。该工具以bibtex的格式按照预定义的规则记录了与PDL科研相关的几乎所有参考文献信息。与之配套的工具可以自动地抽出和生成某篇论文相关的参考文献，不仅给参考文献维护和论文写作带来了很大的方便，更重要的是也为新手调研提供了便利。此外，个人积累同样重要。不仅是知识的积累，还有想法和灵感积累。最好手头有个东西能将你的感悟记下来，用不了一年，你就发现你有一大笔财富。■



刘立坤

清华大学计算机博士生，主要研究方向为存储系统、分布式存储和分布式系统。llk04@mails.tsinghua.edu.cn