# [怎样做研究（一） 刘挺](http://m.blog.csdn.net/blog/angela_615/5383448)

几年前，我写了一套胶片，题目是《怎样做研究》，多次在实验室内部给学生们做报告，也曾对外讲过一次，听众反应良好。也有网友读过这套胶片，给我来信称有所收获。然而，胶片中的文字毕竟只是提纲携领，无法充分阐述我的想法，为此，借周末一点闲暇，把《怎样做研究》写成一篇文章，与师友切磋。

什么是科学

科学是分科的学问，客观地说，是起源于西方的。中国只有经验科学，典型的如中医。我的母亲是学中医的，我从小就对中医耳濡目染，生了病，妈妈就会请他的老师来，一贴小药下去，我的病就好了。因此，我对中医一直是很信服的。然而，近些年来，中医多受批评，发展也越来越缓慢，究其原因，中医不是科学，或者说只是经验科学，而非实证科学。中药的成分以及生化功效不曾用实验进行深入的分析，望闻问切的诊断方法完全凭经验而无法量化，阴阳五行的理论似是而非，祖传秘方的传承方式与知识共享的现代思维背道而驰。因此，尽管中医有诊治的整体观和方剂的个性化两大优点，但其停留于经验层面，而迟迟不能进入科学的殿堂，因此在现代社会中的发展必然步履维艰。

中医不是科学，那到底什么是科学呢？科学（自然科学）是人们用来认识和改造自然世界的思维武器，科学研究可以分为基础研究（理论研究）和应用研究（技术研发）。

基础研究

万事万物皆有其规律，掌握并且利用这些规律就能够为人类造福，这些规律是隐蔽在纷繁复杂的现象背后的，要识破大自然的奥秘，读懂上帝的天书，非要下一番深入观察和探究的功夫不可。以揭示规律为目的的研究活动属于基础研究，从事这些活动的学者是科学家。规律不是被创造出来的，而是早已存在的，人们只有认识规律的权利，而没有创造规律的可能。

从根本上讲，推动基础研究的也是人们在生产生活中的一些实际需要，但是随着基础研究的深入，理论已经成为一个庞大的体系，理论研究早已开始按照它自有的逻辑独立发展，而不必时时刻刻联系实际需要，比如著名的歌德巴赫猜想，可能在百年之后，发现其有重大的应用价值，但是目前到底有什么用，谁也说不清楚。理论的价值在今天这个非常讲求短期功利的社会中常常被忽视，现在有一种倾向认为只有产生实际经济效益的科研工作才有价值，这种极端化的观点显然是错误的，我们必须承认并高度尊重理论研究者的成就。

理论研究的直接动力是科学家的好奇心，以及他们对科学荣誉的渴望。越是单纯的科学家越有希望发现真理，他们的科学探索有点像迷宫探宝或者海边拾贝，伟大的科学家都是没有丧失童趣的人，他们在实验室里是宁静而愉快的，他们是乐此不疲的，很多在常人看来难以忍受的寂寞在他们看来却是一种幸福。越是找不到答案，越是激发探索的热情，在一次次的失败中积累着烦闷与紧张，在终于取得突破后兴奋异常。与此同时，也必须承认科学荣誉也是激励科学家们前进的重要动力，只要别把荣誉看得高于真理，货真价实的荣誉仍然是值得追求的。

理论上的突破对应用研究产生持续不断的推动力，在模式识别领域，神经网络、支持向量机、条件随机域等等机器学习技术不断出现，每当一项理论出现，应用研究者们争相将其应用于自己的研究课题中，于是基于神经网络、基于支持向量机、基于条件随机域的某某研究就成为一个标准的论文题目。首先把某项理论应用于某个实际课题的研究工作应该说还是具有一定的创新性的，毕竟用一个新的思路、新的模型去观察了一个旧的课题，HMM在语音识别上的成功应用就是一例。有人比喻说，理论工具仿佛是锤子，实际课题好比是钉子，一个新的锤子被打造出来，大家都借用过来砸一砸自己手头的钉子，确属常理。不过，需要注意的事，如果拿一个硕大无比的汽锤去砸一个纤细的大头针就荒诞可笑了，不注意思考问题与理论的适配关系而盲目跟风的事情在学术界也是司空见惯，比如我们就曾用HMM试图解决词义消歧的问题，而每个多义词的词义跟它前后一两个词并没有紧密的关系，因此词义消歧貌似和词性标注一样属于线性序列标注问题，其实是有根本差别的。

应用研究

我们是搞计算机的，计算机是一门应用科学，应用科学是由应用驱动的。时至今日，数学定理和物理学定律似乎已经被先哲们发现的差不多了，因此整个科学界中纯粹搞理论研究的人越来越少，很多大学教授都和工业界有着密切的联系，很多大企业也开办企业研究院，这些导致应用科学的研究如火如荼。最近，国家863设立了一个“中文为核心的多语言信息处理”重点项目，总经费7000万，这在多年前的大陆语言处理界完全是不可想象的。

应用驱动，也可以说是市场驱动。市场是一个精灵古怪的家伙，搞应用研究的人如果对市场的未来没有一个基本准确地判断，往往会导致选题上的偏差。二十年前，国内一些研究者开始研究汉字手写输入技术，开始人们觉得从键盘输入汉字很困难，手写输入一定有前途，但是很快，拼音输入法大面积普及，而且拼音输入的速度远比在手写板上输入汉字快得多，于是汉字手写输入套件根本卖不动，前景黯淡。有人开始犹豫，有人开始转向搞印刷体汉字识别等，但忽然有一天，集成了手写功能的商务通大量热销，人们忽然发现原来在手持设备上由于键盘太小，输入不便，给手写功能留下了很大的应用空间。一直专注于手写识别的汉王公司也借着商务通的热销而把多年的科研成果成功地产业化了。再举一个例子：5年前，我认为以图像为输入的图像检索没有什么应用价值，问这些技术的倡导者，他们也只说能够在数码相册中可以找到一些应用，但近来听了微软一些学者们的演讲，他们提到可以用手机拍下一个植物的图片，传回服务器，在大量植物图片库中检索，找到最相似的植物，并给出植物的名称，特点等。哈哈，这对于我这个五谷不分的人来说实在是太有帮助了，可见对于一项技术是否有用实在要仔细思考，不要早下断言。

技术和市场是一个互动的关系，有人认为技术严格地从用户的现实需求出发，这个观点总的来说没有错，但是忽视了技术创造需求的一面。大多数用户往往并不了解技术发展到了什么程度，他们提不出需求来，这时技术专家们需要把技术和产品做出来给人们看，刺激、引领用户的需求，比如数码相机，5年前我想大多数用户和我一样并没有淘汰胶卷相机的强烈要求，但当数码相机进入市场后，人人都意识到：原来我需要这个东东。

在市场与技术的互动中，总的来说，还是市场在引导和拉动技术的发展。市场需要的是产品，产品往往集成了多项技术，因此一项被市场接受的产品能够推动多项技术的进步。比如搜索引擎，它拉动了自然语言处理、并行计算、海量存储设备、数据挖掘等等多项技术的发展。最近中国计算机学会设立了王选奖，在中国真正有市场眼光，能够发明一项技术，拉动一个行业的计算机专家，王选是第一人。怎样根据市场选择研究方向，设计产品，调整技术形态，我在后面还有详细阐述。

科学技术的力量

科学技术的力量是巨大的，爱因斯坦给出的公式E=M\*C2，C是光速啊，质量乘以光速的平方，这是多么巨大的能量啊，爱因斯坦的理论直接导致了原子能的利用与开发。基因图谱的发现以及后基因组时代对基因图谱的深入分析必将为人类征服疾病提供一条崭新的解决道路，通过对损坏的基因进行修复，将使无数患者得以康复，无数家庭重拾幸福。互联网的发明，把全世界连为一体，过不了多久，石头里也会嵌入芯片，在这个世界上有生命的、无生命的各种物质之间都可能进行通讯，人们的生活面貌已经彻底改变了。

当然，科学也是双刃剑：原子弹爆炸了，核战争始终威胁着人类；在对基因组这套上帝给出生命密码没有全面理解以前，任何盲动都可能导致基因污染，以至于玩火自焚；互联网上的虚拟生存让人们感到更加孤独。

# [怎样做研究（二） 刘挺](http://m.blog.csdn.net/blog/angela_615/5383675)

研究的层次

研究是分层次的，很多大科学家在晚年登上了最高层，比如钱学森在80年代倡导思维科学，他对整个科学技术体系进行了重新分类。在中国的大学里，分为一级学科，二级学科等，我就处在计算机科学技术一级学科下面的计算机应用技术二级学科下。二级学科的带头人称为学科带头人，二级学科下面一个研究方向的带头人称为学术带头人，我就被指定为学术带头人。

我的研究方向是信息检索，信息检索下面又有子方向，比如文本检索、文本挖掘、跨语言检索、跨媒体检索等，子方向下面设立具体的科研课题，比如文本挖掘中的多文档自动文摘课题，针对一项课题又有不同的解决办法，基于事件抽取与集成的多文档文摘就是利用一种具体的解决问题的方法。

总结来说，就是6个层级：  
A. 一级学科  
B. 二级学科  
C. 研究方向  
D. 子方向  
E. 课题  
F. 基于某种方法对课题进行的具体研究

君子思不出其位，我是学术带头人，因此主要在思考C类的问题，也就是和信息检索相关的问题。一个学院的院长通常会思考A类的课题，学科带头人或者说是一个博士点的点长是要考虑B类问题的。一个人对相关的方向或学科有所了解，对自己的研究工作是很有好处的，只有看清了整体的学科面貌，才能知道自己处在那个位置上，自己未来的方向在哪里。我在读博士以及在微软做副研究员的时候，只看到E类问题，想到最多的是F类问题，因此你让我提一个新方向，让我对一项技术进行预测，我茫然无知。后来担任院长助理，负责学院的成果转化，需要了解学院里各个方向的发展状态，使我的视野开阔了一些。尽管我凡事不求甚解，但是喜欢总结归纳，因此对信息检索与其它学科的关系有了更多地认识，这对后来的选题很有帮助，特别是在应用研究方面，心里比较有底。

学科好比一棵大树的树根，研究方向如同树干，具体的课题就是枝叶了。和学科中各个方向都相关的研究课题是最基础的研究课题，比如在人工智能中，各类机器学习算法是图像识别、语音识别和语言理解等各个方向都离不开的，机器学习技术提高一步，好比树根抬高了一寸，各项应用技术也都跟着进步，因此越是基础的研究，越会对业界产生较大较深远的影响力。不过，基础研究的突破比较难，而在某个应用课题上不考虑一般情况，只考虑具体需要，成功的可能性大。枝叶上的课题做多了，经过合并同类项，就会发现比较共性的基础课题，比如我们在做问答系统、多文档文摘、例句检索等课题时发现复述(paraphrasing)是一个共性的问题，于是把复述单拿出来展开专门的研究，如此，可以越做越深。

学者的层次

研究有层次，学者也有层次，大致可以分为：  
A. 大家（剑客）：提出问题  
B. 专家（侠客）：解决问题  
C. 学徒：修修补补  
D. 抄袭者：抄来抄去  
E. 搞伪科学的人：弄虚作假

A类是大家，站得高，看得远，他们往往能够前瞻性地提出某个学科领域中的若干重大问题，最著名的是希尔伯特的23个问题，对数学界影响深远。提出问题其实也是解决问题的一种方式，只不过他们是在很高的层面解决问题，类似一个软件系统分析员，他把一个复杂的工程问题分解为若干个有机联系的子问题，然后宣布只要这几个子问题解决了，整个大问题也就解决了。至于这几个子问题到底怎样解决，或者说相应的子系统到底怎样开发，他就不管了。胡乱地提问题并不难，小孩子也会向大人提出各种各样有趣的问题，有的大人也答不出来，问题的关键在于在适当的时候提出适合当前学术发展阶段的关键性课题，这绝对不是一般人能够做到的，这是需要具有对整个领域全面深入的理解才行的。

B类是专家，是在某个研究方向上有专长的人，他们沿着大家指出的方向探索前进，提出全新的方法体系来解决问题。比如在机器翻译领域中，日本长尾真教授提出了基于实例的机器翻译方法，从一个全新的视角看待机器翻译问题。专家经验丰富，能够自由地驾驭课题，稳步地推动课题的进展。

C类是学徒，就是我们这些普通的研究人员了，这部分人的注意力在具体的课题上。学徒们还没有宏大的视野，没有捕捉全局战略要点的本事，也还没有在一个研究方向上提出原创性的解决之道，他们跟在拓荒者后面捡拾麦穗，他们负责对科学大厦修修补补。他们一会儿听说了一个新的机器学习方法，赶紧在自己的课题上试一下；一会儿发现了一个以前忽略了的新的特征，立即想方设法把这个特征提取出来；一会儿为了参加一个技术评测，耐心地调一调系统参数；一会儿为了发表一篇论文构造出一个试验来。我们每天的研究活动差不多都是在这样进行的，很多时候在原地打转转。

我这样描述学徒们的工作情景丝毫没有贬低的意味，在达到专家的水平，证悟研究真谛以前，跌跌撞撞、浑浑沌沌是在所难免的。只要遵守诚信之道，不抄袭，不造假，点点滴滴的贡献对科学界也是有帮助的。从更高的要求看，学徒的目标应该是成为专家，应该时常静下心来想一想，自己的工作是否有价值，是否有新意，揣摩一下大家们、专家们到底是怎样思考问题的，在不断地反思与实践中向上迈进。

D类学者根本算不上学者，他们为了评职称等目的，对别人的论文进行抄袭拼凑，他们是思想的窃贼，对学术界毫无贡献可言。

E类学者不仅仅是做贼了，他编造伪科学，毁坏科学界在公众中的形象，他们是科学界的公敌。

以上的分类也只是为了讨论的方便，在各类之间并没有明确的界限，我只是依次谈出我心中做学问的境界而已。

在人类已知的世界和未知的世界之间有一条动态边界，科学家就站在这条边界上，他们是挑战未知世界的勇士，他们每向前迈出一步，就意味着整个人类的已知世界向前拓展了一步，由此足见科学工作的艰难和科学家的伟大。

研究又好比爬山，一座座山峰如同一个个研究领域，大家已登峰造极，一览众山小，把东南西北各条山路上的沟沟坎坎，把此山与他山之间的距离关系看得清清楚楚。隔行如隔山，隔行不隔道，在一个领域做到顶尖的学者已入化境，一通百通，你把另一个领域的问题讲给他听，他往往也能够很快地抓到要害。专家已到半山腰，看不到山的全貌，但是他找到了一条通往山顶的道路，并一步一步地向上攀登着。学徒还没有进入山门，他们一会儿仰望山顶，一会儿看看山腰，在山脚下绕来绕去找不到门径，费力不少，却并没有缩短与山顶的距离。

# [怎样做研究（三） 刘挺](http://m.blog.csdn.net/blog/angela_615/5383679)

怎样选题

前文曾提到科学研究的层次，并分了6个层级。此处所说的选题指的是从C到E三个层次上的选择问题，即：C. 研究方向、D. 子方向、E. 课题。选择研究方向是实验室(Lab)主任们需要重点思考的事情，选择子方向是研究小组(Group)的组长们需要重点思考的事情，选择课题是研究生们需要重点思考的事情。

选择太多，很容易让人困惑，要想理出一个头绪来，需要一些基本的原则。微软的许峰雄来访时谈到了他选择课题的三个标准：有足够的兴趣，能成为世界第一，能赚钱。（！）兴趣，这个原则是非常重要的，我赞同，获得国家最高科技奖的“黄土之父”刘东生院士是搞地球环境科学的，经常在野外作业，按常人推断，这该是多么枯燥艰苦的工作啊，但他说：“枯燥？不！因为经常有新发现，其中的乐趣难以形容”。我坚信任何一个成功的科学家的直接工作动源都是兴趣，而不是意志。（2）成为世界第一，不容易，但是应该作为一种判断标准，如果某个领域已经非常成熟，很难有什么创新了，或者大牛云集，已经打破头了，则应该有所回避。（3）赚钱，许峰雄是在工业研究院中工作，比较注重实用，因此他强调了“赚钱”，我是在工科大学里工作，也比较偏重应用，因此是赞同“能赚钱”这个标准的。不过，“能赚钱”不等于立即赚钱，5年、10年，20年后能够赚钱的研究课题都是值得关注的。

谈谈我选择课题的一些体会：  
1、 要有实际需求  
一个课题必须有实际需求，可能是现实的需求，也可能是潜在的需求；可能是直接的需求，也可能是间接的需求，总之是的的确确被人们所需要的。据个反例，比如自动文摘，自动文摘是我的博士论文课题，但是实际应用需求始终不清楚，自动文摘的结果用于编辑出版，质量肯定无法保证，用于帮助人们快速浏览资料吧，Google提供的包含查询词的简单的Snippet就起到了这个作用，因此，至今基于全文分析的单文档自动文摘到底用到哪里，仍然不清楚，这方面的研究已经有50多年的历史了，仍然是不死不活，总是找不到应用就无法得到政府和企业界的持续性支持，以往的付出成为鸡肋。我觉得单自动文摘不是一个好课题，目前阶段多文档文摘，或者说对某个题目的自动综述分析是非常好的题目。

2、 有较大的未知空间  
以手写体汉字识别为例，市场上已经大面积应用了，在研究上就不宜再展开。

3、 与自己以往的工作有关联  
如果你觉得自己的研究领域太窄，或者竞争对手太多，或者自己缺乏兴趣，则可以适当扩展研究方向，但最好是相关性地扩展，比如从自然语言处理(NLP)扩展到信息检索(IR)，IR要用到NLP的技术，这种扩展是从底层技术到应用系统的扩展，很自然。再比如从图片检索扩展到视频检索，只是处理对象有变化，很多原有的技术优势仍然能够发挥。如果跳跃性太大，比如搞NLP，忽然发现做数据挖掘有前途，于是单纯地转向数据库中数据挖掘，和文本处理完全脱节，这种做法一方面无法发挥既有的技术积累，另一方面也让同行感觉你不够专注，不容易得到认可。最要命的是有的人根本就没有自己的方向，什么课题都敢接，这样的人可以一时间让人觉得风风火火，经费也很充足，但过不了多久就会摔落下去，因为缺乏积累，学术形象不清，公鸡下蛋，干了自己不擅长的事情，在学术圈还怎么混？

4、 有可能得到国家的支持  
对于资深学者，他选定一个课题后，可以写出立项建议，去说服政府或军方支持他的工作，从而填补国家空白，成为国内这个方向的先驱。哈工大的杨孝宗老师借鉴CMU在wearable computing方面的研究成果，在国内率先提出穿戴计算机的概念，坚持多年，就获得了军方的认可。对于刚出道的年轻人，无力直接影响政府，那只有自己预先判定一个几年后可能成为热点的方向，先走一步，做出一些成绩来，等到大气候适宜的时候，由于他已经取得了一定的成果，也有可能被认可为这个领域的先行者，得到国家的支持。

课题的类型

对一个课题的类型要有一个判断，是研究型的还是开发型的，如果是研究型的，要组织博士生们来攻关，鼓励大家大胆尝试，提出创见；如果是开发型的，要更多地召集硕士生们来做，强调利用一切现有的技术手段把技术或系统做到实用可靠。这两者要分的比较清楚，既不能通过各种打补丁的方法，或者说一大堆小技巧来对付研究型的课题，因为那样是做不出突破性进展的，也不能在开发类课题上总是异想天开，尝试还很不成熟的技术。

如果是研究型课题，还要区别是基础研究还是应用研究，基础研究的结果不能直接被用户使用，类似重工业，应用研究的结果最终用户直接就能够用上，类似轻工业。对于基础研究，可以抛开具体应用的约束，专注于一些科学原理技术原理的突破。对于应用研究，则需要考虑用户的需求。

课题还有长期(long rm)和短期(short term)之分，长期研究的课题往往难度大，研究结果难以预料，短期项目则比较好预测，可以速战速决。

# [怎样做研究（四） 刘挺](http://m.blog.csdn.net/blog/angela_615/5383680)

在一个具体的题目上作研究，应该遵从怎样的程序呢？我觉得可以概括为“螺旋式深入”，也就是在“阅读”，“思考”，“实验”，“写作”，再阅读。。。这四个阶段的时间分配可以根据实际情况灵活调整，刚进入课题的研究生阅读调研花费的时间要多一些，而在一个课题上已经开展了一两年工作的人则可能增量式地阅读资料，阅读时间自然比起步时少一些。专门用于思考、设计、推演的时间可能并不多，但思考是渗透在其它三个阶段中不断进行的，因此总的思考时间并不少。实验中编程的时间应该尽可能短，用更多的时间进行实验数据的分析。写作是常常被中国的研究生忽略的环节，写作的时间要足够长。收集资料，了解别人的工作，找出问题所在，针对性地提出自己的创意，用实验验证自己创意的正确性，总结归纳，撰写论文，发现新的问题，再收集资料，如此反复，这是研究活动的大致流程。

怎样阅读资料

收集资料、阅读资料是从事研究工作的第一步，但是如何收集、阅读资料却很有学问，初学者如果没有得到足够的指导，常常走很多弯路。

1、 阅读重要的论文  
目前互联网上的信息量太大了，对每一条信息的重要性、可靠性的判断是一个人采集信息的关键环节。如果判断一篇论文是否重要呢？Google Scholar给出的引用数是一个有效指标，很多学者都引用的文章往往就是有价值的论文。有的同学觉得看中文论文容易，于是把自己能够查到的中文论文一网打尽，反复阅读，但是很多发表在三流刊物上为了评职晋级而炮制的论文完全没有阅读的价值，白白耽误了时间。即使是英文论文，国外一样有滥竽充数的文章，这样的论文引用数肯定低，用引用数可能很容易地把这样的论文淘汰掉。

计算机领域的顶级会议论文非常重要，在NLP领域有ACL，在IR领域有SIGIR，在机器翻译领域有MT Summit，这些顶级会议的论文质量很高，内容很新，应该高度关注。期刊上的论文是一个作者或机构一个阶段的研究成果的总结，通常质量较高，但由于审稿及编辑出版的周期很长，因此内容不够新，适当关注即可。NLP领域的CL，机器翻译中的MT，信息检索领域的IP&M和JASIST等都是很好的期刊。进入一个领域，必须立即了解该领域有哪些顶级的国际会议和国际期刊。

2、以作者为线索理清脉络  
阅读论文一定要注意论文的作者是谁，研究机构是哪里，以作者为线索理一理就会发现全世界搞你这个方向的也就那么几个、十几个研究机构、研究者，以后就跟踪这些人的研究工作即可，还能够发现该作者的研究工作的演进脉络。如果拿到一篇文章就读，读完了也不知道作者是谁，时间长了，就会感到晕头晕脑，不知道从哪个期刊或会议上就会冒出一篇相关文章来，让你防不胜防。

3、 阅读最新的论文  
学术发展很快，要集中尽力阅读近5年，特别是近3年的论文，对于5年前的论文，只看引用率最高的经典文章即可。

4、 抓住论文的要害

读完一篇论文必须了解哪些关键内容呢？我觉得应该包括以下方面：作者为什么要做这项工作？要解决的是一个什么问题？作者在解决问题时遇到了怎样的困难？为了解决他的困难他提出了什么样的解决办法？试验结果是否可能真的证明他的方法好，数据是否充分，有没有和别人的工作，别的方法进行对比？你认为他的方法是否新颖，你从中学到了什么？该方法有哪些不足，你是否立即有了新的改进方案？如果有立即记录下来。带着上述问题，抓住要点，做好记录，一篇长文就会像庖丁解牛一样轰然倒下。

5、 批判式阅读

真理越辩越明，我们读的是一篇学术论文，不是《圣经》，不能带着崇敬的心理去阅读，要像一个审稿人那样带着批判挑剔的心理阅读论文，在阅读中不断地找出论文中的问题，选题上的，方法上的，实验上的，表述上的，并不断地通过积极独立的思考给出自己认为见解。只有这样，资料才能够为你所用，而不会成为你的包袱。有的同学读资料，越读越丧失信心，发现别人做得太好了，自己的想法都被别人做完了，资料全读完了，自己也准备换课题了，这是失败的读法。

中国的研究生要有信心，不要被国外所谓的名家吓住。中国的科研水平在快速提高，科研人员的素质也在快速提高。一位美籍华裔企业家在一篇文章中写道：“可不幸的是，除了很少顶尖学校的博士外，大部分博士所做的研究课题都是陈旧或者没有意义的。”不知道顶尖高校的含义是什么，但是我觉得我们的研究生要对自己的国家有信心，对自己的学校有信心，对自己的倒是有信心，对自己有信心。只要我们掌握正确的研究方法，广泛阅读国外最新的研究成果，大胆尝试自己人为正确的方法，充分释放我们的聪明才智，我们就丝毫不用对国外的研究工作顶礼膜拜。在科学研究上，欧美人从内心里是瞧不起我们亚洲人，我们中国人的，以至于欧美归来的学者们也以欧美为样板来评估我们教育科研体制，只要和美国不一样就是大错特错了，中国高校的教师们都是在误人子弟。我奉劝每一位研究生建立不崇拜权威，不崇拜欧美，只服从真理的独立思维模式，大胆质疑大胆批判，只有这样才能不死于他人之言下，才能有活脱脱的自己。

# [怎样做研究（五） 刘挺](http://m.blog.csdn.net/blog/angela_615/5383683)

怎样思考

1、把问题定义清楚

有的同学做了很长时间的课题，还没有把问题定义清楚。以自动文摘问题，好像就是把一篇文章中的核心内容提取出来吗，还怎么定义，其实不然。文章是什么样的文章？议论文、记叙文、还是说明文，还是包括小说之类的文学作品？文章的来源是哪里？是《人民日报》等用语非常规范的文章，还是网络文体，比如blog或BBS上的文章？文章有没有长度上限，10万多字的博士论文是否需要提取摘要？能否对多篇话题相同的文章一起做摘要？再说摘要，适用于阅读的报道性文摘，还是用于判定主题相关性的指示性文摘？有字数限制还是比例限制？是否根据用户的需求有所偏向？对文摘的连贯性有没有要求？怎样评价一篇文摘的质量。只有把问题的初始状态（此处为原文）和问题的终止状态（此处为文摘）搞得一清二楚才敢说问题定义清楚了。“知止而后定”，问题定义清楚了，把子立好了看清了，心也容易静下来，再怎么做工作始终围绕一个目标，这样的工作才不会左右摇摆，才有意义。

2、 思维逻辑要干净

思维逻辑啊，思维逻辑！我越来越觉得这是一个人从事科研工作的最重要的素质。我的学生有两种，大多数人的思维是非常干净的，而确有少数同学的思维是混乱的、粘滞的。思维逻辑混乱的同学需要通过不断的自我认识，找出自己的误区，否则在前进的过程中会步履维艰。打一个比方，人脑好比一台计算机，阅读资料是数据输入，加工处理则要靠思维逻辑，思维逻辑如果混乱就仿佛是程序有Bug，输入的数据再充分再正确也无济于事，甚至由于处理能力有限，输入的阅读，处理起来越乱。

有的同学还钻牛角尖，抓住一个很细节的无关紧要的问题不放；有的同学思维的跳跃性非常大，一会在问题的高层思考，还没有讨论清楚，忽然又跳到一个底层的细节上；有的同学听不进别人的意见，满脑子只有自己的声音；有的同学不懂得矛盾对立统一的道理，总想把问题绝对化，造成自己很大的困扰；有的同学做惯了开发，总是想着怎么把一个系统实现，而不是把注意力集中在创新上；有的同学不善于剖析问题，分析数据，不懂得先分析再综合的思维方法，总是在问题的表层打转转，始终无法深入。。。。。。

计算机大学本科专业科中的不少内容在今后的工作中都不能够直接地排上用场，但是有一位在医科大学教计算机课程的老师向我反映，她叫医科大学的学生编写程序非常困难，因为思维逻辑完全不对，医学需要大量的死记硬背，而很多病症都是模模糊糊，需要大量经验，而计算机的思维是非常严谨周密的。教管理学院学生的，教人文学院学生学习计算机的老师们也都有自己不同的感受。这是什么原因呢？原因就是计算机本科专业科，比如离散数学（包括数理逻辑）、算法与数据结构、高级程序设计等课程给计算机专业的学生奠定了一些思维逻辑上的基础，思维方式在无形地对一个人未来的工作发生着重要的影响。

我有一个同事是从理科转过来学计算机的，我和他就明显不同，他很少编程，但每次做了一个试验后，他习惯于花大量的时间对数据进行分析。而我对数据分析不够重视，思维倾向于怎样巧妙地把搞出一套方法来，实现一个系统，解决一个问题。理科偏重于解释世界，工科偏重于实现系统。我们只有通过不断的内省，发现自己的思维方式，善加利用，有所修正，才能够顺畅地开展科研工作。

3、分析与综合

分析与综合是两把思维的利器，一定要好好运用。通常一个问题来了，我们感到无所适从，不要着急，请先使用“分析”这把“刀”，把问题划分为若干子问题，子问题之间的关联越少说明这一刀砍的越是合理，如果实在非要“连着骨头带着肉”，也没有关系，但是要记录下子问题之间的照应关系。子问题如果很容易解决了，就是本原问题，不用再分，如果还是比较复杂，可以进一步分析，得到一些“孙子”问题。经过深入分析，一个貌似强大的问题已经被我们看得清清楚楚，每个本原问题都比较容易找到解决方法了，研究者的精神也可以放松一下。

但是，搞研究不是做工程，不能满足于用打补丁的办法解决一个具体问题。在分析之后，还有在拿起另一把武器“综合”，“综合”的作用是合并同类项，比如通过对子问题1、4、5的考察觉得这个问题象是一个球体，通过对子问题2、3的考察觉得这个问题象是一个圆形平面，经过归纳则可以给这个问题建立起“半球体”模型，很可能就接近真实情况。归纳后可以演绎一下，看看灵也不灵，那就是用测试数据进行测试了。

概括地说，分析是分类并考察每一类的特征，分析是显微镜，帮我们看清了问题的每一个细节。综合是尽可能地找出统一的模型概括各类现象，统一的模型可能是多个模型的融合，但最好不是简单拼接，而是激光焊，在分子层面把多个模型融为一体。分析决定深度，综合决定高度，缺一不可。

顺便给出一个观点：普遍认为国内存在着低水平充分研究的现象，比如搞搜索引擎，从很少有人涉猎到一下子冒出很多家来，都在搞搜索引擎，而且大多数都处于起步阶段，难分伯仲。大家都觉得研究空间狭窄，竞争激烈，但又不知道怎样解决这个问题。其实对一个课题进行深入分析，把大课题分解为小课题，各家不搞大而全，或者即使全，但在“全”中都有自己明确的重点，比如专注于搜索引擎中的分布式计算问题，或专注于跨语言检索问题，或专注于问答系统等精准搜索，等等。对于规模小一点的组，甚至可以专注于更细的问题，比如问答系统中的问题分类，跨语言搜索中的查询翻译及扩展等。我们知道大树的根，越深的地方分差越多，覆盖的泥土空间也越大，研究也是这个道理，都浮在表层，就会感到空间狭小，如果深入下去空间就大了，不容易撞车。而且在一个细分的问题上，可以集中优势兵力作出突破性的成果来，又因为研究同一个细分问题的学者相对较少，研究工作的积累也相对不足，你动一动就可能做出新的成果来。因此，我的建议是：深入、深入，再深入。

4、 创新思维

呜呼哀哉，我的弟子中真正有创新能力的寥寥无几啊。我一直在思考如何挖掘和培养学生们的创新能力，在此谈谈自己的想法。

创新是科研工作的灵魂！“不创新，无宁死”，每个科研人员都应该有这种决心。科学界没有“省级运动会”，“全国运动会”，只有“奥运会”。你说你是中国首次提出“狭义相对论”的人，毫无意义。那么，创新就应该是在世界上内第一个提出某个想法的人，如果你的想法在地球的某个角落里已经有人提出了，那就不是创新，那就是重复，是浪费科研资源，浪费人力物力。人生短暂，大多数芸芸众生都在干着日复一日的重复劳动，有幸成为科研工作者，可以表现一下自己独特的想法，何其快哉，如果拾人牙慧，又何其痛哉。  
创新这件事没有固定的套路，如果有，就好像炒股票有了确定的获利方式一样，大家就都能赚钱，都能创新了。我很喜欢岳飞的一句话：“运用之妙，存乎一心”。史书记载：岳飞英勇善战，受到宗泽的赏识和器重。一次宗泽召见岳飞，说：“尔智勇才艺，世良将不能过，然好野战，非万全计。”因向飞传授作战阵图。飞说：“阵而后战，兵法之常，运用之妙，存乎一心。”宗泽听了以后，深为赞赏。搞研究象打战一样，固定的阵法战法也有，但真的想取胜需要“奇兵”，所谓“以正合以奇胜”。怎么出奇，完全靠指挥员的心思一转。

什么样的人擅长创新的，我觉得首先是那些平素喜欢天马行空胡思乱想的人，孔子有一个非常有才具的学生叫子贡，但孔子说他只能问一知三，说他不如颜回，颜回能够问一知十。触类旁通，在自己的识体系内利用各种相似性建立关联，就为创新创造了条件。“草圣”张旭的草书是从公孙大娘剑舞中悟出的，因为张旭通过多年苦练对书法的基本技巧已经炉火纯青，再想发展在书法本身上已经很难汲取新的营养，最后他从剑舞中悟到了获得了新的草书结构，艺术上峰回路转，又上层楼。

有不少同学，做研究只看和自己的课题最相关的文章，如果是一个新的领域，文章很少，他就会感到很苦闷，他更不知道从领域之外后的灵感了。搞文本检索的人，要了解一下图像检索的知识，从中可以获得启发，反之亦然。语言模型原来就是在语音处理中使用，现在却成了文本处理领域最成熟的理论方法。类比、嫁接从来都是创新的重要手法，视野不开阔，只知道自己眼皮底下的一点点东西，始终跳不出自己给自己设下的思维陷阱，是无法提出新想法的。

创新还必须从自己独特的体验中来。以诗歌为例，诗反映的就是一个人独特的生命体验，因此最忌讳用一些大众常用的泛泛的词汇，比如形容长江大河，不能说“浪涛汹涌”，要说“惊涛拍岸，卷起千堆雪”。搞研究也是如此，要用你自己的眼睛去观察数据，发现别人没有发现的特征，发现别人没有发现的故障点，找到别人不曾用过的观察视角重新观察你的研究对象，人与人的阅历不同，观察和思考问题的方式各异，如果你能够再重复调研的基础上，以“我”为主，把自己的原始体验经过归纳总结表达出来，那么即使是很小的一点进步，因为与众不同，那也是一个有价值的创新。何况，在你这一个课题上，全世界能够有个人坚持不懈地干上两年三年，并不多，就怕你没有自我，只要坚持你自己，表达你自己，即便资质差一点，也一定能够有所创新。

创新有种种，开辟一个新的领域，提出一个新的问题，是大的创新。在研究生阶段不容易做到，可以先扎扎实实地做一些方法层面的创新。方法上的创新也有大小之分，“模型创新”就属于比较大的创新，“模型”永远不等于“实际”，模型是对实际的最大限度的逼近，对于相同的输入，好的模型能够给出与真实情况更接近的输出。比如信息检索中有向量空间模型，这个模型因简单而常用，但是他没有考虑词项之间的关联，如果能够提出一种新的模型，把词项之间的项目制约关系也考虑进去，就有可能获得和更好的效果。这种创新还比较直接，如果能够彻底推翻向量空间模型，提出类似LSI（隐性语义索引）之类的模型，则是更大的创新。再比如长尾真提出基于实例的机器翻译，就是思维一转，对翻译过程看成实例匹配，而不是查词典和调序，这属于比较大的创新，也是我认为很有意义的创新。

如果提出新的特征，或新的特征抽取方法，或采用别人在该问题上没有用过的机器学习方法等，只要能够说出道理来，也都是创新，只是不那么激动人心而已。

创新思维是求异思维，不是求同思维，高人出手，变化多端，无所不用其极。要从东西南北，上下左右去观察事物，如果面对的事物太庞大，压扁了再处理行不行？切成碎片再处理行不行？烧化了变成水再处理行不行？在地面上实在处理不了，运到海底行不行，送到空间站上行不行？图像处理中有从空域到频域的变换，在空域里剥离不了的噪声到了频域里很容易分辨出来；在词义消歧中，bank不知道是“河岸”还是“银行”，放到宏观上下文一看，比如知道这边文章是金融类的，那它十有八九是“银行”。

创新是要“悟道”，很多时候要靠直觉，直觉就是大脑的并行计算，它把各种信息综合在一起，给出答案来。因此，你的灵感来了，要赶紧抓住，我有时有了好点子，身边没有纸笔，就感谢写到手机里，生怕忘记。

# [怎样做研究（六） 刘挺](http://m.blog.csdn.net/blog/angela_615/5383692)

怎样做实验

技术科学是实验科学，一项技术提出来，到底好不好使，“实验是检验技术的唯一标准”。怎样用实验去验证一个想法是否正确呢？第一步是收集数据，第二步是根据你提出的算法编写程序，运行程序，获得实验结果，第三步是对实验结果进行分析，得出结论。

1、 收集数据

收集到的数据必须真实可信，数据就是“证据”，有的人专门搜集对自己有利的证据，比如你要做一个问答系统，需要收集一些问题，那么最好的做法是从某个实际运行的网站获得真实的用户提问，而不应该自己提一些问题。自己提问，因为研制者自己知道什么样的问题能够处理，什么样的问题处理不了，所以你以及你所在小组的同学们提的问题往往会容易一些，以至于不利于激活技术难点。

收集到的数据要有足够的规模，具有统计意义，有的学者用几十个句子做实验数据，在提倡大规模真实文本处理的今天，这样的玩具性实验是不被业界认可的。从研究到开发，到产品，再到市场，本来是一个漫长的过程，但是由于今天的NLP和IR学术界越来越重视大规模真实网络数据的使用，以及很多产品直接在互联网上提供服务，从互联网上来到互联网上去，从一开始实验环境就是真实的，就是大规模的，因此从研究到市场的距离被猛然拉近了。这是我们这个行当目前的一个突出优势，必须坚持。

如果你所从事的研究已经有现成的国际标准评测数据，那么就应该要用标准数据，以便和国际同行比较交流。如果你的研究领域比较新，没有标准数据，也没有关系，只要你自己采集的数据真实充分，同样可以说明问题。

2、 编写程序

编程是研究人员的基本功，研究人员不应该在编程上有什么障碍，应该能够做到有了想法就能够快速实现一套相应的程序代码。你写的程序是否与你设计的算法完全吻合，是一个需要探讨的问题，有时候你觉得你设计了一个天才的算法，一定能够获得10%以上的指标提升，但实验结果令你大失所望，这是你不要立即心灰意冷，首先要自己检查你的程序，没准是某一行代码出了问题呢。在编程代码方面，还应该注意的一个问题是积累工具。由于你可能长期在一个课题上开展工作，因此非常有必要积累一些常用的代码，比如对语料进行预处理的代码等，这样可以节约很多时间。

用于研究的程序不一定太追求可靠性和速度，因为它的核心目标是验证算法的合理性。但这并不意味着可以乱写程序，造成频繁出错，弄得自己心烦意乱，本来注意力在算法验证上，这样一来不得不把注意力放在调试程序上面。程序的速度也必须考虑，如果每运行一次都需要一天一宿，每当你有一个新的想法，即使是一个小小的改进，都等到第二天早晨才能看到结果，你的研究一定会被这蜗牛般的程序速度拖累得奄奄一息。

3、 分析结果

经过一番努力，你的程序调通了，实验结果出来了，你的伟大设计是否真的卓越无比呢？我敢说，10次有9次都会不如人意，怎么会这样？没办法，因为你在做创新性的工作，“你向前一小步，就是人类的一大步”，这注定是一件艰难的工作。

如果你的实验结果达到了预期的效果，说明你触摸到了真理，可喜可贺。但不能停留在实验验证上，要给出理论分析，为什么你的方法能够取得比别人更好的结果，说不出让人信服的道理来，仍会受到同行的质疑。实验结果再好，仍然达不到100%，做错了的都是哪些数据，具有什么样的特点，可以对这些数据进行细分类，分别找出解决办法，在以后的工作中继续改进。如果实验结果不佳也不用沮丧，因为你已经知道“研究中不如意事十之八九”，我就常常碰到这种情况，特别是我给研究生们出了新点子，他们按照我的想法作了实验，结果效果不佳，我多少会觉得有点尴尬。遇到这样的情况，说明这一仗没有打好，对敌情估计的不够充分，有点像“三打祝家庄”，头两会不了解祝家庄的机关埋伏，只好积累经验，第三回再打。有时，实验结果神奇地超出国际最好水平10%以上，遇到这种情况，几乎可以断言你的实验做错了，或者你的实验条件跟别人并不完全相同，切勿盲目乐观。

前面讲过，创新应该是在世界范围内取得最佳结果，因此你的实验结果要在同等条件下和国际上的最好水平进行比较，方能证明你的成功。很多刚开始搞研究的人，往往犯一个致命的错误，就是儿童式的直线思维，例如：<问题>有人求我画一辆汽车；<工具>我找了一支铅笔和一张打印纸；<结果>我画了一辆奔驰（完）。你为什么要用铅笔，而不用毛笔，不用油画笔，为什么不用宣纸，不用油画布？别人用的是什么工具，有没有人用过油画笔和油画布？如果有，他们的作品是否比你的好？你为什么画了一辆奔驰，而不画宝马？你怎么知道别人喜欢奔驰，不喜欢宝马？比较和思辨是科学的翅膀，没有翅膀，就只能在地上爬行。我找到了一支铅笔，因此我就用铅笔画了；我熟悉铅笔的运用；我看到我的一个同学画了一幅铅笔画很好看，因此我也用了铅笔，等等，这些都不是理由，你必须在比较中思辨，在思辨中给出令人信服的理由。

最后还要问问自己，根据你的实验是否具有可重复性，也就是说是否任何一个学者在地球的任何一个角落只要按照你的实验方法重新做一遍，就能够得到相同的结果。真理应该是放之四海而皆准的，如果你的实验不能被重复，它的价值就会被打上一个大大的问号。

怎样写论文

要写好论文，首先要对写论文的意义有足够的认识。不少工科大学的研究生对工程开发很有兴趣，对做实验也还可以，就是在写论文这个环节上打不起精神来，总是草草应付。其实，“立言”与“立功”同样重要，多年后你开发的软件早已被淘汰，但是你在论文中表述的思想精华，却能够持续地对业内同行产生影响。科学的大厦需要每个研究者添砖加瓦，而科学的砖瓦就是“学术论文”，你再才华横溢，再巧思独运，如果不能以学术论文的形式把你的成果展示给整个学术界，你的学术贡献仍然无从体现。一个工程项目可能殷泽一时一地，一个学术思想却能够影响深远。认为写论文是“虚”的，干项目，干产品是“实”的，这种想法是错误的。

认识到了发表论文的价值，就有助于激发撰写论文的激情。科学论文是严谨的，学理工的人也比学文科的人更显冷静，但是写学术论文和写小说一样需要热情。你做了两年的工作，面对难题，几番挫败，终于找到了一条解决之路，你的喜悦从内心深处涌起，快乐需要与知音分享，此时此刻，你的家人由于不懂你的专业，已经无法在深层次理解你的成功，何不拿起笔来，写一篇论文，把你的成果告诉全世界的同行。

写论文很象编程序，作者就是程序员，每个概念就是一个变量，每行语句都是一条代码，变量的定义要清楚，代码的可读性要强。自己写完了论文，最好请导师和同学帮助挑错，这是在做测试；根据大家的意见要反复修改，这就是在调程序，调程序要细工慢火，一点点地熬，不厌其烦。审稿人相当于软件认证机构，最终的运行是在每个读者的大脑中进行的。如果论文写得不好，审稿人和读者看到一半就无法跟上作者的思维了，这就属于死机了。如果我们拿出调程序的态度调文章，用软件工程的思想管理写文章的全过程，那论文的录取率一定会大幅度攀升。

我有一个学生，写一个摘要，总共四五个句子，他要闭门谢客，整整写一个下午，反复斟酌调整，下过这样的功夫，论文的命中率肯定高。一般人都缺乏耐心，论文写完了就想投出去，不愿意看第二遍。其实“反复修改”是决定论文质量的最关键的一环。我认识一个论文高手，他写完论文就放在抽屉里，过几周拿出来再看再改，如此反复多次再投出去，命中率很高。写论文细活，是妙龙绣凤，是雕梁画栋，不是挖煤打铁，不是担水挑粪，没有细密的心思，没有绝佳的耐力，没有创造艺术品的雅趣是做不好的。

编程序，脑子里要想着用户的需求，写文章，也不是内心独白，要始终在心里有一个读者模型。很多初学者喜欢自说自话，概念A不定义清楚，就开始用A描述别的事情，或者把在A的基础上才能够理解的概念B早早地抛了出来。你自己知道A是怎么回事，不能就认为读者们也会理解，作者要找到“导游”的感觉，始终想着读者走到了哪一步，如果觉得某个环节很重要，读者不一定能够完全明白，还有换一种说法再作解释，绝对不能自己连蹦带跳地往前跑，把读者扔在半路不管了。顾客是上帝，读者是我们的上帝。

不要认为自己以前没有写过学术论文就打怵，其实学术论文也是一种议论文，我们在中学里都学过议论文的写作，议论文要有“论点、论据、论证”三个要素。一篇学术论文的论点通常是：“我提出的方法M比现有的其它方法在问题Q上更有效”。为了证明你的论点，需要“摆事实，讲道理”，事实就是实验，光有事实没有理论分析也不行，难以让人信服，二者缺一不可。很多人说自己的英语不行，影响了论文的发表，其实英语不是最重要的问题，最重要的是论证行文的内在逻辑，论证要丝丝入扣，清澈见底，不能枝杈横生，云山雾罩。好的论文还能起承转合，层层深入，带领读者一起走过九曲回肠，直至柳暗花明，让同行们读罢拍案叫绝。

论文怎么才能越写越好呢，关键还是要多实践，在实践中不断总结经验。美国的孩子从小就自己去图书馆查资料，针对一个题目写长篇大论，长大了，他们写出的学术论文冗长，但的确逻辑清楚，文字干净。我们的同学有的语文没有学好，不但布局谋篇不擅长，语法也常有错误，特别是有的还把网络上的语言习惯搬了上来，别字很多，排版凌乱，让人一看就不舒服。耐着性子看看其中的内容吧，看了半天也发现不了论文的要点，论文没有经过精心的组织，简单把一些实验堆砌在一起，啰里啰唆，如同梦呓，这样的论文在审稿人那里，极可能在30秒之内被枪毙。审稿人往往都是业内专家，专家都非常忙，抽空审阅一下论文，如果你不能再论文的开头用最简洁清楚的话语描述清楚你的问题以及你的创新性的解法，就剩下在收到退稿信后大声疾呼：“（审稿人）根本就没读懂”。行有不中，反求诸己，抱怨没有用，还是在投稿前下足功夫吧。

# [怎样做研究（七） 刘挺](http://m.blog.csdn.net/blog/angela_615/5383693)

在事业上每10年就是一代人，我们这一代（70年前后出生）是承前启后的一代。我们读书的年代没有互联网，与国际的交往也非常少，听说过ACL/Coling就不错了，都没想过去参加。在这种相对封闭的条件下成长起来的人，要想在一个点上做出国际领先的成果，是非常困难的。博士毕业后，多数人都承接前辈，成为课题组的青年掌门，作为掌门，要管理团队，要跑项目，事务性的工作太多，在点上的钻研不够。台湾苏克毅对于大陆青年学者过多地陷入管理事务深表惋惜，但是国情如此，一时不容易改变。我想我们这一代的使命应该是在老一辈的基础上把国家的科研工作体系进一步完善起来，在“面”上把握住研究方向，带领80年前后出生的一代人走上国际舞台，在他们中间发现优秀的学术人才，为他们创造一切可能的条件，支持他们在“点”上真正做出世界水平的发明创造，成为世界级的优秀学者。

怎样成为优秀的学者

1、 基本功  
无论干哪个行当，基本功都是至关重要的，任何令人眼花缭乱的高级技巧都无非是一些基本招式的组合及变型。基本功不扎实，就会常常感到捉襟见肘，按下葫芦起了瓢；相反，如果基本功扎实，则可以左右逢源，不受牵绊，日有所进。做研究的基本功包括数学基础、编程能力、专业基本知识、英文阅读、创新能力、写作能力、英文口头交流能力、组织能力和社会活动能力。我一口气列出了9项能力，而且基本上是按照一个学者成长过程在不同阶段能力需求的先后次序开列的。

刚入门时，有的同学由于是从外专业转到计算机专业来的，因此可以接着课程学习的时间补一补离散数学、概率论、线性代数等方面的数学基础。否则一旦进入了课题，发现有的公式看不懂，再回头看数学书，由于心情急迫，就直接翻到相关章节阅读，又看不懂，发现需要前导知识，于是下决心啃整本数学书，啃了头两张，又觉得进度太慢，心里发慌，如此反复，数学基础就越补越夹生了。因此，一定要在课程学习阶段把相关的数学基础打牢，并不需要记住每个公式，但是基本的概念和原理要非常清楚，进入课题后，如果遇到看不懂的公式，能够很快地查书解决即可。

编程能力必须尽快过关，否则影响科研速度。编程不能抱着一本编程的书，死学其中的样例，而应该积极地从老师或师兄那里领到很小的编程任务，比如文本的预处理，或者一个演示的界面等，有了实际任务，也就有了压力和动力，同时也便于融入研究小组，得到老师和师兄的指点，这样进步会很快。

导师会帮助你选一些对课题有帮助的研究生专业课，这些课程务必要认真学习。由于很多选修课没有考试的压力，我发现学生们学得不够认真，基础没有打牢，进入课题后很多概念不清楚，还需要再学习，耽误了时间。什么是功底，那就是对专业知识的一点一滴的积累，将来进入学术界，别的学者提到一些基本概念如果你都不知道，或者模模糊糊，就贻笑大方了。人生每个阶段应该完成的事尽量要在那个阶段完成，如果欠了账，将来想补，则要偿还三倍、五倍的利息，也未必能够补回来。人到中年，猛回头，发现自己的基础如同蜂窝，到处都是漏洞，这样的材料就注定无法做栋梁了，那时悔之晚矣。

每个人都有一定的英语阅读能力，但是阅读能力的强弱差别很大。阅读能力弱的人读英文好似雾里看花，总觉得隔了一层，似懂非懂，读得很慢，却仍然抓不住重点，读完了，一周不看就忘了，仿佛是没有读过的新文章。阅读能力弱的原因有两个，一个是语言问题，另一个是专业基础知识不足的问题，因此除了多读之外，专业知识的积累对阅读的深入和速度也会有帮助。

创新能力、写作能力，我在前面说的较多，不再赘述。在此谈谈英语口头交际能力。这个能力是在参加国际会议时才会深切地感到它的重要性的。在国际会议上，各个国家的学者操着不同的英语相互交流，你的听说能力更不上就成了一个聋哑人，异常尴尬。和“说”相比，更困难的是“听”，常常出现一个亚裔学生作完报告后听不懂听众提问的情况，我自己也是如此，印度英语、日本英语、爱尔兰英语，我的天，你必须竖起耳朵，运用你全部的语音信号处理能力和背景知识加推理能力才能够懂个大概。开国际会议，中国人常常聚堆，这更障碍了你提供英语能力的机会，应该主动找外国人攀谈，不要估计自己的面子，要不断地挑战自己，走向国际舞台，而不要退缩在一个角落里。现在的研究生比我们那时候得到了更好的英语训练，只要拿出自信来，即使第一次出国，不少学生就已经具备了和外国人直接交流的能力。实际上，外国人和你说上一两句，就能够判断你的英语水平，他们也会根据你的水平调整他的用词和语速，再加上你们的谈话一般限定在特定的技术话题，因此交流的难度已经大大降低了。更深层次的交流是能够了解外国人的文化，进而使谈话范围突破学术，和外国人交上朋友。

有了上面的能力，你可以成为一个合格的“学术单兵”。个人能力强，先做好一个单兵很重要，但随着经验的增长，你开始指导一些本科生或低年级的研究生开展工作了，两三个人就是一个团队，组织能力强的人可以把自己的团队带的生龙活虎，“嗷嗷叫”，反之则团队内产生摩擦，反而降低效率。高年级的研究生毕竟不是老师，没有老师的权威，也不能总通过“打小报告”对自己的“小弟”构成威慑，他应该通过自己的为人、才华吸引低年级的同学，首先得到他们发自内心的尊重，让他们意识到跟随自己能够学到知识，然后再用憧憬、鼓励、恳谈、告诫等管理方法激发他们的热情，带动他们跟随自己一道工作。根据他们的能力和时间合理地安排任务，随时给予指导，并从生活上对他们进行关心，和他们成为朋友。

要成为大学者，必须在学术界，以及相关的企业界、政界建立自己的人脉，从而有机会整合更大的资源，做出更大的事业，因此社会活动能力对于大学者是必不可少的。综合型的人才跨国个人学术原始积累阶段后，会成为学术组织者，此时他们的社交才能会得到充分的发挥。

2、 学术诚信

我个人是很赞成学术打假的，今天的学术界越来越受到商业的影响，虚的东西太多，需要有“方舟子”这样的人，震慑那些造假蒙事的所谓学者。

作为研究生，要严格避免做犯以下错误：抄袭（包括在文章中混淆自己和他人的工作）；一稿多投（包括一篇文章略加修改就另投他刊）；编数据，刻意裁减数据，或改数据。我的邻居原是一个学院的副院长，他在编教材时被指整页抄袭了别人的书，受到原著者的追究和学校领导的批评，书生心窄，很快得了肺癌去世了。还有一位同学，把一篇英文论文的主要内容翻译过来，作为其论文的部分内容投往了国内一家重要期刊，结果被该刊物认为是剽窃，申辩无效，她受到学校的通报批评。在治学问题上，不能存在丝毫的侥幸心理，否则后果非常严重，不但个人学术前程被阻断，也使导师和你所在的研究机构蒙羞。

也不要做自吹自擂，比如：在文章中说自己“第一个提出了某某方法”；不要打击同行，对同行的工作要客观而委婉地评价，绝不能不负责任地随意针砭；不要没参与工作，随便挂名。

3、学术勇气与耐力

年轻人一定要有学术勇气，敢想敢干，绝不能墨守成规，绝不能迷信权威。权威往往受到已有成见的束缚，他们忙于社会活动，不在研究第一线，他们的观点未必正确。因此，如果你得到了与权威不同的结论，并确信是正确的，要敢于坚持，当仁不让于师。

学术界有很强的马太效应，喜欢锦上添花，不喜欢雪中送炭。在你没有被业内认可以前，你的能力和成绩很可能会超过你获得的回报，在怀才不遇之际，个别学者会踏上脱离学术圈开展独立研究的道路，窃以为不可，要做大事，需要耐住寂寞，脚踏实地地工作，学会与人合作和交往，迟早破壳而出。

4、进入科学共同体  
在你具备了一定的基础后，可以通过发表论文，参加会议的方式进入学术界（科学共同体），这叫做入行了，出道了。在会议上宣读论文，听完别人的报告后当众提问或提出建议，在会间休息时主动结识知名学者，多多结交同龄的研究生和青年学者等等，都是进入“圈子”的途径。在公众场合，既要虚心地向成了名的“剑客”“侠客”们请教，又不必卑躬屈膝，令人生厌。“弱国无外交”，提高自身的研究水平是本，学术交往是末，你自己的工作做得出色，别人也乐于和你交往，反之，自己的工作稀松平常，得不到专家们的认可，只知交往，适得其反。

5、满怀研究乐趣

世界上有一种职业，别人花钱让你做你喜欢的事情，那就是“研究”。经常有激动人心的发现，充分表现自我，其中的乐趣“如人饮水，冷暖自知”。

科学家静心从事研究工作时发出阿尔法脑电波，据说和僧人入静时发出的脑电波是一样的，对人的身心健康很有好处，因此很多学者非常长寿，中文信息学会的创始人钱伟长先生今年97岁，用他的名字设立的首届“钱伟长奖”颁给了92岁的“亚伟速录机”的发明人唐亚伟先生。

如果你喜欢旅游，那么更要从事科学研究了，学术会议常常在风景如画的地方举行，阳光沙滩海浪，学者们在阳伞下轻松交谈，没有政界的险恶，没有商场的喧嚣，何等惬意。学者自然不如企业家富有，失去一些财富，换来的是自由。

今天是初九了，我的春节假日彻底结束，啰里啰唆，匆匆忙忙地了这许多文字，能够读到此处的读者一定很辛苦了，希望其中的一两个观点能够对您有所帮助。

（完）