

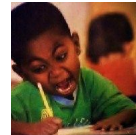


做失败科研 的10个方法

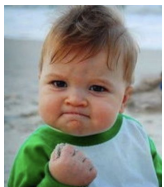
Kang Liu and Xianpei Han



5 Years (May be More) for Ph.D. Students



1st year



3rd year



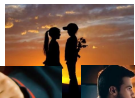
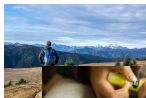
反正时间还早



5th year



2nd year



4th year



反正时间还早



延期



positive sample



negative sample



Classification

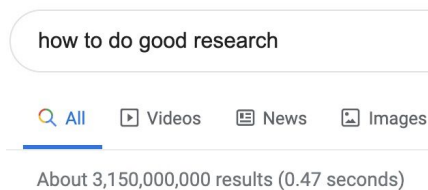


city
classifier



$p(\text{city}) = 85\%$

Positive Examples



施一公院士: 如何做科研



周志华教授: 如何做研究与写论文?



刘洋老师: 如何写NLP论文?



刘知远老师: 如何写一篇合格的NLP论文?



车万翔老师: 如何做一个精彩的学术报告

15 Steps to Good Research

1. Define and articulate a research question (formulate a research hypothesis).
[How to Write a Thesis Statement](#) (Indiana University)
2. Identify possible sources of information in many types and formats.
Georgetown University Library's [Research & Course Guides](#)
3. Judge the scope of the project.
4. Reevaluate the research question based on the nature and extent of information available and the parameters of the research project.
5. Select the most appropriate investigative methods (surveys, interviews, experiments) and research tools (periodical indexes, databases, websites).
6. Plan the research project.
[Writing Anxiety](#) (UNC-Chapel Hill) [Strategies for Academic Writing](#) (SUNY Empire State College)
7. Retrieve information using a variety of methods (draw on a repertoire of skills).
8. Refine the search strategy as necessary.
9. Write and organize useful notes and keep track of sources.
[Taking Notes from Research Reading](#) (University of Toronto)
[RefWorks](#)
10. Evaluate sources using appropriate criteria.
[Evaluating Internet Sources](#)
11. Synthesize, analyze and integrate information sources and prior knowledge.
[Georgetown University Writing Center](#)
12. Revise hypothesis as necessary.
13. Use information effectively for a specific purpose.
14. Understand such issues as plagiarism, ownership of information (implications of copyright to some extent), and costs of information.
[Georgetown University Honor Council](#)
[Copyright Basics](#) (Purdue University)
[Plagiarism: What It Is and How to Recognize and Avoid It](#) from Indiana University
15. Cite properly and give credit for sources of ideas.
[MLA Bibliographic Form](#) (7th edition, 2009)
[MLA Bibliographic Form](#) (8th edition, 2016)
[Turabian Bibliographic Form: Footnote/Endnote](#)
[Turabian Bibliographic Form: Parenthetical Reference](#)
[RefWorks](#)

Negative Examples

how to do bad research

[All](#) [Images](#) [Videos](#) [News](#)

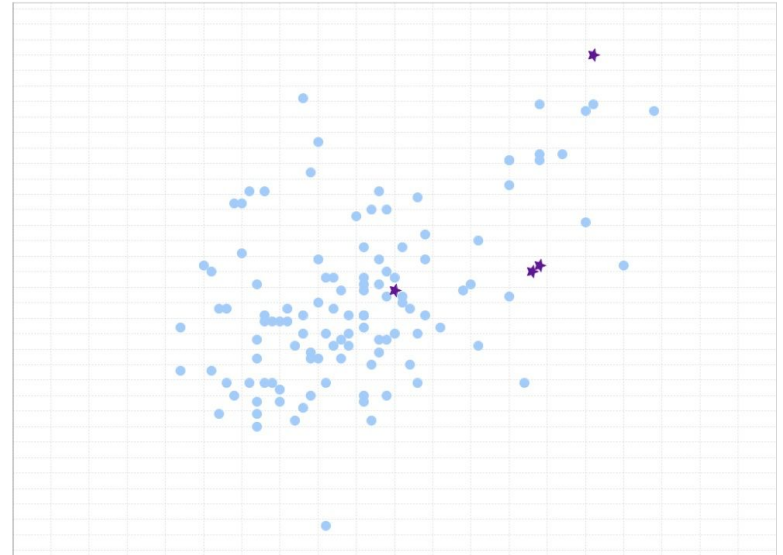
About 581,000,000 results (0.40 seconds)

VS.

how to do good research

[All](#) [Videos](#) [News](#) [Images](#)

About 3,150,000,000 results (0.47 seconds)



So...



2006-Now
14 Years

+



2005-Now
15 Years

=



29 Years

失败的10010个方法

1. 科研不就是发论文吗？
2. 选题当然选热门的！
3. IDEA天启论
4. 拿来主义，投机取巧！
5. 我的IDEA特别好，为什么要改？
6. 论文不就是个实验报告吗？
7. 导师为啥老说：你用一句话说清楚你的Idea？
8. 我的模型实验效果那么好，论文还能不中？
9. Reviewer纯外行，啥也不懂
10. 博士论文=Bag of 小论文

对待科研



Idea



创新性评估



写作



发表

1. 科研不就是发论文吗？



隔壁师兄都发了好几篇论文了...

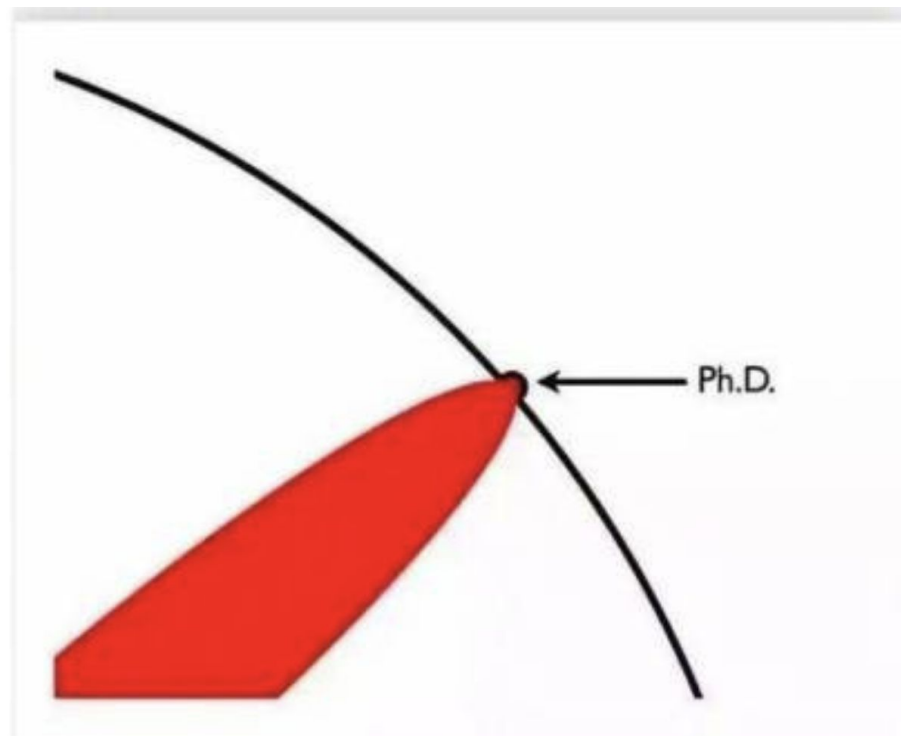
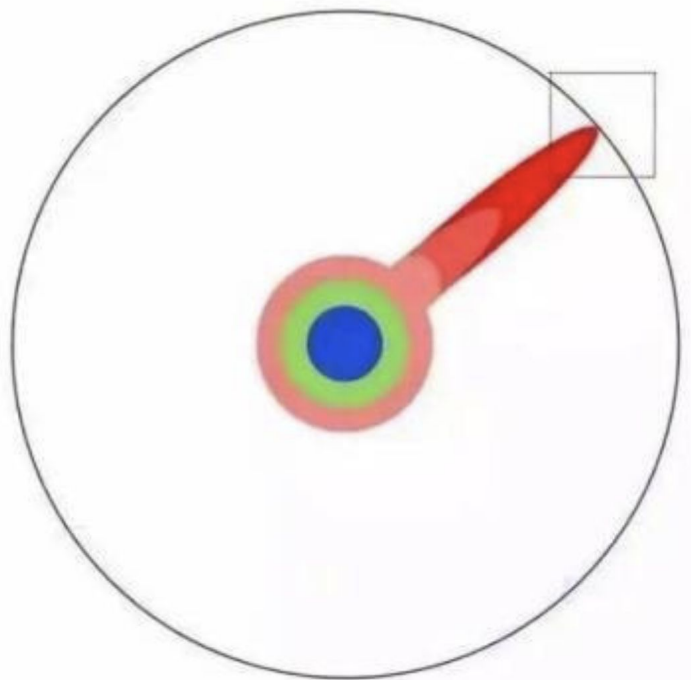
连晚来的师妹都发了好几篇论文了...

离毕业就差3篇论文了...



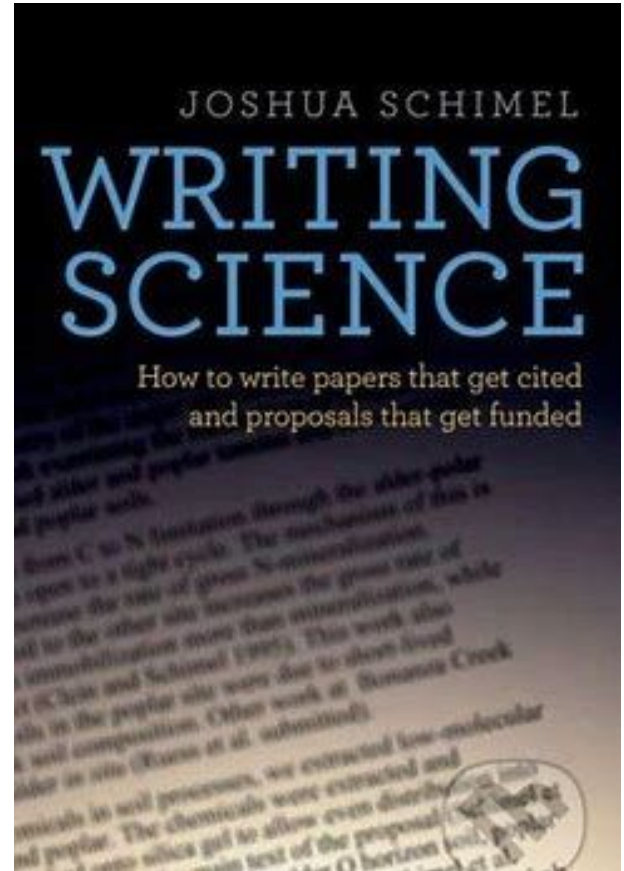
Publish or perish(发表或惩罚)!!!

博士生的科研



成功研究到底是啥？

But “publish or perish” is about surviving, not succeeding. You don't succeed as a scientist by getting papers published . **You succeed as a scientist by getting them cited .**



如何才能被引用？

以问题为导向，而不是以论文为导向

- 解决重要的问题
- 用简洁明了有意思高性能的方法
- 成功不是靠你发表论文的页数来决定的，而是靠他们的影响
- 论文是对你解决问题的阶段性成果的总结

为人民服务

为读者服务

- 论文让人一看就懂--最好让你奶奶都能看懂
- 实现起来及其简单--最好是实现好了别人即插即用
- 用了你的论文，腰不酸了，腿不疼了，连发表论文都更容易了



2. 选题当然选热门的！

技术妖魔化：三天一个“重磅”，一周一个“突破”，一个月一次“突破人类”



自媒体轰炸与大数据生存

原籍 清华大学社情与舆情研究所所长
清华大学公共健康中心特聘教授

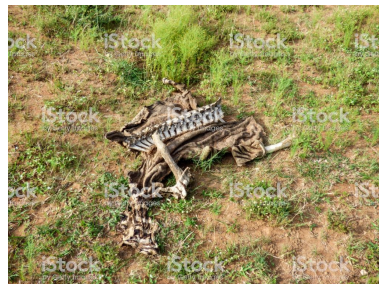
黄花菜都凉了



报考研究生的时候



选题的时候



要出成果的时候

一味追求速度使得每个问题浅尝辄止

冷静！

你应该做xxx方向

做的人还不是很多，避免踩踏事故

应该选NLP特有话题

好评测（评测是个大问题）

有数据集，有米下锅（没有的话自己创造）



导师

再冷靜！

避免Incremental Work

- 老問題、老框架、小調整

嘗試新問題

- 數據、評測、方法

嘗試新理論

- 理論證明、做別人沒做過的事情



新問題

新理論

新方法

身邊總是有些人



不斷挖洞給你跳

3. IDEA天启/神谕论

IDEA降临的时刻
到了!

- 牛顿的苹果掉下来
- 睡觉想到的苯环结构



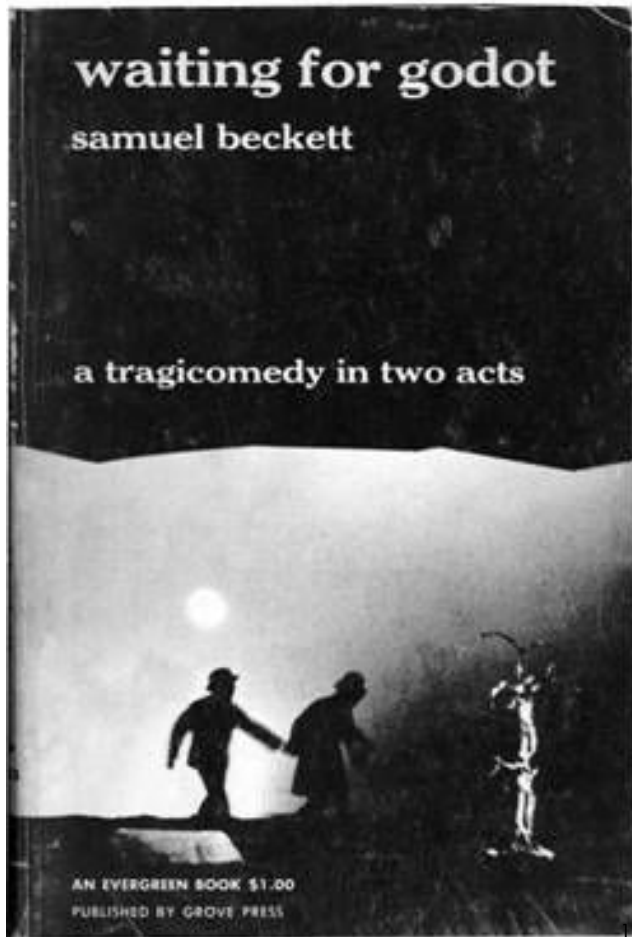
等待IDEA的神谕降临时

通常会：

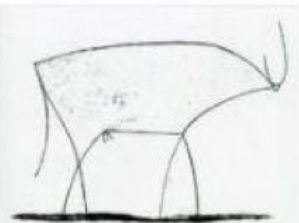
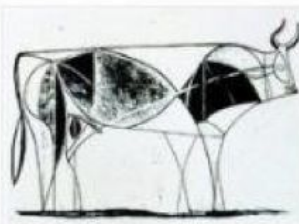
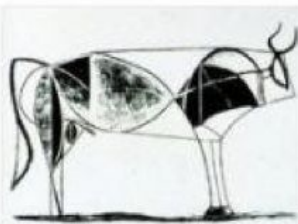
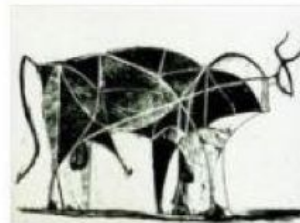
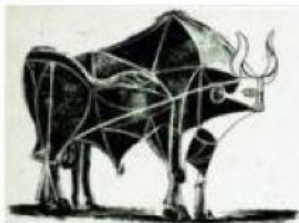
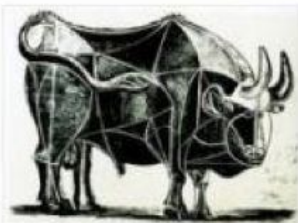
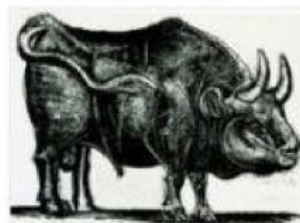
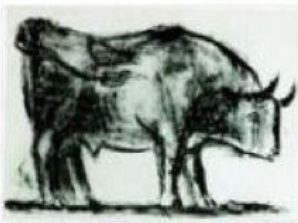
- 在苹果树底下等待苹果成熟
- 睡觉等待
- 上上网、打打游戏、聊聊天
- 向导师汇报：最近没有进展、实验效果不好、还没有想出来好的idea...

偶尔会：

- 看看论文
- 做做实验



即使你是 毕加索



正确的态度是--采矿

IDEA的发现过程更像是采矿

- 明确采什么--目标主题
- 勘探矿脉--大量阅读论文、综述
- 试采 -- 快速原型实验、数据观察
- 大规模开采 -- 实验、论文、迭代

谁看的准, 谁挖的深, 谁的想法就更好、更深入。



4. 拿来主义，投机取巧！

论文发表的第一步：大力出奇迹



捷径：拿来主义



图像处理领域的某个方法



语音领域的某个方法



NLP相关任务的某个方法



Transfer

task

自己的任务

1. 分类
2. 序列标注
3. 结构预测
4. 生成
5. 排序

1. 数据不足
2. 领域自适应
3. 多任务学习
4.

捷径：矩阵填空法

问题

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 6 | | | | | 1 |
| | 7 | | | 6 | | | 5 |
| 8 | | | 1 | | 3 | 2 | |
| | | 5 | | 4 | | 8 | |
| | 4 | | 7 | | 2 | | 9 |
| | | 8 | | 1 | | 7 | |
| | | 1 | 2 | | 5 | | 3 |
| | 6 | | | 7 | | | 8 |
| 2 | | | | | | 4 | |

方法



正确的打开方式

抽丝剥茧，目标任务本质/特有的关键问题是什么？

例如：

实体知识抽取的核心任务是抽取实体关系三元组

- 实体识别+关系抽取
- 传统方法很难利用统一模型完成上述任务，我们是不是可以在这方面做点工作？
- 我们利用端到端模型自动生成实体关系三元组 (ACL-2018)

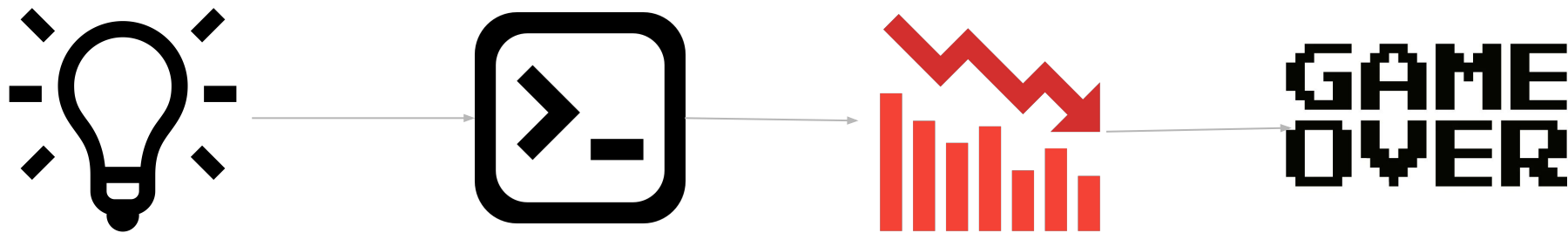


5. 我的IDEA特别好，为什么要改？



想象中的从IDEA到论文的过程

现实中的从IDEA到论文的过程



你们都不理解我idea的精妙之处，
连数据集和任务都出问题了！

IDEA不Work才是现实中的常态

- IDEA必然不work, 我们的核心是如何提升不work的价值
- 最终的论文远远不同于最早的idea
- 快速实现一个最小的可行模型, 迭代改进
- 每一周结束时的模型与开始时的模型应当非常不同
- 错误分析+借鉴可借鉴的一切思想

一份计划的全过程



这是最好的想法

好吧，这比我想象中难

完成这个任务需要花些时间

这令人厌恶，感到无聊

(深夜里思考的灵魂)

我做了但是遇到困难，没有我想象中那么糟糕

这时候结束是正确的，因为下次再做的时候我还能学习

* 从我的朋友莫林·麦克休那里“偷”来的

6. 论文不就是个实验报告吗？

Y任务是一个重要任务。在本文中我们第一次将X方法应用于Y任务，取得了Z%[显著的]性能提升。

...

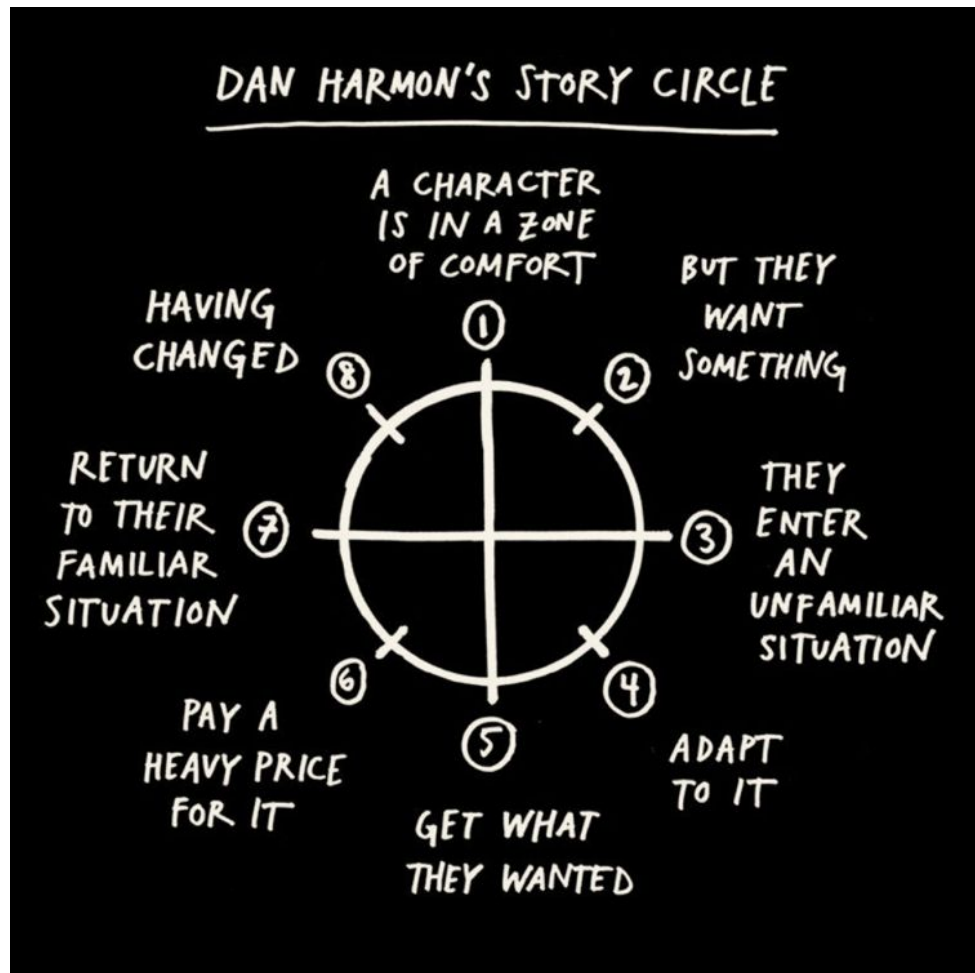
本文的核心贡献如下：

1. 第一次将X方法应用于Y任务
2. 取得了并不(划掉)显著的性能提升

| | | | | | |
|--|--------------|------|------|------|--|
| 课题 | 斜坡的启示 | | 实验名称 | 研究斜面 | |
| 班级 | 五 () 班 | 指导教师 | | 实验时间 | |
| 实验小组长 | | | 组员 | | |
| 实验材料 | 弹簧秤、小车、光滑的木板 | | | | |
| 实验过程： 1、组装器材。 2、通过测力计将小车提起，用多少力，记录在书上。 3、在桌面立一木块，从桌面到木块依次斜搭准备好的三块光滑木板，通过测力计沿斜面上拉动 小车，分别记录小车被抬高的高度，移动的距离及用力多少 | | | | | |
| 我们的发现：斜面可以 _____，并且在斜面其他情况相同的条件下， | | | | | |
| 实验结论：斜面可 _____ 情况相同的条件下， 越小越省力。 | | | | | |



没人喜欢听技术
没人喜欢看公式
但是所有人都喜欢听故事
所以：
论文应当是个故事！



论文就是童话--我听过的关于写论文的最好指导 (from 知乎)

1. 有一条巨龙抓走了公主 (你的问题为什么值得研究)
2. 巨龙是多么多么多么难打(强调你的研究的重要性)
3. 王子提着一把金光闪闪的剑而不是破斧子烂长矛登场 (你的方法好在哪里, 别人不行在哪里)
4. 王子是如何打败巨龙(你的方法简介)
5. 从此王子和公主幸福的生活在一起。(解决了问题)



7. 导师为啥老是说：你用一句话说清楚你的idea



学生

我这篇文章中传统方法通常做法是.....，他们的工作存在1.xxx问题；2.xxx问题,.....，我们这篇文章在xxx网络中引入了xxx特征以及xxxloss, 用以解决xxx的问题.....

你用一句话说清楚你的idea



学生



导师

你以为审稿人是怎么审稿的？



领域专家

其实他们是这样.....

他们可能不是领域专家



他们给每篇文章只留了~~120~~30分钟



他们可能审稿的时候情绪不稳定



他们的目标大部分情况下是



看懂不是目标, 写完 reviews才是目标

快速让审稿人理解到你的贡献才是关键

一句话说清楚idea

- 奥卡姆剃刀原则
- 题目，或者是文章中最关键的一句话
- Idea更清楚
- 降低审稿人理解的成本
- 如果你的导师和同学都需要你解释半天才明白你的贡献，那么这肯定没说清楚



8. 我实验效果那么好，文章还能不中？



性能超过State-of-the-art 5个点！！！！！！

你可能忽略了一件事情...

性能的提升和你的idea有个毛关系？



For Example

Contribution1:xxxxx

Contribution2:xxxxx

实验只说明达到了SOTA

这就完了?

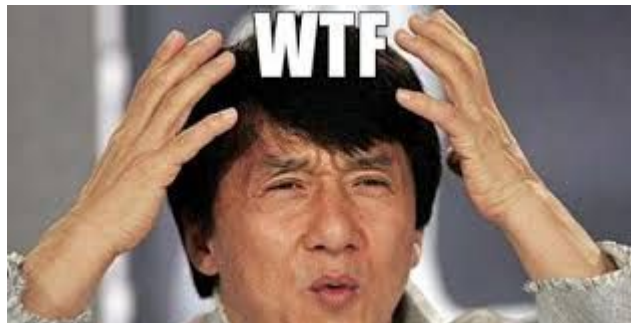


Contribution1:xxxxx

Contribution2:xxxxx

实验只说明达到了SOTA

实验说明3和4



失败了很多次以后，我终于觉醒了

1. Dataset and Evaluation
2. Experimental Settings
3. Our model vs. SOTA
4. The effectiveness of Contribution 1
5. The effectiveness of Contribution 2
6. Discussion or Ablation experiments
7. Parameter Influence

实验一定要充分，ACL(8页)，实验部分应该最少3页

9. Reviewer纯外行，啥也不懂



我看Reviewer似*B，料Reviewer看我亦如是

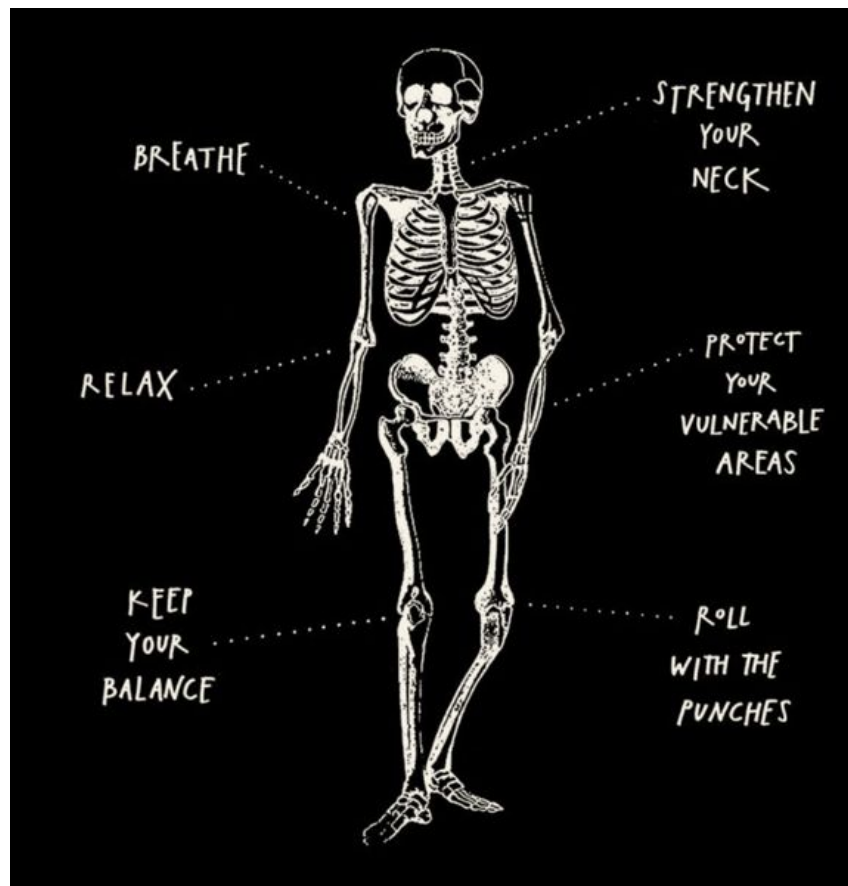
牢牢记住：

- Reviewer是免费劳动
- 没人喜欢看论文，特别是烂论文
- 你自己都瞧不上自己的论文，不过人家身上流的可是你的血...

这货的写作水平和展示的数据太流(shā)弊(bì)了，哥不得不提前下班，匆匆回家，然后花时间思考下人生。。。哥建议我们还是成立个基金，以买单审稿人审阅时可能需要的红酒，哥上火呀。请勿发表，并建议锁定该作者的电子邮件ID，避免此人日后继续投稿。。。

-- 《史上最狠的审稿意见》及《对审稿人伤害最深的烂论文》

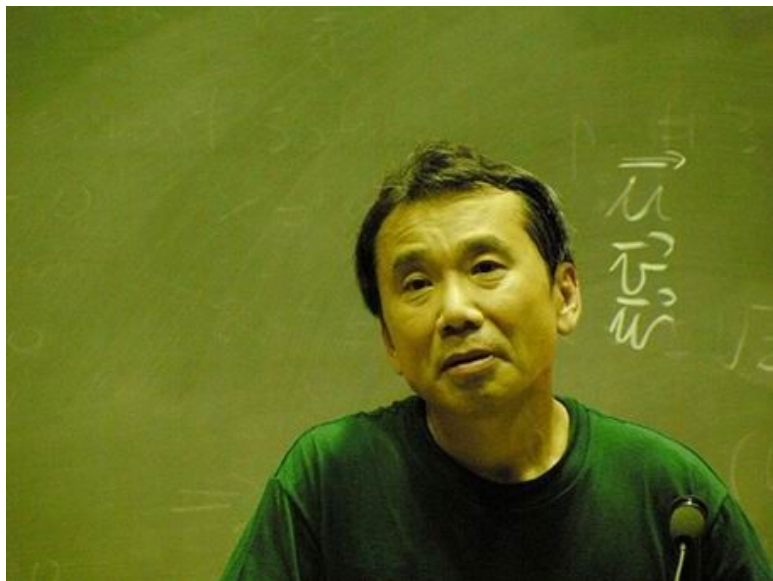
学会挨打



正确态度：看Reviewer如男/女朋友

如果Reviewer(包括其他人)针对一个地方提出修改意见，类似于“**你自己错在哪儿了吗？**”

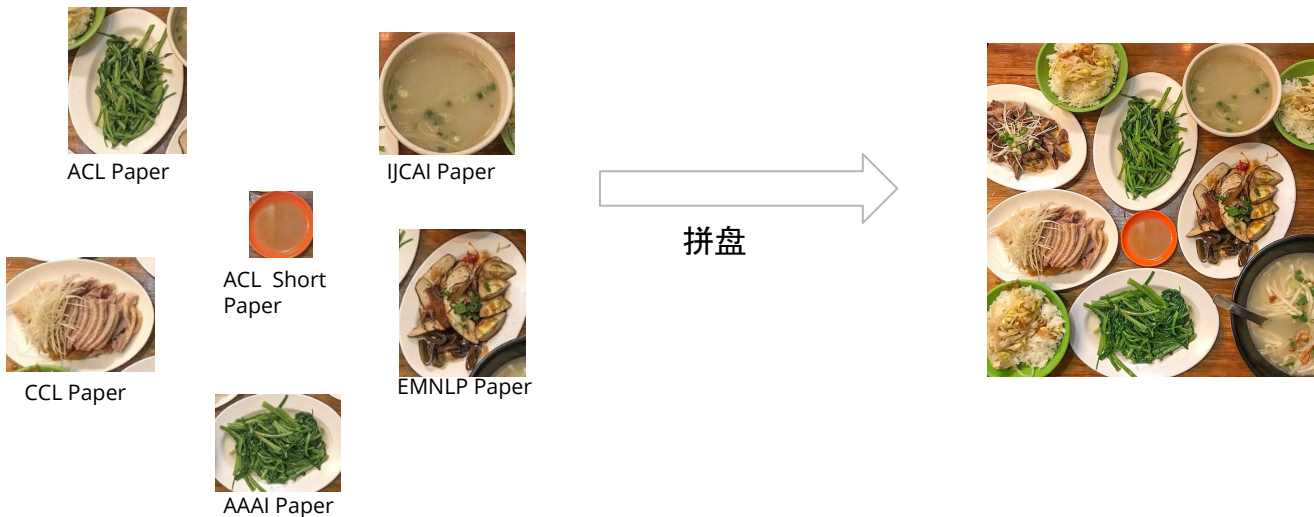
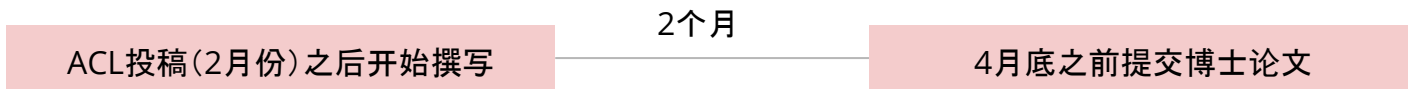
- 都是你的错，错的原因一定是你
- 修改意见不一定是对的
- 但是有意见的地方一定有问题



重要的是修改这一行为本身。下定决心要把这里修改的更好，静心凝神在书桌前坐下，着手修改，这种姿态便具有无比重要的意义

10. 博士论文=Bag of 小论文

博士论文撰写的一般时间(7月份毕业)



xxx的关键
技术研究

问题

大学博士学位论文学术评议书
(评议书请用黑色墨水笔书写或直接套打印, 不得粘贴)

| | | | |
|------|------------|------|------------|
| 论文题目 | [REDACTED] | 学生姓名 | [REDACTED] |
| | | 学 科 | |

对论文的学术评语(论文选题及成果对学科或产业发展的作用和意义, 从论文科研工作中反映作者的基础理论和专门知识属何水平, 作者对本课题范围内的国内外发展动向、重要文献资料是否有较全面的了解和评述, 论文有无错误, 总结写作水平如何):

很遗憾, 这是一篇博士论文, 如果是硕士论文, 我评价还是可以的。作者首先在 [REDACTED] 题目的设计上, 做了一些有难度的问题, 利用神经网络 [REDACTED] 到去年, 确实也算有一定的探索价值。但总的来说, 这篇论文在基本点, 距离非一个五年博士训练所能达到的岗位要求, 距离两百多篇引文, 八十多页的论文, 距离对本领域的国际研究动态、发展水平、基本问题的叙述, 距离对基本问题的叙述, 都看不出作者对自己五年博士期间從事行業有什麼深入了解, 更不用說國際的見解了。自己工作的事情, 完全是一個實驗報告的水平, 這個博士論文, 我完全看不出作者對自己從事的問題的認識程度。此外, 作為自己工作一部分的附帶品, 作者本人只是第二作者, 把別的人的工作放入自己的博士論文, 一篇論文申請個學生申請學位, 是不合適的。這一點, 導師在同意這篇論文送審時, 是有責任的。總之, 作為一個從2016年起帶博士生的教授, 我完全不同意這一論文達到博士論文的基本要求, 建議轉為碩士學位。

(未详尽处接背面)

| | | | |
|-------------|----|-----------------|-----------|
| 评 阅 编 号 | | 是否 博 士 生 导 师 | 是 |
| 专 业 技 术 职 称 | 教授 | 评 阅 意 见 填 写 日 期 | 2018/4/28 |

...论文系统性不足...

博士毕业



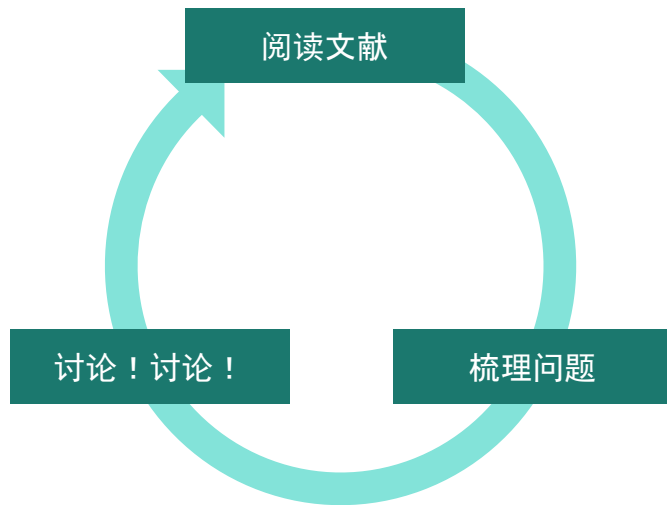
- ACL x篇
- EMNLP x篇
- NAACL x篇
-

未雨绸缪

从读博开始明确研究目标

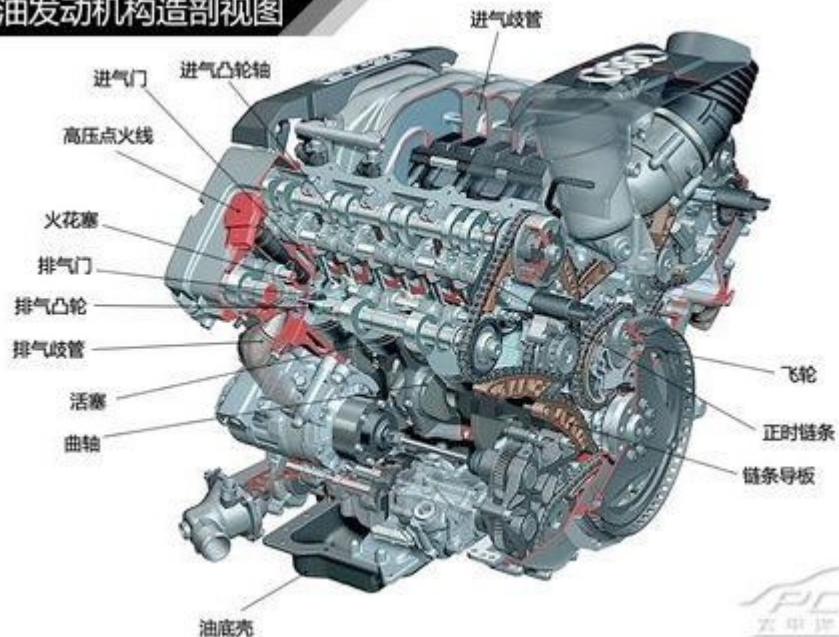
不求处处开花，只求一点突破

重视开题和中期



亡羊补牢

汽油发动机构造剖视图



大马力汽油发动机开发关键技术研究



建造一款大马力汽油发动机所需的关键技术有：

1. 气缸
2. 火花塞
3. 曲轴
4.



本论文解决的是：

1. 曲轴的设计与构建
2. 火花塞的设计
3.

结束语

失败的教训千篇一律，成功的经验万里挑一

请大家谨记我们的教训



失败的100另外90个方法

TO BE CONTINUED...

谢谢