

# 像玩电子游戏一样做科研

清华大学计算机系数据库组

李国良

# 科研动力

---

- ◎ 学位
- ◎ 就业发展
- ◎ 父母家人
- ◎ .....
  
- ◎ 兴趣、投入

# 谁痴迷过电子游戏?

---

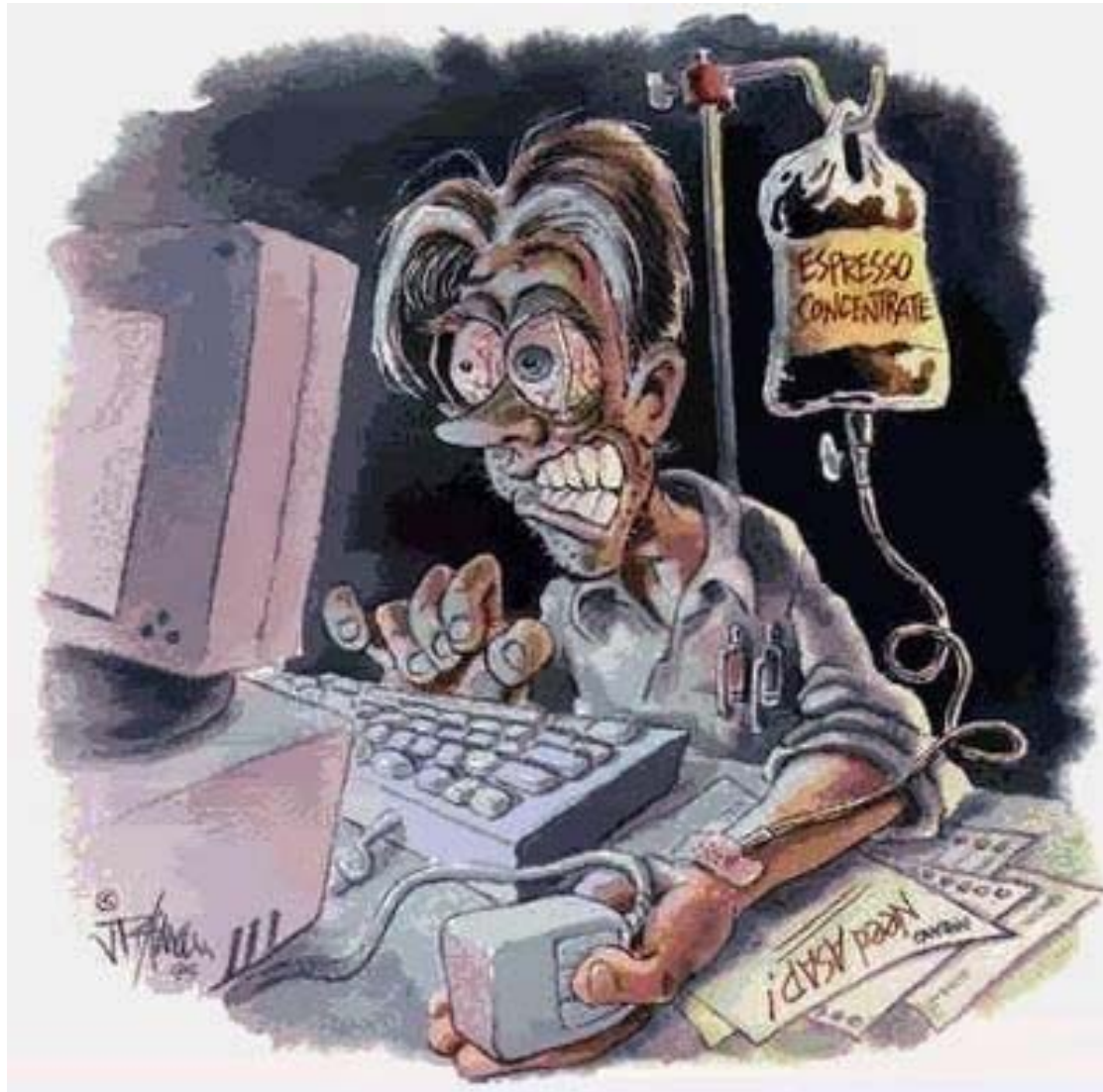


# 谁痴迷过电子游戏？

---



# 谁痴迷过电子游戏?



# 谁痴迷过科研？

- ◎ 晚上10点多下班路上遇到prof tan
- ◎ 一直听说tan老板每天公车上下班，在车上都读paper，今儿眼见为实
- ◎ 真是一位勤恳朴素负责的好老板啊



# 如何玩好游戏？ 如何做好科研？

---

◎ 痴迷

◎ 玩喜欢的游戏

◎ 勤加练习

◎ 攻略秘籍

◎ 切磋对战

◎ 天王

◎ 感兴趣

◎ 选合适的方向

◎ 多读多写

◎ 多交流讨论

◎ 多投稿

◎ 大师

## 上瘾

# 如何找合适的科研方向？

---

- ◎ 问老板、师兄师姐
- ◎ 自己多读论文、多交流
  - 读三大会议论文
  - 多听报告
  - 多参加会议
- ◎ 交流访问
  - 公派出国
  - 有一定科研基础再出去交流
  - 了解科研动态和领域牛人



**Michael Stonebraker**



# 好科研方向应具备的条件

---

- ◎ 要有实际需求 - 重要
- ◎ 能吸引人 - 新颖
- ◎ 有较大的未知空间 - 非显而易见
- ◎ 实验室有好的相关积累 - 基础

# 如何读论文？

---

## ◎ 找方向时

- 广度优先
- 看idea，不用注重细节

## ◎ 找到感兴趣方向时

- 深度优先
- 看细节
- 看实验、数据集
- 能够实现别人代码
- 抓住论文的要害
- 批判式阅读

# 读哪的论文?

---

## ◎ 近3-5年三大会议的论文

- SIGMOD
- VLDB
- ICDE

## ◎ 经典的论文

- 索引: B-tree、R-tree
- Join 算法: Merge-join、Hash-join
- 各个方向的经典算法: knn、TA

# 如何判断一个问题的价值？

---

## ◎ 老问题

- 问题已经明确定义
- 目标：更好的方法

## ◎ 新问题

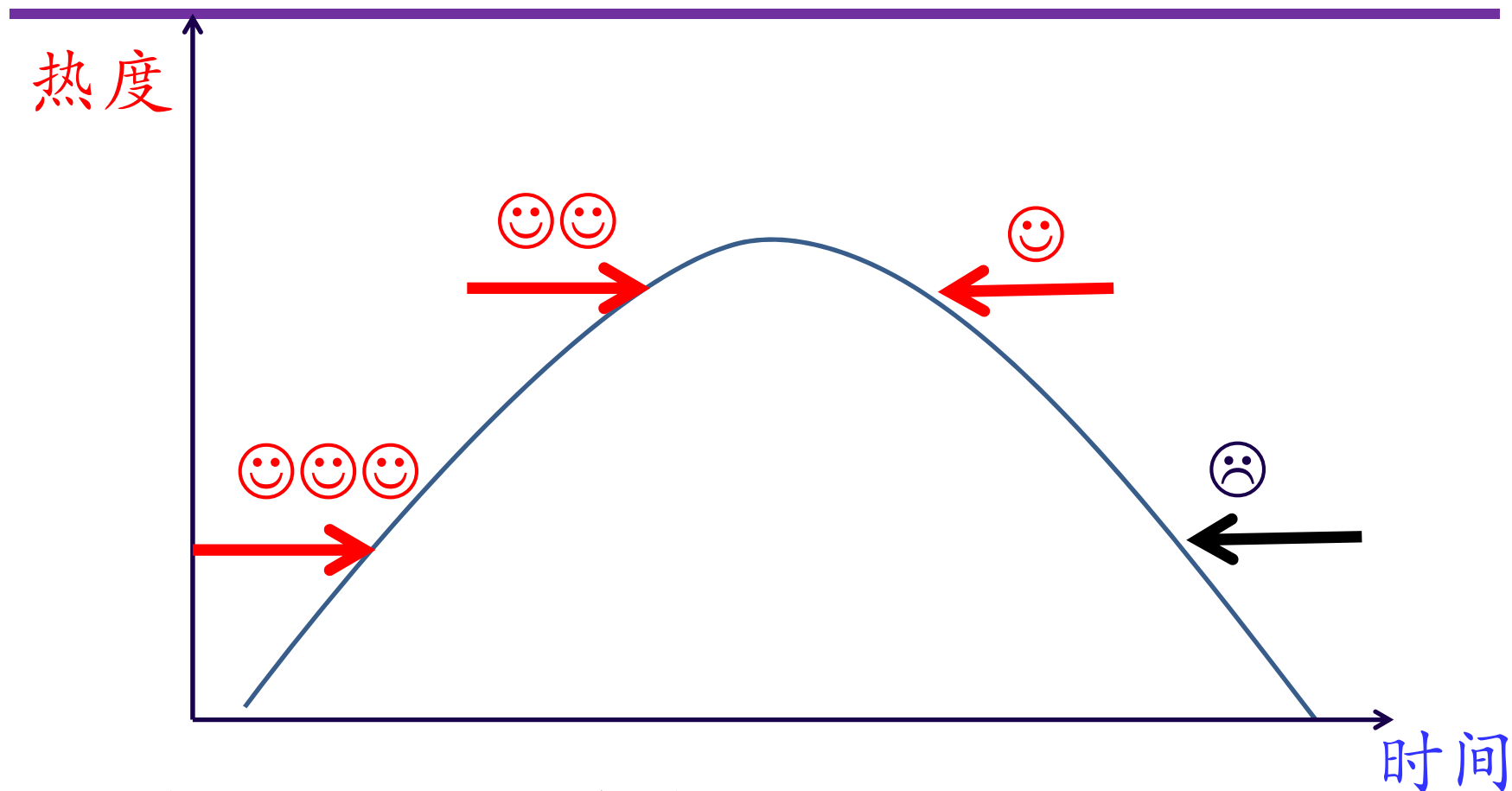
- 目标：解决没人研究过的问题
- 条件
  - 有真实的应用背景
  - 能获得真实的数据集
  - 结果好衡量，如果需要衡量效果（quality），最好有 benchmark

# 如何做老问题？

---

- ◎ 更快、更高、更强 (Faster, Higher, Stronger)
  - 提高性能
  - 提高准确率
  - 提高通用性 - 去掉参数或者假设
  - 充足的实验比较
- ◎ 应用到新的环境
  - 新硬件、新场景
  - 从RAM到Disk
  - 增量算法、并行算法

# 科研问题的生命周期



如何判断问题是否过时？

三大会议（SIGMOD、VLDB、ICDE）论文多少

# 如何找新问题？

---

- ◎ 多想、多联系实际
- ◎ 读交叉学科论文
- ◎ 新问题需要的条件
  - 问题意义 - 令人信服
  - 应用背景
  - 可以解决
- ◎ 当前新问题例子
  - 众包 - Crowdsourcing
  - 大数据管理 - big data

# 如何研究一个问题？

---

- ◎ 多想多借鉴
- ◎ 创新 - 不能重复别人工作
- ◎ 发散性思维
- ◎ 抽象能力
- ◎ 深入分析
- ◎ 有毅力



# 做研究需要具备哪些素质

---

- ◎ 脚踏实地（努力）
  - 不能熊瞎子掰棒子
- ◎ 规划能力（远见）
- ◎ 了解相关工作，做好实验（知彼）
- ◎ 认清论文质量（知己）
- ◎ 多总结，多借助于工具（总结）
  - Latex、SVN、gnuplot
  - 总结经验教训

# 做研究大忌

---

- ◎ 没有兴趣
- ◎ 不了解前人的工作
- ◎ 浮躁、急于求成、不求甚解
- ◎ 马虎
- ◎ 懒惰

# 科研三个阶段

---

## ◎ 三个阶段

- 茫然 - 没idea
- 辨别 - 有idea, 但不知道idea是否可行
- 贯通 - 能够衡量idea价值, 能帮别人想idea

## ◎ 小论文 vs 大论文

- 先小后大
- 直接投大论文

# 如何衡量一个工作的价值

---

- ◎ 问题新颖程度
- ◎ 方法创新程度
- ◎ 技术深度
- ◎ 实验结果好坏
- ◎ 论文撰写功底（逻辑结构、英语）

# 如何写论文？

---

- ◎ 八股
  - 模仿
- ◎ 摘要
- ◎ 引言
- ◎ 相关工作
- ◎ 技术细节
  - 3-4个创新点
- ◎ 实验
- ◎ 结论

# 摘要

---

- ◎ 论文的高度总结（精华）
- ◎ 问题：what（解决什么问题？）
  - 背景意义
- ◎ 动机：why（为什么写这篇论文？）
  - 比别人更快，更好，更通用
  - 别人没研究过的问题
- ◎ 方法：how（介绍基本idea）
- ◎ 贡献：主要贡献
- ◎ 实验：实验效果

# 引言

---

- ◎ 摘要的扩充版，论文的总结版
- ◎ 解决什么问题
- ◎ 有什么应用
- ◎ 研究的动机
- ◎ 基本idea
- ◎ 创新点（含实验结果）
- ◎ 文章结构

吸引评审人，几乎决定文章能否被录取

# 注意事项

---

- ◎ 突出创新点
- ◎ 突出技术深度
- ◎ 明确说出和现有方法区别
- ◎ 细心检查
- ◎ 实验可重复
  
- ◎ 形成自己的风格



# 文章结构比英语更重要

---

- ◎ 结构清晰
  - Leading text, Highlight
  - 一定要自包含
  - 贯穿全文的例子
- ◎ 图文并茂
- ◎ 提纲挈领
- ◎ 言简意赅
  - 避免重复文本

让读者更容易理解

# 相关工作

---

- ◎ 引用相关论文，说出区别
- ◎ 多夸奖，承认别人的贡献
- ◎ 保证语言正确
- ◎ 不要抄袭别人语句

# 结论

---

- ◎ 总结文章贡献
- ◎ 不要与引言重复
- ◎ 和引言不同的语调
  - 引言 - 读者还不知道技术细节
  - 结论 - 读者已经看完了论文

# 论文录用

---

- ◎ 总结经验，继续研究下一个问题
- ◎ 准备presentation
  - 吸引人
  - 容易理解
  - 不要太多技术细节
  - 不要太多文字
  - 让人记住你的工作

# 论文投稿与被拒

---

- ◎ 伤心1-2天就让他过去
- ◎ 不要气馁
  - 任何人的论文都可能被拒
  - 接受率只有16%左右
- ◎ 找原因，根据reviewer意见修改（如果需要）
- ◎ 看清自己论文的分量

---

在科学上没有平坦的大道，只有不畏劳苦沿着  
陡峭山路攀登的人，才有希望到达光辉的顶点

--- 马克思

**相信自己， 坚持， 坚持， 再坚持！**