日常积累

浙江大学 陈为

目录

- 收集与检索
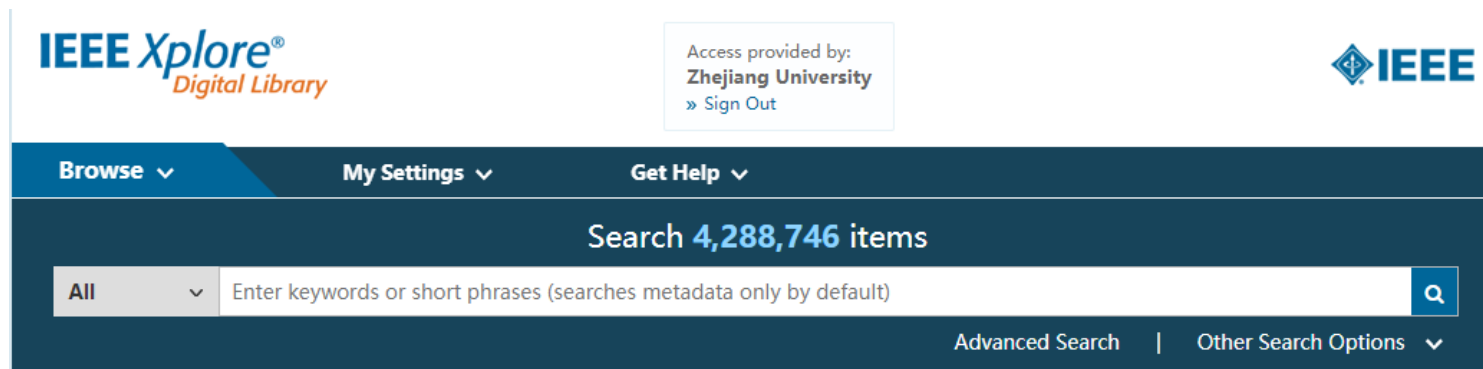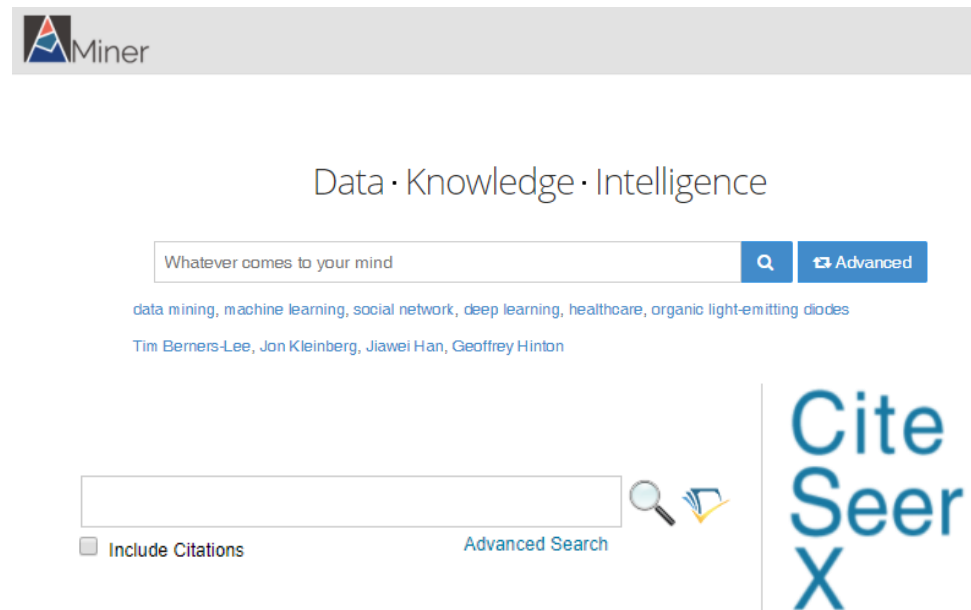
- 阅读与归纳

- 讨论与交流

- 记录与回顾

1 收集与检索

文献收集

- 确定文献收集的范围
  - 确定研究主题，提取主要关键字
  - 根据对领域发展状况和前人工作的了解，确定检索年份、期刊类型等
  - 从文献分布情况出发，确定过滤条件，如影响因子，引用数量等
- 准备文献清单（统计表）
  - 记下主要范围和研究主题所涉及的主要问题，然后带着这个目的去读，去找寻相关资料和回答问题

学术搜索引擎

- Google Scholar
- Arnetminer
- Microsoft' Academic Search
- Elsevier' Scirus
- CiteseerX
- IEEEXplore
- ACM Digital Library

与领域相关的资源

- 相关会议文章

- 相关期刊文章

- 大牛、小牛的主页
  - 了解一个领域最顶尖的研究者的关注方向

以可视化为例

- 相关会议
  - IEEE VIS，Euro VIS，Pacific VIS

- 相关期刊
  - TVCG，JOV，CG&A, CGF
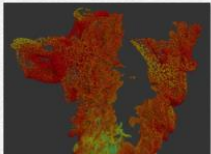
- 大牛，小牛主页
  - 马匡六，Jeffrey Heer等等
  - Jian Zhao等等

搜索基本原则

- 善于搜索，善于筛选、取舍，在浩瀚知识库中寻找到有用的信息

- 注意搜索策略，合理选择关键词

- 采用"高级检索"协助文献的甄选

- 非正式资料不宜用作正式的参考文献

滚雪球式搜索策略

- 书籍
  - 权威书籍的索引条目（index）以及版权页中的主题（subject）可作为学术查询的关键词
- 论文
  - 找到强相关、高价值的论文后，查找论文的参考文献与引证该论文的文献进一步阅读
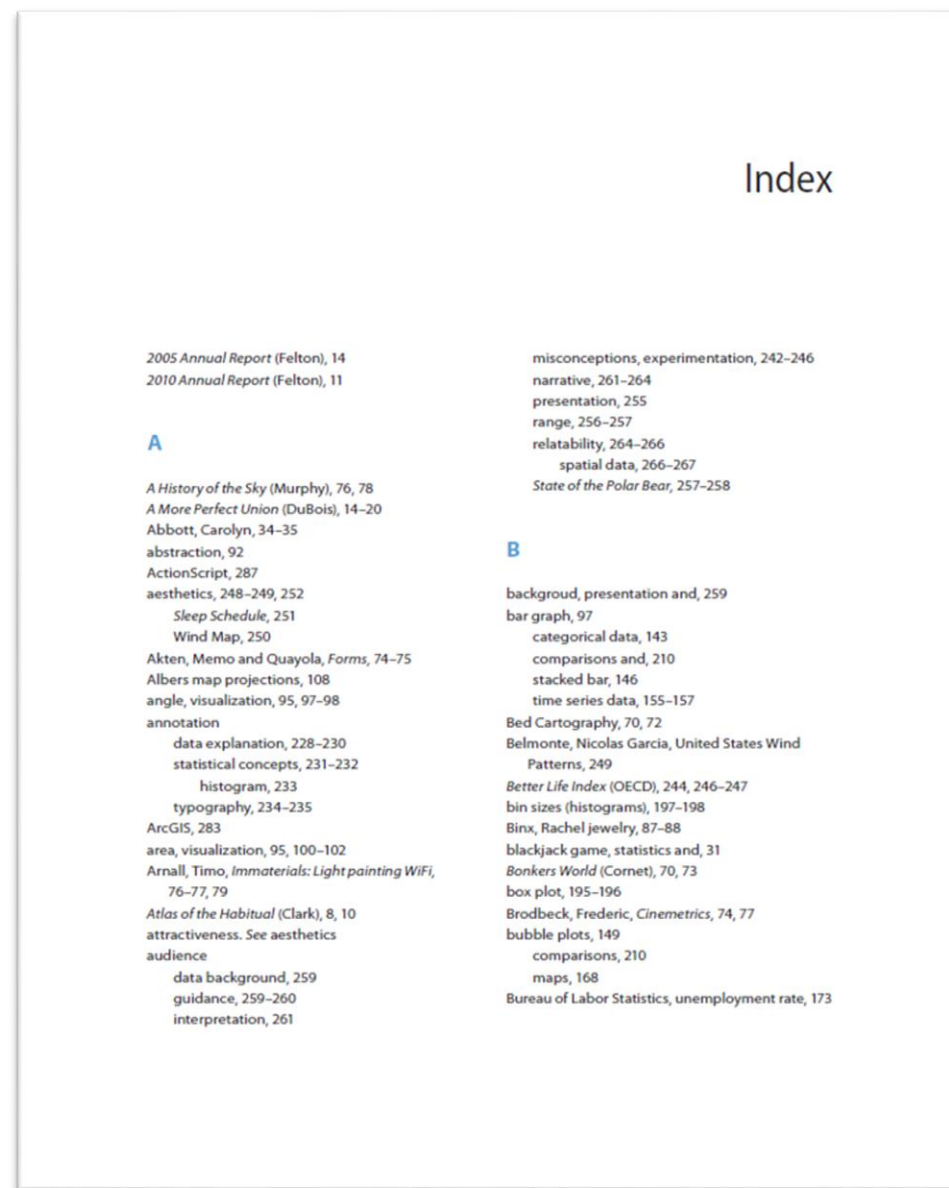
Index

2 阅读与归纳

文献的一般结构

文献的一般结构

文献的一般结构

引言：介绍论文的背景、研究问题、研究动机、主要贡献等等内容，引出论文的主体，引导读者理解全文

文献的一般结构

相关工作：介绍与论文相关的工作，对这些工作进行总结并与论文提出的方法进行对比

文献的一般结构

Title

Abstract

1 Introduction

2 Review of Previous Work

3 A new Method for …

   3.1 Framework of …

   3.2 Major Components of …   → 方法：介绍论文方法的细节

      3.2.1 Feature Extraction

      3.2.2 SVM with a New Kernel Function

   3.3 Time and Space Complexity

4 Experiments

   4.1 Comparing with Previous Methods

   4.2 Parameter optimization

   4.3 Discussions

5 Conclusions

文献的一般结构

Title

Abstract

实验：介绍对方法进行验证的实验细节，过程和结果，并与已有方法进行对比，证明新方法的优势。

文献的一般结构

总结：对全文工作的总结，包括工作的不足、未来工作等等

方法一——咬文嚼字阅读法

逐字逐句，从头到尾细细品读

方法二——见坑就跳阅读法

Title

Abstract

1　Introduction

2　Review of Previous Work

3　A new Method for …

　3.1　Framework of …

　3.2　Major Components of …

　　3.2.1　Feature Extraction

　　3.2.2　SVM with a New Kernel Function

　3.3　Time and Space Complexity

4　Experiments

　4.1　Comparing with Previous Methods

　4.2　Parameter optimization

　4.3　Discussions

5　Conclusions

| Related work 1 |
| Related work 2 |
| Related work 3 |
| Related work 4 |

| SVM |
| Kernel Function |
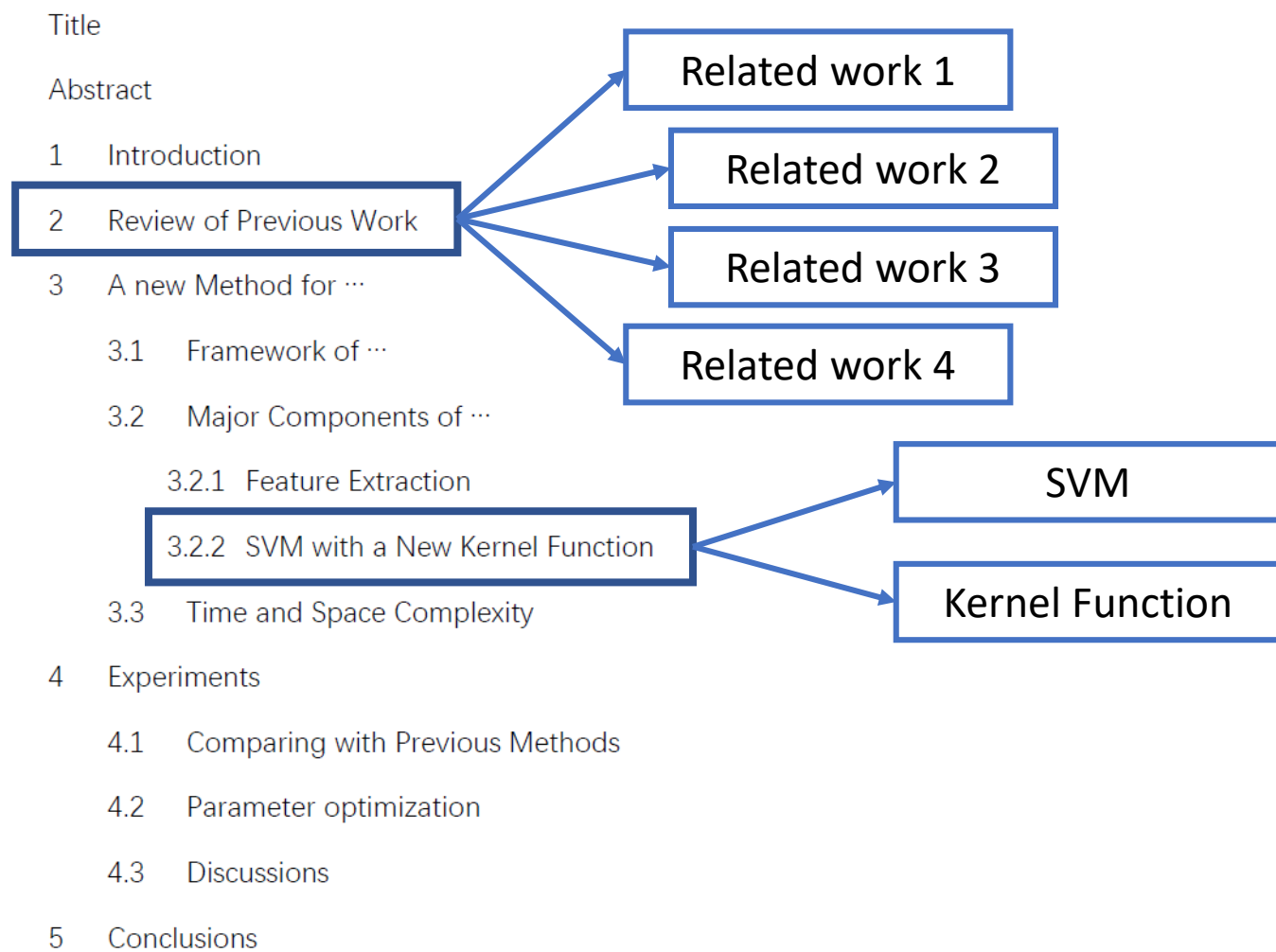
方法三——两步阅读法

- 第一遍：快速浏览全文，做大致了解
  - 研究问题
    - 是否相关？
    - 先自己思考可行的解决方案
  - 假设条件、主要想法、解决方案

方法三——两步阅读法

- 第二遍：细致阅读，仔细推敲
  - 精读文章中最精华的部分
  - 泛读细节部分
  - 略过无帮助的部分

阅读重点的分布

Title

Abstract

1   Introduction

2   Review of Previous Work

3   A new Method for …

   3.1   Framework of …

   3.2   Major Components of …

      3.2.1 Feature Extraction

      3.2.2 SVM with a New Kernel Function

   3.3   Time and Space Complexity

4   Experiments

   4.1   Comparing with Previous Methods

   4.2   Parameter optimization

   4.3   Discussions

5   Conclusions

往往需要精读

时间不允许的情况下可以考虑略过
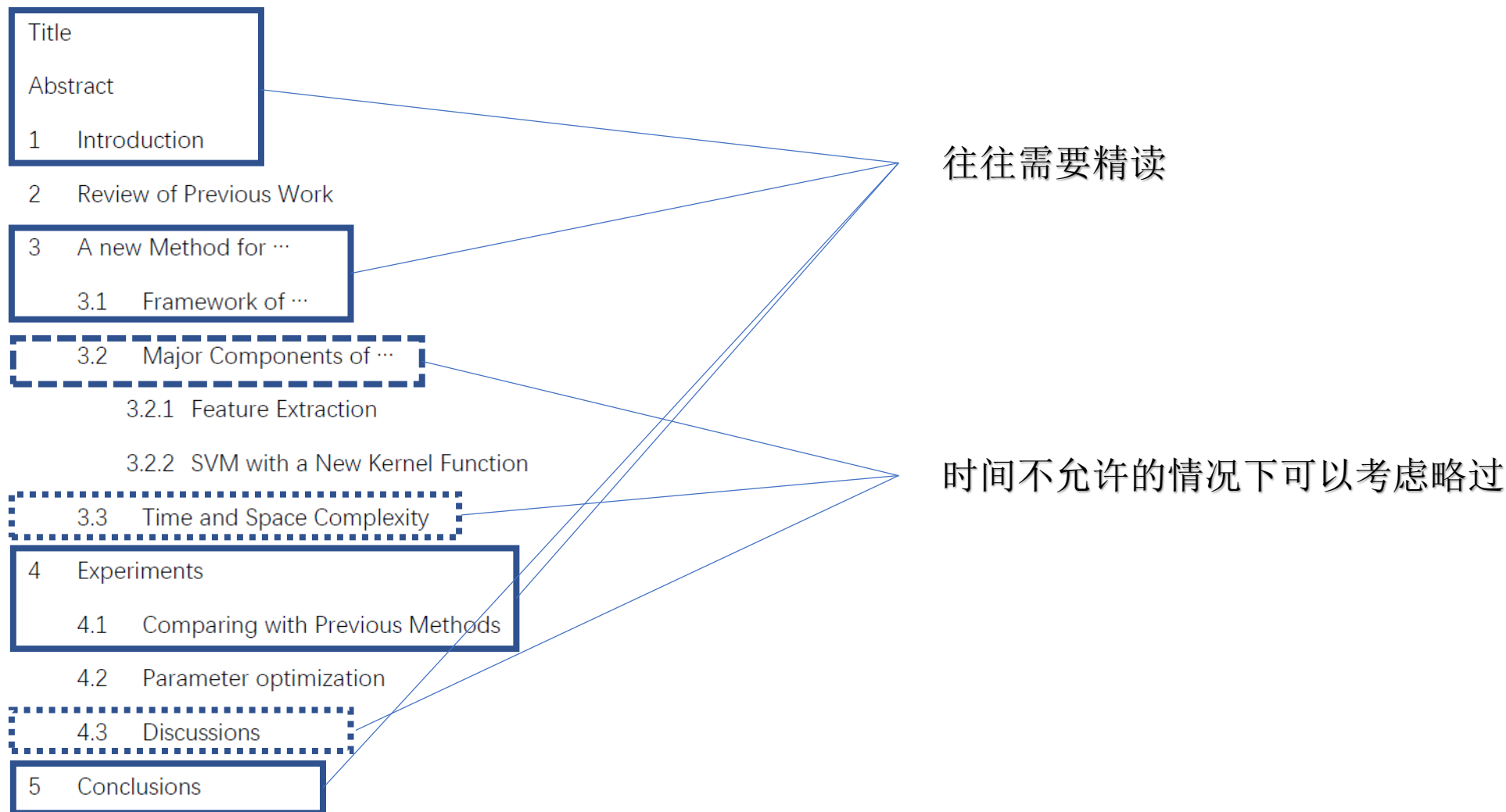
正确的做法——30%阅读，70%思考

- 带着问题去阅读

  - 这篇论文主要表达了什么？写作目的是什么？做出了什么贡献？

  - 这篇论文的成果/观点是如何体现的？论据有哪些？

  - 相关工作有哪些？作者站在了哪些巨人的肩膀上？相比较已有研究成果，作者的创新点在哪？

正确的做法——30%阅读，70%思考

- 批判性思考
  - 以带有批判性的眼光思考论文中的研究问题
  - "鸡蛋里面挑骨头"
  - 假设条件是否成立？
  - 提出问题是否合理？
  - 解决方法有没有缺陷？

正确的做法——30%阅读，70%思考

- 创造性思考

  - 用带有创造性的头脑思考针对这一问题的解决方法

  - 除了作者给出的方案，有没有更好、更新或是不同的解决方案？

不同的场景，不同的方法

- 训练写作能力时
  - 找几篇范文，咬文嚼字的读，学习论文写作方法

- 需要深入了解某一个领域时
  - 从某篇高引论文开始，见坑就跳的读，广泛的了解相关工作

- 能力成熟之后
  - 分两步读论文，了解领域的最新进展，启发自己的研究思路

文献综述

- 对一个研究领域的阶段性总结

  - 在一个领域中主要的研究者

  - 最新的进展和发现

  - 现有研究工作中明显的漏洞

  - 领域中现存的争论

  - 未来的研究方向

Writing a Summary

-shorter than the original ✓

-includes only the main points ✓

-is objective ✓

-includes in-text reference ✓

-in the writer's own words! ✓

文献综述

- 例子：A Survey of Traffic Data Visualization
  - 交通数据可视化的任务
  - 交通数据的处理方法
  - 交通数据的可视化方法
  - 交通数据可视化的作用

"文献综述"与"文献回顾"

- *文献回顾* 倾向于show文献，而 *文献综述* 倾向于use文献

- 共同点
  - 都是从某一方面的专题研究论文或报告中归纳出来

- 不同点
  - "文献回顾"在于把相关文献客观地归纳报告；
  - "文献综述"要求对文献资料进行综合分析、归纳整理，使材料更精练明确、更有逻辑层次，并带有专门、全面、深入、系统的论述。

什么时候读文献综述

- 初入一个研究领域
  - 利用综述文献的过程了解一个领域的现状

- 每次撰写文章时
  - 总结与文章相关的领域的现状，突出文章的贡献

- 对一个领域研究非常深入时
  - 总结领域的研究，对未来的研究方向进行展望

对个人的影响

- 学知识
  - 通过搜集文献资料过程，可进一步熟悉科学文献的查找方法和资料的积累方法；在查找的过程中同时也扩大了知识面；

- 打基础
  - 查找文献资料、写文献综述是科研选题及进行科研的第一步，因此学习文献综述的撰写也是为今后科研活动打基础的过程；

- 提能力
  - 通过综述的写作过程，能提高归纳、分析、综合能力，有利于独立工作能力和科研能力的提高。

归纳相关工作

- 对与文章相关的数个方向分别进行总结

- 例子：LineUp: Visual Analysis of Multi-Attribute Rankings

  - Spreadsheets

  - Point-Based Techniques

  - Region-Based Techniques

  - Line-Based Techniques

  - Ranking Visualization Techniques

归纳相关工作

- 一般套路
  - 第一段：说明有哪些方向与本文相关
  - 分节讨论各个方向，在每个方向中：
    - 某类方法的做法（总）
    - 某类方法中的工作的具体做法（分）
    - 某类方法的优势，但是**与本文方法相比的劣势在哪**（总）

归纳整理

- 根据最初的统计清单，对收集到的文献进行分类整理
  - 按算法进行归类
  - 按数据进行归类
  - 按研究问题进行归类
  - ...
- 根据分类结果对文献进行归纳总结
  - 用清晰、简短的引言，说明综述的大纲，并简要说明论点顺序
  - 论点与论据之间要有清晰的联系，在开头或结尾应有简短的总结
  - 对于与论点有分歧的文献，应仔细斟酌，不可直接忽视

归纳整理

- 两步阅读法中的第一步

  - 通过阅读摘要，引言和结论部分，快速了解每篇文章的内容

  - 大部分细节都可以跳过，只抓重点

- 文献管理工具

  - Mendeley

  - EndNote
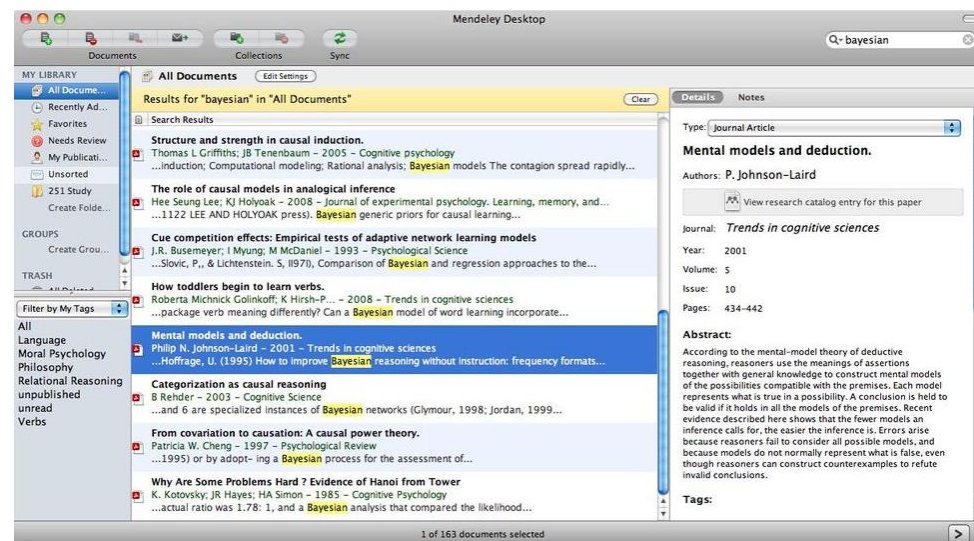
归纳相关工作

- 常见误区

  - 仅仅对相关的论文进行罗列，而不进行任何比较

    - 分类并比较和对比某一领域不同作者不同的观点

  - 为了使自己的工作看上去更好，而将别人的工作写的一无是处

    - 首先要肯定别人的工作，然后适度的点出最要紧的缺陷即可

交流　　　　　　　　3 讨论与交流

组会

- 目的
  - 多个方向的交叉碰撞
  - 提升团队凝聚力
  - 锻炼口头沟通技能
- 参与人
  - 全组同学

讨论班

- 目的
  - 汇报最新的国际学术进展
  - 深度研讨科研主题
  - 实战模拟会议报告
- 参与人
  - 关注某个子领域的部分同学

讨论班

- 几种形式
  - 论文宽读
  - 论文精读
  - 论文共读

2021-11-01 Autoaugment + 对比学习中的数据增强 昨天 23:31

2021-10-11 Deep learning for entity matching 10-11 19:49

2021-09-27 Continual Learning 09-27 19:41

20210922Transferable Contrasive Network for Generalized Zero-Shot Learning 09-22 19:30

2021-09-16 Product Manager in ML & Federated Domain Generalization 09-27 17:04

2021-09-09 Machine Unlearning 09-09 16:29

2021-09-02 Privacy in learning: Basics and the Interplay 09-05 21:53

2021-08-26 奇怪的Python小知识 08-26 17:47

2021-08-19 Semi-Supervised Learning

2021-08-12-Enterprise-Strength Federated Learning: New Algorithms, New Pa... 08-13 14:31

2021-07-15 Decentralized Domain Adaptation 07-15 16:52

2021-06-04 SYNTHESIZER & MLP-mixer 06-09 19:37

2021-06-03 混合精度训练 Mixed Precision Training 06-03 20:57

2021-5-28 HDF5在Pytorch中的应用 05-28 19:17

2021-04-30 自监督学习 Self-Supervised Learning 06-03 11:14

2021-4-22-数据增广技术：Mixup 04-29 19:34

2020-12-14-域泛化(Domain Generalization)+Meta-Learning 2020-12-27 14:00

2020-12-7-如何进行论文写作 2020-12-27 14:02

2020-11-20 机器学习实验管理 + Sacred 教程 2020-11-20 17:02

头脑风暴

- 确定一个主题或一个方向

- 给出详细的背景介绍，列出相近的相关工作

  - 否则参加头脑的风暴的同学没有办法理解你的需求

- 抛出一些开放式的问题供大家讨论

- 列出自己在研究中的难题，集思广益

研讨班的时候该做什么？

- 不要让大脑闲下来
  - 理解
  - 思考
  - 创造性的思考
- 提问
  - 不懂的地方
  - 感觉有问题的地方
  - 有更好的想法的时候

和大牛的交流机会

	主题	长度	成功率
听报告	对方定	~30min	比较稳
会后提问/搭讪	对方相关	~30s	有机会
邮件	自己定	已阅	极低
投稿评审意见	自己研究方向	500+ words	100%
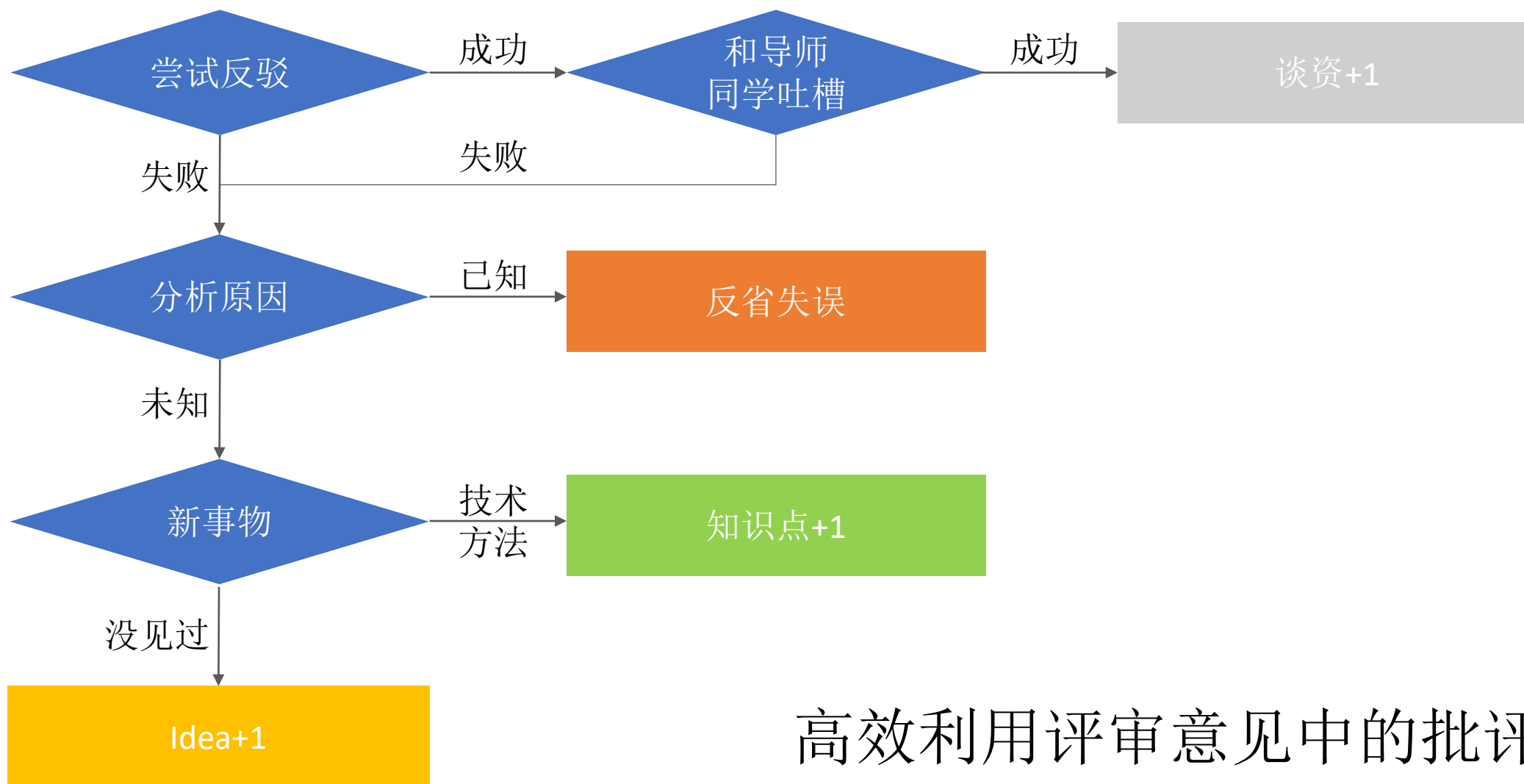
高效利用评审意见中的批评

4 记录与回顾

多记笔记

- 一有想法就记下来

- 将自己的想法和其他思路结合

札记之功必不可少,如不札记,则无穷妙绪,如雨珠落大海矣。

-章学诚在《文史通义》

爱迪生的笔记

- 爱迪生不懈地记录和绘制他在进行发明的探索时的所有步骤。在他1931年逝世以后留下了3500本笔记本

- 他的天才之中极为关键的一点就在于用文字记录下所有工作
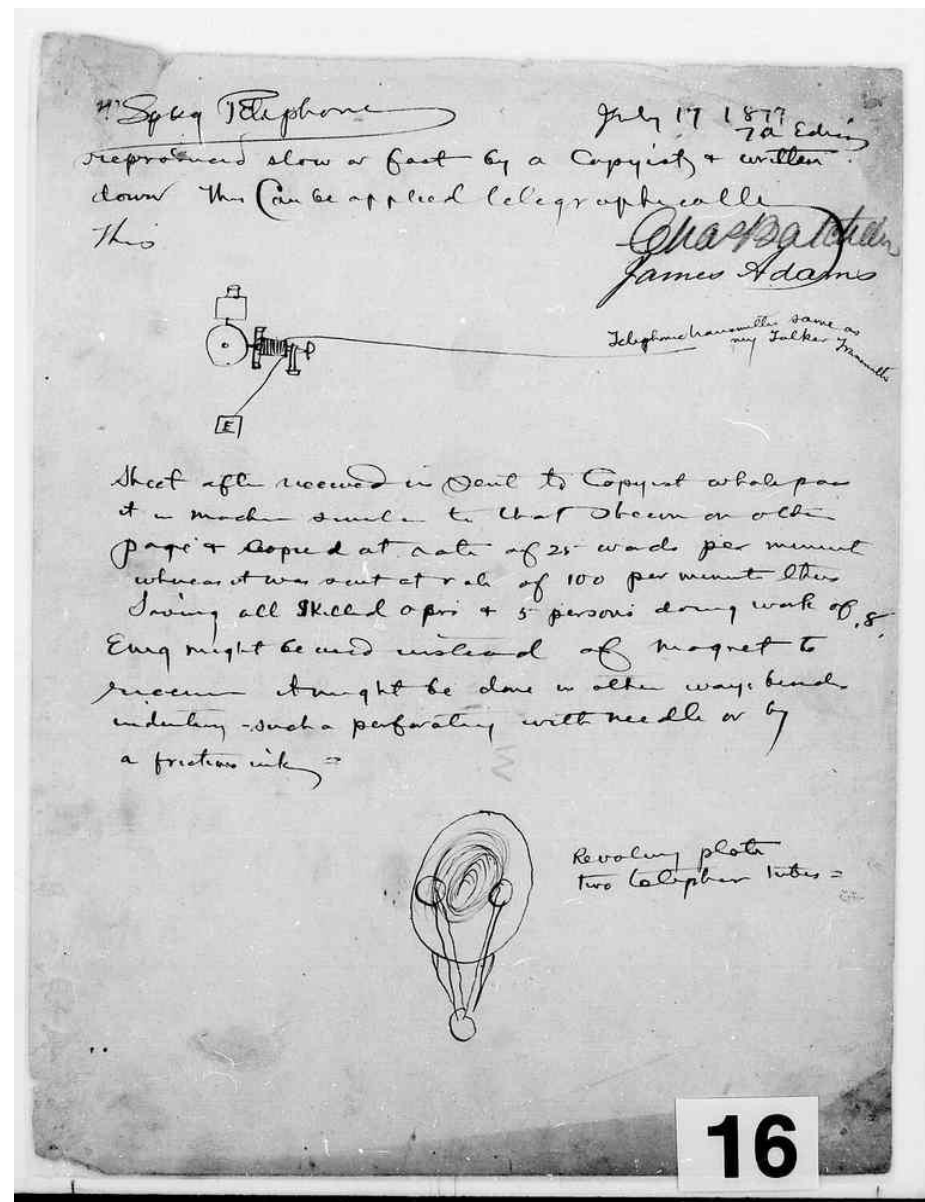
被誉为当代"文化昆仑"的钱钟书先生，其知识之渊博、古文功底之深厚、著述中引用之广博，令人叹为观止。常人所不知的是，钱钟书的博学，不仅与他的天才有关，更与他的勤奋有关。据钱钟书夫人杨绛回忆，钱钟书做笔记的习惯是在牛津大学图书馆读书时养成的，历时十余年的时间。钱钟书的全部外文笔记本共一百七十八册，还有打字稿若干页，全部外文笔记共三万四千多页。中文笔记和外文笔记的数量，大致不相上下。此外还有"日札"，共二十三册、二千多页,分八百零二则。（如果按每页300字计算，钱钟书读书笔记的总字数达2100万；如果按每页400字计算，总字数将达2800万以上，可谓是创造人类做笔记字数的记录了）
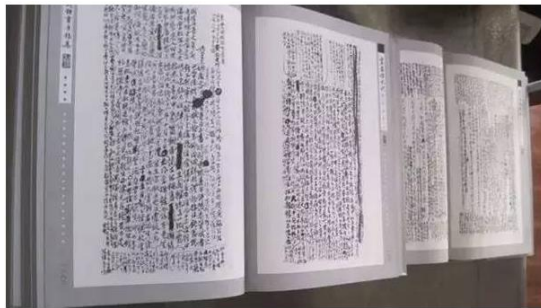
钱钟书先生的笔记

回顾

- 检查日历

- 有任何时间就检查接下来要做的事列表

- 按时检查项目列表、等待处理的事列表、可能某天会去做列表

- 成功的重要因素: weekly review
  - 每周回顾所有悬而未决的事项

周报

- 人员：硕士研究生以上的成员

- 面向：导师、同学、自己

- 时间：每周日晚上12点之前

- 平台：e.g., 浙江大学语雀

周报

- 内容
  - 本周最重要的工作的进展、反思和讨论
  - 每周五篇论文精读的心得
  - 短期（1周内）的规划、进展、困难和拟解决方案
  - 中期（1学期内）的规划、进展、困难和拟解决方案
  - 长期（1年内）的规划、进展、困难和拟解决方案

Weekly Report of WU Yingcai

weekly report(2005-7-1-----2005-7-7)

[Quote] [Modify] [Delete]

Jul 8th, 2005, 9:21am

Knowledge:
I solidified my core knowledge by continue reading some papers relevant to sampling and reconstruction theory in visualization fields.
Besides that, I began to read the book, "Non-photorealistic Rendering" by Gooch & Gooch, and made an overview of what I should learn.
Here are what I will study and present in two weeks:
Lecture 1: Material (pencil, pen and ink, painting techniques)
Lecture 2: Feature Edges: (Silhouettes, Boundaries, and Creases), artistic Line drawing and shading
Lecture 3: NPR in Visualization, and current research direction.

Skill:
For presentation skill:
In this week, I gave my last lecture about sampling and reconstruction theory, besides that I also presented a paper, and I feel that my presentation skill is greatly improved although still not good.
Practice makes perfect, I will continue to practise.

For writing skill:
I rewrote several papers' abstracts, and realized that my writing skill is also not good, and so I will keep rewriting every paper's abstract that I read.

Research:
I read several papers as follows:
1.   "Human Factors in Visualization".
2.   "Effective and Efficient Delivery of 3D Visualization in the Clinical Environment". (From viewpoint of referring physician)
3.   "Importance-Driven Volume Rendering" (IEEE Visualization 2004)
4.   "An Evaluation of Reconstruction Filters for Volume Rendering." (IEEE Visualization 1996, relevant to sampling theorem)
5.   "Illustrative Context-Preserving Volume Rendering" (EuroVis 2005)

I also considered future directions of our vascular project, and the followings are my initial ideas about how to improve the project. They are only the initial versions, and I will polish them every week.
1.   Improve the Transfer Function in our system.
1.1.   Frankly speaking, our system is kind of hard for people to use. (For human factors)
1.1.1.   There are too many widget windows, especially for TF design, according to the human-based design methodology, this will distract people, and impose cognitive overload, and hence I will minimize the widget window in near future.
1.1.2.   Since the parameter space is so huge, especially for 2D or more dimensional TF, it is difficult for people to select suitable arguments. To minimize the parameter searching space and make the TF easier to use without excluding human

1.1.2.   Since the parameter space is so huge, especially for 2D or more dimensional TF, it is difficult for people to select suitable arguments. To minimize the parameter searching space and make the TF easier to use without excluding human interaction with computer, in future I will investigate some possible AI techniques that we can use to tackle this issue.
1.1.3.   In order to make the TF design more intuitive, I will add some History navigation component to our project, in other words, a list of previously explored display TF parameters will be always visible to users.
1.2.   Besides that, we are also not satisfied with the system's performance.
1.2.1.   For the accuracy issue, actually, Our TF is 2D, the first dimension is for scalar value which stem from kindlemann's master thesis, while the 2nd dimension is for pre-filtered volume data. Our approach is that we initially get the 1st dimension TF by the semi-automatic method, and then adjust the 2nd dimension TF in other widget window, and finally get the 2D TF which we apply to final image. However, the kindlemann's method for the 1st TF is not good for vascular, in most cases, it cannot find the estimated threshold isovalue. So I want to improve it, and now I am just thinking about how to improve.
1.2.2.   For the speed issue, our system is too slow, and it cannot achieve interactive rates when we render 256*256*256 data. I think this may be partially due to the VTK library. In future, I will try some other libraries, such as The VolPack Volume Rendering Library, to see whether it can speed up our system. And besides that, I think that VTK is kind of overstaffed, and too complicate, and not easy to integrate our own codes to it.
1.2.3.   I am also not satisfied with the remaining noises existing in the final vascular image, and so we will launch further investigation to find a better filter which can generate new volume data served as the 2nd dimension in our TF. (maybe we can consult with Wilbur, and incorporate with him)

2.   Non-photorealistic rendering is a hot topic in recent years' computer graphics conferences, and it has numerous advantages, so we want to apply some NPR techniques to our system.
2.1.   We can use NPR techniques to increase our system's interactive rates
2.2.   We can use NPR to enhance our comprehensibility of the volume data.
2.3.   We can use NPR to achieve Context + Focus effects
2.4.   We can combine NPR and photorealistic rendering techniques to implement LOD effects.

3.   Add Context + Focus techniques to our system. (still thinking about)
3.1.   By distortion lens, such as fish eye distortion.
3.2.   By Transfer Function. We can make the parts we want more opaque, while the other parts more transparent.
3.3.   By NPR techniques.

4.   Use some sampling and reconstruction techniques to avoid aliasing in our system.
4.1.   Use pre-filtering techniques to filter out high frequencies.
4.2.   Use super-sampling techniques to increase sampling rates.
4.3.   Use nonuniform sampling techniques.

5.   Increase our system's interactive rates by exploiting GPU.

It seems that my suggestions above are just something relevant to implementation; however, through our further investigation, discussion and implementation, hopefully we can find some novel techniques to tackle these above issues.
I will further investigate the promising research directions of visualization not only confined to vascular visualization and think about them carefully.

日报

- 频繁交流进展

- 适用于赶投稿ddl等特殊时期

- 需要及时收到反馈

其他记录载体

- 小组Wiki/博客
  - 有趣的文章
  - 数据集
  - 软件资源
  - 组会报告总结

开个博客记笔记

http://jackieanxis.coding.me/

http://geekplux.com/

http://www.wytiny.com/

Coding

- 安装版本控制工具
  - WinCVS
  - SVN
  - Git
- 代码保持优雅规范