

# 大数据分析实践

## How to Write Academic Paper

Qiong Zeng (曾琼)

[qiong.zn@sdu.edu.cn](mailto:qiong.zn@sdu.edu.cn)

**Research** is *creative* and systematic work undertaken to increase the stock of *knowledge*.





# Course Outline

6 Personal assignments



6 Group assignments



上课日期	授课内容	实验内容	周次
20240905	课程入门、大数据探索式分析	/	第一周
20240912	课程实践项目介绍、项目组队测试、项目经验谈	项目成员集结	第二周
20240919	科研实践入门、数据采样与降维	项目管理工具制定项目计划、Pandas数据采样实践	第三周
20240926	数据质量管理	Pandas数据质量实践	第四周
20241003	/	/	第五周
20241010	众包与电子表格	电子表格实践	第六周
20241017	可视化设计	可视化设计实践	第七周
20241024	统计分析方法与工具	统计方法实践	第八周
20241031	中期汇报（论文+项目进展）1	中期进展报告	第九周
20241107	中期汇报（论文+项目进展）2	BERT实践环境配置	第十周
20241114	机器学习方法与工具	BERT实践	第十一周
20241121	人机交互方法与工具	Canis/Cast/Libra实践	第十二周
20241128	普适计算	手机移动数据采集与分析	第十三周
20241205	大规模数据分析系统	SPARK实践	第十四周
20241212	如何撰写项目论文	大项目收尾	第十五周
20241219	项目结题报告1	大项目验收	第十六周



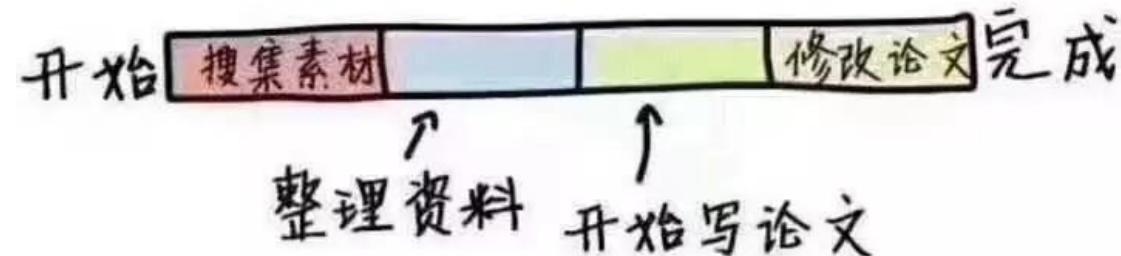
# Reminder



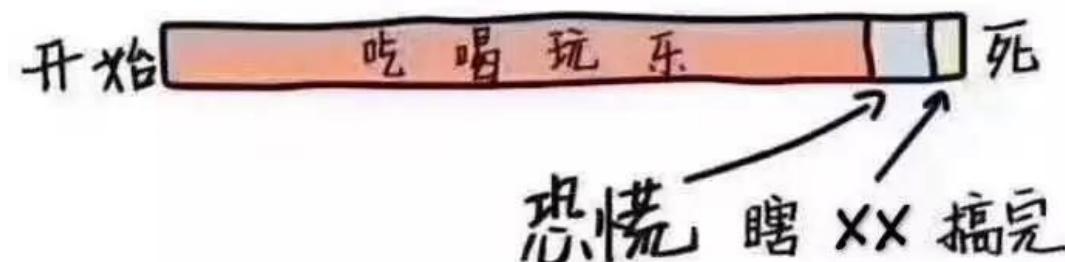
Dec. 19	Group Name	Dec. 20	Group Name
1	你知道的大数据分析实践是一门好课	7	古灵井盖队
2	你不对我队	8	花开复队
3	发际线总和我作队	9	strong队
4	汪汪队	10	这是我们队
5	星穹列车队	11	你们无敌了队
6	我们做的都队		



## 计划的论文时间安排



## 实际的论文时间安排



*Usually students torture themselves by waiting until the last minute to write their papers and by not knowing what they are doing.*

Describe the reasons for paper writing

**Why: paper writing**

作答

# 杜绝学术造假、剽窃、抄袭、不良引用等



国家自然科学基金会

李静海主任

李主任，

昨天收到贵委2019年11月22日寄首都医科大学函，称请单位调查所谓我发表的论文涉嫌存在不端行为。

在我很早就支持贵委（包括贵委的监察委员会）的工作、此事涉及完全是国外经费国外地点国外人物的老旧文章、去年方舟子等已经在热潮中明确我实验室毫无不端的情况下，贵委把函件寄给我新任职不久的单位，不仅不合适，而且说明有些人没有良心。

对于贵委这一不良行为，我先回复，然后提出建议，以便贵委的监督委员会改邪归正，做自己应该做的事情，而不是成为中国学术不端者搅浑水企图的帮凶。

回复：

1) 现在活着的、在科学一线的华人科学家，对改善中国学术规范贡献最多的，如果说本人，恐怕懂得中国科学界的人会有异议的人会很少，也就不奇怪贵委、科技部都请我做过一些相关事情，千人计划请我任其道德委员会主任、北京大学曾请我任学术道德委员会副主任（主任为校长）；

2) 贵委明知本人对科学研究规范的贡献，也知道些有人因此仇恨我，不仅不捍卫我，而在事情已经非常明确的情况下给我刚就任的单位发函，是严重的工作不负责任，而且事实上起到了配合中国学术不端坏分子的作用。贵委的人员咨询过我有关武汉大学李红良学术不端，贵委自己的工作进展缓慢，武汉大学在2018年9月下旬希望我参与调查李红良论文造假事件，2018年11月17日武大邀请我任李红良事件独立调查组组长，我因太忙并未答应。2018年11月22日盖章“武汉大学学术委员会秘书处”的咨询函，由电子邮件寄达我参与主编的《知识分子》微信公号的电子邮箱。不出24小时的2018年11月23日，很多人的电子邮箱收到匿名造谣攻击我的信件。

方舟子在2018年12月就此发表过三篇文章，读者众多，其结论非常清楚：抄袭实验室没有任何造假或其他不端：

<http://xys.org/xys/netters/Fang-Zhouzi/blog/raoyi.txt>

<http://xys.org/xys/netters/Fang-Zhouzi/blog/raoyi2.txt>

<http://xys.org/xys/netters/Fang-Zhouzi/blog/raoyi3.txt>

如果贵委有谁看不清楚方舟子仔细的分析，应该自己好好学习、天天向上。

3) 贵委提出的两篇文章 (Liu et al 2004 和 Liu et al 2009) 都是我在美国的实验室的研究论文。也就是说，贵委调查美国公民用美国经费在美国进行的研究，贵委的工作人员是否配合国内学术不端人员以批我搅浑水的心情太急迫，已知囫囵吞枣管外国的事情，是否这些工作人员太滑稽？何况，这两篇文章，如方舟子所分析，完全没有我实验室不端的问题。

建议：

1) 贵委应该有效、有胆魄地彻底调查武汉大学医学院李红良 17 年如一日明目张胆的造假；

2) 贵委应该严肃调查中国科学院上海生命科学研究院生化细胞所研究院裴钢于1999年，用贵委三项经费 (39630130, 39625015 和 39825110) 支持其发表的论文 (Ling et al., PNAS 96:7922-7927)。该篇论文的图3、图4、图5是不可能真实的，只有造假才能产生（众所周知 GPCR 需要七重跨膜区域才有功能，裴钢号称只要5重跨膜，而且居然两个GPCRs都是这样的，出了3个同样错误的图）。贵委20年不触及这一问题，现在这是实名举报，贵委不要推卸躲避，洗刷中国科学院因这篇造假文章选出院士的耻辱。

3) 今年中国科学院上海药物研究所的耿美玉研究员作为通讯作者的文章 (Wang et al Cell Research 29:787-803)，号称其发明的药物 GV971 能够通过肠道菌群治疗小鼠的阿尔兹海默症。这篇文章，不造假是不可能的。现实名举报，请贵委做些好事，为中国科学界洗刷耻辱。

饶毅

(代表个人，不代表任何单位)

2019年11月28日

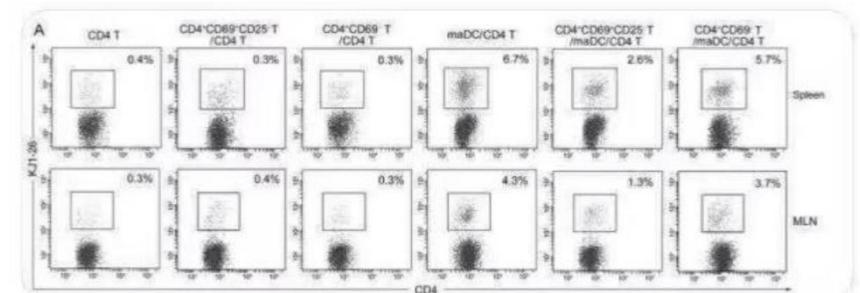
Ling K, Wang P, Zhao J, Wu Y-L, Cheng Z-J, Wu G-X, Hu W, Ma L and Pei G (1999) Five-transmembrane domains appear sufficient for a G protein-coupled receptor: functional five-transmembrane chemokine receptors. *Proceedings of the National Academy of Sciences USA* 96:7922-7927.

Wang X, Sun G, Feng T, Zhang J, Huang X, Wang T, Xie Z, Chu X, Yang J, Wang H, Chang S, Gong Y, Ruan L, Zhang G, Yan S, Lian W, Du C, Wang D, Zhang Q, Lin F, Liu J, Zhang H, Ge C, Xiao S, Ding J and Geng M (2019) Sodium oligomannate therapeutically remodels gut microbiota and suppresses gut bacterial amino acids-



Elisabeth Bik  
@MicrobiomDigest

This morning I am working on a paper by a big-name professor who is a Chinese Academician and president of a top tier Chinese university. But look at this flow cytometry image. If you find something, keep on going.... #ImageForensics Level: Advanced



- **Paper Writing and Practices**
- **Writing tools**

# Basic Structure



**How many sections are there in a paper?  
What are they?**

# Basic Structure

- Title, Author
- Abstract
- Introduction
- Related Work
- Background (requirement analysis, task formulation, etc)
- Algorithm (technical details)
- Experiments (implementation, results, evaluation, etc)
- Conclusion and Future Work
- Acknowledgement
- References
- Supplementary Materials

## Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco

Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, Jeffrey Heer

**Abstract**—There exists a gap between visualization design guidelines and their application in visualization tools. While empirical studies can provide design guidance, we lack a formal framework for representing design knowledge, integrating results across studies, and applying this knowledge in automated design tools that promote effective encodings and facilitate visual exploration. We propose constraints as logical facts in a domain-independent knowledge base (Draco). We show how hard constraints must be satisfied (e.g., shape, size, and color) and soft constraints (e.g., aesthetic principles or perceptual effectiveness). We demonstrate how to construct increasingly complex models by combining learned directly from the results of graph-based experiments.

**Index Terms**—Automated Visualization Design, Perceptual Effectiveness, Constraints, Knowledge Bases, Answer Set Programming

## Clarity of Exposition

### 1 INTRODUCTION

Visualization designers benefit from familiar main under consideration and principles of effective visualization [37]. Although designers can learn these principles from books, research papers, and experience, they do not always follow them in their practice [6, 42]. Automated design tools [37] address this problem: they use formally-encoded knowledge to promote effective visualizations. However, these systems are incomplete and continually evolving. In order to support new knowledge, they must be able to experimental results or compare different theories used to elaborate and refine these bodies of knowledge. Visualization researchers regularly publish empirical study results of how people decode and interpret visualizations. However, new results often make their way into automated design tools even though our knowledge is evolving, we lack a formal model for representing and acting upon this knowledge. Automated design systems [37–39, 65] do not consider the idiosyncratic representations of design knowledge. To address this problem, we seek to make formal models of design knowledge a shared resource for the visualization community.

We present Draco, a formal model that represents visualizations as sets of logical facts and represents design guidelines as a collection of hard and soft constraints over these facts. Draco can systematically enumerate all possible visualizations that satisfy these constraints, and then suggest the best ones based on a user's preferences. We implement Draco using Answer Set Programming, a domain-independent constraint reasoning language. We show how Draco can go beyond these systems by adding new constraints concerning data and a user's primary task. Instead of manually specifying weights, we learn them from two independent graphical perception studies [30, 51]. We compare the learned visualization model to a hand-tuned model, demonstrating improved automated design suggestions.

## Quality of References

## Reproducibility

We implement Draco using Answer Set Programming, a domain-independent constraint reasoning language. We show how Draco can go beyond these systems by adding new constraints concerning data and a user's primary task. Instead of manually specifying weights, we learn them from two independent graphical perception studies [30, 51]. We compare the learned visualization model to a hand-tuned model, demonstrating improved automated design suggestions.

## Novelty

# from overview to details

We view the constraints in Draco as the starting point of an evolving knowledge base of design considerations for researchers and tool designers to extend and use. Hard constraints must be satisfied (e.g., shape

- Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Bill Howe, and Jeffrey Heer are with the University of Washington. E-mails: domoritz, ctwang, glnelson, haldenl, billhowe, jheer@uw.edu.
- Adam M. Smith is with the University of California Santa Cruz. E-mail: amsmith@ucsc.edu.

accelerate the transfer of research knowledge into practical tools. Researchers can also use Draco to systematically sample, enumerate, and reason about the design space of possible visualizations, or to concretely compare different design models. We make Draco available as open source software with supporting tools, documentation, and examples at <https://uudata.github.io/draco/>.

## 2 RELATED WORK

Draco builds on prior work on automated visualization design systems, visualization specification languages, and constraint programming.

# Two Things are Necessary

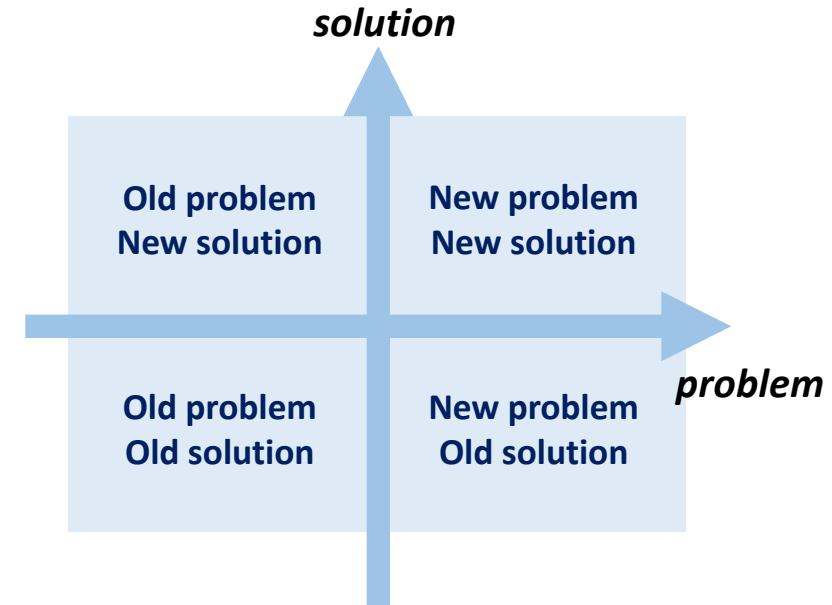
- Content

- Style

- Without good content, style is irrelevant
- Without good style, content is not enough
- Papers can be **rejected** for poor writing

# Content

- This is your great **new** idea
- The whole **problem** might be new
- Or you may have a new way of doing a **step**
- Or you may put existing ideas together to solve a problem in a new **way** (this is weaker)
- If the problem is not new, your method must be an **improvement**



# Abstract and Paper

- The abstract is a **self-contained** document
- The reader may not have the paper, just the **abstract**
- The abstract should tell the reader **whether to read the paper** or not
- It should **summarise** the key contributions of the paper, and other important content

# Abstract



- Primary impression! Important!
- Goals: *identify the research topic, describe the novelty of the presented work, and identify the benefits and advantages of that result*
- Updated during the writing process
- 6-12 sentences (100-200 words)



# Abstract

There exists a gap between visualization design guidelines and their application in visualization tools.

## 背景问题

While empirical studies can provide design guidance, we lack a formal framework for representing design knowledge, integrating results across studies, and applying this knowledge in automated design tools that promote effective encoding.

## 进一步阐述现有工作及问题难点

We propose modeling visualization design knowledge as a collection of constraints in conjunction with a method to learn weights for soft constraints from experiments.

## 方法

Using constraints, we can take theoretical design knowledge and express it in a concrete, extensible, and testable form: the resulting models can recommend visualizations and easily be augmented with additional constraints or updated weights.

## 该方法的优点

We implement our approach in Draco, a constraint-based system based on Answer Set Programming (ASP).

## 实现

We demonstrate how to construct increasingly sophisticated automated visualization design systems, including systems based on weights learned directly from the results of perception experiments.

## 实现的效果

Abstract Showtime!

**15 min:  
write your abstract with 6-8 sentences**

作答

# Abstract



There exists a gap between visualization design guidelines and their application in visualization tools.

While empirical studies can provide design guidance, we lack a formal framework for representing design knowledge, integrating results across studies, and applying this knowledge in automated visualization recommendation.

We propose a new approach to address this gap.

Using constraints,

extensible, and testable form: the resulting models can recommend visualization designs and can easily be augmented with additional constraints or updated weights.

We implement our approach in Draco, a constraint-based system based on Answer Set Programming (ASP).

We demonstrate how to construct increasingly sophisticated automated visualization design systems, including systems based on weights learned directly from the results of graphical perception experiments

## 10 min: Discussion

# Introduction

- Don't be afraid to partly **repeat the abstract**
  - The abstract is a **separate** document
- State what problem you are solving in a **straightforward** non-technical way
- Who are the potential **users** of your idea?
- Why is the problem of **current interest**?

# Introduction

- Explain what your **contribution** is to the topic
- Explain clearly how what you have done is **different** to previous work
- Summarise the content of the **rest** of the paper, making clear which parts are **novel**

**It can be viewed as an extension of the abstract**



# Introduction



- **What is the problem? Why is it important?**

**10 min reading  
Paragraph 3 in Section 1**



# Introduction



- What is the problem? Why is it important?

8 min discussion  
Talk About logical structure



# Introduction



## Contributions

What are the contributions, precisely, of this paper? What does the reader potentially gain from this paper?

- 新问题新方法: Does the presented algorithm(s) provide a novel visualization of some sort? i.e., images never before seen. Does the presented work provide new interaction techniques not previously published?
- 新问题老方法: Does the approach provide new insight in some way?
- 老问题新方法: Does the technique described result in faster performance times than previous algorithms and if so why?
- 老问题老方法: Does the work take advantage of the graphics card hardware in some way? For example, using texture-mapping hardware, or hardware-accelerated interpolation?



# Related Work

Research = Re + Search

**Organize, summarize, analyze**

- Include all **major** similar work
- Do **not summarise** each piece, but **analyse** it
- Show why each reference did not solve your problem:
  - solved a **related** problem
  - gave an **inefficient, limited or inappropriate** solution
- Either you must be the **first** person to work on a new problem; otherwise, it is not good enough to be **different**: Your work must be **better** in some way
- Compare the references to your work, showing how your ideas are **better**



# Related Work

- Research = Re + Search
- Organize, summarize, analyze
- Choose *right* papers to demonstrate: Make sure all due **acknowledgements** are made
- Each previous publication gets 2-3 sentences description
- How to refer: use names not numbers, “*Chen et al. present an algorithm for XX*”

**5 min:**  
**write 2-4 related areas**

## Related Work Showtime!

**10 min:**

**write your basic points in each area**

**List 2-4 papers for each area**

**e.g.**

**XX proposes an automatic way to XXX.**

**However, describe its disadvantages....**

作答

# Background & Method

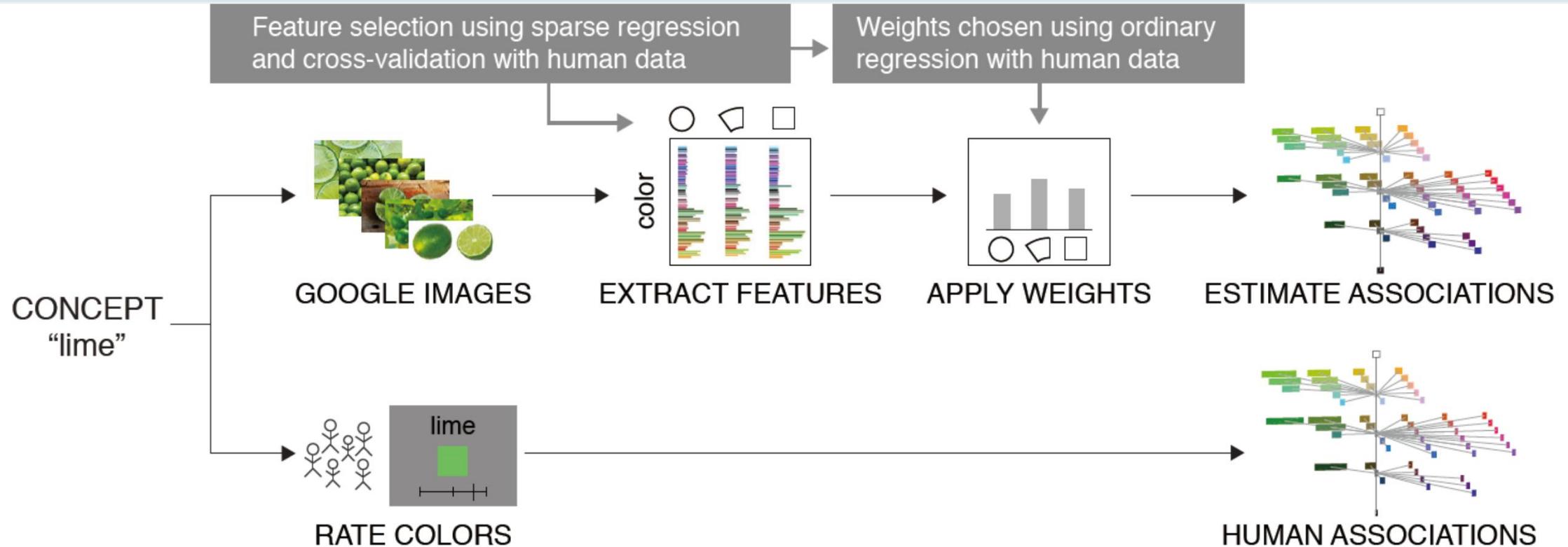


Figure 3. Illustration of our pipeline for automatically extracting color distributions from images. The bottom flow (concepts to color ratings to human associations) describes the slow yet reliable direct approach using human experiments to determine ground-truth associations. The top flow involves querying Google Images, extracting colors using a variety of different methods (*features*), then weighting those features appropriately to obtain estimated associations. Deciding which features to use and how to weight them is learned from human association data using sparse regression and cross-validation. Once the model is trained, color-concept associations can be quickly estimated for new concepts without additional human data.



# Background & Method

- Overview: description with diagram if necessary
- Algorithm subsections: from overview to detail
- Parameters or optimization solvers

Importance:

- Captions of figures and images should be self-contained
- High quality figures, especially for cg and vis

**10 min:**  
**Draw your framework**



# Experiments

- Implementation details
- Qualitative and quantitative evaluation/results
- Case study/User study



# Conclusion and Future Work

- Summarize the novelty of the presented work, why it is so good and what is new.  
→ a summarization of introduction
- Future work: short, about 3 sentences

# Recap

## **title: XXX**

### **abstract**

- problem and challenges
- how do others to solve this problem
- your basic idea, advantages and framework
- your implementation
- your results

### **introduction**

- an extention to claim problem and challenges
- an extention to claim references
- an extention of your basic idea
- your framework
- your evaluation
- your contributions

### **related work**

- area 1
  - your viewpoint
  - supportive references and discussions
- area 2
- ...

### **background**

### **algorithm**

- overview
- step 1
- step 2
- parameter analysis

### **experiments**

- implementation details
- performance
- comparison, if necessary
- quantitative and qualitative analysis

### **conclusion and future workS**



- Paper Writing and Practices
- Writing tools



# Writing Tools



- LaTeX (Online: overleaf)
- Word
- Dropbox
- Google documents → sharing ideas for co-authors
- 石墨文档



# What is TEX?

TEX是一种功能强大且十分灵活的排版语言

- TEX系统提供了 300 + 600 多条基本的排版命令
- TEX是目前公认的数学公式排版最好的排版语言
- TEX是免费的
- TEX的名字来自大写的希腊字母  $(\tau, \epsilon, \chi)$ ，意思是“科技” 和 “艺术”

艺术

科技





# TEX的优缺点

## Advantages



- 高质量的输出: TEX以排版质量为首要目标
- 超常的稳定性: 系统极少崩溃
- 良好的通用性: Windows, Linux, MacOS 都支持
- TEX文档是纯文本文件: 占用空间很小
- TEX是可编程的(宏命令编程语言)
  - 你可以用很少的命令来完成一些复杂的工作
  - 或通过定义新的命令来得到特殊的效果



# TEX的优缺点

## Disadvantages

- 命令繁多，不便记忆
  - 在使用中学习，通过网络寻求帮助，人性化的编辑器
- 错误难找，慢慢积累经验
- 使用不直观
  - 要编译后才能显示效果
  - 每次修改后都要再次编译
  - 目前已有一些所见即所得的扩展，如：LyX, TeXmacs



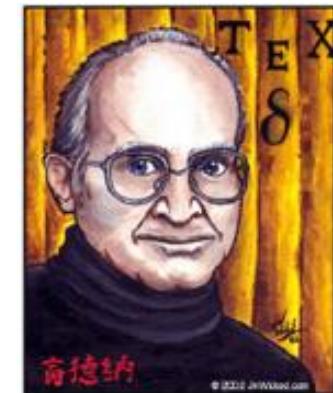


# The Development of TEX

- TEX的第一版于1978年面世

70年代末，斯坦福大学计算机系教授D. E. Knuth (高德纳，现代计算机科学先驱，1974年获图灵奖) 在看到其巨著 “The Art of Computer Programming” 第二卷的校样时，对排版的低质量感到无法忍受，决定开发一套高质量的排版系统，于是就有了TEX

- 1982 年版的 TEX 已经是一个十分稳定的版本
- 1990 年推出 3.1 版，并宣布不再更新  
(只修正 bug)
- 最新版本号：3.14159265 (2014 年)



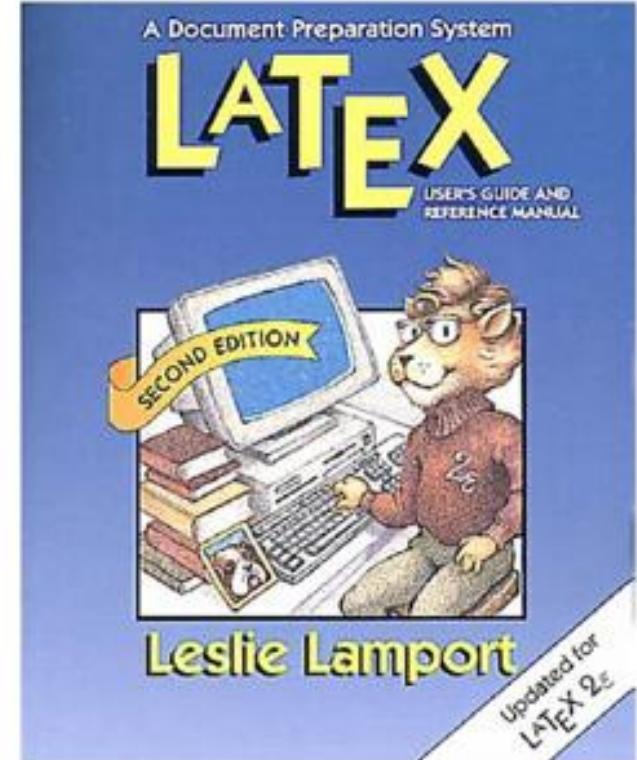


# What is LATEX?

- 1984 年前后，Lamport 开发了 LATEX
- LATEX 提供了一套针对文档逻辑结构的控制序列，让 TEX 变得更好用，普通用户即便不是很了解 TEX，也可以在很短的时间制作出高质量的排版结果，因此迅速在计算机科学、数学及相关学科领域得到推广应用



The LATEX Project





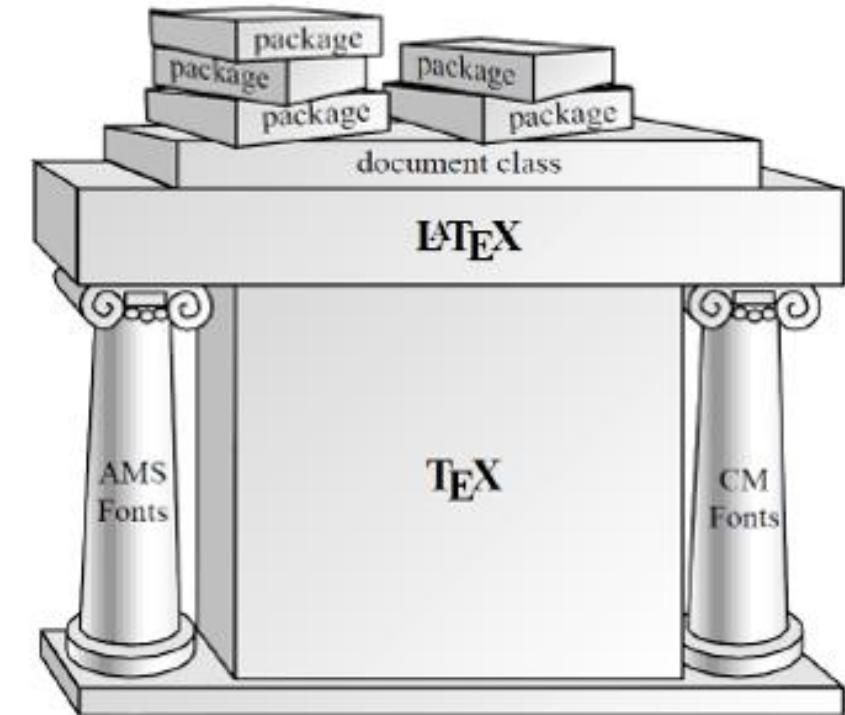
# 软件下载与安装

## TEX的发行版

- Windows 系统: TexLive (推荐), MiKTeX, CTEX 套装 (不推荐)
- Unix/Linux 系统: TeXLive
- Mac OS 系统: TeXLive, MacTeX

## 下载和安装

- TexLive : <http://tug.org/texlive/> (大而全, 全部宏包)
- 编辑器: WinEdt, TeXworks, TeXmaker, TeXStudio, vim, emacs, ...





# LATEX 文稿的排版过程

## ① 编写源文件

tex源文件为纯文本文件，以 .tex 为扩展名，可以使用任何文本编辑器编写，如：WinEdt, EditPlus, Vi, Emacs, ...，推荐 WinEdt →专门针对 tex 开发，提供许多便捷功能，有助于提高排版效率

## ② 编译

用 pdflatex (英文文档) 或 xelatex (中文文档) 编译，生成相应的 pdf 文件





# LATEX 源文件的基本框架

- LATEX 源文件：正文 + 命令 + 注解
- 排版命令(简称 命令)：反斜杠开头的字符串
- 注解符：百分号 %
- 文档类型：\documentclass{...} (论文, 书籍, 幻灯片, 海报)
- 环境: \begin{...} 开头, \end{...} 结尾

```
\documentclass{article} % 指定文档类型
% 导言区：全局设置，宏包调用等
%
\begin{document}
% 正文部分
Hi, this is my first \LaTeX{} file.
\end{document} % 结束
```



# LATEX 排版命令

- 方括号中的是可选的(称为选项), 花括号中的参数是必需的

例: 一些常用命令

- \documentclass, \title, \author, \date, \usepackage
- \begin{环境名}, \end{环境名} → 组成一个环境

- 定义新命令

```
\newcommand{新命令}{命令内容}  
\renewcommand{已有命令}{命令内容}
```

```
\renewcommand{\epsilon}{\varepsilon} % $ \epsilon $ → ε
```



# 文档类与导言区

```
\documentclass[选项]{文档类}
```

- 位于源文件的最前面，用于指定文档的整体结构和布局，必须且只能选一种：
  - 常用文档类: article, book, beamer, ctexart, ctexbook, ctexbeamer
  - 常用选项：
    - 10pt(缺省值), 11pt, 12pt → 指定基本字体的大小
    - letterpaper(缺省值), a4paper, a5paper, ... → 指定纸张的大小
    - 单双面选项: oneside, twoside, openright, openany
    - 数学公式: leqno (公式编号在左边), fleqn(靠左显示行间公式)
- 导言区：\documentclass 和 \begin{document} 之间的区域
  - 导言区用于放置全局控制全局控制命令，如：调用宏包，设置页面大小, ...
  - 放在导言区的命令对整个文档都起作用



# 宏包：LATEX功能的扩展

宏包调用方法(只能出现在导言区)

```
\usepackage[选项]{宏包名}
```

- 如果宏包不带选项，则可以多个一起调用，如：

```
\usepackage{amsmath, amssymb, amsfonts}  
\usepackage[pagebackref]{hyperref}  
\usepackage[numbers, sort&compress]{natbib}
```

- 常用宏包：
  - geometry, fancyhdr, natbib, float, caption
  - amsmath, amssymb, amsfonts, amsthm, ntheorem, bm, mathtools
  - xcolor, graphic, subfigure, epstopdf
  - longtable, colortbl, tcolorbox, mdframed
  - algorithm, algpseudocode, listings



# 插图：graphicx宏包

```
\usepackage{graphicx}
```

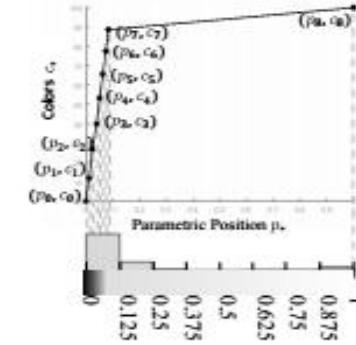
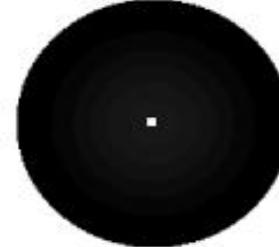
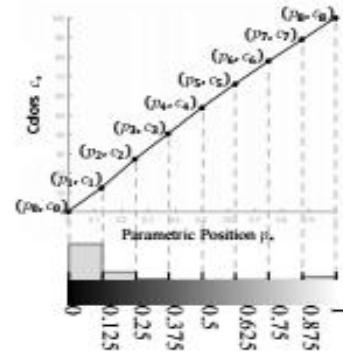
```
\includegraphics[选项]{图形文件名}
```

- 支持的图片格式: pdf, jpg, png(pdfLaTeX 和 xeLaTeX 编译)
  - eps 格式的图片 → epstopdf 宏包, 自动将 eps 转换成 pdf
- 图形文件名中可以含路径
- 常用选项有
  - Width, height → 指定图形的宽度和高度(若只指定宽度或高度, 则按比例缩放)
  - scale → 缩放因子, 如scale=0.8

```
\includegraphics[scale=0.2]{tiger.png}
```

```
\includegraphics[width=2.5cm]{tiger.png}
```

```
\includegraphics[width=0.3\textwidth]{tiger.png}
```



(a)

(b)

```
\begin{figure}[t]
\centering
\setlength{\belowcaptionskip}{-4mm}
\includegraphics[width=1\linewidth]{figures/fig2.pdf}
\vspace*{-5.5mm}
\caption{An illustration of \texttt{colormap} definition based on control points and linear mapping function. By shifting parametric position of control points, we can produce a new \texttt{colormap} by linear interpolations. Please see this figure under a high-brightness display setting.}
\label{fig:cpnts}
\end{figure}
```



# 标题部分：标题, 作者, 日期, 脚注等

```
\title{标题}  
\author{作者}  
\date{日期}  
\maketitle
```

- \title, \author, \date 只提供生成标题所需的信息，并不生成标题！
- 生成标题的命令: \maketitle → 这个不可省略
- \date{} → 留空则不输出日期  
    如果省略 \date 命令，则自动加上当前日期
- 标题部分的角注: \thanks{...}  
    必需放置在 \title, \author 和 \date 的参数中(即{}内, 可多次出现)
- 如果标题过长, 可以用 \\ 强制换行
- 若有多个作者, 可用 \and 隔开(横排), 也可以用 \\ 隔开(竖排)



# 摘要

```
\begin{abstract}
```

```
    . . . . .
```

```
\end{abstract}
```

Book 文档类没有摘要



# 章节自动编号

- 每类章节都被赋予一个层次号：层次号越小，级别越高

层次号	章节命令	层次号	章节命令
-1	\part	2	\subsection
0	\chapter	3	\subsubsection
1	\section	4	\paragraph
		5	\ subparagraph

- book 类中的自动编号的章节有：
  - \part, \chapter → 独立编号
  - \section, \subsection → 关联编号, 如 2.1, 2.2.1 (节的编号前带有章的编号)
- article 类中的自动编号的章节有：
  - \part, \section → 独立编号
  - \subsection, \subsubsection → 关联编号

# 章节举例



## Part I 篇 (Part) 的标题

1 节(Section)的标题  
1.1 小节(Subsection)的标题  
1.1.1 子节(Subsubsection)的标题

article

## Part II 章 (chapter) 的标题

1.1 节(Section)的标题  
1.1.1 小节(Subsection)的标题  
子节(Subsubsection)的标题

book



```
\begin{document}  
  \maketitle  
  \input{01_intro}  
  \input{02_related}  
  \input{03_methods}  
  \input{04_discussion}  
  \input{05_conclusion}  
  \acknowledgments  
  \bibliography{coloropt}  
  \end{document}
```

```
@article{plos17opt,  
  author = {Nunez, Jamie R. AND Anderton, Christopher R. AND  
    Renslow, Ryan S.},  
  journal = {PLOS ONE},  
  publisher = {Public Library of Science}  
  doi = {10.1371/journal.pone.0199239}  
}
```

10 min:

**Download a LaTeX template from a target conference or journal**

**Google: bibtex**



# 两个重要概念：分组和环境

- 分组
  - 有些命令只对其参数起作用,如 `\textbf{abc}`
  - 有些命令对后面所有的内容都起作用, 这些命令通常也称为声明, 如`\bfseries`
  - 可以利用大括号(即分组)来限制声明的作用范围

```
This is \textbf{bold face} style.\\" This is \bfseries bold face style.\\" This is {\bfseries bold face} style. ⇒ This is bold face style  
This is bold face style  
This is bold face style
```

- 环境: 某些具有特定格式的内容需要放在相应的环境中, 如表格、数学公式等

```
\begin{环境名}  
...  
\end{环境名}
```

- Document 是LATEX的一个最基本环境, 一篇文档有且只能有一个document环境



# 列表环境

## itemize, enumerate, description

itemize → 带相同的标签

enumerate → 条目标签为自动编号的符号

description → 人工指定各条目的标签

- itemize 列表环境

```
\begin{itemize}
    \item [标签] 条目内容
    ...
\end{itemize}
```

- 缺省的标签与层数有关，分别为： ■, -, \*, ·
- 也可通过选项标签来指定标签○
- 不要标签： \item[]



# 列表环境

## itemize, enumerate, description

- enumerate 列表环境

```
\begin{enumerate}
  \item[标签] 条目内容
  ...
\end{enumerate}
```

- 缺省标签
  - 第一层: 阿拉伯数字后跟圆点: 1. 2.
  - 第二层: 圆括号包围的小写拉丁字母: (a) (b)
  - 第三层: 小写罗马数字后跟圆点: i. ii.
  - 第四层: 大写拉丁字母后跟圆点: A. B.
- 高级列表功能: list 环境, enumitem 宏包

# 行内公式



- 行内公式：与普通文本混合排版
- 三种实现方式：

```
\begin{math} ... \end{math}  
\( ... \)  
$ ... $
```

上面三种方式是等价的，推荐最后一种方式，例：

勾股定理  $a^2 + b^2 = c^2$  也称商高定理



# 行间公式

- 行间公式：包括单行公式和多行公式
  - 单行公式 → 一个公式，独占一行
  - 多行公式 → 多个公式，每个公式独占一行。
  - 行间公式可以编号，也可以不编号
  - 给公式编号时，一般采用 自动编号，也可以人工编号。
  - 在输入多行公式时，对应的代码中不能出现空行！

单行公式

$$(a + b)^2 = a^2 + 2ab + b^2$$

多行公式

$$\begin{aligned}(a + b)^3 = \\ a^3 + 3a^2b + 3ab^2 + b^3\end{aligned}$$



# 排版基础：换行，分段，分页

- 换行：自然换行(若需 强制换行, 可使用 \\ 或 \\linebreak)
  - 一般情况下, 不建议使用强制换行
- 分段：一个空格或 \\par
  - 建议使用空行进行分段 → 简洁直观
- 分页：自然分页, 若需强制分页, 可用 \\newpage, \\clearpage 或者 \\pagebreak
  - 一般情况下, 不建议使用强制分页
- 行间距, 行间距伸展因子 \\baselinestretch 或伸展命令 \\linespread

```
\renewcommand{\baselinestretch}{1.2}
\linespread{1.2}
```
- 段落间距与段落缩进：用自动设定的即可，英文每节的第一段首行不会自动缩进



# 排版基础：水平间距

- 强制空格: \f
- \quad → 产生一段宽度为 1em 的水平空白
- \quadquad → \quad的两倍
- \ → 大约为 \quad 的  $3/18$
- \hspace{宽度} → 产生指定宽度的水平空白
- \hspace\*{高度} → 若要在行首产生一定的空白，则需要使用此命令
- \hfill → 根据排版需要插入空白,撑满整行
- \phantom{文本内容} → 水平空白的宽度等于文本内容的总宽度



# 排版基础：垂直间距

- `\smallskip` → 垂直空白高度为 $3pt\;plus\;1pt\;minus\;1pt$
- `\medskip` → `\smallskip` 的两倍
- `\bigskip` → `\smallskip` 的四倍
- `\vspace {高度}` → 产生指定高度的垂直空白
- `\vspace* {高度}` → 同 `\vspace`, 主要同在页面的顶部
- `\vphantom {文本内容}` → 垂直空白的高度等于文本内容的总高度



# 特殊字符

- 有10个字符被赋予了特殊用途,需要使用相应的命令才能输出

字符	#	\$	%	{	}	~	-	^	&	\
命令	\#	\\$	\%	\{	\}	\~{}	\_{}	\^{}	\&	\textbackslash

- 符号 “>” , “<” , “|” 被定义成数学符号, 只能用在数学模式中, 若要在普通文本中输出, 需使用相应的命令

字符		<	>
命令	\textbar	\textless	\textgreater

- 引号与连字号

字符	‘	,	“	”	’	-	-	—
命令	\` (倒引号)	\,	\``	\” 或 \“	\`{}	-	--	---



# 特殊字符

- TEX标识符，重音符号等

字符	TEX	LATEX	LATEX 2 $\epsilon$	AMS-LATEX <sup>1</sup>
命令	\TeX	\LaTeX	\LaTeXe	\AmS-\LaTeX

ò	ó	ô	ö	ő	ō	ò	њ
\o	\'o	\^o	\"o	\~o	\=o	\.o	\u{o}
њ	њ	њ	њ	њ	њ	њ	
\v{o}	\H{o}	\r{o}	\t{oo}	\b{o}	\c{o}	\d{o}	

字符	§	¶	†	‡	©	£
命令	\\$	\P	\dag	\ddag	\copyright	\pounds

- 更多特殊字符参见：The Comprehensive LATEX Symbol List
- 可使用 WinEdt 或 TeXFriend 中提供的便捷方式



# 字体命令

- TEX标识符，重音符号等
  - 这些命令在遇到新的同一类命令之前一直起作用
  - 若想限制声明的作用范围，可使用分组或环境
- 每一个字体声明都有对应的带参数的字体命令
- 字体命令：只对其参数中的文本起作用，如：  
`\bfseries 这是黑体 } ⇔ \textbf{这是黑体}`
  - 族：`\textrm{...}, \textsf{...}, \texttt{...}`
  - 形状：`\textup{...}, \textit{...}, \textsl{...}, \textsc{...}`
  - 系列：`\textmd{...}, \textbf{...}`
  - 采用缺省值：`\textnormal{...}`, 对应的声明为：`\normalfont`
  - 强调命令：`\emph{...}`, 对应的声明为：`\em`
    - 缺省的西文字体：中等权重、直立的罗马字体



# 一些注意事项和排版建议

- 中文和英文 之间，中文与数学公式（变量）之间 → 用空格隔开
- 英文标点后面留空格
- 各种环境的开始和结束 → 独占一行
- 自然分段 → 建议使用一个空行
- 自然换行 → 慎用强制换行 “\\”
- 标点符号一致 → 全部采用中文标点，或者全部采用英文标点
- 数学公式中的标点 → 必须用英文标点
- 单个回车 → 编译时作为一个空格处理
- 多个连续的空格 → 编译时作为一个空格处理
- 编译出错时如何终止编译 → x
- 如何清空辅助文件



# **Run the template Fill in each section**

<http://math.ecnu.edu.cn/~jypan/Latex/index.html>

# Thank You

