

# 软工论文阅读理解

## 第一章 总论

南大软院DevOps组内部训练资料

匡宏宇

# 大纲

- 课程相关概念定义
- 论文基本特征
- 论文基本结构
- 论文评价指标
- 如何（有效）获取论文

# 论文的定义与作用

- 论文（Paper）的定义
  - 进行各个学术领域的研究和描述学术研究成果的文章
  - 探讨问题，进行学术研究的一种手段
  - 描述学术研究成果，进行学术交流的一种工具
- 论文的作用
  - 总结（阶段性）科研成果
  - 学术交流的途径
    - 同行评审与网络传播
  - 毕业、项目申请、个人能力评价的客观标准（之一）

# 本课程所关注的论文类型

- 计算机领域（以SE为主）内的会议（Conference）论文，期刊（Journal）论文，*学位论文 (Thesis)*
  - 会议论文一般又分为长文（regular paper）、短文（short paper）和研讨会论文（workshop paper）
    - Tools Demo, Doctor Symposium, ERA, Replication...
  - *内容关注点*
  - *论文长度*
- 会议与期刊论文以英文为主，学位论文以中文为主
  - 本次研讨以会议论文作为切入点，原因在于其问题定义明确，内容范围可控，行文逻辑相通，高水平论文易得

# 大纲

- 课程相关概念定义
- 论文基本特征
- 论文基本结构
- 论文评价指标
- 如何（有效）获取论文

# 论文的文体与文风

- 文体：说明文
  - 科技类写作（Technical Writing）：重视名词，对形容词和副词的使用非常谨慎
- 文风：准确清晰，简洁流畅，客观表达主观意见
  - 一般使用结构简单的短句
    - 不使用修辞手法和复杂的句式（如设问句、倒装、复杂的定语从句等）
  - 重视工作的引用（Reference）与图表的编号（Caption）
    - “未标明引用的、连续十五字相同的文本即可视作抄袭”
  - 很少使用被动语态，以“We”开头（中文略有不同）
  - 在表达意见、描述程度时，会表达这些意见是明确的、个人的、有限的
    - “We trust” -> “We think/argue/believe”; “guess” -> “suggest”
    - “None of” -> “To the best of our knowledge, there is no ...”

# 大纲

- 课程相关概念定义
- 论文基本特征
- 论文基本结构
- 论文评价指标
- 如何（有效）获取论文

# 论文的逻辑结构

- 基本逻辑结构：“Hammer-Nut”
  - “Nut”：问题定义
  - “Hammer”：解决方案
  - 如何评估Hammer-Nut的效果：实验设计与结果分析
- **问题定义**
  - 特定背景下，基于若干假设，严格定义的特定问题
- 解决方案
  - 描述解决方案的思路与框架，一般不强调具体实现细节
- **实验设计与结果分析**
  - 客观性、针对性、显著性、实用性、局限性



# 论文的章节结构

- 标题 (Title)
- 作者 (Author)
- 摘要 (Abstract)
- 绪论 (Introduction)
- 背景与相关工作 (Background and Related Work)
- 提出的方法 (Proposed Approach)
- 实验设计 (Experimental Setup)
- 结果与讨论 (Results and Discussion)
- 可信威胁 (Threats to Validity)
- 总结与未来工作 (Conclusion and Future Work)

评估 (Evaluation)  
评估系统 (Evaluated Systems)  
定量分析 (Quantitative Analysis)  
定性分析 (Qualitative Analysis)  
局限性 (Limitations)

# 论文逻辑结构在章节中的对应

- 问题定义
  - 标题、背景与相关工作
  - 摘要、绪论、总结与未来工作的主要部分
- 解决方案
  - 提出的方法
  - 摘要、绪论、总结与未来工作的一小部分（一句或一段，对思路的抽象）
- 实验设计与结果分析
  - 实验设计、结果与讨论、可信与威胁
  - 摘要、绪论、总结与未来工作的最后部分

# 大纲

- 课程相关概念定义
- 论文基本特征
- 论文基本结构
- 论文评价指标
- 如何（有效）获取论文

# 计算机领域论文具有不同的评价指标

- 论文最常见的评价指标：EI，SCI，核心期刊
  - EI多为会议论文，SCI多为期刊论文
- 文史类常见指标体系
  - 核心期刊、党报/新华社社论、专著
- 理化生医类常见指标体系
  - 基于影响因子的中科院SCI分区（共四区）
- 计算机类无法适用于上述指标体系
  - 计算机类顶尖论文以会议为主，SCI分区不合理
  - CCF论文推荐列表

# CCF论文推荐列表



- CCF针对计算机不同领域的会议和期刊分别进行打分
- 每个领域内按照业内共识分为A、B、C三个等级
- 软工领域A类的会议和期刊
  - 会议：ICSE FSE ASE
  - 期刊：TSE TOSEM
  - 相关领域A类期刊与会议：**TOPLAS** OSDI PLDI POPL SOSP OOPSLA
- 软工领域重要的B类会议与期刊
  - 会议：ICPC, ICSME, ISSTA, SANER, RE
  - 期刊：JSS, JSEP, EMSE

# 大纲

- 课程相关概念定义
- 论文基本特征
- 论文基本结构
- 论文评价指标
- 如何（有效）获取论文

# 论文获取

- 名字已知：直接搜索引擎或学术搜索
- 名字未知：利用dblp数据库
  - 遍历已有会议（A、B类）
  - 跟踪领域内热门学者
    - 快捷搜索：在搜索引擎中输入“dblp+学者全名”

# 论文阅读的四种方式

- 略读
  - 目标：定位该论文是否在兴趣范围之内
  - 主要阅读论文的标题、作者、摘要、总结及绪论倒数第二段
- 泛读
  - 目标：了解论文的主要内容、关注点与技术创新点
  - 主要阅读论文的绪论，并快速浏览方法实现与实验评估
- 精读
  - 目标：掌握论文的研究思路与逻辑，判定论文的实验可重复性
  - 浏览全部章节，理解提出的方法，关注实验的设计与评估效果
- 反复回顾
  - 目标：理解研究领域在方法假设、实验设计、遣词造句、机构关系上的“潜规则”，梳理技术发展脉络，对未来发展趋势有所预期
  - 深入研究方法的切入点、实验的设计思路、局限性，以及背景与相关工作



# 作业： 挑选一篇论文

- 要求：

- 来源：本年度的ICSE、FSE、ASE
- 请选择你最喜欢的一篇
  - 最好每个会议都有候选论文
- 尽可能选择论文标题比较长的paper
  - 更贴近于标准的Hammer-Nut结构
- 请在钉钉群个人文件夹中上传选好的论文和简要的笔记文件，包含你对论文摘要、绪论（部分）、总结与未来工作这三块的理解