

浅谈如何做软件 engineering 研究

江 贺
2017-8-6



1 / 1.1 软件工程背景

软件工程：将系统化的、严格约束的、可量化的方法应用于软件开发、运行和维护中。

指出软件过程跨越了生命周期的不同阶段。

强调了对系统化的、严格约束的方法的需求

强调了量化的重要性（不能量化，就不能管理）



软件工程

软件工程是一个跨学科的学科。它从技术问题（数据或操作系统）跨越到语言问题，再到社会问题和心理学。其意味着对软件开发方法的改变，需要用**科学的方法**来研究

什么是科学的方法

- 1) 科学方法就是人类在所有认识和实践活动中所运用的全部正确方法。
- 2) The scientific method is a body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge.

1 / 1.3 软件 Engineering 研究方法分类

理论研究

经验研究

调查法

一种通过收集来自于人或者与人有关的信息，来描述、比较或者解释人们的知识、态度和行为的方法。

案例研究

一种经验性探究方法，它通过多个证据源来调查在真实环境下当前软件工程现象的一个实例。

实验

一种经验性探究方法，它操控一个变量，在其他变量不变的前提下，通过对主体施行不同方案，来观察影响。

问卷调查

问卷可以是纸质的，也可以是某种电子形式如网页或者邮件。基本方法是将问卷发出去，等待应答者的反馈。

注意：问卷的设计不能过长，应精选问题。

用尽办法让应答者用心答题。

预设选项+自由填空。

访谈

让访谈者处理问卷调查而不是由应答者自己填写，更有优势：

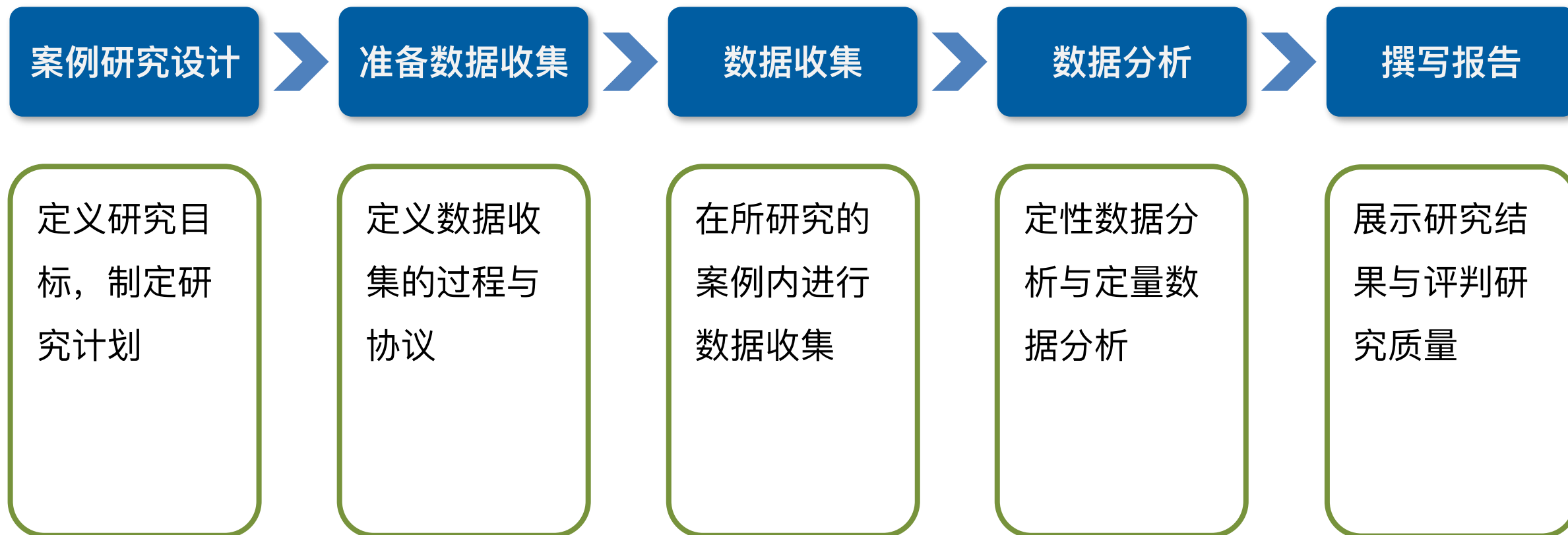
访谈调查会得到比邮件调查等更高的回复率。

访谈通常会降低“不知道”和“无应答”的数量。

访谈者不仅可以观察，还可以提问。

1 / 1.5 案例研究

案例研究5个主要的过程步骤



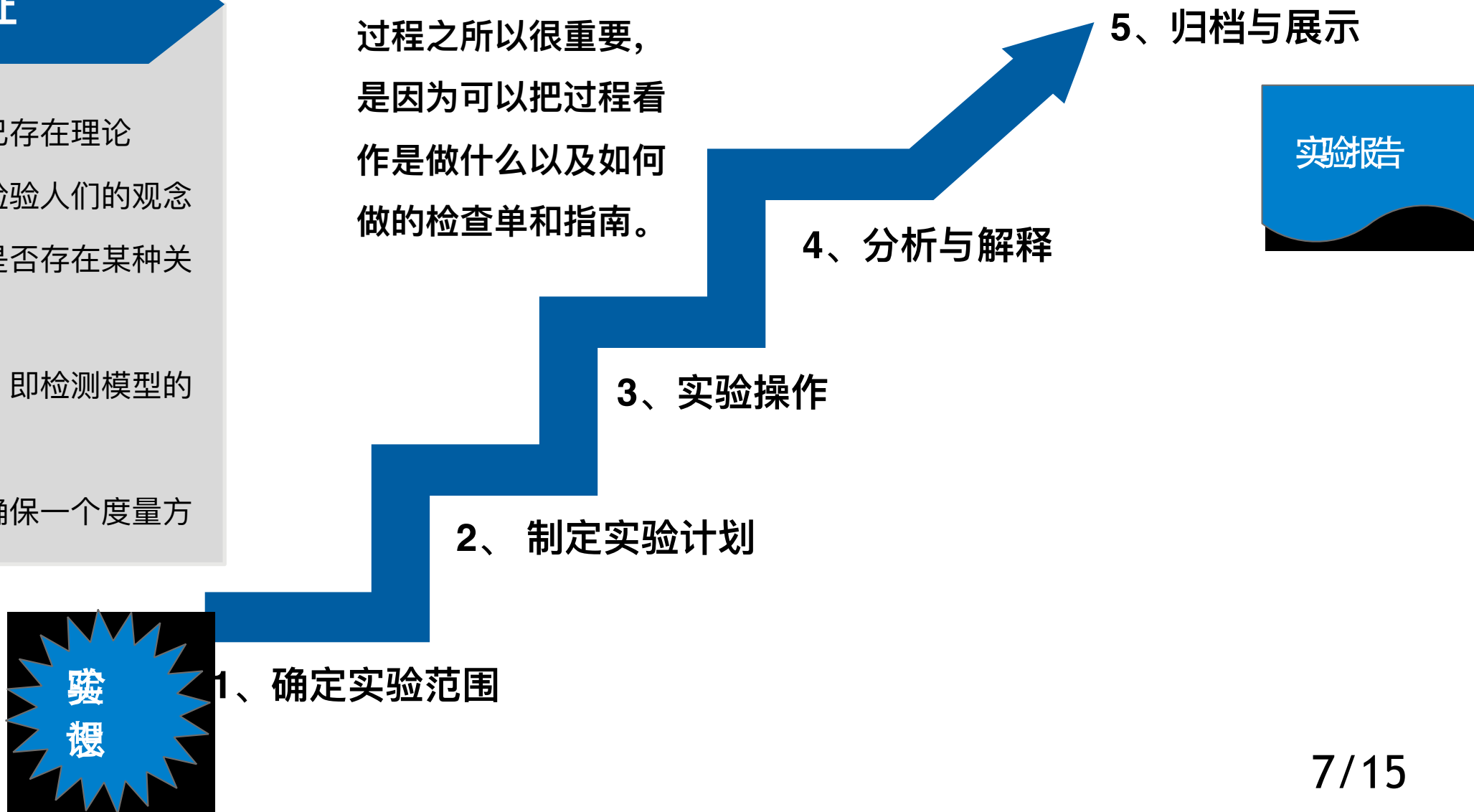
Ma W, Chen L, Zhang X, et al. How do developers fix cross-project correlated bugs?: a case study on the GitHub scientific python ecosystem[C]//Proceedings of the 39th International Conference on Software Engineering. IEEE Press, 2017: 381-392. 6/15

实验特征

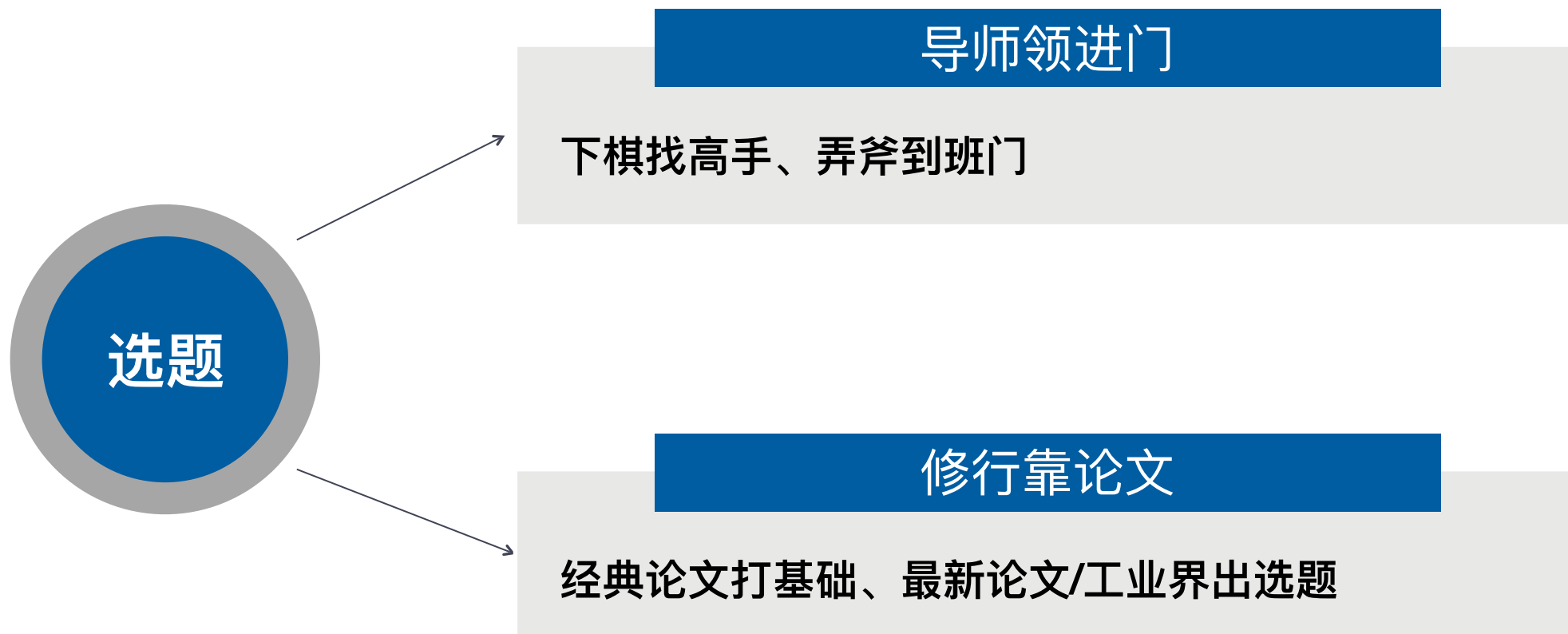
- (1) 确认理论，即检验已存在理论
- (2) 确认传统观点，即检验人们的观念
- (3) 探索关系，即检测是否存在某种关系
- (4) 评估模型的准确性，即检测模型的准确性是否符合预期
- (5) 验证度量方法，即确保一个度量方法。

实验过程概览

过程之所以很重要，是因为可以把过程看作是做什么以及如何做的检查单和指南。



2 / 2.1 如何选择候选研究方向



2 / 2.2 如何判定好的研究方向

定题

```
graph LR; A((定题)) --> B[潜力股]; A --> C[分治可控];
```

潜力股

工业界新出现的挑战、顶会顶刊上新发表论文等

Tips: 中国工业界的应用驱动的研究是国际学术界喜闻乐见的。

分治可控

一个研究方向可以分解为若干相关联、自主可控
(数据可获得/构建、现有结果有提升空间) 的题目

Tips: 构建数据集是软件工程领域的重要贡献，也是提高影响力的重要方式！

Tips: 10%以上的指标提升是顶会顶刊论文的潜在要求。

Jiang H, Li X, Yang Z, et al. What causes my test alarm?: automatic cause analysis for test alarms in system and integration testing[C]//Proceedings of the 39th International Conference on Software Engineering. IEEE Press, 2017: 712-723. 9/15

2 / 2.2 如何判定好的研究方向

定题

潜力股

工业界新出现的挑战、顶会顶刊上新发表论文等

Tips: 中国工业界的应用驱动的研究是国际学术界喜闻乐见的。

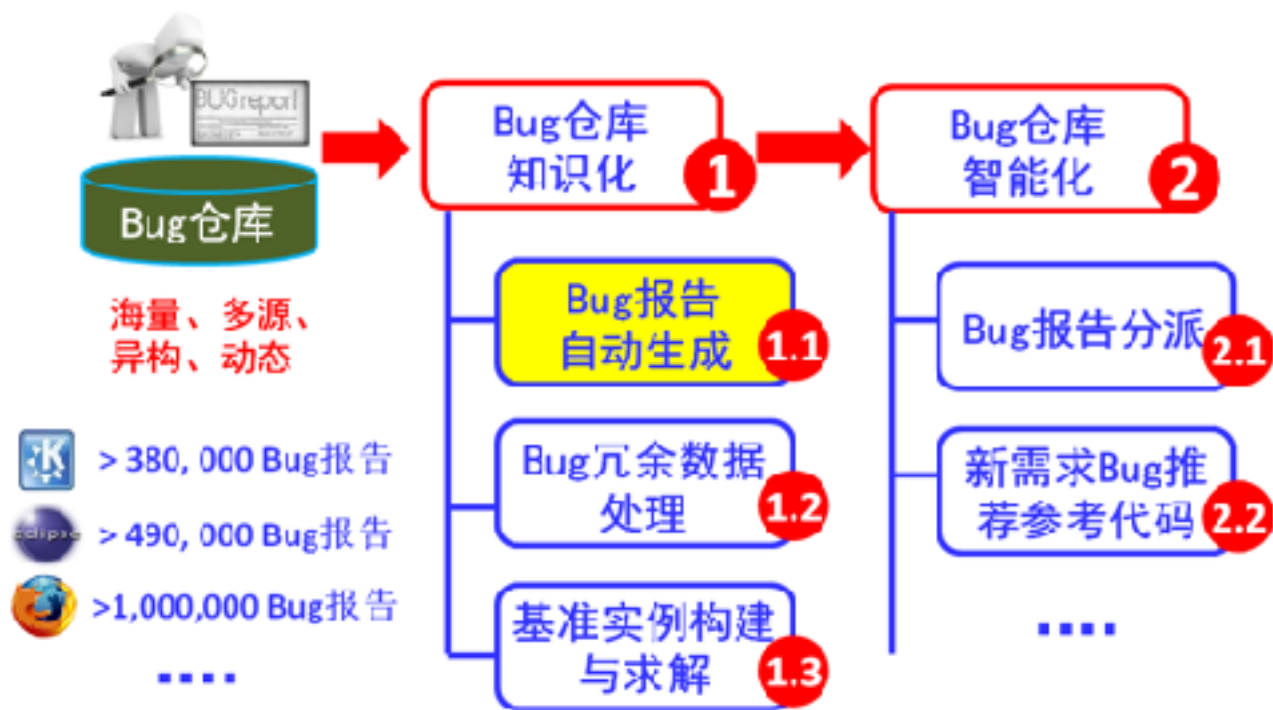
分治可控

一个研究方向可以分治
(数据可获得/构建、

Tips: 构建数据集是重要方式!

Tips: 10%以上的指

Jiang H, Li X, Yang Z, et al. What causes
for test alarms in system and integration
International Conference on Software En



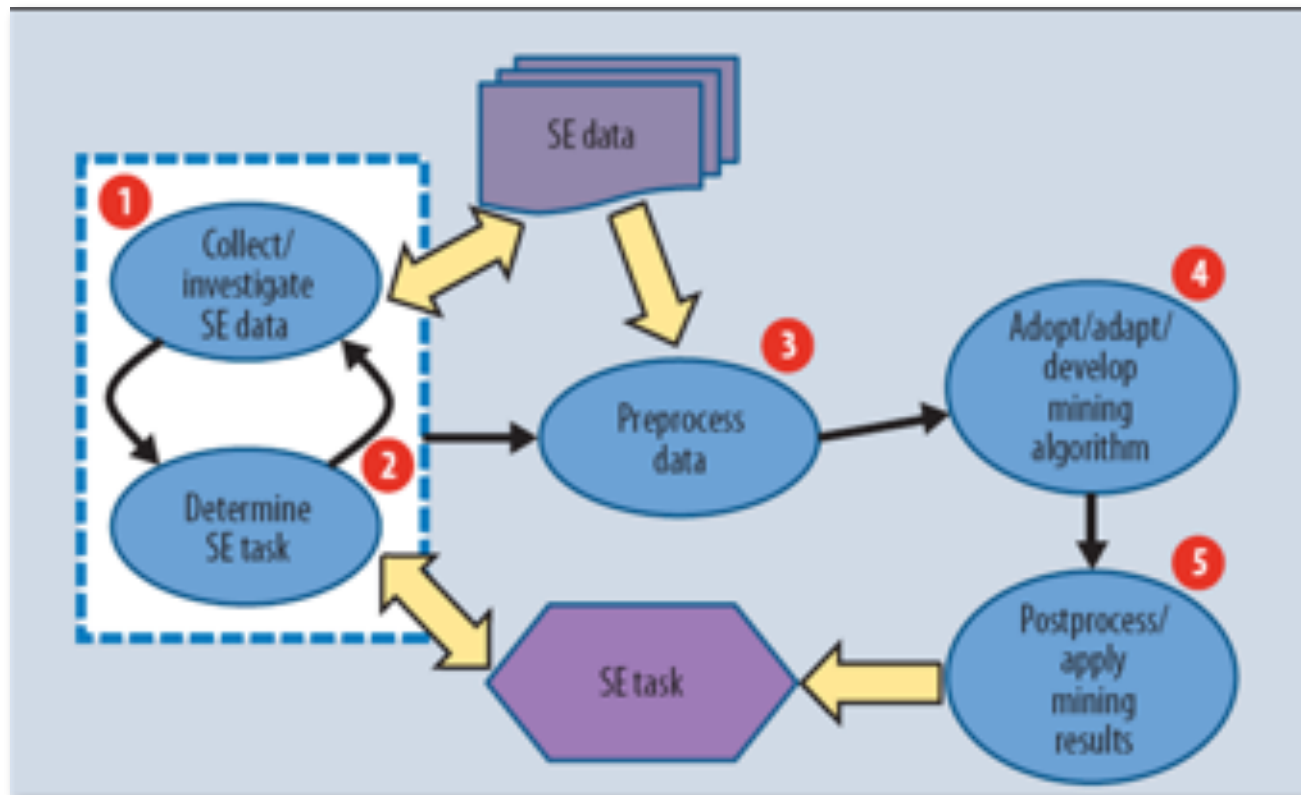
2 / 2.3什么是好的idea?

- 新鲜、有软工特色
- 外行看了有道理
- 内行看了有水平
- 可操作、有效果

Tips: Idea 有趣有道理，再动手实验。

2 / 2.4怎么提出好的idea?

➤ 普适方法



2 / 2.4怎么提出好的idea?

➤ 他山之石

e.g., 深度学习做缺陷定位、代码推荐、API推荐、重复bug报告检测、bug报告摘要

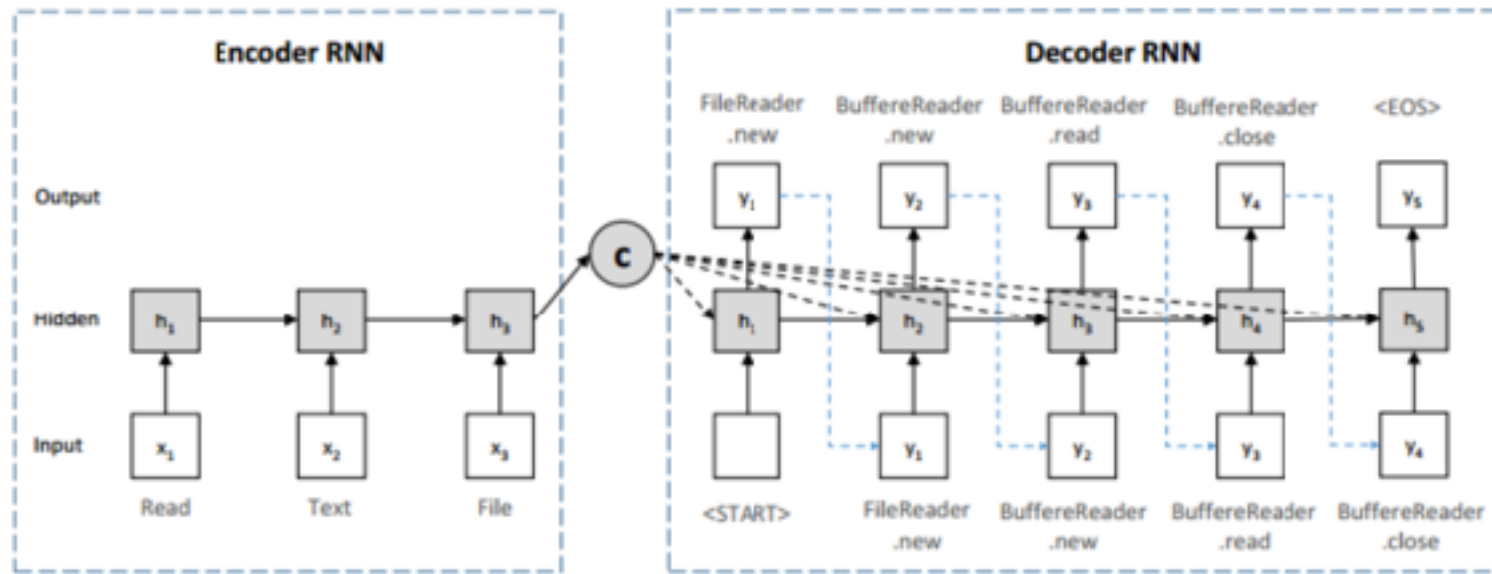


Figure 2: An Illustration of the RNN Encoder-Decoder Model for API learning

Tips: 他山之石+软工特色才能珠联璧合。【天下武功、唯快不破】

2 / 2.4怎么提出好的idea?

➤ 众智驱动

StackOverflow、GitHub

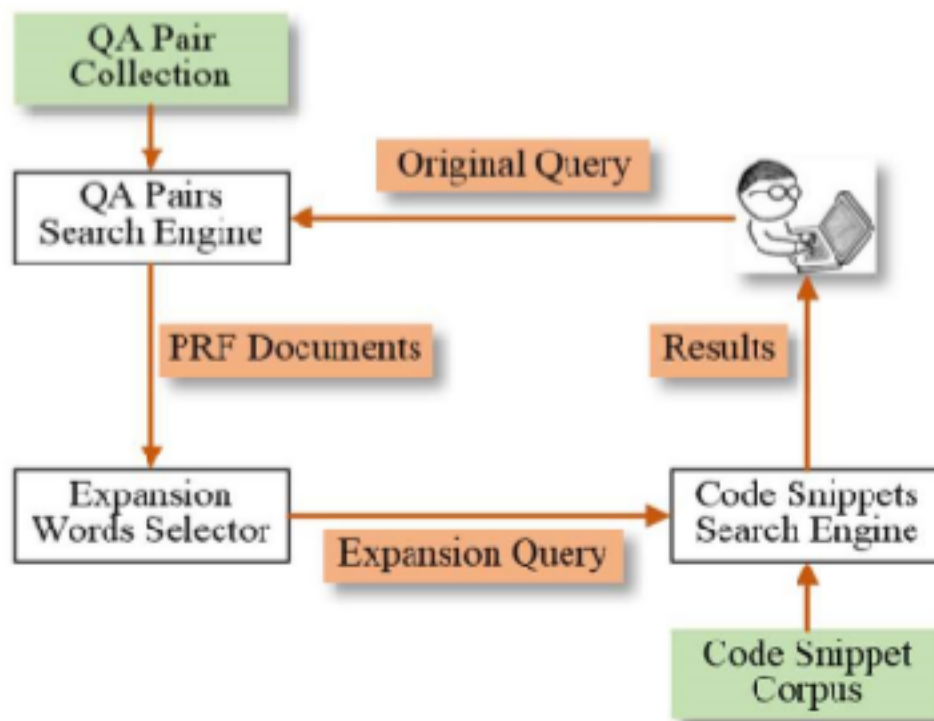


Fig. 1. Overall structure of QECK based Rocchio's model

➤ 逆向思维

• 非欧几何学的创立

《几何原本》：五条公设中第5条：若一直线落在两直线上所构成的同旁内角和小于两直角，那么把两直线无限延长，它们将在同旁内角和小于两直角的一侧相交



高斯



罗巴切夫斯基



波尔约

19 世纪，高斯、波尔约及罗巴切夫斯基等人一反过去人们试图从其他公设、公理及定理来证明第五公设的做法，而从打破“第五公理成立”这一潜规则着手，从而发展起第五公理不成立的新几何学，创立了“非欧几何学”。



在非欧几何的基础上，黎曼创立了黎曼几何，成为相对论的数学基础。

2 / 2.4怎么提出好的idea?

➤ 逆向思维

· 非欧几何

《几何原本》：
同旁内角和小于两直角
角和小于两直角

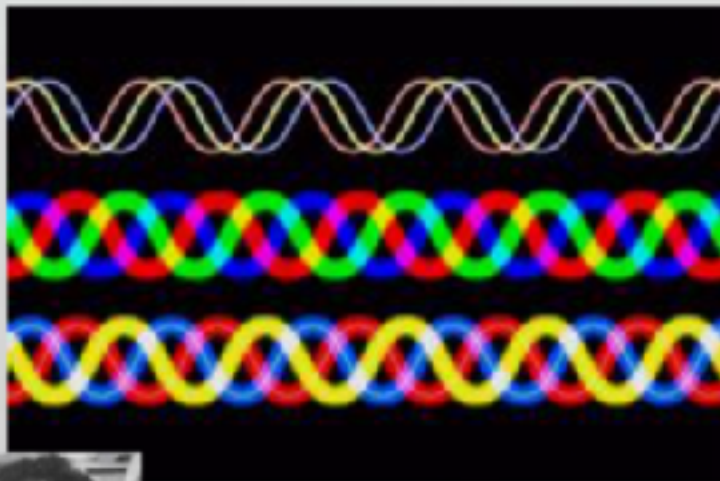


高斯



· 量子力学的创立

光是波还是粒子？



对光子的认识，存在

- 波动学说
- 粒子学说

这两种学说在长达百年之久的争论中，交替占据上风。往往一个新的实验结果的诞生，就导致这两种学说的地位的转变。

先后有海森堡、泡利、薛定谔等著名科学家参与了这场旷世辩论，甚至包括大名鼎鼎的爱因斯坦。

波粒二象性！



德布罗意

Summary of Take-Home Tips

Tips: 中国工业界的应用驱动的研究是国际学术界喜闻乐见的。

Tips: 构建数据集是软件工程领域的重要贡献，也是提高影响力的重要方式！

Tips: 10%以上的指标提升是顶会顶刊论文的潜在要求。

Tips: Idea 有趣有道理，再动手实验。

Tips: 他山之石+软工特色才能珠联璧合。【天下武功、唯快不破】

A close-up photograph of a person's hand holding a white rectangular card. The person is wearing a dark navy blue suit jacket. The card is held in the palm of the hand, with the fingers curled around it. The word "THANKS!" is printed on the card in a bold, blue, sans-serif font. The background is a plain, light-colored surface.

THANKS!