

总体要求：后台数据库的设计实现为主，前端界面仅需要能 list 各个表的数据即可

3 人/小组，要求有明确的分工，禁止划水和摸鱼

数据库不应少于 5 个表，要尽可能体现键、外键等

评分原则：工作量、复杂度、难度、新颖性

以小组形式汇报，上交汇报 PPT+设计文档+源代码（暂定）

- 1、参考 Github 网站，假设利用网络爬虫技术抓取 github.org 数据，相应的本地数据库的设计
- 2、Stackoverflow 网站，假设利用网络爬虫技术抓取 github.org 数据，相应的本地数据库的设计
- 3、
- 4、假设以从 Stackoverflow 网站爬取的数据为基础，进行代码分析、数据挖掘等方面的研究工作，相应的本地数据库的设计【对 Stack Exchange Data Dump 进行逆向工程，设计相应的本地关系行数据库】
- 5、逆向设计本地数据库，[Extraction and Mining of Academic Social Networks | AMiner](#)，并将数据导入【在此基础上补充 1-2 个保存进一步研究分析临时数据结果等的关系】
- 6、以从 Github 下载的 zip 包源代码为初始研究对象、以生成函数调用图维研究目标，相应数据库的设计。【参考 FlowDroid.】
- 7、参考和借鉴 Github，假设要实现一个软件版本控制系统，设计其后台数据库（暂不考虑代码合并、回滚等功能）
- 8、以从 Github 下载的 zip 包源代码为初始研究对象、假设以“**实现一个相对完整的源程序特征统计工具**”为研究目标，相应数据库的设计。
- 9、以从豌豆荚、google Playstore 等下载的 apk 为初始研究对象，以“**利用代码扫描等识别、搜集 APP 的权限信息，构建 APP-权限网络**”为研究目标，相应数据库的设计。
- 10、以[专利数据服务试验系统 \(cnipa.gov.cn\)](#)上的“中国发明专利申请公布全文文本数据”和“中国发明专利授权公告全文文本数据”为研究对象，参考其数据手册和样例数据，设计相应的本地关系数据库，为后续专利数据的分析和挖掘奠定基础。
- 11、以[专利数据服务试验系统 \(cnipa.gov.cn\)](#)上的“中国发明专利申请公布全文文本数据”和“中国发明专利授权公告全文文本数据”为研究对象，参考其数据手册和样例数据，设计相应的本地关系数据库，为后续专利数据的分析和挖掘奠定基础。
- 12、以[专利数据服务试验系统 \(cnipa.gov.cn\)](#)上的“美国专利申请公布全文文本数据”和“美国专利授权公告全文文本数据”为研究对象，参考其数据手册和样例数据，设计相应的本地关系数据库，为后续专利数据的分析和挖掘奠定基础。
- 13、假设通过分析研究 google Playstore 上 App 的用户评论来研究安卓应用质量等问题（可以替换成 JD 上的、淘宝上的商品级用户评论），设计相应的本地关系数据库存储相应数据。
- 14、基于 GH Archive 数据的逆向工程（设计相应的本地关系数据库存储相应数据）
- 15、JD 数据库？
- 16、Taobao 数据库？