

## 实验一：需求获取

### 实验目标：

- 软件需求的抽取与分类

### 实验基本思路：

- 选定一个开源项目，确定可能的信息来源，获取有效信息，使用特定技术获取潜在需求，并进行分类。所选项目要求如下：
  1. 必须是知名开源项目，如 Eclipse、VSCode 等；
  2. 有丰富的项目文档；
  3. 有确定的问题或缺陷跟踪管理系统（如 Bugzilla、GitHub Issue Tracker、Jira 等）；
- 具体过程如下：
  1. 确定项目；
  2. 明确信息源；
  3. 获取数据；
  4. 分析需求；
  5. 对需求进行分类。

### 思路一

JetBrains 的 IntelliJ 是一个著名的 IDE，其代码托管在 GitHub 上，项目地址为 <https://github.com/JetBrains/intellij-community>。

在这个项目中，有大量的 Pull Request (PR)。PR 是协作开发流程中的一种协作方式，某个开发者可以通过 PR 向项目管理者发起请求，将自己的完成的代码变更合并到项目中。因此，在许多情况下，可以将 PR 看作是一个需求。

因此，本项目的一种思路是：对 PR 进行分类，可以利用自然语言处理和机器学习技术。

类似的资源还包括：

- Eclipse 的缺陷报告数据 (<https://bugs.eclipse.org/bugs/>)
- Visual Studio Code 的 Issue 或 PR (<https://github.com/microsoft/vscode/issues> 等)

### 思路二

Stack Overflow 是开发者进行问答的著名网站，上面存在大量的有关软件开发的优秀问答。例如，我们通过其“IDE”标签进行搜索，就能够找到大量的有关 IDE 的问答 (<https://stackoverflow.com/questions/tagged/ide>)。这些问答也可以作为 Smart IDE 的潜在需求。

因此，思路二是通过爬虫抓取这些问答，然后对这些问答进行分析，从而获取可能需求。其中，可能的分析包括：将每个问答作为一个需求，直接对问答进行分类。对大量的问答文本进行词云分析，提取出若干关键词作为潜在需求。然后再对这些关键词进行分类（聚类）。除了“IDE”，也可以选用其他 Smart IDE 相关的标签进行搜索获取数据。

### 思路三

Eclipse 已经是一个非常成熟的 IDE，其网站 (<https://www.eclipse.org/eclipse/>) 上拥有非常丰富的数据。这些数据可能存在于其文档、论坛、邮件列表、Wiki 等资源中。因此，也可以从这些资源中获取潜在的需求，然后进行分类。

### 实验得分：

- (1) 基础得分，完成即可获得（共 80 分）
  - a) 数据获取：30 分
  - b) 需求分析：30 分
  - c) 需求分类：20 分
- (2) 主观评价（共 20 分）：实验内容和实验报告评价。
- (3) 每个人的得分（个人得分满分为 100）：
  - 小组得分\*小组人数\*个人分配所得比例。

### 实验提交：

1. 每个小组在 GitHub 上构建一个项目仓库，所有能够反映实验内容的材料（代码、数据、实验报告等），直接提交到项目仓库中。
2. **实验报告为必须提交的内容**，实验报告中必须给出：小组成员列表、小组成员成绩分配比例。实验报告命名：EX1.pdf。
3. 只需将项目仓库的地址发送到邮箱即可，邮箱：hren@smail.nju.edu.cn。邮件标题：软件需求工程\_EX1。
4. **实验截止日期：2020 年 11 月 1 日 24:00**（以项目仓库记录的提交日志时间为准）。
5. 如果不希望公开项目信息（包括个人信息），建议以私密仓库的形式建立项目仓库。

注意，实验报告请包含以下必须内容：

1. 小组成员以及得分分配；
2. 实验目的
3. 实验方法
4. 实验结果及效果分析
5. 结论