

# [CS209A-23Spring]期末项目(100分)

问题设计: Yida Tao

演示&数据: 邱逸伦、吴晓峰

评价: 陈秋江

Git & 代码风格: 闫云翔

提前提交: 5月23日23:55pm(第15周周二)。

最终提交: 5月30日23:55pm(第16周周二)。

最终截止日期后逾期投稿，恕不受理。

## 背景

在软件开发的过程中，会出现很多问题。开发人员可能会求助于问答网站，发布问题并寻求答案。

[Stack Overflow](#)就是这样一个面向程序员的问答网站，它属于[Stack Exchange Network](#)。Stack Overflow是一个用户提问和回答问题的平台，通过会员和积极参与，对问题和答案进行向上或向下的投票，并以类似wiki的方式编辑问题和答案。使用Stack Overflow的用户可以获得声望积分和“徽章”；例如，一个人在某一问题上获得“up”的投票或某一问题的答案，就可以获得10个声望点，并可以因其有价值的贡献而获得徽章。用户可以通过声望值的增加来解锁新的特权，比如投票、评论，甚至编辑其他人的帖子(来源: 维基百科)。

在这个最后的项目中，我们将使用Spring Boot开发一个web应用程序，用于存储、分析和可视化Stack Overflow问答数据w.r.t. [java编程](#)，目的是了解与java编程相关的常见问题、答案和解决过程。

## 基本要求(60分)

在Stack Overflow上，与Java编程相关的问题通常都标记为[Java](#)。你可以使用这个[java](#)标签来识别与java相关的问题。一个问题和它所有的答案和评论一起被称为一个[线程](#)。

对于Stack Overflow上与[java](#)相关的[线程](#)，我们感兴趣的是以下几个问题。

### 答案数量(15分)

- 有多少比例的问题没有答案? 答案的平均和最大数量是
- 多少? 答案数量的分布情况是怎样的?
- 

### 接受答案(15分)

- 有多少百分比的问题可以接受答案(一个问题只能有一个接受答案)?  
问题解决时间的分布情况(即问题发布时间与被接受答案发布时间之间的持续时间)是什么?
- 未被接受的答案(即未被标记为接受的答案)比被接受的答案获得更多赞数的问题比例是多少?
- 

## 标签(15分)

- 哪些标签经常和java标签一起出现?哪些标签或标签组合获得的赞数最多?哪些标签或标签组合获得的浏览量最多?

## 用户(15分)

- 许多用户可以参与一个帖子讨论。这种参与的分布是什么(即在一个帖子中发布问题、回答或评论的不同用户的数量)?哪些是经常参与帖讨论的最活跃的用户?
- 

你的web应用程序，在浏览器中打开，应该能够以适当的可视化方式回答这些问题。设计web应用程序并选择合适的可视化是你的工作，这样用户就可以轻松地使用你的应用程序来获得他们想要的答案。

## 实现的要求(25分)

### 数据收集和存储(10分)

你应该从Stack Overflow收集适当的数据来回答上述问题。请查看[Stack Overflow官方REST API文档](#)，了解收集不同类型数据的REST API。

- 为了使用其完整的REST API服务，你可能需要创建一个Stack Overflow账户。
- API请求受[速率限制](#)。请仔细设计和执行您的请求，否则您可能会很快达到您的每日限额。

Stack Overflow REST服务的连接有时可能不稳定。所以，请尽快启动数据收集!

•

在Stack Overflow上，有超过100万个线程被标记为java。你不必把它们都收集起来。然而，你应该收集至少500个线程的数据，以便从数据分析中获得有意义的见解。

建议你使用数据库(例如PostgreSQL, MySQL等)来存储数据。然而，如果你将数据存储普通文件中也是可以的。

## Web框架(10分)

你应该使用Spring Boot作为web框架。

## 前端(5分)

前端功能，如数据可视化和交互控件，可以用任何第三方库或框架在任何编程语言(如JavaScript、Java、JSP、HTML、CSS等)中实现。

## 高级要求(12分)

### 经常讨论的Java api(8点)

哪些Java api(例如类、方法)在Stack Overflow上被频繁讨论?要回答这个问题,你可能需要从线程内容(包括帖子、答案和评论)中提取代码片段,并进一步识别类名或方法名。

### REST服务(4分)

你应该构建一个web服务来回答上述问题,以便用户可以使用RESTful api来获得他们想要的答案。web服务可能包括本文档中定义的问题,或者你也可以定义新的问题。然而,你的web服务应该提供至少3个不同的RESTful端点来回答3个不同类型的问题(例如,获取<https://your.rest.server/java/answers?Status=accepted>将返回用java标记的问题的所有接受答案的数据)。

### 文档(3分)

你应该提供一份书面报告,描述你为这个项目收集的数据。书面报告还应该介绍项目的架构设计,以及重要的类、字段和方法。最后,你的报告应该突出你从数据分析结果中获得的见解,例如,Java编程的哪些主题被问到最多,等等。

## 团队合作

我们鼓励你们在这个最终项目中团队合作。最好的团队规模是2人,而3人或只有1名学生的团队也被允许。但是,规模为3的队伍不能只由CS学生组成。此外,规模为3的团队将获得项目分数的九折优惠,因为每个学生的平均工作量减少了。只有一个学生的团队将不会获得奖金,因为他/她不需要进行沟通,而沟通是昂贵的,但对团队合作至关重要。

请尽快找到你的队友,并在这个表格中填写你的团队信息:【腾讯文档】CS209A-23S-项目组队<https://docs.qq.com/sheet/DQ0FLdmN2TEhickJF?tab=BB08J2>

## 演示

我们提供了一个简单的演示,可以[在这里访问](#)。

注意,这个演示的数据是带有rust标签的Stack Overflow线程,这意味着你不能直接重用这个数据。

## 提交

请在截止日期前向Sakai提交一个名为“studententid - name - project .zip”的zip文件。提交的zip文件应包括两部分:

1. 项目文件夹,包括运行项目所需的所有源代码和其他相关文件。
2. 书面报告(.pdf格式)。

## 演讲

每个团队都应该在5月24日(第15周)或5月31日(第16周)的实验会议上展示你的项目。**你只能和同一实验课的同学组队**。另外,所有团队成员必须出席项目演示(演示不到场会扣分)。

要在5月24日(第15周)进行展示,你的团队需要在提前提交日期(5月23日)之前提交项目。在第15周提交并展示项目的团队将获得1分的**整体**课程成绩奖励。

此外,在第15周表现良好的团队将有机会在第16周的讲座(周二)中展示项目。这样的团队将获得**最多1分**的**整体**课程成绩奖励。

## 评价

- **功能:**每个团队必须在实验室会议期间展示项目(见上文),我们将检查你是否已经完成了现场所需的功能。
- **版本控制:**你应该使用GitHub来管理你项目的代码变更(有关如何使用git的更多细节,请参见实验室1)。你应该至少提交2次。在截止日期之前,你在GitHub上的远程仓库应该设置为**私有**,这样其他人就不会看到你的代码。
- **编码风格:**你应该注意编写可读和可维护的代码。如何使用CheckStyle达到这个目的参见实验1。截止日期后,你可以将你的GitHub仓库设置为**public**,我们将根据google\_checks.xml检查你的提交是否减少了CheckStyle警告。