

# [CS209A-23春季]最终项目（100分）

---

问题设计：陶怡达

演示和数据：邱一伦，吴晓峰 评价：陈秋江

Git与代码风格：严云翔

提前提交：5月23日晚23:55（第15周的星期二）。最后提

交：5月30日晚23:55（第16周的星期二）。

在最后截止日期之后提交的文件将不被接受。

## 背景介绍

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来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网络。Stack Overflow作为一个平台，供用户提问和回答问题，并通过会员资格和积极参与，对问题和答案进行向上或向下投票，并以类似于维基的方式编辑问题和答案。Stack Overflow的用户可以获得信誉积分和"徽章"；例如，一个人在一个问题或一个问题的答案上获得"向上"的投票，就可以获得10分的信誉积分，并且可以因其宝贵的贡献而获得徽章。用户随着声誉的增加而解锁新的特权，如投票、评论，甚至编辑他人的帖子（来源：维基百科）。

在这个毕业项目中，我们将使用Spring Boot开发一个网络应用程序，用于存储、分析和可视化Stack Overflow问答数据，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被标记为java。你可以使用这个java标签来识别与java相关的问题。一个问题它的所有答案和评论一起被称为一个线程。

对于Stack Overflow上的java相关线程，我们对以下问题感兴趣。

### 答案的数量（15分）

- 没有任何答案的问题的百分比是多少？答案的平均数和最高
- 数是多少？答案数量的分布情况是怎样的？
- 

### 接受的答案（15分）

- 有多少百分比的问题有接受的答案（一个问题只能有一个接受的答案）？
- 问题解决时间（即问题发布时间和被接受的答案发布时间之间的时间）的分布是什么？
- 有多大比例的问题的非接受的答案（即没有被标记为接受的答案）比接受的答案获得了更多的赞成票？

### 标签（15分）

- 哪些标签经常与java标签一起出现？哪些标签或标签组合
- 收到最多的加注？哪些标签或标签组合收到最多的浏览量
- ？

### 用户（15分）

- 许多用户可以参与到一个主题的讨论中。这种参与的分布是什么（即在一个线程中发布问题、答案或评论的不同用户的数量）？
- 哪些是经常参与主题讨论的最活跃用户？

你的网络应用程序在浏览器中打开后，应该能够**正确**回答这些问题。

**可视化**。你的工作是设计网络应用程序并选择适当的可视化，以便用户可以舒适地使用你的应用程序来获得他们想要的答案。

## 实施的要求（25分）

### 数据收集与存储（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的官方REST API文档](#)，了解收集不同类型数据的REST API。

- 你可能需要创建一个Stack Overflow账户，以便使用其完整的REST API服务。
- API请求是有**速率限制**的。请仔细设计和执行您的请求，否则您可能很快达到您的每日配额。
- 与Stack Overflow REST服务的连接有时可能不稳定。所以，**请尽快开始收集数据!**

在Stack Overflow上有超过100万个标记为java的线程。你不必全部收集它们。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。

建议你使用一个数据库（如PostgreSQL、MySQL等）来存储数据。然而，如果你将数据存储在普通文件中也是可以的。

### 网络框架（10分）

你应该使用[Spring Boot](#)作为Web框架。

### 前台（5分）

前端功能，如数据可视化和互动控制，可以用任何编程语言（如JavaScript、Java、JSP、HTML、CSS等）和任何第三方库或框架实现。

## 高级要求（12分）

### 经常讨论的Java APIs（8分）

Stack Overflow上经常讨论哪些Java API（如类、方法）？为了回答这个问题，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，并进一步确定类名或方法名。

### REST服务（4分）

你应该建立一个回答上述问题的 *网络服务*，以便用户可以使用RESTful API来获得他们想要的答案。该网络服务可以包括本文档中定义的问题，或者你也可以定义新的问题。尽管如此，你的网络服务应该提供至少3个不同的RESTful端点，回答3种不同类型的问题（例如，GET <https://your.rest.server/java/answers?status=accepted>将返回所有被接受的问题的答案数据，这些问题被标记为java）。

## 文件（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你项目的架构设计，以及重要的类、字段和方法。最后，你的报告应该强调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得最多，等等。

## 团队合作

我们鼓励你以团队的形式来完成这个期末项目。首选的团队规模是2人，但也允许3人团队或只有1名学生的团队。但是，3人团队不能只由CS学生组成。此外，3人团队的项目分数将有90%的折扣，因为每个学生的平均工作量减少了。只有一名学生的团队不会得到奖金，因为他/她不必进行沟通，而沟通对于团队合作来说是昂贵而关键的。

请尽快找到你的队友，并在此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。

请注意，这个演示的数据是带有锈迹标签的Stack Overflow线程，这意味着你不能直接重复使用这些数据。

## 提交

请在截止日期前向Sakai提交一个名为 "StudentID-Name-Project.zip" 的压缩文件。提交的压缩文件应包括两个部分：

1. 项目文件夹，其中包括所有源代码和运行项目所需的其他相关文件。
2. 一份书面报告（.pdf）。

## 演示文稿

每个团队应在5月24日（第15周）或5月31日（第16周）的实验课上展示你的项目。**你只能与同一实验课的人组队**。此外，所有团队成员都必须出席项目展示（如果你没有出现在展示中，将被扣分）。

为了在5月24日（第15周）展示，你的团队需要在提前提交日期（5月23日）前提交项目。在第15周提交并展示项目的团队将获得1分的课程总成绩奖励。

此外，在第15周表现出色的团队将有机会在第16周的讲座（星期二）上展示该项目。这样的团队将**最多获得1分**的课程总成绩奖励。

## 评价

- **功能性**：每个团队必须在实验课上展示项目（见上文），我们会在现场检查你是否完成了所需的功能。
- **版本控制**：你应该使用GitHub来管理你的项目的代码变化（关于如何使用git的进一步细节，请参见实验1）。你应该至少做两个提交。你在GitHub上的远程 repo应该在截止日期前设置为**私有**，这样就不会有其他人看到你的代码。
- **编码风格**：一路走来，你应该注意编写可读和可维护的代码。关于如何使用CheckStyle来达到这个目的，请看实验1。在截止日期之后，你可以将你的GitHub repo设置为**公众号**，我们将检查您的任何提交是否按照以下方式减少了CheckStyle的警告  
google\_checks.xml。