

[CS209A-23Fall] Final Project 评分标准 (100分)

数据收集 (10分)

- 至少500条数据，数据都是有 `java` tag的 (5分)
- 数据存储 (5分)

话题热度 (20分)

- 不少于10个不同的话题 (2分)
- 不少于3个不同的“热度”指标，每个4分 (12分)
- 可视化能够比较不同话题的热度 (6分)

Bug热度 (20分)

- 从Stack Overflow raw data中提取出error与exception的信息 (5分)
- 合理的热度指标 (5分)
- 可视化能够比较同一类errors或exceptions (5分)
- 可视化能够比较不同类errors或exceptions (5分)

相关话题 (20分)

- 合理的话题相关度指标 (5分)
- 根据输入，返回相关的一系列话题 (5分)
- 可视化能够比较不同话题的相关程度 (10分)

REST服务 (20分)

- 单个话题热度接口 (4分)
- 多个话题热度 (需排序) 接口 (4分)
- 单个bug热度接口 (4分)
- 多个bug热度 (需排序) 接口 (4分)
- 相关话题 (需排序) 接口 (4分)

日志 (10分)

- 记录用户日常行为 (5分)
- 记录用户或服务器异常行为 (5分)

其它注意点

可视化

好的可视化应准确、有效，清晰、简洁地传达信息。大家需要根据数据的性质选择最合适的呈现方式，以确保用户能够正确地解读图表，理解数据的含义。同时，可视化应兼顾美观、交互性等。因此，对可视化部分的评分会综合考虑以上方面，而不是仅仅实现功能就可以拿满分。

指标

什么算是“合理”的指标？这是一个开放性的问题。大家需要了解Stack Overflow的数据特点并**自行设计**合理的指标来衡量热度、相关度等概念。比如，对于“multithreading”这个话题，统计所有包含“multithreading”标签的threads的平均浏览量，就是一个合理的衡量“热度”的指标。然而，如果去统计有多少用户名包含“multithreading”，并以此衡量此话题的热度，这就不太合理。

数据获取

数据获取是offline的：所有数据要用RESTful API收集，并预先存储在本地。项目演示时，用户发送请求给服务器端，服务器端应从本地数据中获取相关信息，进行分析统计，再将结果返回给用户；而不是当场发送REST请求向Stack Overflow服务请求数据。

数据分析

数据分析是online的：演示时的数据分析过程应该是服务器根据用户请求当场进行运算（e.g., 根据用户请求，服务器从数据库读出相关数据，计算后得到结果，返回给前端，前端以图表方式展示），而不是预先算出结果，存好一张图片或静态网页，演示时直接在网页端展示静态内容。若采取后一种做法，项目分数扣除20分。