

[CS209A-25Fall] Final Project Grading Scheme (100 points)

Data Collection (10 points)

- **Collect at least 1,000 Java-tagged threads** from Stack Overflow using the official REST API.
(5 points)
- **Persist data locally** in a structured database (e.g., PostgreSQL, MySQL) or well-organized local files like json. (5 points)

Data Analysis (60 points)

1. Topic Trends (15 points)

- Topic identification: **5**
- Time dimension: **5**
- Meaningful measurement of "activity": **5**

2. Topic Co-occurrence (15 points)

- Topic identification: **5**
- Correct co-occurrence logic: **10**
- **Deductions:**
 - N too small or too large: **-2**

3. Common Pitfalls in Multithreading (15 points)

- Identifying pitfalls for multithreading: **10**
- Pitfall aggregation (find the common ones): **5**
- **Deductions:**
 - Only analyze tag, without thread body content: **0**
 - No text analysis (e.g., use AI generated common pitfalls): **0**
 - Limited or wrong approach for identifying pitfalls

4. Solvable vs. Hard-to-Solve Questions (15 points)

- Identify solvable questions: **3**
- Identify hard-to-solve questions: **3**
- Identify at least 3 meaningful contributing factors: **9**
- **Deductions:**
 - Shallow or wrong analysis
 - < 3 factors (**-3** each)

RESTful APIs (10 points)

- Available REST API: **5**
- Returns valid JSON with correct result: **5**

[CS209A-25Fall] 最终项目评分标准（100 分）

数据收集（10 分）

- 使用官方 REST API 从 Stack Overflow 收集至少 1,000 个 Java 标记的线程。
(5 分)
- 将数据本地持久化到结构化数据库（例如，PostgreSQL、MySQL）或组织良好的本地文件中，如 json. (5 分)

数据分析 (60 分)

1. 主题趋势 (15 分) 主题识别：5

-
- 时间维度：5 对"活动"进行有意义的测量：5
-

2. 主题共现（15 分）主题识别：5

-
- 正确共现逻辑：10
- 推论：
 - N 太小或太大：-2

3. 多线程常见陷阱（15 分）识别多线程陷阱：10 分

-
- 陷阱汇总（找出常见的）：5 分
- 推论：
 - 仅分析标签，不分析线程体内容：0 分 无文本分析（例如，使用 AI 生成的常见陷阱）：0 分 识别陷阱的方法有限或错误：0 分
 -

4. 可解问题与难解问题（15 分）识别可解问题：3 个

-
- 识别难解问题：3 个 识别至少 3 个有意义的贡献因素：9 个
-
- 扣除：
 - 浅层或错误分析
 - <3 因素（每个-3 分）

RESTful API（10 分）

- 可用的 REST API：5
- 返回正确结果的 JSON：5

- **Deductions:**

- Returns non-JSON (e.g., HTML, plain text): **-5**
- Poor API design (e.g., confusing resources and parameters): **-3**
- < 2 REST API: **-5 each**

Visualization & Insights (20 points)

For each of the 4 question:

- Visualization: 3 points (good, fair, poor). A good visualization must be **accurate, clear, aesthetically appropriate, and user-friendly** to answer the question.
- Insights: 2 points (good, poor). You should discuss the insights during presentation.

Framework & Compliance (Deduction-only)

- Not using Spring Boot: **-40**
- Precomputed/static results: **-20**
- Use other language (e.g., Python) for data analysis: **-20**. Yet, it's okay to use Python **only for** data collection with Stack Overflow REST API.
- AI-generated analysis: **0 points for the affected question(s)**

- 扣除项：
 - 返回非 JSON 格式（例如，HTML、纯文本）：-5
 - API 设计不佳（例如，资源与参数混淆）：-3 < 2 REST API：每个-5
 -

可视化与洞察（20 分）

对于每个问题：

- 可视化：3 分（好、一般、差）。好的可视化必须准确、清晰、美观且用户友好，以便回答问题。
- 洞察：2 分（好、差）。你应在展示时讨论洞察。

框架与合规（仅扣分）

- 不使用 Spring Boot：-40
- 预计算/静态结果：-20 使用其他语言（例如 Python）进行数据分析：-20。然而，使用 Python 仅通过 Stack Overflow
- REST API 进行数据收集是可以的。
- AI 生成的分析：受影响的问题（如有）不得分：0 分