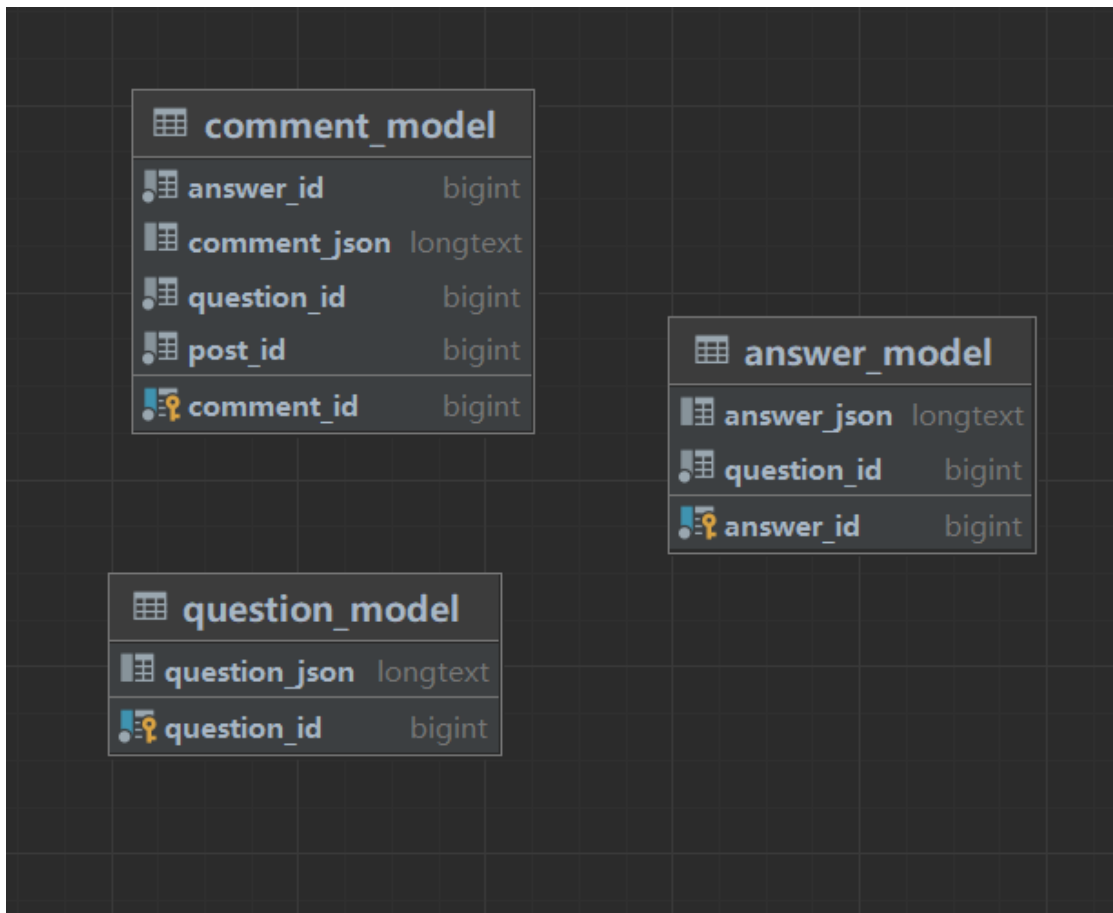


1. Describes the data you collected:

收集了 1038 个 thread。收集的问题是 2022 年 1 月 1 日到 2023 年 1 月 1 日前按时间早晚倒序取最晚的那一部分 thread。



存储方法是使用 MySQL 数据库。以 id 信息和 json 字符串的方式建表存入数据库中。表定义可见上图。

2. introduce the architecture design of your project

本 project 的代码分为两部分。一个是访问 Stack Exchange API 并存储的脚本。另一个是分析数据库数据并通过网页可视化的 MVC。渲染器使用 Thymeleaf。MVC 使用 Service 访问数据库对数据进行分析。MVC 返回的页面

是动态的，会根据数据库数据变化返回相应结果页面。

3. important classes, fields, and methods.

访问 StackExchange 的 API 的脚本是 **GrabScri.java**。它使用 hibernate orm 存储数据。一次运行得到整个 thread 是可能的，但 GrabScri.java 需要多次运行依次获得问题、答案和评论。需配置 Application.java 运行脚本。

WholeController.java 是本 project 设置的唯一一个 Controller。**WholeService.java** 读取数据库数据并给出每个问题的结果以供 WholeController.java 使用。

WholeController.java 为 **index.html** 配装数据。并响应请求返回 index.html。

4. highlight the insights you obtained from the data analysis results

答案数量问题

1. 没有答案的问题的百分比

0.3882466281310212

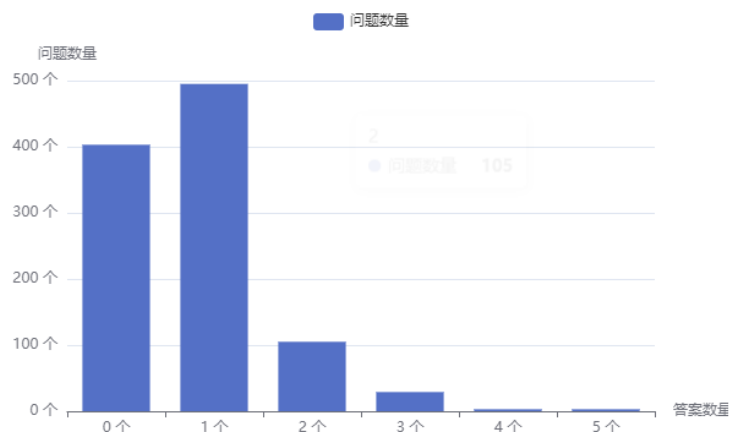
2. 每个问题平均有多少答案个数 最多有几个答案

平均有 0.7890173410404624个
最多有 5.0 个

上图显示接近 40%的问题没有答案。有热门问题含最多 5 个答案。

3. 问题答案数量分布情况 下面map的key是答案数量种类， value是问题个数

{0=403, 1=495, 2=105, 3=29, 4=3, 5=3}



上图显示问题答案数量分布情况。可见大部分只有 0 个或 1 个。有更多参与者和答案数量的问题占小部分。可见 StackOverflow 上的问题会比较特殊，大多只获得一个回答。没有更多解法。

已接受答案问题

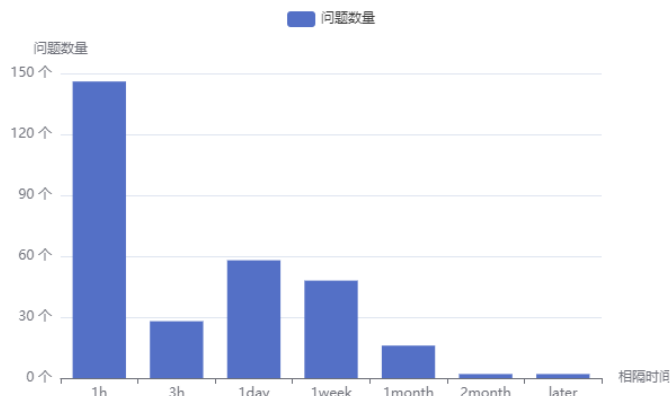
1. 有接受答案的问题的百分比

0.28901734104046245

上图显示有接受答案的问题只占小部分。

2. 从问题创建到出现接受答案的时间间隔分布 下面map的key是小时数， value是问题个数

{0=112, 1=34, 2=22, 3=6, 4=7, 5=5, 6=6, 7=3, 8=2, 9=3, 11=2, 12=5, 13=3, 14=3, 15=2, 17=3, 18=2, 19=2, 20=4, 21=1, 22=1, 23=1, 24=3, 25=4, 26=1, 28=2, 29=1, 30=1, 100=1, 109=1, 112=1, 114=1, 115=1, 117=1, 125=2, 134=1, 146=1, 155=1, 174=1, 188=1, 195=1, 215=1, 275=1, 333=1, 335=1, 354=1, 364=1, 366=1, 369=1, 382=1, 394=1}



上图显示时间间隔有很大一部分在 1 小时以内。这些问题可能比较有意义，比较容易。能够吸引解决问题能力强的人在短时间解出。

3. 在有接受答案的问题中 含有比接受答案更高的投票数的答案的百分比

0.42333333333333333334

此图显示有较多有接受答案的问题在问题提出者接受答案后仍会出现更被接受的解法。猜想最早出现的解法可能比较可以解决问题提出者的燃眉之急，后面提出的解法可能更加正确。

标签问题

1. 出现次数较多的标签组合

[[spring-boot]=167, [android]=143, [spring]=124, [spring, spring-boot]=61, [maven]=43, [kotlin]=36, [hibernate]=33, [android-studio]=28, [spring-data

Related Tag WordCloud



上图显示出现次数较多的标签组合, 大多跟 spring 框架有关。
可以看出此框架的热门和被关注。

2. 投票数多的标签组合

[[spring-boot]=79, [spring]=46, [android]=42, [spring, spring-boot]=24, [gradle]=17, [hibernate]=17, [maven]=15, [spring-security]=14, [spring-data-android]=9, [java-17, concurrency]=8]

Related Tag WordCloud



3. 浏览数多的标签组合

[[spring-boot]=39475, [spring]=21015, [spring, spring-boot]=12781, [android]=11919, [spring-security]=9606, [maven]=8261, [spring-boot, spring-security]=3446, [javascript]=3323, [lombok]=3236, [intellij-idea]=3144, [security]=3092, [spring-data-jpa, spring-boot]=3049]

Related Tag WordCloud

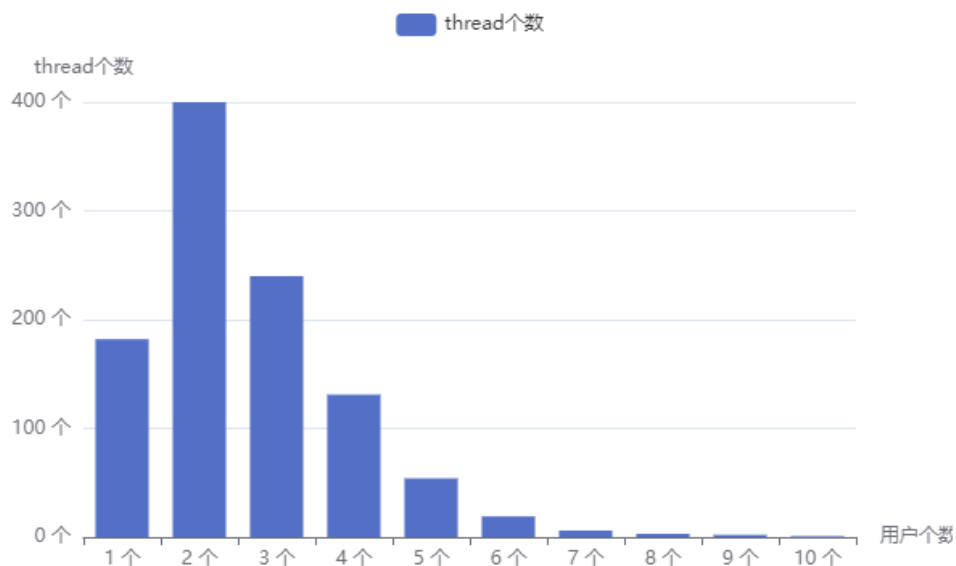


上面两个词云与出现次数词云有较多重合。可见出现次数与获得投票数和浏览数正相关。

用户问题

1. 每个thread有多少用户参与的分布

{1=182, 2=400, 3=240, 4=131, 5=54, 6=19, 7=6, 8=3, 9=2, 10=1}

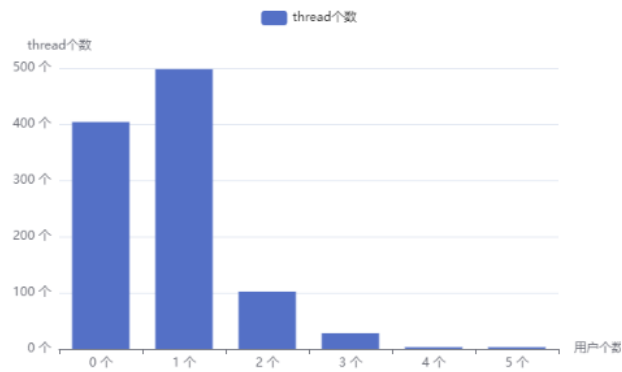


上图显示每个 Thread 的人数参与情况，可见大多在 1、2 和 3 个。猜想是 stackoverflow 比较活跃，问题提出较多。许多人来此网站寻求帮助，但得到多人关注的问题较少。

2. 对问题1分别根据答案和comment角度进行统计

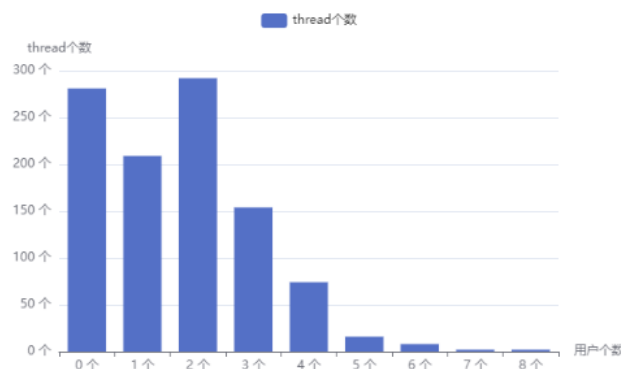
2.1 根据答案角度

{0=404, 1=498, 2=102, 3=28, 4=3, 5=3}



2.2 根据comment角度

{0=281, 1=209, 2=292, 3=154, 4=74, 5=16, 6=8, 7=2, 8=2}



上图显示每个 Thread 在回答和评论上的参与人数。其中答案的人数分布图与每个问题的答案数量分布图类似。评论在一个 Thread 中会出现更高的上限。可见人们更倾向于评论来参与 Thread 的交流。

3. 哪些用户活跃较多 (根据发的问题, 回答, comment数量之和)


[16376827=54, 11002=50, 5772882=47, 17949945=42, 112968=41, 139985=30, 642706=30, 1180351=29, 6395627=27, 466862=26, 52,

下图词云根据发的问题, 回答, comment数量之和排列大小 名词表示用户id



上图显示最活跃的 20 个用户。

About Products For Teams [Log in](#) [Sign up](#)



g00se
Member for 1 year, 11 months Last seen this week

[Profile](#) [Activity](#)

Summary

REPUTATION
2,676
top 2% this month
2.7k 2.1k 1.6k
AUG NOV FEB MAY

Top tag: java
Next privilege: 3,000 Rep.
Cast close & reopen votes

BADGES
2 4 8
Newest: java
Next badge: Excavator 0/1

IMPACT
~29k people reached 4 posts edited
1 helpful flag 71 votes cast

Answers
View all 265 answers

Score	Activity	Newest
5	StringBuilder not being initialized to the...	Jul 14, 2021
5	Trying to find the height and width of ...	Aug 12, 2021
4	How to format date to 01-JAN-2021 us-...	Jul 16, 2021
4	Regex: Word with parenthesis (out and in)	Jul 5, 2021
4	How to round up to the next hundred i...	Jul 25, 2021

Questions

Score	Activity	Newest	Views
0	Correct options to package dependen...	Aug 31, 2021	
0	Tomee does not use MySql datasource	Jul 2, 2022	
0	Entity bean field not updating	Jul 23, 2022	

上图是 id 为 16376827 的用户。

总结：

在 StackOverflow 上提出问题，60%概率得到回答。许多问题只有单个或没有回答，只有少数问题得到关注，含有多个回答。问题分布情况是多数有非常特定的条件，少数是多数人遇到的并更有价值的问题。

含有接受答案的比例占少部分，大约占 30%。可见答案效果大多不能接受。但可能是有答案有效但问题提出者没点击接受的原因。

Spring 框架火热，经常于 java tag 一同出现。

用户会倾向使用评论这种不是非常正式的方式参与 Thread 讨论。