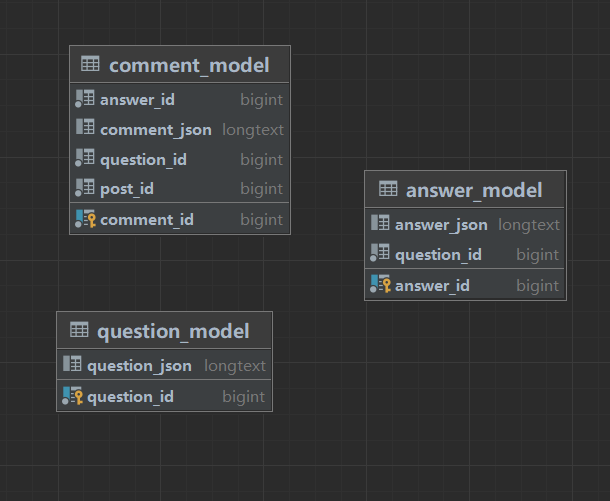
1. **Describes the data you collected:**

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。

1. **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。

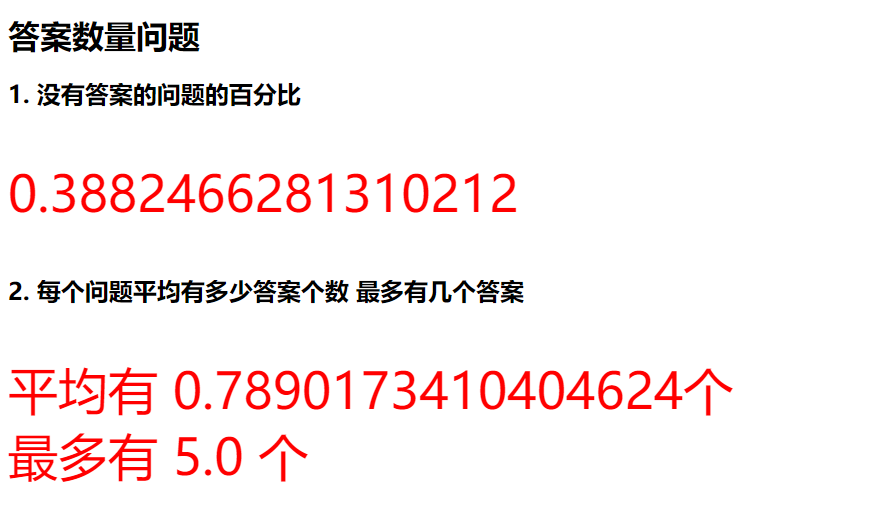
1. **important classes, fields, and methods.**

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。

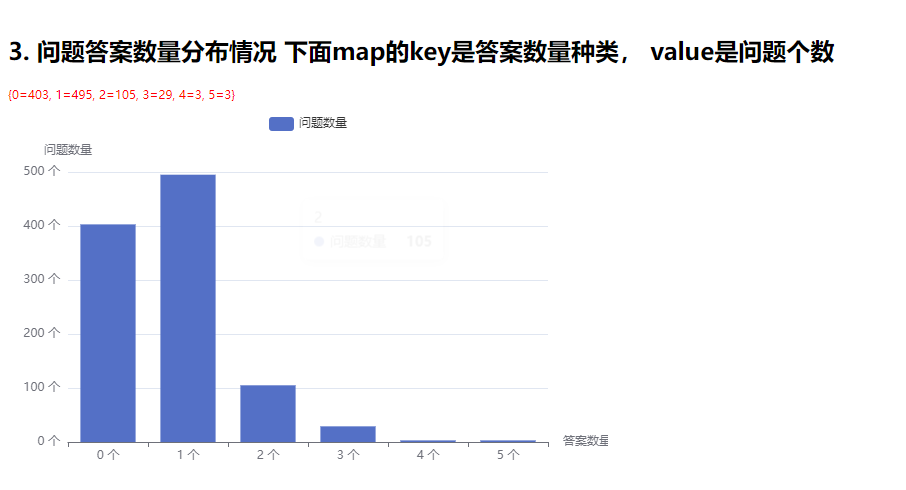
**WholeController.java**是本project设置的唯一一个Controller。**WholeService.java**读取数据库数据并给出每个问题的结果以供WholeController.java使用。

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

1. **highlight the insights you obtained from the data analysis results**

****

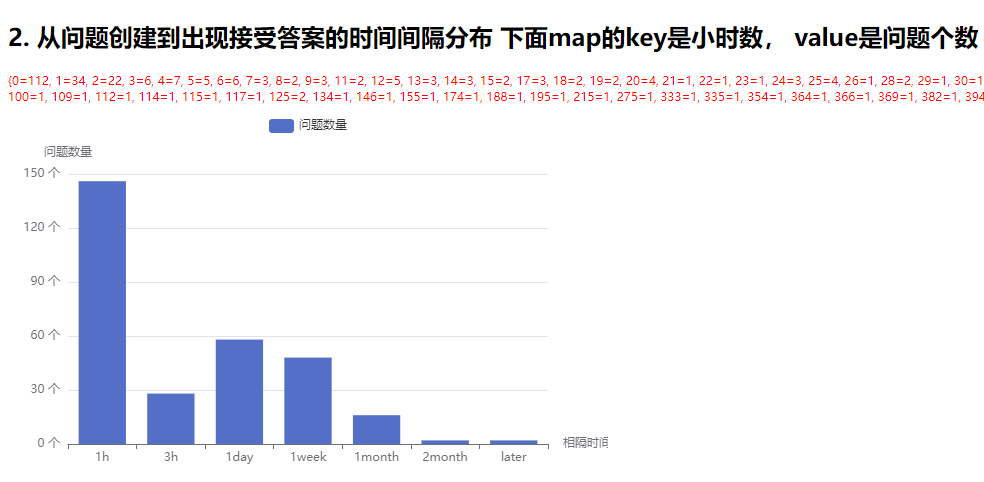
上图显示接近40%的问题没有答案。有热门问题含最多5个答案。



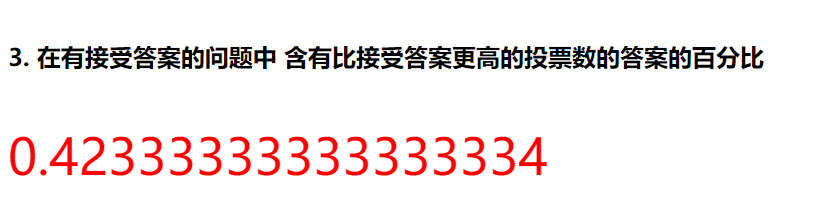
上图显示问题答案数量分布情况。可见大部分只有0个或1个。有更多参与者和答案数量的问题占小部分。可见StackOverflow上的问题会比较特殊，大多只获得一个回答。没有更多解法。



上图显示有接受答案的问题只占小部分。



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。



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。

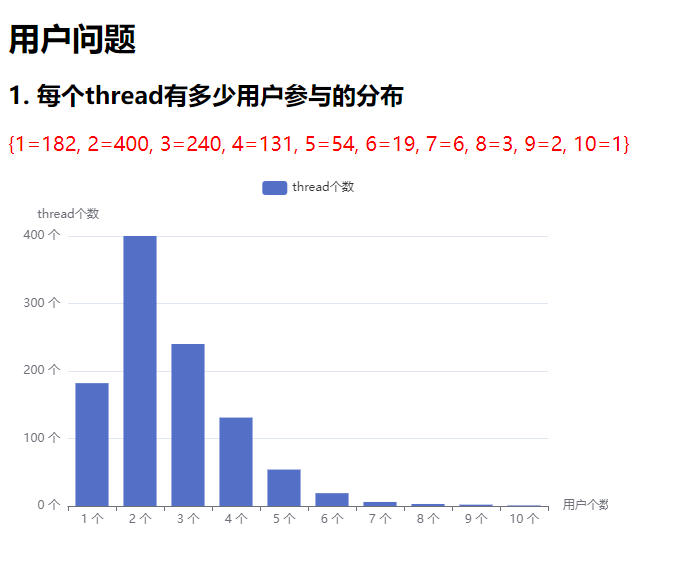


上图显示出现次数较多的标签组合，大多跟spring框架有关。可以看出此框架的热门和被关注。

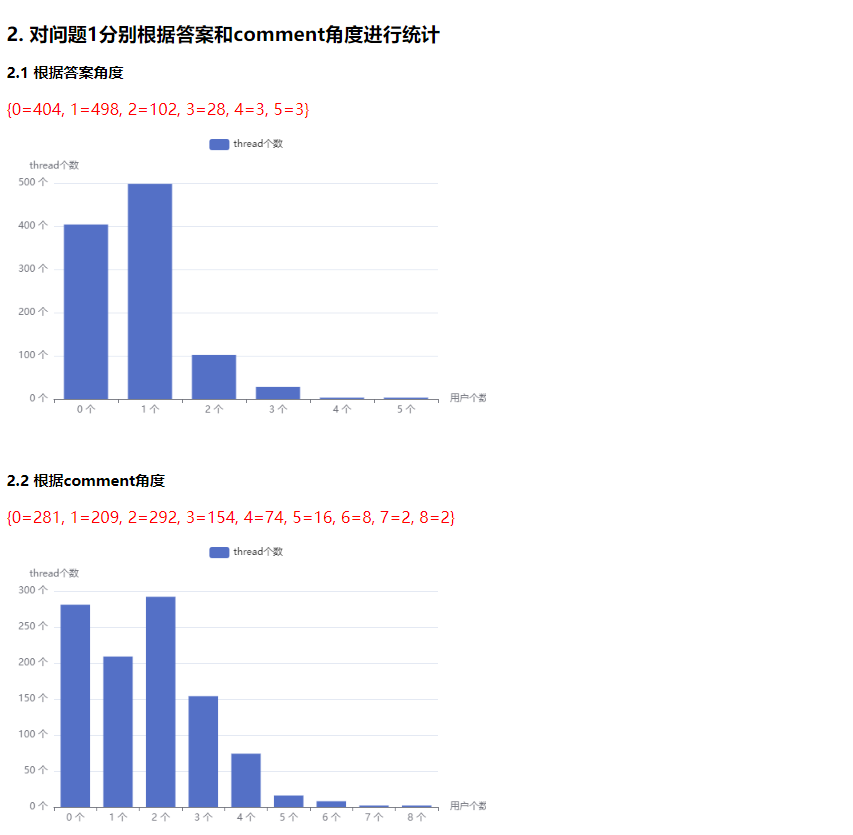




上面两个词云与出现次数词云有较多重合。可见出现次数与获得投票数和浏览数正相关。



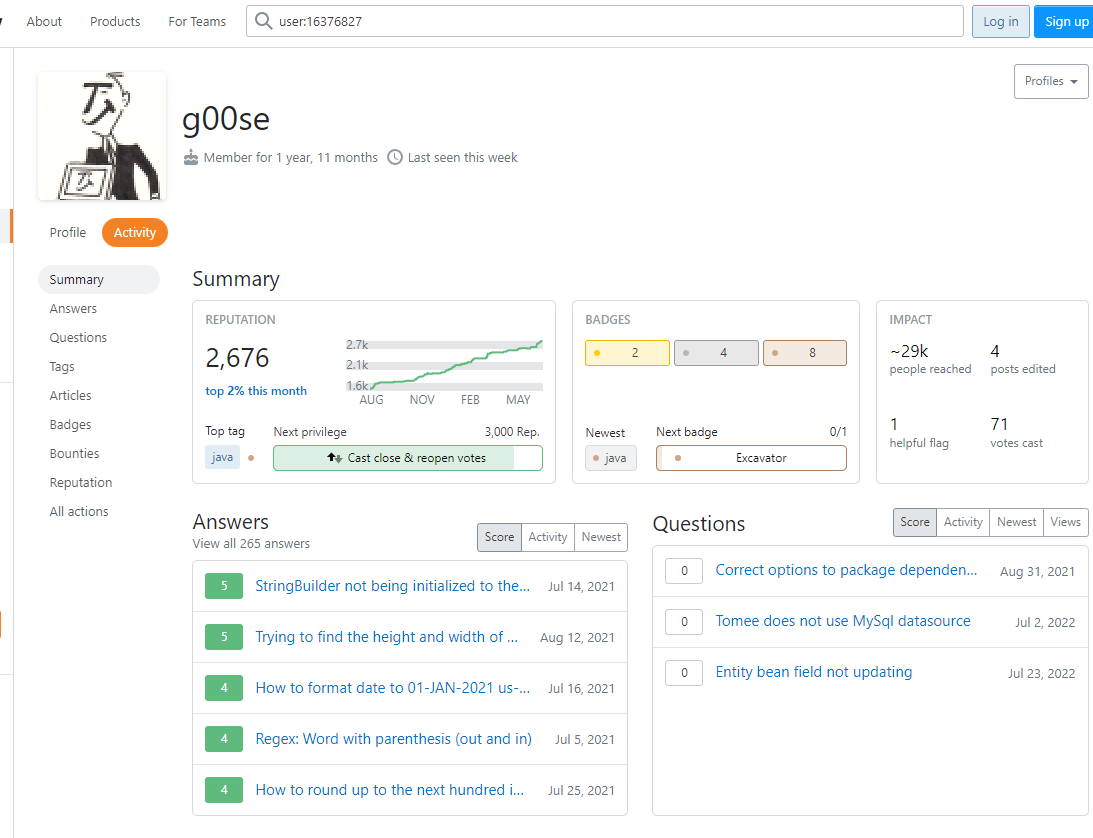
上图显示每个Thread的人数参与情况，可见大多在1、2和3个。猜想是stackoverflow比较活跃，问题提出较多。许多人来此网站寻求帮助，但得到多人关注的问题较少。



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。



上图显示最活跃的20个用户。



上图是id为16376827的用户。

总结：

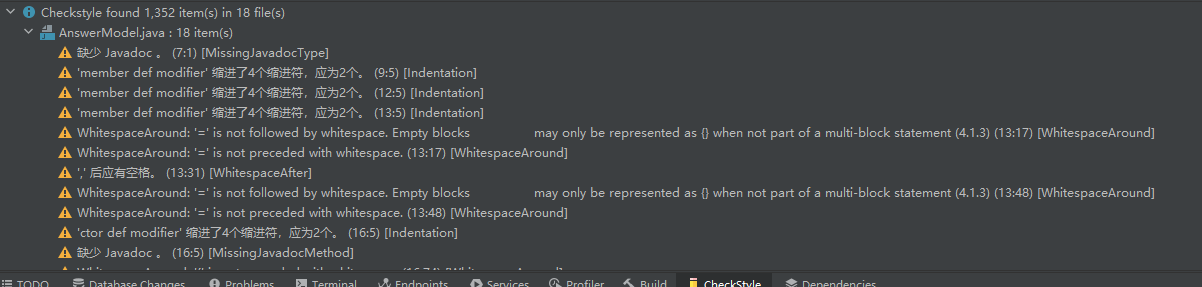
在StackOverflow上提出问题，60%概率得到回答。许多问题只有单个或没有回答，只有少数问题得到关注，含有多个回答。问题分布情况是多数有非常特定的条件，少数是多数人遇到的并更有价值的问题。

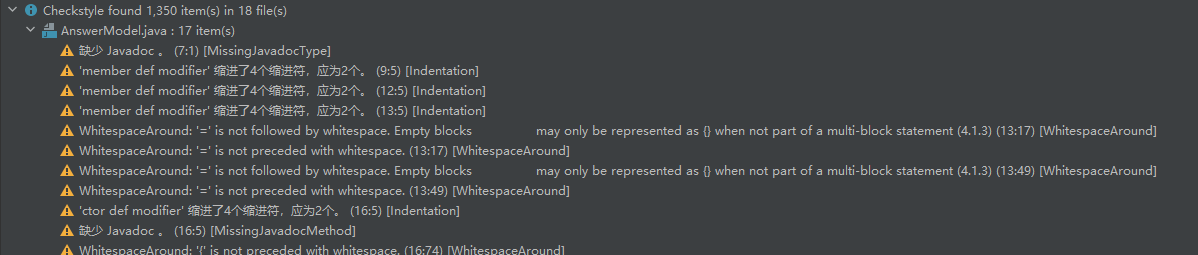
含有接受答案的比例占少部分，大约占30%。可见答案效果大多不能接受。但可能是有答案有效但问题提出者没点击接受的原因。

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。

CheckStyle





Git

