用户如何修改技术问答网站上的答案？

Stack Overflow的一个实例研究

王少伟，会员，陈泽勋（彼得），会员，IEEE和Ahmed E.Hassan，会员，IEEE

摘要-为了确保共享知识的质量，Stack Overflow鼓励用户通过基于定量测量的徽章系统修改答案（例如，在修改超过500个答案后授予徽章）。先前的研究表明，徽章可以在Stack Overflow时（例如，增加用户参与）积极引导用户行为。然而，由于一些研究表明，某些用户可能通过游戏激励系统获得奖励，但是很少有人知道与修订相关的徽章是否会对修订的质量产生负面影响。在本研究中，我们分析了从2377692Stack Overflow答案中收集的3871966个修订记录。我们发现：1）与正常天数相比，用户在授予徽章的天数上进行了比平时大得多的修改；25%的用户在收到第一个与修订相关的徽章后没有再进行任何修改。2）在一天内进行比平常更多的修订，增加了此类修订被回滚的可能性（例如，由于不希望的或不正确的修订）。3）如果用户在一天内进行多次修改，则更可能愿意做一些对文本的修改以及一些小的修改。我们的发现得到了Stack Overflow community的认可，它们强调了对当前徽章系统进行更改的必要性，以便在修订的质量和数量之间提供更好的平衡。

索引项-叠加溢出、激励系统、徽章、答案修订f

# 1        引言

技术问答（Q&A）网站改变了开发者在网络上寻找信息的方式。问答网站正在成为一个重要的和流行的知识共享和学习平台。当遇到问题时，用户通常通过在这些问答网站上发布问题（例如，Stack Overflow、Quora和msdn Forum）来寻求其他开发人员的帮助。例如，截至2018年9月，Stack Overflow是最受开发者欢迎的问答网站之一，拥有超过1600万个问题、2500万个答案、6800万条评论和5000万月访问量。  
 然而，在问答网站上提问和回答问题可能并不总是简单直接的。例如，答案可能缺乏对一些重要概念或参考文献的解释，这反过来可能使这些答案难以理解。此外，答案可能包含不正确的信息或错误的代码片段。因此，问答网站面临的一个重大挑战是确保其内容的质量。

基于以上事实，Q&A网站开发了几种机制来确保其内容的质量（例如，审查问题和修订，以及修订问题和答案）。Stack Overflow的一个主要机制是使用徽章系统来鼓励用户修改答案。用户将根据量化指标获得徽章（例如，通过修改超过500个关于Stack Overflow的答案）。此类徽章旨在鼓励问答网站上用户的积极贡献（例如，提高内容质量）。然而，这种与修订相关的徽章只考虑修订的量化措施，而不考虑其质量。

先前的研究表明，徽章系统可以积极引导问答网站上的用户行为，例如[3]，[4]观察到徽章可以提高用户参与的整体水平。另一方面，一些先前的研究表明，激励系统可能不会总是在问答网站上以积极的方式驱动某些用户[5]，[6]，例如，用户可能会为了利润而积极地与系统进行博弈。

因此，在本研究中，我们希望探讨徽章系统如何在Stack OverFlow中控制用户的修订行为。例如，当用户即将收到徽章时，他们的修订活动是否会发生变化？我们也有兴趣调查这种用户行为改变对修订质量的潜在威胁。例如，在一天内进行更多的修订是否会降低修订的质量（例如，增加此类修订被回滚的可能性）？更好地理解徽章奖励可能有助于问答网站设计师改进当前的徽章系统（例如，重新制作一些徽章或创建新的徽章以减少低质量的修订次数）。

为了了解徽章系统的有效性，以及鼓励某些修订活动是否会对修订质量产生负面影响，我们研究了3871966个答案修订，基于2008年8月至2017年3月在Stack Overflow时创建的2377692个答案。这些答案涉及280617个独特的用户。我们研究收集的数据，以了解用户如何随时间修改答案。更具体地说，我们研究了用户的修订模式（例如，用户是否会在很长一段时间内突然或不经常修改答案），尤其是当他们即将收到修订相关的徽章时。此外，检查修订的内容可以帮助我们更好地了解修订徽章的效果。因此，我们调查了修改答案的哪些部分（即文本块或代码块），以及这些修改的根本原因。我们还研究了回滚修订以及与此类回滚相关的因素。特别是，我们解决了以下研究问题：

·RQ1：当他们即将收到徽章时，徽章是否会随着时间的推移而改变用户修改的次数？

在授予徽章的日子里，用户进行的修改（统计上显著）要比正常日子多。此外，获得与修订相关的徽章（即徽章持有人）的用户比未获得与修订相关的徽章（即非徽章持有人）的用户执行峰值修订的可能性高出17倍（即，在某一天执行的修订次数比平常多）。最后，25%的用户在收到第一个与修订相关的徽章后没有再做任何修改。

·RQ2：用户如何根据内容和大小修改答案？

当用户在一天内进行许多修订时，他们更可能执行小的和文本的修订。用户更有可能对其他用户的答案执行更简单的修订（即，文本更正和代码格式）。

·RQ3：哪些因素与答案回滚相关？

在一天内进行大量修改会增加修改被回滚的可能性（即被早期的回答者拒绝）。通过一项定性研究，我们发现，由于不需要的代码/文本格式、不正确的代码修订以及其他与文本相关的修订，答案修订经常被回滚。

简而言之，我们观察到，当用户在一天内进行更多的修改时（例如，在追踪徽章时），他们通常会执行更多的文本和小的修改。但是，在一天内进行比平常更多的修订会增加由于不希望/不正确的修订而回滚修订的可能性。

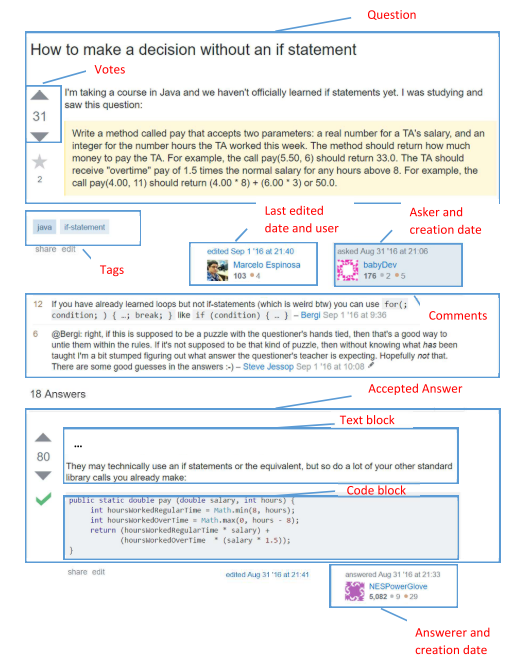
根据我们的发现，Stack Overflow网站设计者可能希望调整他们当前的徽章系统，以改进答案修订过程。例如，Stack Overflow网站设计者可能会考虑合并某些规则，以便在修订的质量和数量之间建立更好的平衡。Stack Overflow可能会考虑更改，他们的徽章只考虑展有效修订，而不是简单地计算所有修订。我们与StackOverflow社区讨论了我们的研究结果，几个社区成员一致认为我们的研究结果强调了对他们当前的一些徽章进行改进的必要性。我们的研究强调了未来研究的必要性，以设计改进修订相关徽章的方法。

本文的结构如下。第2节介绍了Stack Overflow的背景、应答修改过程和当前的标记系统。第3节描述了我们的研究问题和数据收集过程。第4节介绍了我们研究问题的结果。第5节讨论了我们从Stack Overflow元forum5收集的反馈以及我们研究的含义。第6节介绍了对我们研究有效性的威胁。第7节概述了相关工作。最后，第8节总结了我们的研究，并讨论了可能的后续研究。

# 2        背景

在本节中，我们简要概述了Stack Overflow、用户如何修改他们的答案，以及通过使用Stack Overflow中的实际示例，当前的Stack Overflow标记系统。

## 2.1       StackOverFlow的问答过程

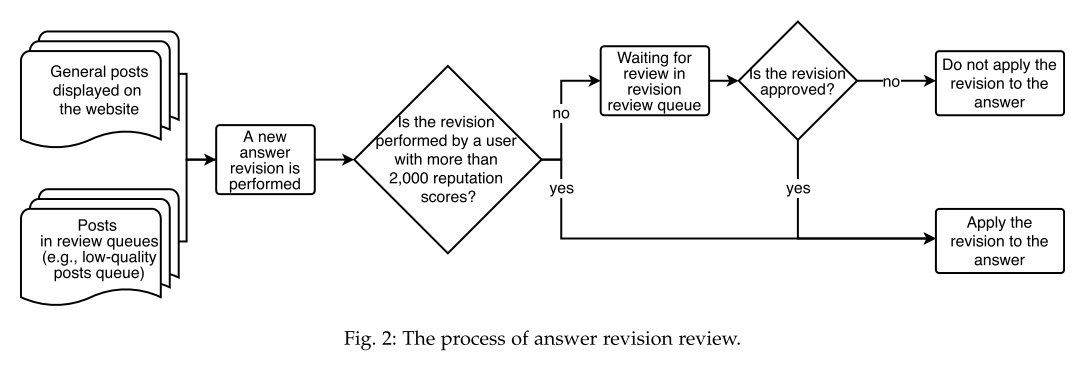
Stack Overflow允许用户注册、发布问题、回答问题、对文章发表评论（即问题或答案）、修改文章、对文章投票以及基于标签搜索或浏览文章。用户可以包含代码片段和其他引用（如URL或图像），以丰富他们的问题。请注意，我们在本文的其余部分中将代码段定义为代码块，而所有其他非代码内容（例如，文本描述、URL）定义为文本块。其他用户可以根据自己的经验回答发布的问题。每个问题可能会收到来自不同用户的多个答案。但是，最初提出问题的用户最多可以将一个答案标记为接受（以表明此特定答案是最合适/最正确的答案）。一篇文章的得分（即问题或答案）表示该文章收到的上下投票总数。图1显示了一个问题及其相应答案的示例快照。

**图1**：一个关于Stack Overflow的示例问题及其可接受的答案。该示例强调了许多可以观察到的细节，例如：用户在2016年8月31日询问“如何在没有if语句的情况下做出决定”，并收到另一用户的回答。然后提问者接受了这个答案。这个答案得到了社区80分。这个问题与标签“Java”和“if语句”相关。

## 2.2       提高  StackOverFlow答质量

Stack Overflow通过两种机制鼓励用户提高文章质量：1）鼓励用户修改文章（即修订过程）；2）鼓励用户审阅文章和修订（即审阅过程）。然而，修订和审查是两个不同的过程。例如，Stack Overflow只允许信誉分数超过2000的用户进行审阅，而Stack Overflow则鼓励每个用户执行修订。

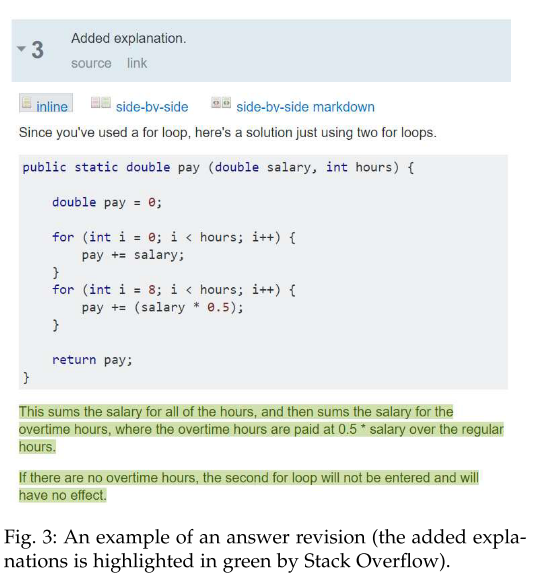
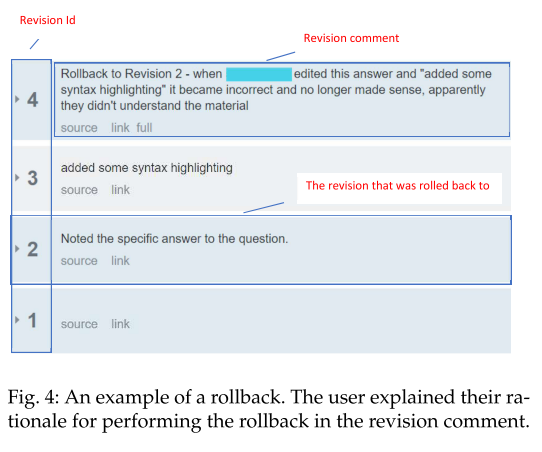
答案审查过程的概述如图2所示。



用户可以查看显示在网站或位于审阅队列中的日志（例如，用于低质量日志的队列），并对日志执行修订。一旦用户执行修订，修订将被添加到建议的修订审核队列中，等待信誉分数超过2000的用户进行审核。如果修改是由信誉分数超过2000的用户执行的，则修改将应用于答案，而无需任何审核过程。

在所有的答案修订中，我们发现只有9%通过了审查过程。换言之，91%的答案修改是由信誉评分超过2000分的用户进行的，因此，修订审查过程不能保证这些修改。因此，在本研究中，我们主要研究答案修订过程，这是解决Stack Overflow问题的主要质量保证过程。图3显示了一个用户修改答案内容以使答案更加准确和简洁的示例。

图4显示了一个回滚6的示例。执行回滚的人提到了他回滚的原因

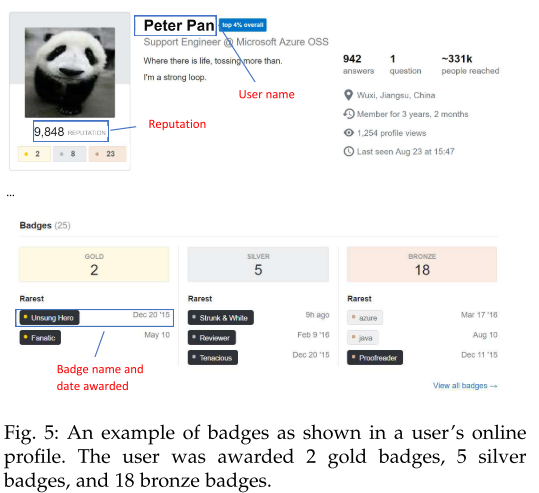
3

答案是：“当[其他用户]编辑这个答案并添加一些语法，强调它变得不正确，不再有意义，显然他们不理解材料”。通过这样一个例子，我们发现，虽然修改答案的主要目的是提高答案的质量，但有时修改可能是不可取的，甚至导致错误的答案。因此，如果不回滚这些错误的修订，那么这些修订后的答案将在Stack Overflow时误导其他用户。由于执行了更多此类不希望的/不正确的修订，其中一些修订仍有可能保留，并且不会回滚。因此，减少或避免这种情况对于Stack Overflow很重要。事实上，Stack Overflow要求对信誉分数低于2000的用户执行的修订进行审核，然后再将其应用于相应的答案。但是，我们仍然观察到Stack Overflow中有大量的回滚修订。因此，我们希望探讨回滚修订的原因。

Stack Overflow为用户提供了一个搜索答案的平台。对于这些寻求答案的人来说，有一个清晰而正确的答案是非常重要的。然而，我们发现，对答案所做的修改并不总是提高答案的可理解性，在某些情况下，修改甚至可能是不正确的（图4提供了这样一个例子）。我们还比较了对问题和答案进行修订的频率，发现对答案（每个答案修订1.6次）的修订比对问题（每个问题修订1次）的修订更频繁。因此，在本文中，我们将重点研究用户如何修改答案以及导致此类修改回滚的原因。简而言之，本文研究了徽章系统的有效性，并为改进答案修订过程提供了见解。

# 2.3使用徽章系统激励答案修订

徽章系统广泛应用于各种在线系统，如学习系统和问答网站[7]、[8]。徽章用作成就、技能、质量或兴趣的指示器。例如，Quora使用了几个徽章（例如，最受欢迎的作者）来激励用户撰写高质量的文章。堆栈交换7下的所有网站（包括技术性和非技术性问答网站）使用相同的徽章系统。在本研究中，我们主要研究堆叠溢位的徽章系统，因为它是一个被世界各地的开发者日常使用的网站。Stack Overflow使用徽章系统来激励用户参与社区。用户可以在完成特定目标（如修改答案）后收到徽章。这些徽章是用户成就、技能或兴趣的标志。徽章有三种不同的颜色：金、银和青铜，这表明了获得徽章的难度。例如，图5显示了用户配置文件中列出的徽章。



我们有兴趣研究与修改答案相关的徽章。表1列出了这些徽章的细节：斯特伦克和怀特（银）、考古学家（银）和副本编辑器（金色）。请注意，徽章“复制编辑器”不包括自我修订、删除的帖子和标签编辑，而“strunk&white”和“考古学家”都考虑自我修订。因此，我们研究RQ1中的自我修订和其他人的修订。然后我们重点研究RQ2和RQ3中其他人的修订。我们将获得修订相关徽章的用户定义为徽章持有者，而从未获得修订相关徽章的用户定义为非徽章持有者。当前的徽章系统鼓励用户根据修改的数量修改答案（例如，修改超过500个答案）。我们有兴趣研究徽章系统如何在Stack Overflow时引导应答修订过程。

表1：授予徽章持有者的徽章。

|  |  |
| --- | --- |
| 徽章 | 描述 |
| Strunk&White（银色） | 修改80个帖子 |
| 考古学家（银） | 修改100个6个月不活跃的职位 |
| 文案编辑（金） | 修改500篇文章（不包括自己或删除的文章和标签编辑） |

# 3         研究问题与数据收集

在本节中，我们将描述我们的研究问题和动机，以及我们如何收集用于回答研究问题的数据集。

## 3.1       研究问题

**RQ1：当他们即将收到徽章时，徽章是否会随着时间的推移而改变用户修改的次数？**

用户生成的内容的质量在Stack Overflow时会有很大的变化。某些内容包含有价值的信息，而其他内容可能包含不需要的垃圾邮件。Stack Overflow的一个重大挑战是确保其内容的质量[2]。因此，Stack Overflow鼓励用户通过徽章系统修改答案（参见第2.3节）。在这个RQ中，我们研究了徽章如何引导修订过程。我们研究用户的修订模式（例如，用户是否在很长一段时间内突然或不经常修改答案），尤其是当他们收到修订相关的徽章时。用户拿到这些徽章后还需要修改吗？我们也有兴趣调查具有修订相关徽章的用户与没有修订相关徽章的用户相比是否具有不同的修订模式。

**RQ2：用户如何根据内容和大小修改答案？**

在这个RQ中，我们调查用户修改答案的哪些部分（即文本块或代码块），以及修改的基本原因。例如，是否执行修订以修复代码块中的错误或优化文本块中的描述？当用户突然进行修改时，他们会修改答案的哪些部分？我们的实证调查将使我们对用户执行的修订类型（内容方面）有更深入的了解，并对修订徽章的有效性提供见解。

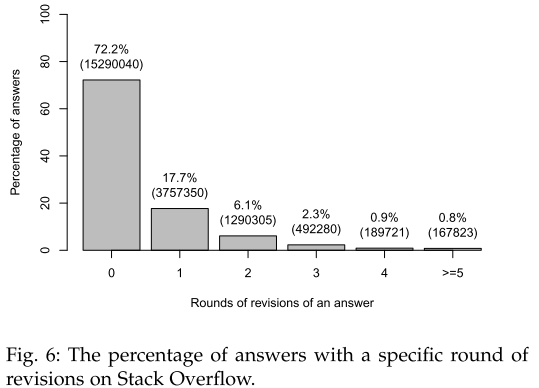
**RQ3：哪些因素与答案回滚相关？**

我们发现，对于Stack Overflow来说，大多数修订都不需要审查（见第2节）。此外，我们发现一些修订被回滚，并且这些回滚的修订可能与不正确的更改相关（参见图4中的示例）。减少或避免回滚修订非常重要。因此，在这个RQ中，我们研究与回滚相关的因素。更具体地说，我们研究回滚与用户在一天内所做的修改数量之间的关系（即，RevisionsPerDay）。我们还研究了回滚背后的原因。我们对回滚的研究可能有助于Stack Overflow设计人员和研究人员提出解决方案，以改进当前的徽章系统并可能减少回滚。

# 3.2从Stack Overflow获取应答修订数据

为了研究Stack Overflow的应答修正并回答我们前面提到的RQ，我们从Stack Exchange8提供的链接下载了Stack Overflow的数据转储，Stack Exchange8是一个关于主题的问答网站网络。

跨不同领域（如规划和教育）。数据转储包含关于文章的详细信息（即问题和答案）。数据转储存储每个日志的所有事件历史记录（例如，正文编辑、日志删除和日志关闭），包括每个事件的日期、触发事件的用户、每个事件的注释以及每个事件后更改的日志。

在我们的研究中，我们使用了2017年3月之前发布的所有数据。下载的数据共有1300万个问题和2100万个答案。图6显示了在Stack Overflow时对答案所做的修改数。从图6中，我们注意到Stack Overflow的答案中有27.8%（5897479）至少有一个修订版，0.8%（167823）至少有5个修订版。

# 3.3数据预处理

Stack Overflow9跟踪的事件有38种类型。在本研究中，我们对“编辑主体”和“回滚主体”事件感兴趣，这些事件与在答案主体上执行的修订相关。“编辑正文”事件表示答案的正文已更改。“Rollback Body”事件表示已将答案的正文还原为以前的版本。我们分别将对答案执行的“编辑主体”和“回滚主体”事件视为应答修订（简称“修订”）和应答回滚（简称“回滚”）。

我们根据以下标准选择我们研究的答案：1）至少存在一年的答案；2）分数大于0的答案。我们选择这样的标准来确保所研究的答案得到了社区足够的关注。最后我们得到了2377692个答案和3871966个相应的修订（17156个回滚正文和3854810个编辑正文）。280617个独特的用户参与了这些修订。

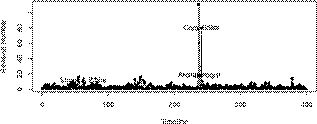


图7：用户随时间变化的修订示例。奖章颁发日在图中标出。用户在授予徽章的日子里进行了大量的修改。奖章奖励日的修订次数分别为16天、18天和80天，分别为Strunk&White、考古学家和副本编辑。峰值的阈值是14天。

# 4         案例研究结果

在本节中，我们将介绍我们的研究问题的结果。我们将从三个部分来讨论每个研究问题：使用方法、实验结果和对我们的发现的详细讨论。

RQ1：当他们即将收到徽章时，徽章是否会随着时间的推移而改变用户修改的次数？

途径：为了了解徽章如何引导用户的修订过程，我们调查了用户的修订模式，一旦他们被授予修订相关的徽章（即徽章持有人），随着时间的推移。更特别的是，当用户即将收到与修订相关的徽章以及获得徽章后，我们会查看修订的数量是如何变化的。

我们比较用户收到徽章（即徽章授予日）和正常日期（即没有徽章授予日）当天执行的修订数。注意，我们只考虑用户至少执行一次修订的天数。我们进行mann-whitney u检验和Cliff’s d检验[9]，以确定正常天数和勋章奖励天数之间的修订次数之间的差异是否具有统计意义以及差异的大小。使用Cliff[9]提供的阈值评估影响大小：d<0.147表示影响大小可以忽略，d<0.33表示影响大小较小，d<0.474表示影响大小中等，否则影响大小较大。为了研究徽章持有者的修订模式，我们使用了三西格玛规则（平均值±3标准差）[10]，该规则广泛用于异常值检测[11]，以识别所有用户的修订峰值数量。

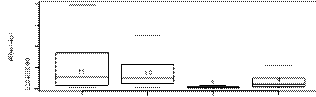
结果：与正常天数相比，徽章持有者在授予徽章的天数上进行了相当多的修改（统计显著）。图8显示了徽章持有者在正常日期和授予徽章日期进行的修订数量的框线图。我们观察到，在正常天数（即图中标记为正常，中值为1）进行的修订数量远远少于进行的修订数量。

图8：徽章持有者在获得徽章的当天（即考古学家、编辑、斯特伦克和怀特）与正常日子（即，没有获得徽章的日子）进行的修订次数。中间值在图中标出。

在授予徽章的日子（即考古学家、文案编辑和Strunk&White，中值分别为11、10和4）。这些观察到的峰值与之前记录的社会心理学现象相一致：当人们知道自己接近目标时，他们往往会加大努力[12]。Mann-Whitney U和Cliff的D结果表明，这两种类型之间的差异显著（值为0.05），且较大（d>0.474）。此外，我们发现，徽章持有者在24.7%（即3150）的徽章授予日内执行了峰值修订。作为一个例子，图7展示了一个用户随时间的变化10。我们注意到，在颁发徽章的日子里，总会有一系列的修改。

25%的徽章持有者在获得一枚徽章后没有做任何修改。此外，我们还检查了徽章持有者收到第一个徽章后的修订活动。我们发现77%的徽章持有者只有一个与修订相关的徽章。在这些只有一个徽章的徽章持有者中，33%的人在获得一个徽章后没有做任何修改。这种现象反映出，有些用户可能会受到徽章的激励来进行答案修改。一旦他们获得徽章，他们就停止修改。

*徽章持有者进行修订的可能性是非徽章持有者的17倍（也就是说，在某一天进行比平时更多的修订）。*为了进一步了解徽章持有者的修订模式，我们将修订模式与从未获得徽章的用户（即非徽章持有者）进行了比较。我们检查非徽章持有者和徽章持有者每天执行的修订数量。在考虑用户至少进行一次修订的天数时，我们观察到，平均而言，徽章持有者和非徽章持有者每天分别进行3.0次修订（差异27.0）和1.6次修订（差异1.9）。为了进一步检验27.0的高方差，我们研究了徽章持有者是否更有可能在特定日期进行大量修改（使用三西格玛规则时，阈值为14）。

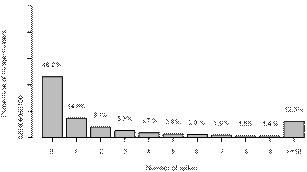
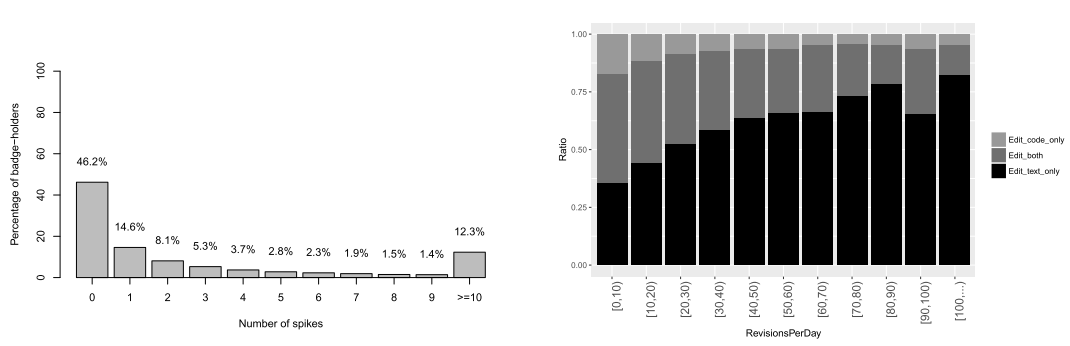
修正尖峰检测结果表明，修正尖峰在白天的2.7%（即73325）上。

图9：在所研究的Stack Overflow数据中（从2017年3月开始）进行x次修订的徽章持有者百分比。46.2%的徽章持有者从未制作过徽章。当他们进行修订时，非徽章持有者的修订峰值在他们进行修订的日期的0.16%（即6751）上。换言之，徽章持有者拥有修订钉的可能性是非徽章持有者的17倍。为了进一步了解徽章持有者是否很少出现徽章，我们绘制了一个图来显示徽章持有者相对于此类徽章持有者的百分比制造的徽章数量（见图9）。我们观察到53.8%的徽章持有者制造了修订钉，而只有1.3%的非徽章持有者制造了修订钉。12.3%的徽章持有者总共至少制造了10个徽章，这表明修订的徽章持有者并不罕见。总而言之，**徽章持有者在即将收到徽章时进行了比平常更多的修改**。一些用户在收到第一个与修订相关的徽章后也会停止修改帖子。我们的发现与之前的一项研究相呼应，该研究发现，徽章在Stack Overflow时控制用户行为（例如，增加参与度）。然而，还不清楚用户在这种高强度的修订活动中会修改什么，以及这种高强度的活动如何影响修订的质量。因此，在下一个RQ中，我们研究用户在修改答案时的变化。

RQ2：用户如何根据内容和大小修改答案？

途径：我们首先进行定量分析，检查当一天内进行的修订数量增加时，修订的大小和类型是否会改变。然后我们进行定性分析，以了解这些修订的根本原因。下面，我们将描述我们用于定量和定性分析的方法。

# 定量分析

我们特别感兴趣的是了解在一天内进行的修订数量（称为每天修订）与类型（即*只编辑代码仅编辑文本*和以及修订的大小）。我们通过汇总修订中添加和删除的字符来计算修订的大小。为了研究这种关系，我们对照每天的修订来检查三种类型的比率和修订的大小，并可视化我们的结果。请参阅附录A部分中的详细说明。

结果：当用户在一天内进行多次修订时，他们更可能进行小的或文本的修订。我们发现，当用户在一天内进行大量的修改时，文本修改是显著的。图10显示了修订类型与用户在一天内执行的修订数量之比。总的趋势是，随着在一天内进行的修订数量的增加，比率也会增加。换句话说，当用户在一天内进行多次修改时，他们修改的文本可能多于代码。

图11显示了针对num的修订的大小, 用户在一天内执行的修订的误码率。总的趋势是，随着每天修订次数的增加，修订的大小会减小，这意味着用户在一天内进行多次修订时，可能会进行小的修订。一个可能的原因是用户更喜欢执行简单的修改，这样他们就可以在一天内执行尽可能多的修改。定量分析的附加结果见附录A节。

# 定性分析

我们进行了一项定性研究，以揭示答案修改背后的原因。为了达到95%的置信水平，置信区间为5%[13]，我们从整个修订数据集（即3871966修订版）中随机抽取384个修订版，并确定了此类修订的理由。为了计算随机样本的大小，我们使用以下公式，其中*n*是人口规模（如3871966），是对应于特定置信水平的z分数（如95%置信水平为1.96），是置信区间（如5%），是人口比例（如0.5）。我们首先检查了随机抽样的修订，没有考虑到任何特殊类型的原因。然后，我们定性地分析了样本数据，并得出了这些修订的一组原因（例如编码规则）。然后，前两位作者手动检查了样本修订，并根据导出的原因对样本修订进行分类。我们用科恩的kappa[14]来衡量评级机构间的协议。我们的kappa值是0.89，这意味着高度一致。在达成一致意见之前，对任何差异进行了讨论。*Ze磷*

在定性分析中，作者还需要阅读答案下的评论，这有助于作者更准确地确定原因。我们对每个原因的修订数量进行了一些基本统计，并将结果可视化。

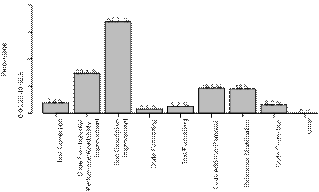


图12：基于随机抽样修订的修订背后的原因分布。66.7%的修改是为了改进答案的文本描述。

结果：用户更容易对其他用户的答案进行文本更正和代码格式设置；用户很少帮助其他用户进行代码更正。表2显示了我们得出的八种原因

8

表2：人工推导的原因概述

（编码规则）。

|  |  |
| --- | --- |
| 修订原因 | 定义 |
| 文本校正 | 修复文本块中的错误。例如解决语法/打字错误问题。 |
| 文字说明  改进 | 改进文本描述，使描述更清晰、简洁。 |
| 文本格式化 | 设置文本格式，以便更好地进行演示。例如更改字体、突出显示文本以及删除/添加空间。 |
| 代码校正 | 修复代码块中的错误。例如修复错误和/或修复打字错误。 |
| 代码功能/性能/可读性改进 | 在功能、性能和可读性方面改进代码。例如更改函数、更改逻辑、添加注释、更改类型和更改变量名。 |
| 代码格式化 | 在格式方面改进代码。例如添加/删除空间和添加换行符。 |
| 代码添加-  去除/去除 | 添加/删除代码段。例如，添加/删除整个代码块或从代码块添加/删除代码行。 |
| 参考修改 | 添加/更新/删除文本块中的引用，如URL、超链接和图像。 |

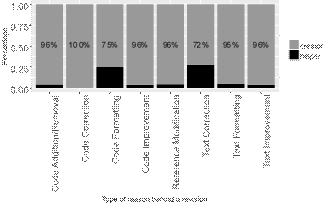
基于采样数据。我们使用这些类型标记每个样本修订，并在图12中可视化分布。请注意，一个修订可能有多个类型，因为修订可能会更改多个位置，并且这些位置可能会因不同的原因而更改。特定类型的百分比计算为我们观察到，最常见的修订类型是文本改进（67.7%），这表明大多数修订都是为了使答案更简洁和清晰。

图13显示了答案创建者（创建者）对其他用户（帮助者）对每种原因类型所做修订的百分比。我们发现帮助者更有可能帮助进行文本更正和代码格式化。另一方面，几乎所有的代码修正都是由应答创建者自己进行的。对于用户（即非创建者）来说，帮助他人改进代码也是很少见的。

我们上述观察的一个可能原因是，纠正代码需要对问题/答案有深入的了解。因此，帮助者可能更难进行代码更正修订。在协作代码审查任务中也观察到类似的现象，审查人员倾向于提供肤浅的反馈[15]。基于这一观察，我们可能建议Stack Overflow考虑鼓励用户执行与代码相关的修改，因为以前的研究已经表明代码片段是高质量答案的一个重要方面[16]、[17]、[18]。

RQ3：哪些因素与答案回滚相关？

途径：我们首先进行定量分析，以了解每天修订的次数与回滚之间的关系。然后，我们进行定性分析，以了解此类回滚的根本原因。

图13：答案创建者（创建者）v.s.对其他用户（帮助者）在基于我们随机抽样和手动标记的修订的每种原因后所做的修订的百分比。

下面，我们将描述我们用于定量和定性分析的方法。

我们使用以下两个步骤构建了我们研究的数据集：

1）我们将回滚修订和回滚修订之间发生的所有修订视为导致回滚的修订（称为未接受修订）。例如，在图4中，回滚将答案从版本4恢复到版本2，然后版本3（介于版本2和版本4之间）被视为未接受的版本。识别回滚的修订。

2）回滚有两种情况：1）用户对自己的修订不满意，因此会将答案回滚到以前的版本；2）用户对其他用户的修订不满意，而答案创建者会将答案回滚到以前的版本。我们将研究重点放在第二个场景上。删除自还原的修订。

经过我们的数据预处理步骤，我们最终得到了9087个未接受的修订及其相应的回滚修订。

与我们在RQ2中使用的方法类似，我们首先进行了定量分析，以了解回滚与在一天内执行的修订数量（即，每天修订）之间的关系。然后，我们进行了定性分析，以了解回滚的根本原因。我们在下面描述我们用于定量和定性分析的方法，以及我们的结果。

# 定量分析

为了理解回滚和RevisionsPerDay之间的关系，我们比较了用户执行未接受（即回滚）修订（称为）的天数和同一用户执行未接受修订（称为）的天数的RevisionsPerDay。*每日修订未接受*

9

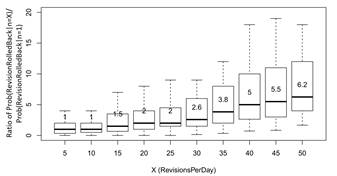


图14：每天修订次数大于（（RevisionRolledBack n>x）和每天修订次数为1（（RevisionRolledBack n=1））的修订次数之比，为0.2%。随着在一天内进行更多修订，回滚修订的可能性增加。1628152、528607、56703、29885、19332、14124、11357、9511、8106和6803修订日期分别为5、10、15、20、25、30、35、40、45和50。如果显著大于，则可能表明每天进行更多的修订更可能导致回滚。

我们计算了一个修订被回滚的概率，给出了每天不同的修订值（基于所有用户）。当用户在特定日期（即）执行修订时，修订被回滚的概率（即（revision rolled back n=n））可以通过以下方式使用引导程序[19]进行估计。对于每一次迭代：1）我们随机抽样（即本研究中=1000）对在上执行的所有修订进行修订（包括替换）；2）我们计算在步骤1中抽样的这些修订中存在不可接受修订的概率（即（n））。我们重复迭代100次。我们使用这些值prob1（n），…，（n），…，（n）来估计（revisionrolledback n=n）。为了进一步了解每天修订次数与回滚修订次数之间的关系，我们计算了（RevisionRolledback n>x）和（RevisionRolledback n=1）的比率，其中是每天修订次数的不同值（即5、10、15、20、25、30、35、40、45和50）。我们用数字表示比率。我们还进行了mann-whitney u检验，以测量（revisionrolledback n=1）和（revisionrolledback n>x）之间的差异是否具有统计学意义。我们做了一个悬崖D测试来测量差异的大小。*丙纶n每天修订我SS每天修订普罗里S普罗里丙纶*100*丙纶丙纶丙纶X丙纶丙纶*

结果：在一天内进行更多修订会增加修订被回滚的可能性（例如，由于修订质量较低），尤其是当修订数量较大（例如，大于50）时。（即3.6）的平均值几乎是（即2.5）的1.4倍。我们进行了曼-惠特尼U检验并计算了克里夫的D。我们观察到*每日修订未接受每天接受修订*

*修正日认可的*并且具有统计显著性（−值0.5），影响大小较大（克里夫D 0.474），这意味着用户执行的修订次数比平时多，通常会执行不可接受的修订。换句话说，在一天内进行更多的修改会增加回滚的可能性。*修正日未被接受的磷<>*

图14进一步显示了当每天修订次数大于（（RevisionRolledBack n>x），其中=5,10,15,20,25,30,35,40,45,50）时，以及当每天修订次数为1（（RevisionRolledBack n=1））时，修订被回滚的概率比率。我们观察到，随着每天修订次数的增加，修订被回滚的可能性增加。如果每天的修订次数大于50次，则回滚的可能性（即1.2%）是每天修订次数为1次（即0.2%）的6倍。此外，Mann-Whitney U检验和Cliff&apos;s D的结果表明（RevisionRolledback n>x）在统计学上高于（RevisionRolledback n=1）（−值0.5），且当大于10时，效果尺寸较大（Cliff&apos;s D 0.474）。*X丙纶X丙纶丙纶丙纶磷<>X*

一个可能的原因是，当一个用户在一天内执行的修订量比平时大得多时，用户可能难以确保每个修订的质量（例如，由于时间有限），从而增加了最终回滚的低质量修订的可能性。

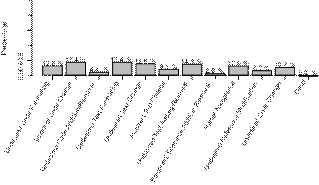


图15：发生定性螺旋状回滚的类型概述。例如，在所研究的未被接受的修订中，17.4%的修订因错误的代码更改而被拒绝。

# 定性分析

为了进一步了解回滚背后的原因，我们调查了单日执行的数字修订与回滚之间的关系。我们采用95%置信水平和5%置信区间，从9087次回退中随机抽取369例进行了定性分析。我们执行了一个轻量级的开放编码过程[20]，[21]来确定回滚背后的原因（有关详细信息，请参见第B.1节）。

10

表3：回滚修订的手动派生原因概述。

|  |  |
| --- | --- |
| 回滚原因 | 定义 |
| 不需要的代码  格式化 | 用户对代码格式进行了不希望的更改，例如添加/删除空格和添加换行符。 |
| 错误的代码  变化 | 用户更改了错误的代码。 |
| 不需要的代码  添加/删除 | 用户添加/删除了不需要的代码，例如添加替代解决方案和删除代码片段。 |
| 不需要的代码  变化 | 用户进行了不需要的代码更改，例如更改命令选项、更改API、重构和编辑注释。 |
| 不需要的文本格式 | 用户更改了不需要的文本格式，例如更改字体、突出显示文本以及删除/添加空间。 |
| 错误的文本  变化 | 用户更改了错误的文本，例如改变了句子的含义。 |
| 不需要的文本添加/删除 | 用户添加/删除了不需要的文本，例如添加其他解决方案和添加工具广告。 |
| 不需要的文本  变化 | 用户做出了不需要的文本更改，如更改段落结构和改写。 |
| 部分承兑 | 修订被回滚，但部分更改仍被接受，并包含在以后的修订中。 |
| 情感句子的添加/删除 | 用户添加/删除了表达个人情感的句子。 |
| 不需要的参考修改 | 用户修改/添加/删除了答案中的引用（例如链接和图像）。 |
| 其他 | 其他修改，如提问者在回答中提出问题。 |

结果：由于不需要的代码/文本格式、不正确的代码修订以及其他与文本相关的修订，应答修订经常被回滚。图15概述了用户回滚应答修订的原因。我们发现，30.2%的答案因为不需要的格式（即12.8%不需要的代码格式和17.4%不需要的文本格式）而被回滚。基于我们的定性分析，用户回滚格式化修订的一个可能原因是，用户不喜欢修订者更改格式，因为这样的修订可能使答案看起来难看（从应答创建者的角度），甚至可能改变应答创建者希望表达的意图（例如，强调或者不强调部分答案）。

图16给出了一个不需要的文本格式示例11。用户在修订注释中提到了他不喜欢这些编辑的原因。用户认为“gvim”看起来很难看，不喜欢“internet”的格式改变。用户也不喜欢命令模式“g”和“t”的格式。

我们观察到，大多数修订（超过38.5%）由于与文本相关的修订（即不希望的文本更改、不正确的文本更改和不希望的文本添加/删除）而回滚。帮助者可能会误解答案，从而做出错误的修改。此外，有些修订可能会使答案偏离原来的目的。但是，当手动检查采样

11。http://stackoverflow.com/posts/24156/revisions

0098-5589（c）2018年IEEE。允许个人使用，但重新发布/重新发布需要获得IEEE许可。有关详细信息，请参阅http://www.ieee.org/publications\_standards/publications/rights/index.html。

回滚，由于修复语法/拼写错误问题，我们没有看到任何回滚。我们还发现，由于错误的代码更改，许多修订（17.4%）被回滚。例如，一个答案创建者回滚了以前的修订版，并提到“我正在还原到我原来的问题文本。最新的编辑实际上使我的代码不正确，如“12”。*<？php表单：：…&apos;不向浏览器回显任何内容*

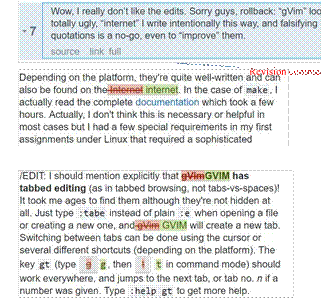


图16：不需要的文本格式示例。

其他结果见附录B部分。

# 5         讨论

在本节中，我们将讨论从Stack Overflow元forum13收集的反馈。我们还讨论了我们的发现的含义。

## 5.1       来自Stack Overflow论坛的反馈

为了了解我们的研究是否发现了Stack Overflow的相关问题，并寻找可能的解决或缓解问题的建议，我们将研究结果发布在Stack Overflow元论坛上，该论坛是Stack Overflow社区成员（即Stack Overflow的开发人员和用户）讨论Inn的Stack Overflow论坛。Stack Overflow的ER工作和策略。Stack Overflow鼓励社区成员在其元论坛上留下反馈，以便Stack Overflow可以根据反馈（例如，功能请求）改进其网站。

截至发稿时，我们的第14位被评为所有徽章相关问题中排名前7%的问题。我们的帖子得到45张赞成票，3张最喜欢的票，666张意见。Stack Overflow元论坛上的中位数upvotes、最喜爱的投票数和视图分别为4、1、146。

12.            https://stackoverflow.com/posts/14756994/revisions

13.            https://stackoverflow.com/help/whats-meta

14.            https://meta.stackoverflow.com/questions/350326/shouldthe-post-revision-related-badges-e-g-copy-editor-and-strunk-whiteb/350426 350426

11

我们的帖子收到了来自Stack Overflow社区的有用反馈。社区成员认为我们的研究触及了Stack Overflow的真正挑战。一位社区成员留言说：“所以对你所做的几乎所有事情都有利率限制，特别是为了避免这样的AB/使用。但由于一些相当神秘的原因，编辑上没有。这导致了很多其他问题，从DOS攻击用户感兴趣的页面视图，到当前编辑审查队列的问题”。该海报认为，任何对修订版的删减甚至都不会引发一些安全问题。另一位社区成员认为我们的发现很有趣（即，“啊，这已经有很多好的研究了。”）他还问这是否是其他类型的徽章（如任务回顾徽章）的问题。

社区成员提供了许多有用的建议，这些建议可分为两大类。首先，社区成员提议对修订使用评级（投票）机制，并将修订的评级视为徽章的一部分。例如，一位社区成员提到：“我一直希望能够对好的编辑进行投票。也许可以使用这些行中的某些内容来帮助确定高质量的编辑。再加上编辑次数，它将类似于标签徽章（例如20次编辑，总分为20分）”。另一位社区成员提到，“就数量与质量而言，在我看来，量化质量的唯一简单方法就是为编辑自己设置投票机制。然后只允许2000个代表用户投票。当然，这样的事情会很复杂，而且可能会产生意想不到的后果”。*>*

第二，社区成员建议在授予徽章时不要考虑回滚修订。例如，“考虑到关于回滚的观点，阻止“数量超过质量”行为的一种方法是，在回滚编辑的同一天提交的任何编辑不计入徽章（或者，在提交稍后回滚的编辑后24小时内提交的编辑不计入，因为这可以是C。”使用编辑历史记录的一次传递来完成。

另一个有趣的建议是添加一个RollbackMessage特性。一位社区成员提到，如果用户的修改被回滚，Stack Overflow不会通知用户，也就是说，“你的文章关注的是回滚，但是作为一个修正主义者（编辑了12433篇文章），我无法知道我的哪些编辑被回滚，所以我不能在这方面提高自己。”我认为，我们应该首先改进编辑回滚时的通信。也可能是为什么编辑最终添加了回滚消息而被回滚的原因。

特点。

简而言之，我们的发现揭示了Stack Overflow机制中的挑战（例如，在授予徽章时，每天进行的修订数量没有限制，修订没有质量控制），并吸引了Stack Overflow社区的注意。未来的研究应该探索改善修订相关徽章的方法。

## 5.2       我们发现的含义和重点

我们观察到，有些用户在获得徽章之前往往会有一个峰值的修改。这一观察结果与Ashton等人的发现一致。〔3〕，其中他们发现，在用户获得徽章之前，用户的活动显著增加。然而，目前与修订相关的徽章只考虑修订的数量，没有考虑修订的质量。Stack Overflow网站设计者可以考虑改进他们的徽章系统，以确保修订质量。下面，我们将详细介绍我们的发现以及对Stack Overflow网站设计者的相应影响。

某些规则可能会阻止用户玩徽章系统游戏，并在用户执行的修订的质量和数量之间提供更好的平衡。我们发现，每天进行更多的修改会增加回滚的可能性，尤其是在进行比平时更多的（例如超过50次）修改时。为了减少回滚的机会，我们建议Stack Overflow考虑开发某些规则，以防止用户在一天内进行比平常更多的修改（例如，通过每天执行100次以上的修改来模拟徽章系统）。例如，Stack Overflow网站设计者可以设置最大数量的每日修订，这些修订可以计算为接收徽章。或者，Stack Overflow可能选择不限制修订，而只计算未展开的修订。另一个建议是采用评级机制来提高修订的质量。请注意，任何规则都可能产生不良的副作用。例如，惩罚回滚或投票修改的质量可能会影响用户参与修改活动。因此，未来的研究应该研究如何确保修订的质量和数量之间的良好平衡，同时最小化对用户参与的影响。

Stack Overflow设计器应考虑通过设计新的徽章或使当前的徽章支持代码修订来鼓励用户执行与代码相关的修订。我们观察到，当用户在一天内进行相对更多的修改时，他们倾向于执行文本修订而不是代码修订。此外，用户更有可能对其他用户的答案执行文本更正和代码格式设置；用户很少帮助其他人更正或改进答案中的代码片段。然而，以前的研究表明，代码片段是高质量答案的一个重要方面[16]、[17]、[18]。因此，我们建议Stack Overflow网站设计者可以考虑鼓励用户执行与代码相关的修改。例如，Stack Overflow网站设计者可以设计与代码修订相关的新徽章，或修改当前徽章以支持代码修订（例如，使正确的代码修订等同于几个简单的文本修订或普通的代码格式修订）。

# 6         有效性威胁

外部有效性。对外部有效性的威胁与我们发现的普遍性有关。在本研究中，我们将重点放在Stack Overflow上，这是开发者最受欢迎的问答网站之一，因此，我们的结果可能不会推广到其他问答网站（例如，Stack Exchange下的非技术问答网站，不关注软件开发），如代码项目15和照片16。到

为了缓解这种威胁，未来可以研究更多的问答网站。

我们在RQ中进行了几个定性分析；但是，无法手动研究所有修订。为了在进行定性分析时尽量减少偏差，我们对所有相关修订进行了具有统计代表性的样本，其95%置信水平和5%置信区间[13]与之前的研究[22]、[23]中所做的相同（我们最终研究了RQ2中的384个随机抽样修订和369个随机抽样未接受的修订）。RQ3）。

内部有效性。对内部有效性的威胁与实验者的偏见和错误有关。我们的研究涉及RQ2和RQ3修订的定性分析。为了减少偏差，每一次修订都由两位作者标记，并讨论差异，直到达成共识。我们还发现，定性研究的利率间一致性水平较高。

在本研究中，我们使用现有的HTML标记“and”来检测答案中的源代码，Stack Overflow建议用户对其文章中的任何代码进行格式化。不保证所有用户都将使用推荐的HTML标记格式化所有代码。这可能会对我们研究的有效性造成威胁。*<pre>><代码>…< /代码>*

构造有效性。一个构造威胁是很难找到能够直接显示修订质量的数据。因此，我们使用回滚，我们认为这是捕获修订质量的一个合理和基本的度量。

# 7         相关工作

在本节中，我们将讨论与我们的研究相关的工作。我们关注三个相关主题：在问答网站上的协作、理解和提高问题质量、理解激励系统。

## 7.1       协同编辑

协作编辑已经在许多领域得到应用，如在线知识数据库编辑[24]、[25]、科学协作[31]、[32]和软件开发[33]、[34]。朱等。研究了关于Stack Overflow的文章（即答案和问题）的协作编辑，并探讨了其对内容质量的好处以及对用户活动的潜在负面影响[24]。他们发现，合作编辑可以提高正面选票的数量，这意味着提高了帖子的质量。与他们的研究结果不同的是，协作编辑提高了Stack Overflow时的文章质量，我们的研究结果表明，在一天内进行比平常更多的修改可能会降低答案的质量（即回滚）。Munteanu等人介绍了一种网络广播扩展的设计，该扩展使用户能够以类似wiki的方式协作编辑由自动语音识别技术生成的文本[35]。Munteanu等人表明这是一种提高转录质量的可行方案[35]。KITTUR等。研究了维基百科上编辑的数量以及他们使用的协调方法如何影响维基百科文章的质量[25]。他们发现，增加更多的编辑与

这篇文章已被接受发表在本杂志的未来一期，但尚未完全编辑。内容可能在最终出版前发生变化。引文信息：doi 10.1109/tse.2018.2874470，IEEE软件工程学报

13 表4：我们的发现和先前研究的发现之间的比较。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 我们的研究 | 前期研究 | 比较 |
| 协同编辑 | 协作编辑并不总是能得到更高质量的答案。在一天内进行比平常更多修改的修订者可能会对答案的质量产生负面影响（即，导致回滚）。 | 朱等。发现协同编辑可以提高栈溢出时的文章质量[24]。 | 我们的研究集中于特定的案例（例如，尖峰修订），同时他们测量了协作编辑的整体影响。 |
| Kittur等人研究发现，增加编辑数量与文章质量的提高没有关系，当文章均匀分布在编辑之间，或者当他们在文章讨论页面上使用明确的交流来协调维基百科时[25]。 | 我们的研究在某种程度上与他们的发现相呼应：不恰当的合作编辑可能不会提高答案的质量。 |
| 理解和提高问答质量  网站 | 在一天内进行比平时更多的修订，增加了修订被回滚的可能性。由于不需要的代码/文本格式、错误的代码修订以及其他与文本相关的修订，应答修订通常会回滚。 | 先前的研究基于文章本身的文本、代码和用户信息，对问答网站上的文章质量进行了检查[2]、[17]、[26]、[27]、[28]。 | 以前的研究从未使用修订信息来检查问答网站上的帖子质量。我们的研究结果表明，修订信息（如修订峰值）也可用于此类研究。 |
| 胡适等人。利用文章编辑历史（例如，评论行为、作者权限）中的文章与其贡献者之间的交互数据来衡量维基百科文章的质量[29]。 |
| 了解激励制度 | 与非徽章持有者相比，徽章持有者极有可能执行峰值修订，尤其是在授予用户徽章的当天。 | 此前的研究证实了激励系统的价值及其对问答网站用户参与的有效性[3]、[4]、[30]。 | 我们的发现与先前的研究相吻合。 |

文章质量的提高，特别是当文章在编辑之间均匀分布，或者他们在文章讨论页面上使用明确的交流来协调时。我们的研究在某种程度上与他们的观察结果相呼应：协作编辑（例如，修订）并不总能提高答案的质量。卡尔沃等。提出了一种支持学术写作协同编辑的体系结构[32]。他们分析了写作活动对成果质量的影响。

这些先前的研究主要集中在研究协作编辑对用户生成内容（如回复帖子）质量的影响，发现协作编辑总体上可以提高文章质量。我们的研究集中在用户如何修改Stack Overflow的答案上，发现在某些情况下（即合作者在一天内进行的修改比平常多），合作者所做的编辑可能会降低文章的质量。此外，我们通过回顾而不是积极的投票来衡量修订的质量，这些投票可能会受到许多混淆因素（例如答案的有用性）的影响。表4强调了我们的研究结果与先前研究结果的比较。

# 7.2了解和提高问答网站的质量

在众包网站（如问答网站）上，用户生成内容的质量差异很大。某些内容可能包含有价值的信息，而其他内容可能包含不需要的垃圾邮件。问答网站面临的一个重大挑战是确保其内容的质量[2]。因此，为了了解和提高问答网站上的帖子质量，人们进行了大量的研究。Asaduzzaman等人对Stack Overflow的未回答问题进行了研究，发现有些问题由于问题太短、不清楚、太难或不相关（与Stack Overflow社区无关）而没有得到任何答案[26]。Rahman和Roy研究了与未解决问题相关的五个方面（即回答拒绝率、最后访问延迟、主题熵、声誉和投票），并基于这五个方面建立了模型来理解未解决问题[36]。哈德森等。调查了问答网站上的帮助者要求澄清的问题的特征（例如，信息缺失、目标不明确、术语不标准）。哈德森等。在不同的网站上观察到这些特征的巨大差异[37]。Ponzanelli等人进行了一项实证研究，以了解一组建议的因素与后栈溢出质量之间的关系[17]。Ponzanelli等人还构建了分类模型，以便在问题发布后立即识别高质量和低质量问题[38]。姚明等。发现答案的质量与其问题的质量高度相关[39]。因此，基于这一观察，他们提出了预测问题和答案质量的算法。同样，Yao等人还提出了一系列算法，根据答案与其问题之间的相关性来识别问答网站上的高质量帖子[40]。哈珀等人。在雅虎上调查了答案质量的预测因素。答案〔41〕。他们发现，为一个答案付钱会带来更好的结果。李等。调查了影响研究答案质量的因素，发现响应者的权威性、响应时间短和回答长度长与答案质量呈正相关[42]。Duijn等人研究了各种文本相关和代码相关因素对问题质量对Stack Overflow的影响，发现问题的代码文本比是最重要的因素[27]。Calefato等人调查影响问题成功的因素（即收到接受的答案），并进行调查，以收集在Stack Overflow上编写问题的指导性建议[28]。他们提供

0098-5589（c）2018年IEEE。允许个人使用，但重新发布/重新发布需要获得IEEE许可。有关详细信息，请参阅http://www.ieee.org/publications\_standards/publications/rights/index.html。

这篇文章已被接受发表在本杂志的未来一期，但尚未完全编辑。内容可能在最终出版前发生变化。引文信息：doi 10.1109/tse.2018.2874470，IEEE

软件工程学报

在Stack Overflow时编写问题的指导原则，例如使用中性情感风格编写问题，并提供示例代码和数据。

先前的研究主要是探讨影响帖子质量的因素，并提出评估问答网站帖子质量的方法。然而，以前的研究在研究职位质量时没有考虑到答案修订所产生的信息。我们的研究结果可能表明，可以利用修订信息（例如，一个帖子是否有一个用户在他/她进行修订的当天所做的修订）来衡量问答网站上的帖子质量。类似于Hu等人之前的研究。[29]提出了三种质量度量模型，利用文章编辑历史（例如，评论行为、作者权威）中的文章与其贡献者之间的交互数据来度量维基百科上文章的质量。

# 7.3了解激励制度

许多研究集中在研究问答网站的激励机制。Cavusoglu等人对栈溢出的激励机制进行了实证研究，为确定激励机制的价值及其提高用户参与度的有效性提供了依据[30]。Anderson等人研究了标签系统在Stack Overflow时如何控制用户行为[3]。他们观察到，徽章可以提高用户在网站上的整体参与程度，指导程度取决于用户离徽章边界的距离。格兰特等人。还观察到，徽章可用于影响用户行为，方法是在授予之前立即证明与徽章相关的用户活动增加[4]。我们在研究中获得了类似的观察结果（例如，在RQ1中，我们观察到用户在获得徽章之前更可能执行峰值修订）。安亭和丘吉尔从心理学的角度分析了社会媒体中的徽章制度，提出了徽章制度的五种社会心理学功能：目标设定、指导、声誉、地位/肯定和群体认同[43]。Hsieh等人。调查了经济激励对问答网站的影响，他们发现支付更多可能会导致

答案的数量越多，但可能不会得到更高质量的答案[5]。简等。研究了经济激励如何影响社会问答服务的不同参与者[6]。他们发现，金融激励能更快地吸引专家的回答，但这种激励也会促使某些用户积极地与系统博弈以获取利润。王等。探讨了如何改进现有的激励系统，以激励快速回答问题[44]。他们建议，问答网站应该改进他们的激励系统，以激励不需要回答的人更加积极。

我们的研究不同于上述的研究，这些研究主要集中在对激励制度的理解上。我们的研究重点是调查用户如何修改Stack Overflow的答案，以及徽章如何影响修订活动。根据我们的发现，我们还提供了一些关于栈溢出设计器如何改进标记系统的建议。

14

# 8         结论

Stack Overflow采用了一种激励系统，激励用户（通过向用户授予徽章）不断提高和保持答案的质量。这种应答修改活动在Stack Overflow时非常常见。我们发现超过25%的答案在最初发布后被修改，这意味着答案修改是Stack Overflow的主要活动。

在本研究中，我们分析了从2377692个答案中收集的3871966个修订版，以了解用户如何修订答案以及这些修订版的影响。我们发现，与正常工作日相比，徽章持有者对徽章发放日的修订（统计意义重大）。我们还发现，在这种峰值期间执行的修订更有可能回滚。此外，用户更可能执行较小的和非代码的修订，特别是当他们在一天内执行许多修订时。此外，我们还与Stack Overflow社区分享了我们的观察结果，后者同意我们的观察结果，并引导讨论提出了改进当前徽章系统的几种方法。

简而言之，当前的Stack Overflow徽章系统的设计是为了确保修订的数量（即，徽章是根据修订的数量等定量措施授予的），但是这样的徽章系统没有考虑修订的质量。因此，Stack Overflow设计人员可以考虑改进他们的标记系统，以便在修订的质量和数量之间提供更好的平衡。Stack Overflow设计器还可以考虑鼓励用户修改代码，方法是设计与代码修订相关的新徽章，或者更改当前徽章，使其支持代码修订。