Predicting Issue Types on GitHub An Empirical Study on Using Multi-Labels for Issues in GitHub

在GitHub上使用多标签的实证研究

**摘要Abstract**：

在Github上，标签用来为提出的Isuse的表示信息和决策服务。之前对标签的研究仅限于简单的数据统计或者对issue的类型的标签建议。

关注标签在软件开发中的应用，重点关注在issue上使用多个和自定义标签。

作者收集了github上的project的issue和label的数据，然后对具有多标签特征的项目的性能进行了定量研究，定性研究了多标签的类别，以及这些标签的使用情况。

该研究得出的结论，可用于问题管理研究，从而有助于开发标签技术，以减轻问题管理的负担。

**介绍 Introduction**：

前人对标签的系统的研究不够深入。第一，没有考虑一个issue上使用多个标签的问题，而标记给issue的不同标签从不同的方面表明了issue的特征或特性；第二，以前的研究没有考虑自定义标签，自定义标签是对应project管理者对管理issue而表达的信息，也是影响issue management标签分析的障碍之一。比如github提供了默认的标签bug，而很多project设计了自定义的标签“type: bug”。

于是，作者根据上述提到的对标签研究的局限性，克服了这些局限性并研究了其他的未知问题。

首先分析了有多少问题有多个标签，以及多个标签如何影响Issue管理。

分析结果显示，90%的项目在其issue上至少使用过一次多个标签。此外，具有多标签问题的项目比没有多标签问题的其他项目（81%）有更高的完成率（68%∼74%).

就关闭时间而言，具有多标签问题的项目比没有多标签问题的其他项目具有更长的平均关闭时间，因为多标签问题通常更难解决。

但是，如果仅比较单标签问题的平均关闭时间，则具有多标签问题的项目的关闭时间（131.9天）比具有单标签问题但没有多标签问题的项目（173.6天）短一个多月。这些结果表明，使用多标签的项目更有效地解决了它们的问题。

接着，研究了project中的自定义标签以及自定义标签的使用情况。定性分析了三个选定项目(vscode, dartlang/sdk, typescript)中的自定义标签。分析结果表明，问题最多有7个∼9个标签，经常使用2个∼4个标签。自定义标签通常在feature、development、issue concept中定义。与issue状态相关的自定义标签被广泛使用，这说明状态相关的标签对推动issue的生命周期运转至关重要。

贡献：

进行了一个实证研究，分析了多标签在Issue管理上的有效性：比较了在no-label、single-label或multi-label时的Issue management效果。结果表明，使用多标签的项目更有效地解决了它们的问题。

分析标签是如何自定义的：与之前对GitHub标签的其他研究不同，研究了标签是如何自定义的。

**研究问题 Research Question：**

Project > Repository, 因为一个project下面可以有多个repository

RQ1 多标签和issue management效果

RQ1.1 多标签在Issue Management中的使用。研究project和issue使用了标签的情况，然后分析了project和issue在标签使用方面对Issue Management的影响。

RQ1.2 使用多标签是否会影响project的issue management 效果？作者分析了是否使用多标签会影响project的Issue management 性能。

RQ1.3 使用多标签是否会提高单个Issue的issue management效果？在软件项目中，可能会有不同数量标签的不同issue。作者将与no-label或single-label的issue相比，multi-label的issue在issue management效果方面是否存在差异。

RQ2.project中多标签问题的研究

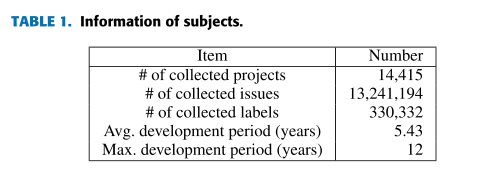
RQ2.1 project中issue的标签分布？研究了有多少标签用于标记issue，特别关注一个issue上有多少标签。

RQ2.2项目中**定义**了哪些类型的自定义标签？分析了在project中自定义标签的种类。

RQ2.3项目中**使用**了哪些类型的自定义标签？即使在project中定义了自定义标签，有很多标签也可能不会在项目中使用。分析了project中使用的自定义标签的种类。

RQ2.4哪些自定义标签一起使用？分析了哪些标签经常在一起使用，以表示有关问题的各种信息。

**数据的信息**：



**分析方法论**：

RQ1：

对于RQ1，

将Issue分为三类：no-label group、single-label group、multi-label group

对于Project：no-label group、single group、multi group

对于RQ1.1 统计了每组的Issue数量和Project数量；

对于RQ1.2

计算了Issue关闭时间 Issue close time 和 Issue 关闭率 Issue close rate

Issue关闭时间是Issue的creation时间和closed时间的之间的持续时间。

对于RQ1.3，计算了每个Issue group的关闭率和关闭时间。

RQ2:

RQ2 调查自定义标签的作用和影响，首先选择了Issue数目最多的前100个project，然后手动选择了10个项目，根据以下条件：

1. 是否是一个积极开发的project吗？
2. 其contributor超过10个吗？
3. 在调查时，是否积极使用issue tracking System？
4. 是否积极使用多标签？

在选择了10个项目后，重新检查了这些项目，并选择了三个具有相对多标签的目标项目。

对RQ2.1 对三个project的issue的标签数目的分布进行了统计。

为了研究RQ2.2、2.3、2.4对自定义标签的研究，使用了定性分析方法Grounded Theory，分两步手动标记标签。首先用标签中常用的一个词来标记每个标签。如果标签不包含任何常用词，将指定一个表示标签的词。如果发现很难理解标签的含义，会参考标签的描述并试图理解其含义。如果描述不足以理解标签，会检查存在该标签的issue以了解其上下文。将第一个标记分配给每个project后，根据第一个标记形成的概念将第二个标记添加到每个标签。第二个标记表示比第一个标记更一般的概念。

对于RQ2.4：首先提取每个issue的标签，将其转换为该issue的一组标签，然后使用关联规则挖掘，最后使用共现次数对生成的关联规则进行排序，得到最常用的一起出现的标签集。

**分析结果**：

RQ1：

从下表可以看出，多标签在issue management中是很常见的。

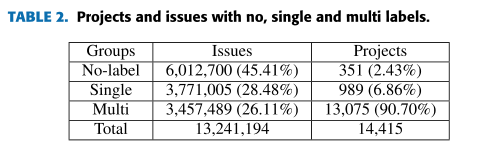
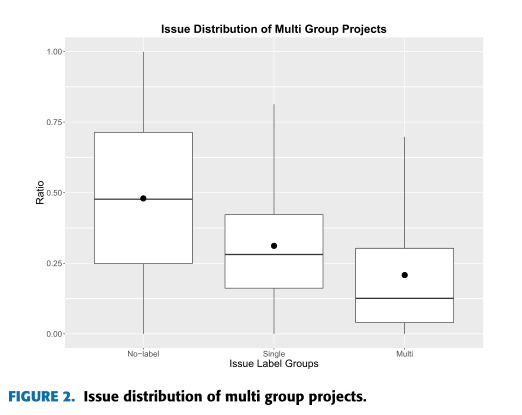


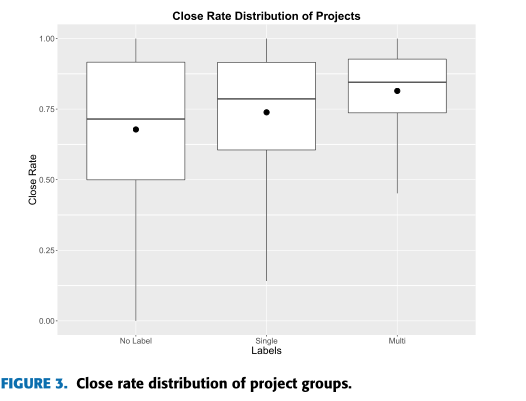
图2表示multi-group的issue分布。对于每个project，计算属于no-label、single-label和multi-label的问题部分。在每个方框图中，水平条表示中间值，点表示平均值。如图2所示，无标签issue平均占issue的比例略低于50%（47.99%）。单标签和多标签issue的平均比率下降，分别为31.18%和20.82%。



RQ1.1结论：

具有多个标签的project的百分比。90%的project至少使用过一次多标签，但只有20%的issue具有多标签。

RQ1.2:



在图3中，no-label、single-label和multi的平均close rate分别为67.80%、73.87%和81.46%。在multi group中，close rate往往更高，表现为朝上的框较短。验证三组之间的差异具有统计学意义（p值<0.05）。结果表明，multi group更有效地管理了他们的问题。与其他两组相比，其close rate分别高出13.66%（无标签）和7.59%（单标签）。

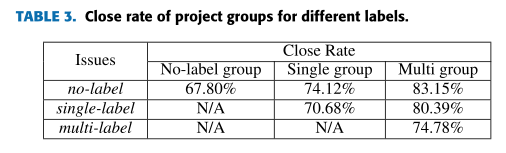
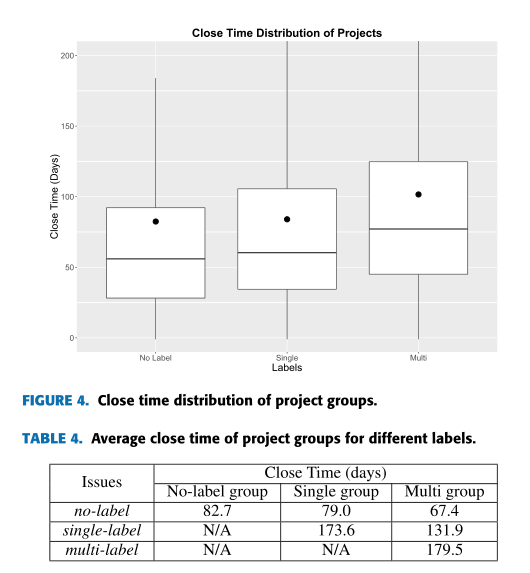


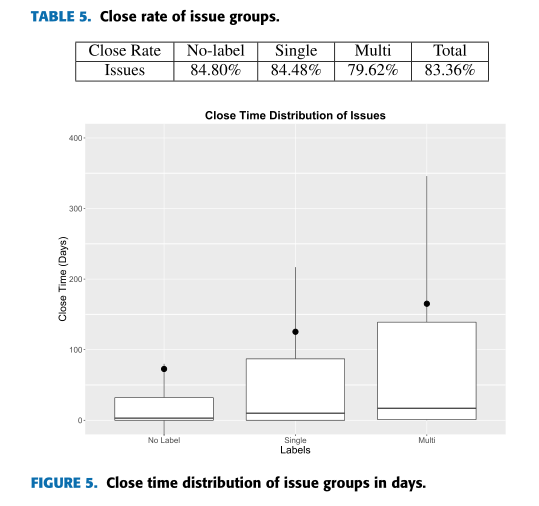
表3显示了每组项目中无标签、单标签和多标签issue的平均关闭率。Multi group无标签issue和单标签issue的平均关闭率分别为83.15%和80.39%。single group project无标率74.12%，单标率70.68%，no-label group project只有67.80%，低于Multi group project。Multi group project更积极有效地处理了无标签和单标签问题。 横向比，纵向比



关闭时间分布

从图4可以看出，multi group关闭时间更长，但是这并不能说明多标签对issue的生命周期管理是低效或者是负面影响的，从表4横向看，single-label在multi group比single group平均关闭时间更长，因此，可以说multi group project已经有效地管理了它们的问题，尽管它们有更耗时的多标签问题，这可能是因为多标签的issue确实更难以处理。

RQ1.3:



水平条表示中间值，点表示平均值

图5中的multi方框图比其他组的方框图扩展得更远。多标签问题如此长的关闭时间是由于它们的困难和复杂性。作者进一步调查了评论的平均数量、标题长度和正文长度。评论更多或标题和正文更长的问题可能表明这些问题更复杂，因此需要更多的讨论。多标签issue需要更多的评论、更长的标题和正文。多标签问题的评论（4.05）多于无标签问题（3.15）和单标签问题（3.56）。多标签issue的平均标题长度（50.48）也长于无标签issue（44.63）和单标签issue（47.03）。

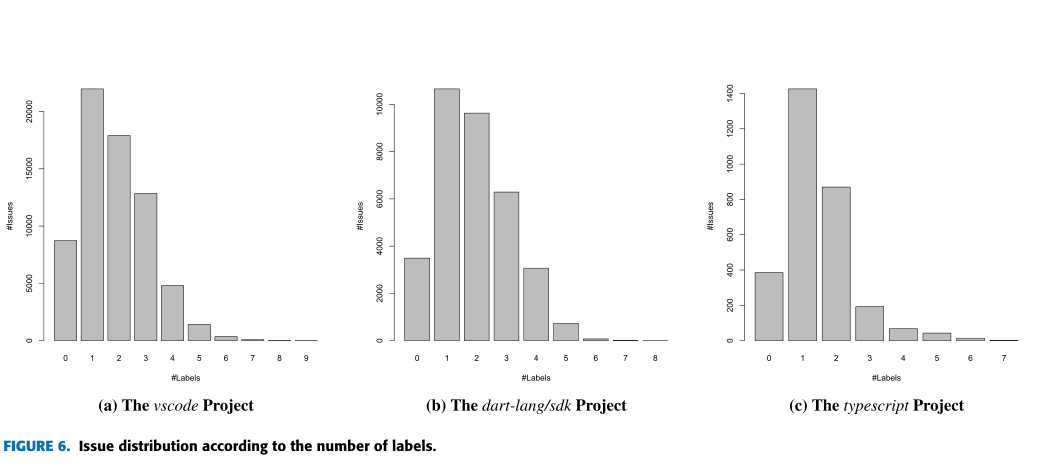
结论：

RQ.1.3具有多个标签的issue的issue management效果。多标签issue的关闭率比其他issue低5%，多标签issue的平均关闭时间约为40- 93天。多标签问题更复杂，也更难解决。

RQ2：

RQ2.1：

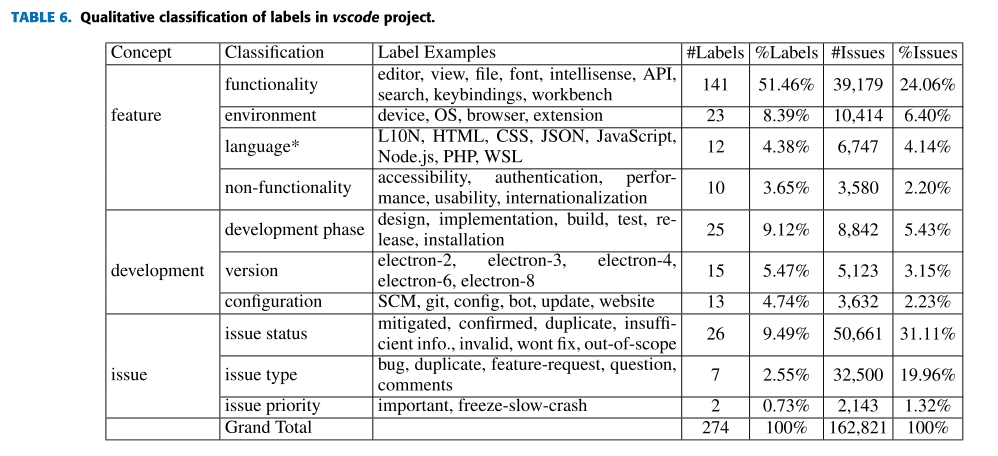
三个项目中issue的标签数目的分布：

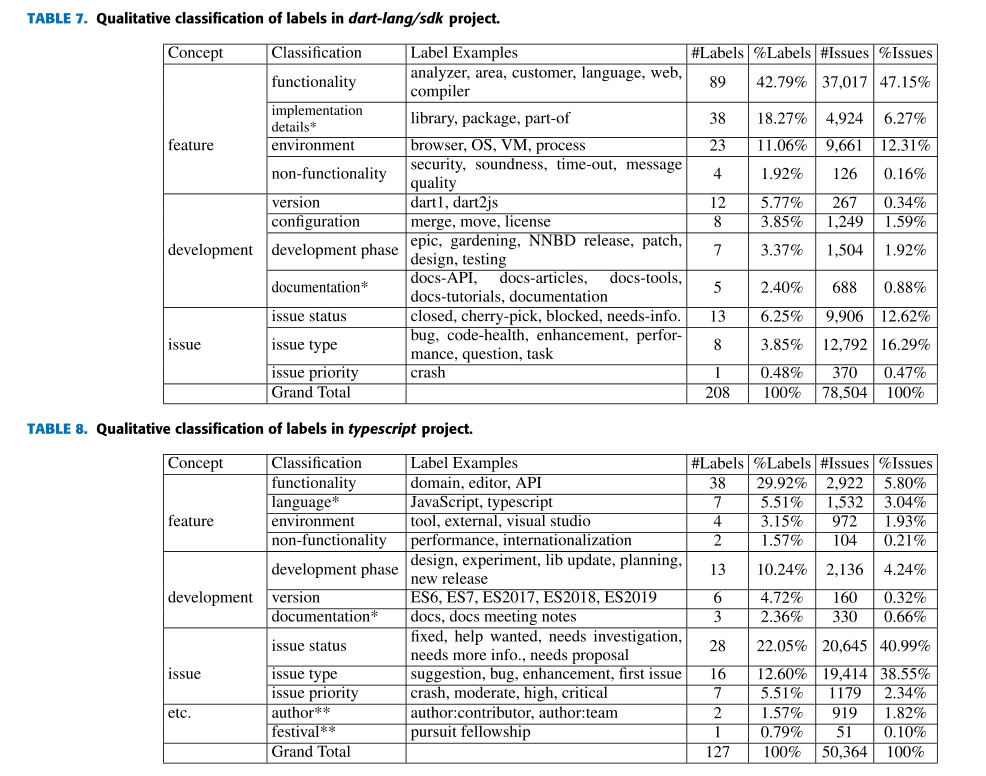


RQ.2.1 issue的标签数量。issue最多有7-9个标签。虽然一个标签最常用于issue，但两个、三个和四个标签也经常使用。相比之下，五个以上的标签很少使用。

RQ2.2：

对三个项目中的标签的定性分析：





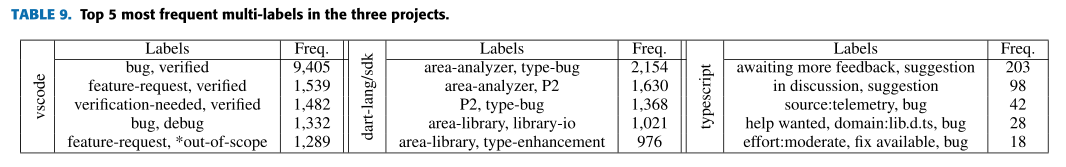
RQ.2.2自定义标签的种类。自定义标签定义在三个概念中，分别与features, development和 issues相关。此外，功能和issue状态是在三个项目中创建许多自定义标签的类别。

RQ2.3：

表6、7、8的#Issues和%Issues说明了自定义的标签的出现的Issue数目和出现Issue的百分比。

RQ.2.3 project中使用的自定义标签类型。所有三个项目都使用“issue status”类别中的标签。关于两个项目使用最多的功能类别中的标签，另一个项目不太使用该类别中的标签。

RQ2.4:



**总的结论**：

首先，在RQ1上的结果显示，大多数项目（约90%）使用了多标签，并且它们显示了更好的issue management性能。虽然总体关闭时间较长，但multi group的关闭率较高，无标签和单标签issue的关闭时间较短。当然，没有确凿的证据表明使用多标签实际上可以提高问题管理性能。然而，可以考虑多标签的使用作为项目有效和有效的issue管理的指标，这是值得研究的。在这一点上，可以考虑什么样的多标签可以用于问题管理。

通过统计，RQ2表明有一个标签的问题数量最多，而有2-4个标签的问题数量高于没有标签的问题数量。基于这些事实，建议为issue添加考虑2∼4个标签。

接下来，RQ2研究了定义和使用了哪些类型的自定义标签。定义了许多feature、development和issue的标签，并且所有三个目标项目都使用了与问题状态和类型相关的标签。这些信息有助于对issue和标签的研究，因为可以将问题和标签聚合起来，并根据这些类别进行分析。

进一步分析了具有不同类型标签的issue的issue management性能，因为这些标签出现在所有三个项目中，并且使用方式类似。带有状态标签的issue的关闭率（79.61%）高于其他issue（72.67%）。这表明使用issue状态标签有助于处理issue，**但自动标签技术通常只关注问题类型**。

此外，所有三个项目中很少使用issue优先级标签，但具有优先级标签的issue的结束时间比其他问题的结束时间（202.4天）短得多（171.9天）。很可能优先级标签主要用于紧急issue。然而，这增加了**分配优先级标签**的重要性，值得开发一种方法来帮助此类标签，并减少开发人员改进问题管理的负担。