

《开源软件设计与开发》课程总结

姓名：朱梦霞 学号：51195100046

1、开源理解

最开始听说“开源”两个字，是在本科阶段做无人机项目时，导师说代码网上有开源的代码可以找到，可以在它们的基础上进行修改。因此最开始对于开源的理解就是，开放的资源，不从中赚取利益，有需要的人就可以自取，可这是为什么呢，现在流行不要专利了吗？我觉得很疑惑。无人机项目进行到后来，我们两三个同学协同编程，为了便于代码的版本管理，导师开始让我们使用 `git`，进行简单的 `commit`，`push`，`pull` 操作，保证每个人每次开始编写代码时，都能够使用当前最新版本的代码，而不至于冲突。但 `git` 和开源这件事，我没有联系起来过，只觉得这两件是分开的事。

读研之后，我频繁听到“开源”这个词，也在课程中，对开源有了更多的了解。一开始的了解让我觉得，开源不就那么回事，觉得身边的人都在夸大其意义，但随着课程的深入，我开始明白开源的意义所在。

开源的全称是开放源代码，最大的特点是开放，任何人都可以得到软件的源代码，加以修改学习，甚至是重新发放，不过是在版权限制范围之内，也就是说，原先我以为开源就没有专利了的想法是错误的。开源系统除了面向普通终端用户之外，还有更重要的一点是面向程序员，促使程序员在开放源代码的基础上进行二次开发利用，产生

新的成果，或者改进原先的成果。由于如今互联网的发达，世界上的所有人都可以通过各种方式，通过互联网取得联系，而开发工作，也可以由世界各地的人才通过互联网上的合作，进行共同开发，即便可能合作团队中的人素未谋面，互相不认识，但是通过开源，通过互联网，在 git 平台上，人们就可以实现这样的远程协同工作，科技是没有国界的，人人都可以为开源社区做贡献，这是一个全民献智慧的过程。由于开源社区的开放性，开放源代码可以被不断更新，得到持续的维护升级，给终端用户以更好的使用体验。

关于开源，在网上看到一句话叫，“不重复发明轮子”，源码被开放后，人们的开发工作不再必须从零开始，而可以在已有“轮子”的基础上进行进一步的创新，减少了许多前期工作，对知识重新利用，效率更高，科技发展与变革的速度也更加之快。

2、开源贡献

在开源社区中，我曾关注了一个开源的日志数据一场分析与实现的项目，项目名为 DeepLog，使用深度学习的方法，识别日志数据中的异常。在项目中我发现自己按照开源的数据集与代码做的实验结果，与开源社区中给出的结果不相同，仔细检查了实验过程，没有发现自己的实验中出现了什么错误，因此在项目中提了 issue，但是始终没有收到作者的回复。

回看 issue 时，发现有人和我的实验出现了相同的问题，也提了相似的 issue，并且也有许多人在 issue 下面追问了相关的问题，但是

似乎作者并没有给出合理的解释，回复也都很敷衍，提问者们的问題仍然没有得到实质性解决，我想这不是一个开源项目管理者该有的态度。而许多大型开源项目，其管理者往往也富有热情，积极响应提问者们的问題，并且不断更新改进源代码，才能保证项目拥有长久的活力。

The processed dataset #13

 Open daphnezmx opened this issue on 12 May · 0 comments




daphnezmx commented on 12 May

+ 😊 ...

Hi, I found that the dataset in the ZIP I cloned which is parsed from the raw hdfs log data, is different to what I processed. I processed the same raw log data according to the log template provided by the method called Spell, but I found that the log sequences are different to what is in the "hdfs_train" file and the other three files. And I also found that there are actually 33 log keys rather than 29 log keys. Did you filter the dataset or not, it confused me these days.

图 1 5 月份提的 issue

data conversion #1

 Open Athiq opened this issue on 18 Feb · 18 comments



Athiq commented on 18 Feb

+ 😊 ...

Do you have script that converts the log files(HDFS files - text) to numbers ??
https://github.com/wuyifan18/DeepLog/blob/master/data/hdfs_train
How did you get the above ?? -- using Spell ?? ... after running the parser i still have text data --- how did you convert to numbers (vectors) ?? --- do you have a script ?? can you please upload ??
<https://github.com/logpai/logparser/tree/master/logs/HDFS>
is this the above data converted to numbers ??
thanks in advance



wuyifan18 commented on 19 Feb · edited ▾

Owner

+ 😊 ...

I use the dataset provided by the author of this paper.
More details please refer to the [web page](#).



sotiristsak commented on 23 Feb

+ 😊 ...

hello. thanks @wuyifan18 for the great job. I think what @Athiq is talking about can be found in the 4.3 paragraph of the published paper. I'm also trying to find out how this could be implemented! any help would be much appreciated

图 2 相似的 issue

3、课程反馈

在课程中，来自企业的老师们通过自身丰富的实践经验，给我们科普了很多东西，教授我们如何开启一个开源项目，如何在开源社区中混出一席之地，如何让大佬们注意到自己，如何在开源社区中做贡献等等，每一堂课都干货满满，有许多具有实用性的建议。

尽管一开始在课堂中就让我们选择一个开源项目，并加入其中，在社区中做贡献，但是似乎真正去做的同学很少很少，我认为主要原因是，对于原先没有接触过开源的同学们来说，一下子进入这些大项目来练习与实践，在如此大型的重要的开源项目中做贡献，起步门槛还是太高了。

我认为在之后的开源课程中，老师们可以创建一个小型的项目，任何起步者都可以接受的小项目，所有人都参与到同一个开源项目中，比如一个刚刚搭建好的，代码量很少的小平台，便于起步的学生们能够基本上都看懂，再让同学们共同思考，在这个平台中，如果我们要实现怎样的一些简单的实际需求，我们需要加一些什么样的功能，通过怎样的一段代码可以实现，或者说，界面如何设计与美化，如何通过代码去实现等等，可以将学生分成后端组，算法组，前端组等等。

这样降低了阅读项目已有代码的门槛之后，同学们会更有热情加入到项目中去做贡献，也会更容易产生成就感，增加对开源的兴趣。

4、参考文献

无。