

权威看台

文章编号: 1672-5913(2017)01-0002-04

中图分类号: G642

开源对软件人才培养带来的挑战

梅宏, 周明辉

(高可信软件技术教育部重点实验室(北京大学), 北京 100871)

摘要: 开源是世界信息技术及产业发展的重要方向, 然而, 面对开源进一步繁荣发展的需求, 开源人才的供应无论在质或量上均存在较大不足。文章分析开源发展及开源人才的现状, 剖析开源人才的能力特点及其培养需求, 对开源软件人才的培养实践进行总结, 并提出对未来教育实践的展望。

关键词: 开源教育; 开源文化; 开源意识; 开源技能

DOI:10.16512/j.cnki.jsjy.2017.01.002



梅宏

开源软件是互联网时代人们在软件领域开展的极为成功的大规模协作实践。开源软件在软件开发效率、质量、安全性、可控性等方面具有的优势极大地推进了世界信息技术及产业的发展。开源软件所取得的巨大成就不仅孕育了新的软件开发方法、改变了软件产业格局和商业模式, 其独特文化和运营机制所产生的影响也远远超出了软件领域。开源软件的成功离不开开源贡献者的创造性劳动,

然而, 面对开源进一步繁荣发展的需求, 开源人才的供应无论在质或量上均存在较大不足。如何面向未来更好地培养开源创新人才是我们急需探讨的问题。软件人才培养体系有必要面向开源的机会和挑战, 我们要厘清现有教育体系存在的现实问题, 把握发展脉络, 积极探索改革, 大力支持开源文化的熏陶和开源人才的培养, 构建基于开源的创新生态环境, 为未来创新软件技术的发展打下坚实的人才基础。



周明辉

作者简介: 梅宏, 男, 教授, 高可信软件技术教育部重点实验室主任, 北京理工大学副校长; 中国科学院院士, 教育部长江学者计划特聘教授, 发展中国家科学院院士, IEEE Fellow, 中国计算机学会会士; 主要研究领域为软件工程和系统软件, 在构件化软件中间件、开发方法学和工具环境等方面取得了系列成果。发表学术论文 150 余篇, 多次获国际会议最佳论文与优秀论文荣誉; 获国家技术发明专利授权 10 余项; 曾获国家自然科学基金二等奖、国家技术发明二等奖、国家科技进步二等奖、中国高等学校十大科技进展等科技成果奖励; 个人荣誉包括: 国家杰出青年科学基金获得者、中国青年科技奖、国家百千万人才工程国家级人选、何梁何利基金科学与技术进步奖、IEEE Computer Society Technical Achievement Award、IEEE Computer Society Meritorious Service Award、国家自然科学基金委创新研究群体学术带头人等; meih@pku.edu.cn。

周明辉, 女, 副教授, 主要研究方向为挖掘软件开发活动数据, 度量软件产品和程序员及其工作文化之间的关系(尤其是开源项目); 在软件工程领域顶级国际期刊和会议 TSE、TOSEM、ICSE 和 FSE 等发表 40 多篇论文, 获 FSE2010 的 ACM SIGSOFT 杰出论文和 COMPSAC 2012 最佳论文; 入选 2012 年教育部新世纪优秀人才计划, 获 2015 年中创软件人才奖; 多次担任国际会议 PC, 如 FSE 2014 Research Demo PC Co-chair, MSR 2016/17、ICSE 2018 的 PC 等; 在研项目包括国家自然科学基金重点项目“软件生命期数据组织、分析及应用”, 973 项目“基于开源生态的网构化软件开发原理和方法”, 以及华为和西门子的咨询项目等; zhmh@pku.edu.cn。

1 开源发展及开源人才现状

开源开发模式为产业模式变革和转型提供了新的途径, 开源软件资源则为 IT 产业发展提供了直接可用的软件技术、工具和产品。开源开发模式的开放性和透明性有助于快速聚集大众智慧, 并有效促进技术和应用生态的形成与发展。

目前有围绕 Linux 操作系统、Android 智能手机操作系统、OpenStack 云平台等著名开源项目形成了众多成功开源生态系统。这些成功开源生态是由广大开源贡献者建立的, 其中的社区形成需要社会提供大规模的潜在用户和开发者群体。例如, Android 在全球拥有超过 14 亿的用户^[1], Linux kernel 有来自世界各地约 1 300 个公司、14 000 人为其贡献过代码^[2]。开源运动发展至今, 软件产品的种类越来越多样、功能越来越强大, 社区的规模也越来越庞大, 相应地, 支撑开源开发的工具和技术体系也越来越难以掌握。新的形势对开源软件的进一步繁荣发展带来了挑战, 对开源参与者提出了更高的要求。在开源软件开发的参与方面, 有研究指出, 近年来在知名开源社区中的长期贡献者的数量和比例有下降的趋势^[3]。调查发现, 在开源产品和技术的应用方面, 国际上有相当多的 IT 公司缺乏掌握开源技术的人才^[4]。有效地培养开源人才, 壮大开源贡献者的队伍势在必行。

2 开源人才的能力特点及培养需求

相对于传统软件开发领域, 开源对软件人才有着不同的甚至更宽阔的要求, 这使得传统的人才培养体系可能需要适应需求并关注开源人才教育的特定方面。

1) 开源模式对开发人员的能力有什么特别的需求?

首先, 开源社区的组织及开发模式有别于传统 IT 企业, 厘清开源开发的特点及其对开发人员的能力需求非常重要。在开源软件开发中, 参与人员的地理分布范围广、背景及文化差异大, 社区的组织相对松散、自主, 开发的过程缺少统

一的、明确的规范, 项目的系统化、标准化的设计和描述文档相对缺乏, 各个项目间的上下游关系复杂, 同类项目间的竞争激烈, 项目本身的流行程度变化迅速。近年来, 开源运动中又融入了商业公司的参与, 由于各方利益相关者可能存在需求冲突并对社区参与者施加影响, 使得开源开发更加复杂^[5]。开源的这些特点需要开源参与者具备理解开源生态和掌握开源技术的开源意识, 具备开拓互联网创新应用的创新意识, 以及具备贴近应用、技术熟练、善于协作的实践能力。然而, 传统软件人才培养体系对开源能力的关注度还远不够, 所培养的学生在这些方面所具备的能力与要求还存在较大差距。根据 Dice.com 和 Linux 基金会的开源技术就业统计报告^[4], 招募开源技术人才已经成为 IT 招聘经理的第一要务, 87% 的招聘经理表示开源人才很难寻觅。企业作为开源技术人才的需求者, 已经率先有所动作。例如, RedHat 公司积极开展开源教育来提高参与者的开发能力^[6], 希望培养更多能够参与开源运动的人才, 以促进开源事业的发展。

其次, 开源社区的形成和维持跟传统 IT 组织 (或项目) 有很大的不同, 特别是参与者的驱动力与传统组织中靠行政力约束员工有很大的差异。根据 Dice.com 和 Linux 基金会的开源技术就业统计报告^[4], 86% 的 IT 专业人士表示开源技术推动了他们的职业发展, 但他们最看重的却并非薪酬和待遇。深入认识开源社区中参与者的动因及社区的成因有助于理解如何建立相应的培育措施来吸引和维持合适的开源人才对社区的参与。开源社区中参与者因为各种各样的驱动力而进行贡献, 如想要从社区中学习或通过社区贡献来建立声誉等。研究表明, 贡献者加入社区时初始的意愿、能力及其环境对他们是否会在社区中长期贡献有统计意义上的重大影响^[3]。目前, 开源社区和一些商业公司正在尝试一些针对上述发现的举措来吸引和维持人们的参与。例如, 为增强贡献者的意愿, GitHub 公司创立了开放、透明的社会化的编程协助模式, 极大地调动了人们的参与热情^[7]; 为营造友好的社区环境, Rails 社区引入了 Highfive 项目^[8], 通过自动化的方法引

导新的贡献者融入社区。

2) 教育体系需要重点关注开源人才培养的哪些方面?

对开源软件开发的学习与实践是增强开源能力的一种重要途径。开源为学习者提供了开放、低成本的跟踪学习机会,但开源的内涵丰富、规则复杂、文化多元和语言差异明显,如何有效地跟踪学习开源仍面临诸多挑战。开源具有开放、自由、分享的基因,引领者鼓励学习者参与其中并做出贡献。开源涉及的内容非常广泛,除了包括开发者关注的技术性内容,如开源软件本身以及开源开发的支撑技术和工具,还包括开源参与的个人、政府、企业需要学习的开源制度、开源规则、开源方法、开源模式、开源文化等。具体来说,开源人才的教育需要包括下述几个方面:

(1) 开源文化教育。对开源的认识首先要从理解开源文化开始。开源最初源自自发的源代码学习与分享,后来发展到来自全球的开发者根据开源社区的规则自由地参与进来,既可以对自己感兴趣的软件项目提供外部贡献,又可以把自己的创新想法发布出去,开发者在项目核心团队的引导下,与全球的开发者共同改进开源软件。开源文化有别于传统的商业开发,其基于互联网的“大众化协同、开放式共享、持续性演化”的开发模式是开源软件的核心。开源文化的内容主要包括开源社区的形成机理和运转机制及其得以持续生存和发展的机制机理等,同时还包括开源历史、开源共识及开源社区的治理(governance)规则等。开源文化的熏陶对个体自我开源意识的形成至关重要。

(2) 开源意识教育。开源作为一种综合了软件创意、生产、分享、使用、培训、创新、营销、生态等内涵的大规模协作活动,其相关意识主要体现为创新意识和开放透明的协作共享意识。创新意识表现在可以在开源技术迭代的基础上,敏锐感知新兴技术的需求并进行快速创造。而协作共享在当前全球分布式开发的趋势下是需要必备的意识。我们需要拓广传统教育在这方面的训练,尤其是全球分布式共享协作思维的训练在传统教育中较少涉及。

(3) 开源技能教育。开源技能一方面体现为开发者传统的编程能力,另一方面体现为开发者对开源技术和工具的使用能力。开源中存在适合各类场景的技术和工具(如分布式版本管理工具 Git 和项目托管平台 GitHub),涵盖开发的各个过程和步骤。开源技能还表现为分布式环境下的协调协作能力。它跟传统的协作开发能力有所差别,例如协作成员可能从来没有见过彼此(因此缺乏经常见面可以达到的基本信任),还可能有时区和语言的差异等。

3 开源软件人才培养探索

培养 Internet 时代的开源人才才是全世界面临的挑战,需要构建适合高校教学特点、基础条件和教师队伍的新型软件人才培养体系和支撑环境。应该说,当前对开源人才的系统性教育体系尚处于探索阶段,离满足开源繁荣发展的需求还有较大距离。已有的探索实践主要涉及以下两个方面,其在深度和广度上还需进一步延展,而系统性教育体系的建设则更是任重道远。

1) 开源软件的推广与教育。

涉及的主要问题是建立怎样的机制进行开源推广?可以从多种渠道开展。首先,需要帮助建立直接接触开源软件和开源社区的通路。开源社区往往离不开学校的贡献,例如, Linus Torvalds 在读研究生的时候开始撰写的 Linux kernel,加州大学伯克利分校的 AMP 实验室所开发并开源的通用并行框架 Spark 等。如何在学校的教学与科研中推进和利用这种实践,使得学生能够接触到开源软件及其创新思维,是教育体系需要首先考虑的问题。

其次,需要在学校读物和基础设施上对开源进行投入。例如,联合国国际开源网络(IOSN)在2003年出版了《自由/开源软件》通论,阐释了使用开放源码软件的优势,并提供了在教育领域使用的服务器和台式机软件资料。这套读物阐述如何创作、发布共用许可证(creative commons license),是推广开源和自由软件的优秀资料和入门“教材”,对社会应用、教育和政

策支持均有重要价值。在基础设施方面,例如,卡耐基·梅隆大学、华盛顿大学、纽约大学等高校联手百度深度学习研究院(IDL)、微软亚洲研究院建立了分布式深度机器学习开源平台 DMLC (<https://dmlc.github.io/>),使得全世界的学者、学生都能够以它为基础学习、研究并参与贡献。

2) 建立以开源为中心的软件工程实践教学体系。

引导学生建立起开源理念,学习开源方法,掌握开源工具,理解开源生态;建立产学研协作的、开源为中心的软件工程实践教学平台,让学生能够在实践环节接触开源、使用开源、分析开源、参与开源;利用学校的相关课程鼓励学生参与并贡献开源软件。例如,联合国教科文组织在 2013 年启动了开源教育管理信息系统,对使用国家没有条件限制。开源教育管理信息系统能够在计算机和移动设备上使用,用以收集、处理、分析和传播教育信息。在目前的教育体系中,已有把开源教育应用到软件工程教育中的实践,如卡内基梅隆大学开设了开源软件课程(编号 17-624),通过学习该课程,学生可以了解开源技术的特点、开源社区的性质、开源软件的开发过程以及开源产业结构的演化,进而具备使用开源软件与技术以及参与开源软件开发的能力。

如何在创新实践中发现、培育、凝聚大规模高素质的开源软件创新人才,需要政府、高校和企业合作建设产学研深度融合的新型人才培养体

系,这是一件重要而具有挑战性的任务,也是一项长期的任务。就高校而言,教育体系的改革就是一件费时而复杂的事,涉及方方面面,唯有符合教育规律、满足开源发展人才培养需求的系统性教育体系的建立,才能使高校成为开源人才甚至开源奇迹的摇篮。

4 结 语

开源软件是人类历史上一次基于群体智慧、利用互联网实现分布式协作的大规模成功实践。开源运动已逐步建立起基于群体智慧的软件开发方法和生态环境,实现了对软件产业的重大变革。这种变革对全世界参与开源的人们数量和质量都提出了挑战,各国教育体系应能随之应变,重视开源软件相关能力的培养,满足新时代的创新需求。

就我国而言,在信息技术领域仍然处于跟跑的位置,特别是长期信息化进程中存在的关键技术“空心化”和技术装备“低端化”的现象仍未得到缓解。应该说,开源软件对我国实现核心基础技术的自主可控带来了“弯道超车”的机会。我们需要大力重视开源,基于开源模式,加快发展并提升我国信息技术领域的核心竞争力。同时,围绕公益性开源生态环境及其基础设施建设,引导高校进行课程体系改革,加强开源软件人才培养。

参考文献:

- [1] Craig S. Amazing Android statistics[EB/OL]. [2016-11-28]. <http://expandedramblings.com/index.php/android-statistics/>.
- [2] Kroah-Hartman G. Presentation on how the Linux kernel developed[EB/OL]. [2016-11-28]. <https://github.com/gregkh/kernel-development>.
- [3] Zhou M H, Mockus A. Who will stay in the floss community? Modeling participant 's initial behavior[J]. IEEE Transactions on Software Engineering, 2015, 41(1): 82-99.
- [4] 2016 Open Source Jobs Report[EB/OL]. [2016-11-28]. <http://go.linuxfoundation.org/download-2016-open-source-jobs-report>.
- [5] Zhou M H, Mockus A J, Ma X, et al. Inflow and retention in OSS communities with commercial involvement: A case study of three hybrid projects[J]. ACM Transactions on Software Engineering and Methodology, 2016, 25(2): 13.
- [6] The importance of open source in education[EB/OL]. [2016-11-28]. <https://www.redhat.com/en/about/open-source-education>.
- [7] Dabbish L, Stuart C, Tsay J, et al. Social coding in GitHub: transparency and collaboration in an open software repository[C]// Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. Seattle, Washington, USA: ACM, 2012: 1277-1286.
- [8] Github hooks to provide an encouraging atmosphere for new contributors[EB/OL]. [2016-11-28]. <https://github.com/rails/rails-bot>.

(编辑:彭远红)