

浅谈开源社区与开源项目的发展与现状

赵炫皓

学号: 1120200603

北京理工大学/ 中国北京市海淀区

zhaoxuanhao0111@gmail.com

摘要

本文从开源的定义谈起, 阐述了开源发展的历史进程, 简要分析世界开源的发展与现状, 论述了国内开源界所存在的问题和隐患, 并以多个具体的开源社区为例, 分析其中的运行模式和商业模式, 提出了开源发展对企业发展和人才培养的影响, 最终对中国开源的未来进行一系列的展望。

关键词: 开源; 开源社区; 模式分析; 发展现状; 未来展望

Brief introduction to the development and status of open source communities and open-source projects

Xuanhao Zhao

Student ID: 1120200603

Beijing Institute of Technology / Haidian District, Beijing, China

zhaoxuanhao0111@gmail.com

Abstract

This article starts with the definition of open source, expounds the historical process of open source development, briefly analyzes the development and status quo of open source in the world, discusses the problems and hidden dangers existing in the domestic open source community, and analyzes the development of open source by taking a number of specific open source communities as examples. The operating mode and business model of open source development, put forward the impact of open source development on enterprise development and personnel training, and finally make a series of prospects for the future of open source in China.

Keywords: Open Source, Open-Source Community, Model Analysis, Development Status, Future Outlook

1 引言

过去的数十年以来, 开源社区和开源软件以一种新型的生产方式, 成为全球信息行业的重要组成部分。目前, 开源在全球范围内飞速发展。据各方数据显示, 中国开源目前正在接近世界先进水平。“十四五”规划明确将开源列入国家总体规划目标当中[28], 从此为中国开源开启了新的篇章。

1.1 开源定义

开源是指开放开发软件的源代码，是从软件行业兴盛起来的一种开放模式[12]，是促进信息交流与技术创新的重要手段。当前技术领域的开源种类繁多，包括但不限于开放源代码，数据等资源，其中开放源码是开源技术中最主要的内容[12]。

使用者将以许可证的方式，在一定的条件模式下，可以相对便捷地获取软件源代码，并自由地使用，修改甚至添加一定的功能以后再发布。开源模式通过集群管理和维护，以开放的方式不断吸收新的智慧，并安排维护者监督，通过互联网共享源码，共同开发，在一定程度上大规模地提高开发效率，并提高了运维社区的活力[8]。

开源还包含了诸多要求与限制，最常见的是要求开发者需要遵守特定的开发规范，并且对于源代码的使用和源代码的传播遵守着公平性和中立性等准则，需要开发者注意使用和维护开源软件的特定规范[22]。

1.2 开源特点

开源工程有以下特点：拥有在技术领域广泛而开放的科学创新理念，具有协同共享的价值特征，具有相对宽松并且自由的开发理念，但是于此同时在使用和传播的过程中需要遵循一定的规范。

开源通过放开源码吸引更多的人才和智慧进入开源社区。开源使用的激励机制有别于日常所常见的基于经济回报所创建的激励机制，更多的是基于维护整个社区，推动社区经济生活整体发展的机制，同时通过有效的监管，避免原创技术的流失和整体工程的烂尾[4]。

简而言之，开源可以通过其本身便利，共享开放的特点，不以经济收益为主要的目的，从而集中智慧，集中原创技术，使得开发者共同维护并且自我更新。

1.3 国际开源发展历史

开源（Open Source）一词最开始来源于黑客社区Debian社长Chris Peterson在上世纪末所提出的一项“自由软件指导方案”[1]。在提出概念的次日，Linux的开发者与创始人就给予其版权许可声明，Bruce Perens创办了第一个开源网站。在自由软件的会议中，开源软件的支持者讨论并确定了开源软件推广并传播的必要性，就此开源一词开始真正走进软件开发行业从业者的生活中[22][12]。

开源的起源与Unix系统相关。自世界上第一台计算机诞生到上世纪70年代中后期，计算机一直是以大型机的形式生产，软件一般并不作为单独的产品进行提供，软件的传播一般依赖硬件的销售进行附赠。因此在这段时间Unix的源码曾向大众开放。到了二十世纪末，随着用户规模和市场不断增大，软件销售出现了巨大的收益，因此软件逐渐从免费开源变成付费闭源的销售模式，这引起了一些美国的软件开发者的不满。

Richard Stallman在80年代发起了自由软件运动，开始创建并维护GNU，越来越多的UNIX开发者转而投向GNU的开发和维护。GNU/Linux此类自由软件的迅猛发展和壮大对UNIX等闭源软件公司带来极大的冲击。而经过多年的维护，自由软件逐渐难以形成商业规模，出现了一定的自由软件发展瓶颈。因此也就出现了后来的开放源码，使得全球的开发者和公司都可以对自由软件进行开发与维护[12][27]。

1.4 中国开源发展历程简述

中国开源自上世纪90年代开始，中国Unix编辑委员会联合全国众多专家和业内人士翻译并出版了Unix相关书籍，并逐步引入UNIX系统。

到了本世纪初，自由软件和开源软件的概念和理念不断在软件行业相关企业和从业人员之间传播。随着互联网时代的到来，UNIX和LINUX的用户在国内激增，此时尚处于开源发展的萌芽阶段[12]。

2010年以来，中国开发者逐渐成为国际开源资源的重要贡献者。中国开源社区纷纷尝试使用开源技术发展和维护自身产品[25]。现如今，中国已经拥有成熟并且商业化的开源模式，并且互联网企业联合成立基金会，填补了中国开源基金会缺失的空白[10]。

1.5 开源发展所面临的挑战

中国开源技术由于起步晚，投入少，目前面临的挑战如下：

开源基础欠发达

我国开源基金会创立较晚，基金会数量较少。由于缺乏行业规范，基金会工作人员的薪酬水平难以得到保证[10]。因此这种管理制度不利于开源生态长远发展。同时我国许多不成熟的开源社区是依赖国外的开源代码或开源社区，总体上还是国外开源社区的衍生社区[27]，因此在建立成熟规范的开源社区以前，还存在着开源断供的风险。

开源管理合规和管理不善的风险

企业和用户在引入开源软件时存在着诸多风险。例如，开源许可协议作为外来物之一，在国内是否可以直接使用，是否受法律保护这一问题并没有明确说明[17]，存在法律规范的隐患。同时开源软件本身也存在安全风险。由于开源软件数量众多，难以统筹管理，因此未妥善管理或存在安全漏洞的源码数量在不断增加，过期和废弃的开源组件非常普遍[17]。目前，木兰许可协议是我国唯一的开源许可协议，容易陷入协议条款纷争，引发知识产权风险。因此健全相关规章制度是未来中国开源社区健康发展的必需条件[12][28]。

开源人才供给不足

开源社区的活力来源于高质量的技术人才梯队。但相比于欧美等一众发达国家，我国受过高等教育的人才中，熟悉并参与开源项目的比例较小[15]，因此开源人才缺乏将成为未来开源社区的一大挑战之一。

2 开源模式分析

2.1 开源社区运作分析(以GitHub为例)

如图1[8]所示，用户通过个人专业知识和思考使用Fork库维护开源代码，通过Pull Request提交给项目团队，社区根据提交的代码决议是否合并入仓库[20]。

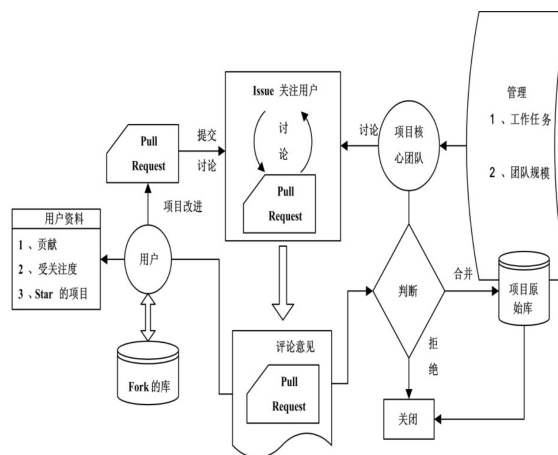


Figure 1: GitHub的社会化项目开发机制

如图2[19]所示，显性知识指的是软件项目，用户将该仓库Fork以后学习的过程便是知识内化。

用户通过提交代码到审核的过程中，人们通过不断完善项目，交换知识从而使得隐性知识的进行内化[25]。从Pull Request到Issue讨论的过程中，所有对项目感兴趣的从业者都可以参与讨论之中，这是知识充分社会化流转的过程[6][26]。

原项目组可以根据审核结果，将得到的显性知识与原项目合并，组合得到新的显性知识，这即是GitHub社区中知识组合的完整过程[19]。

2.2 开源项目量化分析

如图3[6]所示，截至2015年，GitHub代码仓库的数量一直呈现指数型上升，并且达到了1600万个仓库之多[6]。借助SPSS数据分析软件，对GitHub社区的数据进行分析，结果如图4[6]所示。本图展示了开源软件在发布之际，将会吸引大量开发者参与。在达到峰值以后，会迅速流失一批开发者，经过时间的推移，开发者的数量会逐渐趋于稳定，最终形成一个成熟的开发社区[23][6][3]。

而对于开发进程而言，数据可以由图5[6]所示。

模块度指模块化度量值，是一种用来衡量网络社区结构强度的方法。模块度值主要取决于网络中结点的社区划分情况。其值越高说明社区结构强度越强[7]。

由图5可以看出一个新软件缺陷出现，若干开发者聚集成为一个小团体，对软件缺陷进行修复。随着修复和开发接近尾声，其模块度也会下降。如此循环往复，直到项目逐渐趋于稳定[25][6]。

2.3 开源软件的商业模式分析(以OpenStack为例)

企业厂商在研发和销售的过程中可以采取三种不同的方式对待开发的项目[28][16]。

完全私有化：指厂商在封闭的状态下自主研发，对于软件的制品有着完整的知识产权。开源服务支持：指在研发阶段开源，开发由开源社区承担，厂商对软件没有完整的控制权，可以任由软件自由传播，厂商所提供的增值服务是利润的来源。开源混合授权模式：厂商在研发阶段独自开发，对结果进行私有和开源双重授权。

OpenStack是如今发展最迅速，商业活动最多的热门开源社区之一，因此研究OpenStack的商业活动可以帮助理解一般开源社区的商业逻辑。

如图6[4]所示，OpenStack开源社区参与的商业模式基本如下：

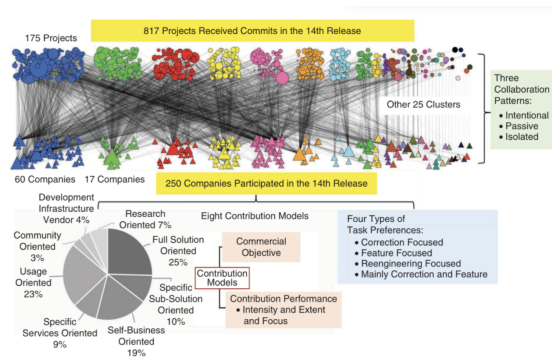


Figure 6: OpenStack商业参与

可以看出，OpenStack的商业参与普遍且广泛。平均每个版本都有100多家公司参与，其中公司做出的贡献已经远远多于志愿者，如图7所示[4][5]。

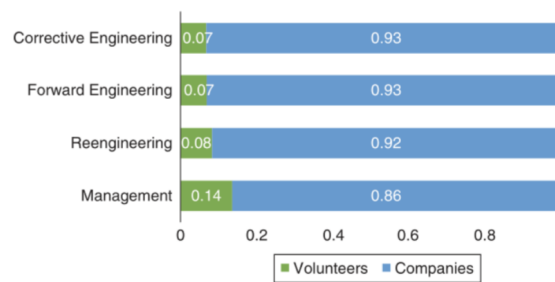


Figure 7: 由公司和志愿者参与项目开发的占比

一些公司通过向用户提供云服务来盈利，还有一些公司在生产环境中直接使用OpenStack或其生态系统内的产品来获利[13]。而OpenStack项目由于积聚多方的帮助，自身项目规模和项目质量得到大幅提高，以此来实现共赢的局面。

如图8所示[16]，基于OpenStack的会员商业组织，可以看出OpenStack参与商业的可能组织形式。

可以看出，OpenStack商业的主要逻辑是通过开源项目，通过允许代码进行自由传播，从而吸引开发者和企业贡献代码或资金，并扩大影响力。还有一些情况下通过提供增值服务（如云服务和咨询服务），推广产品来增加企业的利润和影响力，以此达到商业上的成功[9][16][14]。

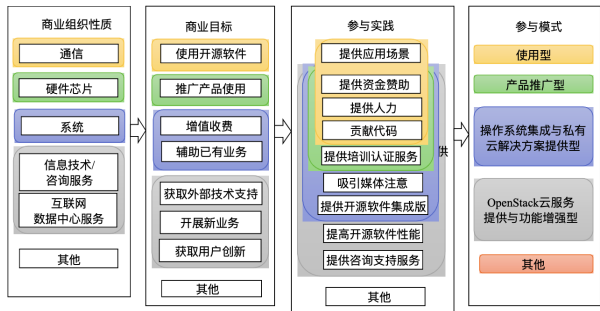


Figure 8: 一般化开源项目商业参与模式

2.4 开源社区治理与维护

同时，公司在发现开源项目的业务目标无法实现或工作量太大时，往往会选择退出开源项目的开发。然而，当一些长期主导开源项目开发公司推出时，开源项目将会失去此前维护和发展的能力，届时将面临极大的风险[7]。例如，一家公司可能会主导一个开源项目的开发，以此来实现利润最大化或迅速实现开源项目的某个功能。某个版本的绝大部分项目可能主要由一家公司主导，这样会造成不利的影 响，并且减少其他企业和用户的工作和创新[16]。

因此，开源社区需要尽可能维护公司的利润与开源社区产品之间的微妙平衡。开源社区在管理项目的同时需要参与商业治理。通过监控一个公司对项目贡献的比重来维护开源项目的健康状况。一旦某个项目的支配度大于50%，开源社区可能需要采取一定的手段来防止项目被垄断或控制的风险[17]。

3 开源软件与社会发展

3.1 开源对于企业发展的影响

当某一项目研发成本较高时，开源项目可以为企业带来最高的成本优势，企业可以通过项目的增值性服务实现利益最大化。当研发成本中等或较低时，开源混合授权模式能够充分利用授权协议的特殊性，有效缓解用户对于创新成果开放的抵触情绪，实现厂商的市场覆盖和利润最大化[16]。

3.2 开源与人才培养

相较于普通的软件开发行业的从业者来说，开源工作者的组织较为分散，在自主开发的过程中缺乏规范性的监督，因此对从业者有较高的专业能力的要求[7]。其次，开源工作者缺乏薪酬待遇的激励，开源组织对待从业者时应当阐明和给出开发者对于社区统计意义上的贡献，以此激励开发者[15]。

因此，开源教育体系应该着重介绍开源社区的基本流程和激励机制，弘扬开源文化来吸引更多的人加入到开源社区当中，同时开展开源意识教育和开源技能教育，培养具有专业素养的开源人才[11]。

具体措施例如：学校与知名开源社区合作，使得学生有机会参与开源项目的开发和维护。高校的读物和基础设施与开源接轨，并构建以开源项目为中心的工程实践[24][18]。

4 国际开源发展对中国开源的启示

开源发展应该着重进行生态建设。目前我国已经在国际开源社区中做出重要贡献，但是还欠缺贡献程度和国际影响力。作为科技强国，我国将会在十四五规划的布局下壮大并发展开源社区[28]。

4.1 健全开源基金会管理机制

开源基金会作为一种成熟而有效的管理机制，可以有效地对开源项目进行监管与维护。目前我国基金会数量较少[17]，规模小且缺乏一定规范。因此除了建立以国家为主导的开源基金会以外，还应该鼓励头部科技企业建设和发展开源体系，建立基金会或开源联盟，共同推动中国开源事业的蓬勃发展。

4.2 推动开源产业的落地和实施

要充分发挥十四五规划对于开源社区的扶持和建设，可在软件信息发达地区强化地方财政对开源社区或开源服务的支持与补贴，落实人才，资金，场地的管理建设。对于积极建设开源社区的企业和团队进行进行技术认定并给予优惠政策[27]。

4.3 引导产业加强开源治理

(1) 制定和认证符合规范的许可证体系

可以出台相关配套政策以规范开源社区和开源流程，对开源产品的引入和运维管理全流程做出明确规范。

(2) 健全发展知识产权联盟，保护开源工程的原创性工作

可以对标开源发明网络之类的知识产权组织，由相关部门和产业出资，对信息产业的细分领域建立专利审查相关工作[27]。

(3) 提高国内开源项目的国际影响

建设独立自主并且强大的开源体系必须坚持走出去的原则，在与国外开源社区开源项目合作的基础上，构建好国内社区建设和人才培养体系，推动开源项目走向世界舞台[28][21]。

5 开源发展趋势以及展望

我国开源项目已经日益接近世界先进水平，但其中还有一定的差距。因此我国应该着重注意开源社区和开源项目的建设，使得十四五计划在开源领域可以顺利落地。同时，开源社区和开源软件具有集中智慧，集中力量的作用，是未来软件行业的一种重要的发展模式，所以成熟的开源社区在我国的复兴之路上可以说是必不可少[1]。加强开源社区的建设对提高我国综合科技实力有着重要的作用。目前，开源软件作为基础软件的重要模式，在全面数字化的今天意义非凡。我国应分析开源社区和开源项目发展的短板和缺陷，借鉴国际开源发展的成功经验，加强顶层设计和统筹协调，弘扬开源开放共享的精神，提高我国开源在国际上的影响力和水平，探索一条符合我国软件发展的独立自强之路[2]。

References

- [1] 《软件和集成电路》编辑部. “传承开源基因释放创新活力2020第十五届”开源中国开源世界高峰论坛云端盛大召开”. In: 软件和集成电路 8 (2020), p. 2.
- [2] 《软件和集成电路》编辑部. “拥抱开源缔造创新模式——2021第16届开源中国开源世界高峰论坛圆满举办”. In: (2021).
- [3] Y. Tian et al. “What Makes a Good Commit Message?” In: (2022).
- [4] Yuxia Zhang, Hao He, and Minghui Zhou. “Commercial participation in OpenStack: Two sides of a coin”. In: *Computer* 55.2 (2022), pp. 78–84.
- [5] Yuxia Zhang et al. “Turnover of Companies in OpenStack: Prevalence and Rationale”. In: *ACM Transactions on Software Engineering and Methodology* (2022).
- [6] 叶培根, 毛建华, and 刘学锋. “基于大数据的GitHub开源社区开源项目量化分析”. In: 电子测量技术 40.8 (2017), p. 6.
- [7] 常静. “百度百科及开源社区参与者的动机与行为关系的实证研究”. In: 华南理工大学 (2010).
- [8] 张劲松, 黄华, and 胡子祥. “知识管理视角下开源社区知识流转机制研究——以GitHub为例”. In: 情报科学 36.7 (2018), p. 7.
- [9] 张宇霞 et al. “OpenStack开源社区中商业组织的参与模式”. In: 软件学报 28.6 (2017), p. 14.
- [10] 张文波. “中国开源软件产业政策研究”. PhD thesis. 中国人民大学, 2009.
- [11] 张晓华 and 李春雷. “以开源项目促进大学生科技创新活动的开展”. In: 华北水利水电大学学报(社会科学版) 27.1 (2011), pp. 142–143.
- [12] 无. “中国开源发展概述,机遇与挑战”. In: 软件和集成电路 2 (2022), p. 10.
- [13] 李小宁 et al. “基于OpenStack构建私有云计算平台”. In: 电信科学 28.9 (2012), p. 8.

- [14] 林利 and 石文昌. “构建云计算平台的开源软件综述”. In: 计算机科学 39.11 (2012), p. 8.
- [15] 梅宏 and 周明辉. “开源对软件人才培养带来的挑战”. In: 计算机教育 1 (2017), p. 4.
- [16] 王宇. “开源创新企业的商业模式与授权机制研究”. In: 科技进步与对策 31.1 (2014), p. 6.
- [17] 王晓冬. “我国开源软件产业面临的突出风险及对策研究”. In: 信息安全研究 7.10 (2021), p. 4.
- [18] 祖文清. “开源系统网络教学平台及其学习评价系统的研究与实现”. PhD thesis. 北京交通大学, 2008.
- [19] 范家宽 et al. “数据驱动的开源贡献度量化评估与持续优化方法”. In: 计算机科学 48.5 (2021), p. 6.
- [20] 蒋竞, 吴秋迪, and 张莉. “开源社区评审过程度量体系及其实证研究”. In: 软件学报 032.012 (2021), pp. 3698–3709.
- [21] 赵国栋 and 黄永中. “开源软件在高校的应用与推广策略研究”. In: 中国远程教育 1 (2007).
- [22] 金芝, 周明辉, and 张宇霞. “开源软件与开源软件生态: 现状与趋势”. In: 科技导报 34.14 (2016), pp. 42–48.
- [23] 陈丹 et al. “开源社区中已有开发者的合作行为分析”. In: 计算机科学 S1 (2016), p. 5.
- [24] 陈丽 et al. “开源模式下我国软件人才培养的挑战和对策”. In: 高等工程教育研究 4 (2022), p. 8.
- [25] 陈光沛, 魏江, and 李拓宇. “开源社区:研究脉络,知识框架和研究展望”. In: 外国经济与管理 043.002 (2021), pp. 84–102.
- [26] 陈晓红, 周源, and 苏竣. “分布式创新, 知识共享与开源软件项目绩效的关系研究”. In: 科学学研究 34.2 (2016), pp. 228–235.
- [27] 隆云滔 et al. “国际开源发展经验及其对我国开源创新体系建设的启示” in: 中国科学院院刊 36.12 (2021), p. 9.
- [28] 韩羽. “打造安全可靠的新型数字基础设施赋能经济社会各领域绿色低碳发展——工信部发布《“十四五”信息通信行业发展规划》”. In: 中国科技产业 12 (2021), p. 2.



Figure 9: 查重率