

开源课程及数据科学导论的开源

朝乐门

中国人民大学数据工程与知识工程教育部重点实验室 北京 100872

中国人民大学信息资源管理学院 北京 100872

摘要 为解决当前课程建设面临的教师备课工作效率低、教育管理部门的激励覆盖面受限以及 MOOC 课程在教师群体关注点上存在盲区等问题,提出了开源课程行动倡议。在此基础上,结合作者自 2017 年以来建设首门示范开源课程——数据科学导论——的经验教训,给出了开源课程的内涵与特征,并探讨了开源课程建设的流程、原则和关键技术。开源课程行动倡议是继 MOOC 之后的一项全新的教学改革思想,为进一步推动教学改革,尤其是在复用教学资源、提升备课质量、分享教学经验、弥补 MOOC 课程的不足方面具有较高的参考价值。

关键词: 开源课程;数据科学;教学改革;课程建设

中图法分类号 TP391

Open-source Course and Open-sourcing Intro to Data Science

CHAO Le-men

Key Laboratory of Data Engineering and Knowledge Engineering, Renmin University of China, Beijing 100872, China

School of Information Resource Management, Renmin University of China, Beijing 100872, China

Abstract A novel initiative called Open Source Courses is proposed to provide an alternative solution for improving the poor efficiency of teachers' lesson preparation activities, expending the limited incentive coverage of education management departments, and satisfying the interests of teachers in MOOC courses. Furthermore, the definition, characteristics of open source courses are described, and the process, principles, and key technologies of open source course construction are proposed based upon the lessons from the construction of the first Open Source Course named Introduction to Data Science. The Open Source Course action initiative is a novel teaching reform idea after MOOC, which has a high reference value for further promoting teaching reform, especially to reuse teaching resources, improve the quality of lesson preparation, share teaching experience, and make up for the deficiency of MOOC courses.

Keywords Open Source Course, Data Science, Teaching reform, Course construction

1 引言

开源课程(Open Source Courses, OSC)是中国人民大学朝乐门老师于 2017 年首次公开提出的教学改革新思想,目前已受到国内外同行的广泛关注。开源课程是继大型开放式网络课程(Massive Open Online Courses, MOOC)之后的一种新的教学改革思想。MOOC 的出现解决了学生关心的问题,即学生之间的公平性,任何学生都可以学习全球最好的教师开设的课程,不论其是在哪一所学校上学。与 MOOC 同等重要的是,OSC 要解决的是教师关心的问题,即教师之间的公平性,任何教师都可以在其所属领域最优秀的教师开源的全套教学资料的基础上开发出自己的特色课程,并将自己的特色工作合并至开源课程中,与开设同类课程的教师们共同维护

一门精品课程,不管该教师是在哪一所学校任教。

本文基于作者在主持全国高校大数据联盟示范开源课程、国家精品在线开放课程数据科学导论、中国人民大学一流本科课程第一批建设课程数据科学导论、北京市教育委员会优质教材《数据科学理论与实践》等过程中总结的建设经验,重点探讨现阶段传统课程建设中存在的挑战、开源课程的定义与特征、开源课程的建设方案以及数据科学导论的开源工作,对教学改革与课程资源建设均具有一定的借鉴价值。

2 传统课程建设中存在的问题

建设“金课”是我国现阶段教学改革的主要任务之一,也是破解传统教育难题的重要途径。近年来,我国课程建设和教学改革取得了长足的进步,但仍面临如下挑战。

到稿日期:2020-09-03 返修日期:2020-11-05

基金项目:国家自然科学基金项目(72074214)

This work was supported by the National Natural Science Foundation of China(72074214).

通信作者:朝乐门(chaolemen@ruc.edu.cn)

(1)教师备课工作的困境。找素材、输入文字、排版等低级重复性劳动耗费了老师们的绝大部分备课精力;技术和环境变化太快,每次都要花费大量的时间来更新课程内容;当有疑惑时,不知道向谁请教,就算知道应该找“他”,但也请不动“他”;当完成了很好的探索和自己的原创性工作时,没有机会向同行展示;备完一门课后,不知道是否还有缺陷,没有人协助改进;花费几年的时间备好一门课,结果上一两次就被叫停,教师的付出无法被有效再利用。

(2)教育管理工作的痛点。教育管理部门尚未解决以下问题:目前,精品课程和优质课程资源的受益者是广大学生群体和极少数来自课程团队的教师,而广大教师难以从中受益;除了精品课程或优质课程的建设团队,其他教师只能看到这些示范课程的“可执行程序级别的最终结果”,根本拿不到真正有价值的东西,如源代码级的可编辑备课材料;教育管理部门关心的是学生,强调如何提升课堂教学质量,但是相对忽略了另一个更深层次的问题,即缺乏一种确保绝大多数教师的备课质量的保障机制,“作坊式备课”导致课程质量参差不齐。

(3)MOOC 课程的局限性。自 2012 年以来,MOOC 的建

设取得了较好的成绩。研究表明,MOOC 将为现有教育系统中的特定领域提供新的支持,这主要是为受过教育的学习者提供支持^[1]。MOOC 的出现解决了学生学习活动的公平性问题,但并未解决教师备课活动的公平性问题。

因此,“金课”建设需要探索一种全新的教学改革,从而为上述教师备课工作的困境、教育管理工作的痛点以及 MOOC 的局限性提供统一的解决方案。本文中开源课程行动倡议的提出为上述问题提供了较好的解决方案。

3 开源课程及其优势

3.1 开源行动

开源行动(Open Source Initiative)起源于开源软件(Open Source Software)。开源软件指按照特定许可协议公开了自己源代码的软件。通常,这种许可协议的版权所有者授予用户向任何人出于任何目的学习、更改和分发该软件的权利^[2]。表 1 列出了开源软件、自由软件和免费软件的区别。相比自由软件和免费软件,开源软件更加强调的是软件开发和维护模式的创新,即以开源社区模式学习、开发和维护软件系统。

表 1 开源软件、自由软件和免费软件的区别
Table 1 Comparison of Open-source Software,Free Software and Freeware

	开源软件	自由软件	免费软件
英文名称	Open-source Software	Free Software	Freeware
实例	FireFox,mySQL,Apache	Linux,ubuntu	Flash Player,Foxmail,Skype Acrobat Reader
代表人物	Eric Raymond	Richard Matthew Stallman	Andrew Fluegelman
主要推动者	开源软件基金会(Open Software Foundation,OSF)	自由软件基金会(Free Software Foundation,FSF)	无
主要特点	属于软件开发模式的创新,以开源社区模式学习、开发和维护软件系统	属于自由主义在软件领域的应用,强调软件的开发、使用、修改、发行和学习活动的“自由”	属于营销模式的创新,通过提供免费软件,获得更多用户的关注并为其提供收费的增强功能或高级服务
基本原则	源定义(The Open Source Definition)中的 10 条原则	自由软件的“四项自由原则”	无统一原则,软件开发可以自己制定原则
许可证	Apache,MIT,BSD 等	GNU 的 GPL/LGPL/FDL 等	CC BY-NC-SA 等
是否受版权保护	是	是	是
是否免费提供	不一定/不要求	不一定/不要求	是
是否提供源代码	是	是	不要求

通常,开源软件应遵循开源促进会(Open Source Initiative,OSI)给出的开源定义(The Open Source Definition)中的 10 项基本原则:免费再发行、提供源代码、衍生软件、作者源代码的完整性、个体或群体的无歧视、工作领域的无歧视、许可证的发行、许可证不得专属于特定产品、许可不可限制其他软件、许可证必须在技术上中立^[3]。许可证协议是开源软件遵循上述基本原则的主要保障,目前常见的开源软件许可证有 Apache 许可证(The Apache Licenses)、MIT 许可证(The MIT Licenses)、BSD 许可证(The BSD Licenses)等。

随着开源软件的成功,开源软件的思想已经深入至硬件设备研发^[4]、科学研究^[5]、情报研究^[6]和农业服务^[7-8]等多个应用领域,并逐渐发展成为一种新的革命。目前,开源行动的发展呈现出从传统软件产品向智慧产品转移的趋势,其关注点也从“代码的开源”进一步深入至“思想的开源”。

3.2 开源课程的定义

开源课程(Open-source Course,OSC)是开源(Open-source)思想指导下的课程建设模式,属于开源行动的一个新应用领域。具体而言,开源课程指按照既定的开源课程的许可协议,以可编辑的原始格式公开全套课程资源,并授予任何参与者对所公开资源进行学习、教学、更改和再发布的权力。图 1 通过发起者、维护者和使用者 3 个重要角色揭示了开源课程的内涵。

(1)发起者。将原始课程提交至开源课程社区平台,作为该课程的“主干课程(版本号为 x=1.0)”,并与维护者和使用者一起共同维护一门课程。在此阶段,“主干课程”的课程资源作者为发起者一个人或一个团队。

(2)维护者。根据发起者提供的“主干课程(版本号为 x)”建立自己的“分支课程”,进一步开发出自己的“特色课

程”,并将其合并到开源平台上的“主干课程”,“主干课程”的版本升级为 $x+1$ 。随着更多维护者的参与或维护次数的增多,“主干课程”的版本将不断升级。同时,“主干课程”的课程资源的作者队伍也会不断壮大。

(3)使用者。可以得到免费、可编辑、可溯源的最新课程。最后,课程资源的作者可能覆盖到开设相关课程的绝大部分教师。

“开源课程”建设中的发起者、维护者和使用者是 3 个不同角色的名称,因此,同一个人(或团队)可以同时担任其中 1 个或者 2 个,甚至 3 个角色。此外,开源课程与课程资源的共享(Source Available Courses)是两个不同的概念。前者代表的是一种课程资源的可持续共建模式,突破了后者所代表的课程资源简单共享的范畴。

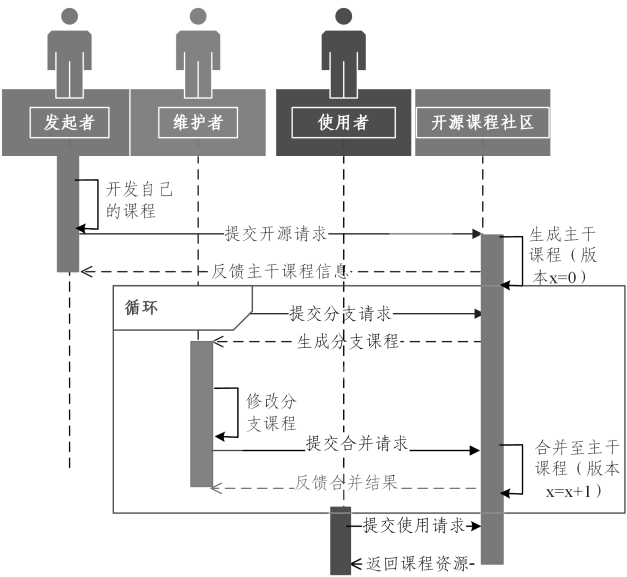


图1 开源课程中的 3 种角色
Fig. 1 3 roles in Open-source Courses

3.3 开源课程的优势

开源课程是一项无歧视的行动。在开源课程社区中,每一位教师都是平等的,不考虑其是否来自名校,也不考虑其是否为刚入职的教师。对于教师而言,参与开源课程的主要益处包括:

- (1)实现教师之间的机会平等。在开源课程中,非名校教师、非资深教师也可以为课程资源的建设做出重要贡献,可以突破传统教育中课程资源建设机会和话语权过度依赖于教师所在学校的知名度而不是教师本人的努力程度的局限性。
- (2)结识开设同门或同类课程的教师。开源课程平台具有开源社区的属性,开源课程的参与者可以结识全国乃至全球范围内开设同门或同类课程的教师,并与他们共同探讨课程建设过程中的心得体会,交换课程教学经验,形成多个院校共同准备同一门课程的机制。
- (3)展示自己的教学成果和经验。在教学过程中,任课教师通常会有一些自己独有的经验和体会,但未达到专业探讨和学术发表的高度,因此被目前的学术出版体系排除在外,难以与同行分享。在开源课程建设机制中,任课教师可以通过

将自己的教学成果和经验置入“分支课程”,合并至“主干课程”的方式来分享给同行。

(4)获得来自同行教师的解答和教学资源。当开源课程的参与者有疑问或缺少教学资源时,可以通过开源课程平台向同行发出请求,很容易得到来自同行的解答和资源推送。

(5)发现自己的教学资源中存在的缺陷。通过将自己的课程资源提交或合并至开源平台,开源课程资源的提供者和维护者可借助更多同行的力量找出自己课程中存在的缺陷,从而改变传统课程建设的盲区。

(6)长久留存和动态更新自己的课程资源。开源课程的参与者将自己的课程提交或合并至开源平台后,通过利用开源平台的版本回溯和动态优化机制,不仅能够确保自己的课程资源不会丢失,还可以实现课程资源的动态更新及方便回溯。

4 开源课程的建设

4.1 开源流程

开源课程的流程如图 2 所示,其中开源课程平台分为服务器端和本地终端。课程资源分为主干课程(C_m)、分支课程(C_b)和本地课程(C_l) 3 种类型,更新内容可进一步分为 4 种,即已识别的更新内容(F_i)、准备提交的更新内容(F_s)、已申请的更新内容(F_a)和已审核的更新内容(F_p)。开源课程的维护和建设的基本流程涉及以下 8 个活动:

- (1)分支。开源课程的维护者(或使用者)对开源课程平台服务器上的主干课程 C_m 进行分支操作,进而生成自己的分支课程 C_b ,并将其放在自己的账户下。分支课程 C_b 存放在开源社区服务器上。
- (2)下载。开源课程的维护者(或使用者)将自己账户下的分支课程 C_b 下载到本地计算机,生成本地课程 C_l 。
- (3)修改。开源课程的维护者根据自己的实际需求对本地课程 C_l 进行增/删/改操作,并将本地课程 C_l 存放在本地计算机中。
- (4)识别。开源课程平台的本地终端将自动识别本地课程 C_l 中发生的变更,包括内容、格式和元数据的变化,并显示给课程维护者。
- (5)选择。课程维护者基于本地终端给出的更新情况详细清单,选择准备提交的更新内容 F_s 。课程维护者提交的更新内容是有选择地自愿提供,并不强制要求提交本地课程的全部更新内容。
- (6)提交。课程维护者将自己准备提交的更新内容 F_s 提交至开源课程服务器端,并生成已申请的更新内容 F_a 。已申请的更新内容 F_a 将自动合并至该用户账户下的分支课程 C_b 。
- (7)确认。由课程发起者及其授权用户对课程维护者提交的已申请的更新内容 F_a 进行审核,并生成已申请的更新内容 F_p 。与分支课程的合并不同的是,主干课程的合并需要由课程发起者及其授权用户进行审核。
- (8)合并。开源课程服务器端将已申请的更新内容 F_p 自动合并至分支课程 C_b 。

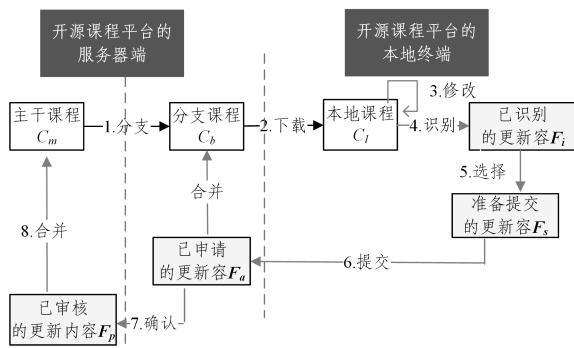


图 2 开源课程的流程

Fig. 2 Process of Open-source Courses

4.2 开源原则

开源课程的建设必须遵循以下 9 项基本原则:

(1)统一许可。开源课程的建设需要在统一许可协议框架下进行,协议内容包括必选项和可选项。其中,可选项可以根据不同课程或课程群的特点进行个性化定制。许可协议中的可选项主要用于开源课程的修改方式和署名方法等。

(2)全套资源。课程教学资源是相互关联和映射的一个完整系统。因此,开源课程的发起者需要提供一门课程的全套资源,而不能仅限于课程 PPT 课件的共享。

(3)自由使用。在遵循开源许可的前提下,开源课程的参与者,包括发起者、使用者和维护者,均享有对课程资源的学习、教学、使用、修改和再发布的自由。

(4)原始文件。开源课程的发起者和维护者需要提供可编辑的原始格式文件,并注明必要的编辑方法说明文档。开源课程中不宜提供 PDF 等难以编辑的文件格式。

(5)平等对待。开源社区平等对待所有参与者。对教师的所在单位、工作年限、社会知名度等不能存在歧视。

(6)社区支持。开源课程的建设需要提供社区支持,开源社区是开源课程的运行平台,用于支持课程开源和协作交流。

(7)特色课程。开源课程应鼓励课程的特色建设,开源课程的维护者可以结合自己的需求和特长开发出分支课程,并将分支课程进一步发展和演化成多个方向的基础课程或多门特色课程。

(8)版本回溯。开源课程平台需要有版本回溯能力,支持开源课程的参与者查看和回溯某课程的历史版本及其演化全过程。

(9)知识产权保护。开源课程的建设过程中必须保护参与者的知识产权。通常,除非授权方宣布放弃版权或开源协议有特殊规定外,开源课程的维护过程一般不改变版权性质及所属关系。

4.3 关键技术

开源课程的建设涉及以下 5 种关键技术:

(1)开源课程的社区平台。开源社区平台是开源课程建设的基础设施。目前, GitHub, Sourceforge, GitLab, Launchpad 等开源社区在软件开发领域得到了广泛应用,但在课程资源的开源应用中效果并不理想,主要原因在于课程资源比

软件源代码更为复杂。因此,开源课程的建设需要面向开源课程的专用社区平台。

(2)课程资源的版本管理。在开源课程的建设中,尤其是课程资源分布式并行编辑时,同一门课程会产生多个不同的版本,而如何识别版本之间的差异性是一个亟待解决的问题。此外,当某一版本的更新存在争议时,如何回滚到前一版本也是需要解决的问题。

(3)更新工作的自动审核。开源课程中对修改内容,尤其是已申请更新内容的审核需要采取自动评估技术,或至少采取人机合作的审核机制。对于格式改动和局部文字修改等更新工作,开源课程应采取全自动评估与审核技术。

(4)离线数字版权保护技术。在开源课程的建设过程中应采取数字版权保护技术,尤其是当教学资源离开开源社区时,也应带有其数字版权信息,进而保护教学资源提供者的知识产权。

(5)课程资源的一致性保障技术。在开源课程建设过程中容易出现课程资源的不一致性,例如同一个课程资源被两个或更多的参与者同时离线修改后产生的多个分支课程资源副本之间可能出现重叠甚至冲突。因此,在将分支课程合并至主干课程时,需要有数据一致性保障技术。

5 数据科学导论的开源

数据科学导论是数据科学专业的核心课程^[9],已受到国内外高校的高度重视^[10],开设其前首先需要解决课程资源的建设问题,如如何准备教学大纲、教案、PPT、讲义、作业、考题等。针对该问题可以采取两种方法,一种是每位教师可以自己闭门造车,从零做起,所有的工作都自己做;另一种做法是采取众包和开源模式,正如朝乐门老师提出的开源课程行动。开源课程行动不等同于传统的课程资源的分享行动,而是将开源(Open Source)思想引入课程资源建设工作,以实现课程资源的不断创新和可持续发展,将教师们从繁重琐碎的基础性备课工作中解放出来,将精力聚焦在课程特色的建设和个性化教学中。

(1)准备工作与正式启动。作为开源课程领域的第一个示范课程,数据科学导论的开源筹备工作始于 2017 年初,并于 2017 年 11 月 18 在微信公众号“数据科学 Data Science”上正式公布。2017 年 12 月 23 日,在全国高校大数据教育联盟“数据科学与大数据技术专业核心课程建设系列研讨会”上正式成立开源课程组,并举行了启动仪式。

(2)开源平台。目前,该课程的资源开源在 GitHub 服务器上,并配有邮件讨论组,支持教师订阅、发帖和退订该课程。此外,本课程受到了全国高校大数据教育联盟的支持。

(3)开源内容。数据科学导论课程的开源内容包括:教学大纲、新开课申请表及课程简介、课程讲义 PPT、课程编程源代码(R 版)、课程编程源代码(Python 版)、原始数据、作业题、教师参考书目、数据科学相关重要资源目录和支撑平台,以及作者联系方式。

(4)开源效果。自 2017 年 11 月正式公布数据科学导论

课程的开源以来,其受到了广大老师,尤其是数据科学导论和大数据导论类课程的任课教师的密切关注。目前,有超过200所高校的教师申请参与数据科学导论课程的开源建设,产生了良好的社会影响,并已成为我国数据科学导论课程建设领域的一个亮点。该课程于2019年被评为国家精品在线开放课程以及中国人民大学一流本科课程第一批建设课程。此外,课程采用主讲人自编的教材《数据科学理论与实践》(清华大学出版社),该书第二版于2019年被北京市教委评为北京市高校优质本科教材^[11]。

(5)经验与不足。目前,数据科学导论课程的开源建设过程中主要存在的问题是,虽然关注的人多,但多数教师习惯于直接下载后开发出自己的课程的利用模式,并未将自己修改后的课程再次上传至开源课程平台。经调查发现,造成此现象的原因有:1)教师虽然在关注或者手备课,但尚未开展或完成数据科学导论课程的教学工作;2)部分教师认为教学资源是自己的智力资本,不愿意公开和共享;3)部分教师认为目前相关政策、制度和技术配套还不够完善,仍处于观望状态。

结束语 从开源课程行动倡议和数据科学导论的开源工作经验来看,开源课程的建设亟需政策和制度保障。结合开源课程自身的规律及我国高等教育发展的实际情况,建议我国教育管理部门优先采取如下对策:

(1)鼓励开源课程的科学研究。开源课程是一种全新的教学改革,具有其独有的特征、规律、问题以及挑战。因此,建议教育部、国家自然科学基金委和全国哲学社会科学规划办公室等有关部门通过专项项目、重大项目招标和面向项目,鼓励科学家和学者们进行有针对性的系统研究,揭示开源课程中的科学问题,如开源课程的核心特征、主要流程、基本原则、平等编辑、社区支持、版本回溯、版权保护、关键技术及保障机制。

(2)制定开源课程的知识协议及行动规范。知识产权和行动规范是落实开源课程的两大突破口和重要抓手。对参与开源课程的教师而言,他们最关注的是自己的知识产权是否得到保护以及其他参与者的行动是否符合规范。为此,建议教育部牵头制定开源课程的知识协议和行动规范。知识协议指关于开源课程中教师的知识产权保护与激励、教育管理部门的监管职责的定义与落实以及知识对象的封装与关联的约定;行动规范指参与开源课程的教师个体、高等学校和教育管理部门的行动指南及基本要求。

(3)提出开源课程倡议及建设一批示范性课程。作为新的教育改革与探索,开源课程的推广和发展亟待一批示范性课程或课程群的出现。因此,建议教育部在提出开源课程倡议的基础上,以新专业和新课程为试点,培育一批示范性课程。示范性课程的建设,不仅能够为开源课程的推广提供范例,而且可以总结经验与教训,进而优化开源课程的教育改革。

(4)推动开源课程的政策支持与鼓励机制。开源课程的普及发展关系着广大教师的切身利益,需要他们的积极参与。为此,教育管理部门需要给予一定的政策支持和建立鼓励机

制,尤其是将开源课程的建设纳入工作量的认定范围,并针对开源课程领域设立优胜劣汰的机制,对开源课程领域的先进人物和示范性课程给予一定的奖励。

总之,以开源方式建设“金课”是一项富有借鉴意义的创新举措。OSC的提出是继MOOC之后的一项全新的教学改革设想,是“金课”区别于“水课”的重要属性,不仅可以保证“金课”建设的质量和速度,而且可以通过“金课”建设来提升教师之间的知识转移与素质提升。开源课程建设还可以为全球教育改革领域做出中国贡献,有助于提升我国在国际教育界的话语权。

参 考 文 献

- [1] REICH J, RUIPÉREZ V J A. The MOOC pivot[J]. Science, 2019, 363(6423): 130-131.
- [2] LAURENT A M S. Understanding open source and free software licensing: guide to navigating licensing issues in existing & new software[M]. O'Reilly Media, Inc, 2004: 4.
- [3] PERENS B. The open source definition [J]. Open Sources: Voices from the Open Source Revolution, 1999, 1: 171-188.
- [4] AIDUKAS T, ECKERT R, HARVEY A R, et al. Low-cost, sub-micron resolution, wide-field computational microscopy using opensource hardware[J]. Scientific Reports, 2019, 9(1): 1-12.
- [5] DWIPRIYOKO E, BON A T B, SUKONO F. Enterprise Architecture Planning as New Generation Cooperatives Research Methods[C]// Journal of Physics: Conference Series. IOP Publishing, 2019: 012094.
- [6] BAZZELL M. Open source intelligence techniques: resources for searching and analyzing online information[M]. CreateSpace Independent Publishing Platform, 2016.
- [7] LOUWAARS N P. Open source seed, a revolution in breeding or yet another attack on the breeder's exemption? [J]. Frontiers in Plant Science, 2019, 10: 1127.
- [8] GIOTITSAS C. Open Source Agriculture: Grassroots Technology in the Digital Era[M]. Springer Nature, 2019.
- [9] CHAO L M, XING C X, WANG Y Q. Unique Curriculums for Data Science and Big Data Technology[J]. Computer Science, 2018, 45(3): 3-10.
- [10] CHAO L M, YANG C J, WANG S J, et al. Data Science Curriculums Around the World: An Empirical Study[J]. Data Analysis and Knowledge Discovery, 2017, 1(6): 12-21.
- [11] CHAO L M. Course Design and Redesign for Introduction to Data Science[J]. Computer Science, 2020, 47(7): 1-7.



CHAO Le-men, born in 1979, Ph.D., associate professor, is a member of China Computer Federation. His main research interests include data science, big data analytics, and knowledge processing on semantic Web.