

挖掘教学大纲的宝藏

开放课程计划利用数据共享改善及回馈教学

尽管收集可获得的在线学术材料的想法日益增多,如从文章和数据集到会议报告和讲座,但有一种学术文献却几乎未得到审核。那就是教学大纲:列出阅读材料、话题以及大学课程期望值的文件。

至少,在今年1月,当美国纽约市哥伦比亚大学的数据科学家、社会学家和数字人文研究人员发起一个叫做“开放课程资源管理器”的工具之前,情况是这样的。它融合了超过100万种公众可获得的课程大纲,并通过一种方便搜索的形式列出了它们的数据。另一个包含至少3倍于当前教学大纲的版本将于2017年1月启动。

这个名为“开放课程计划”(OSP)的工具背后的团队希望促进大学课程更加公开化。他们认为,这样做有助于教科书作者、教务工作人员和课程开发者,并对有效教学材料设计作出反馈,而这很大程度上在常规学术工作评估中被忽视。

“教学大纲是由学者编写的应该被共享的最重要的材料,然而它们却未被广泛共享。”马萨诸塞州剑桥市哈佛大学学术交流办公室哈佛开放获取计划主任 Peter Suber 说,他也在 OSP 顾问委员会工作。“它们反映了关于什么值得一教的严肃学术判断。”

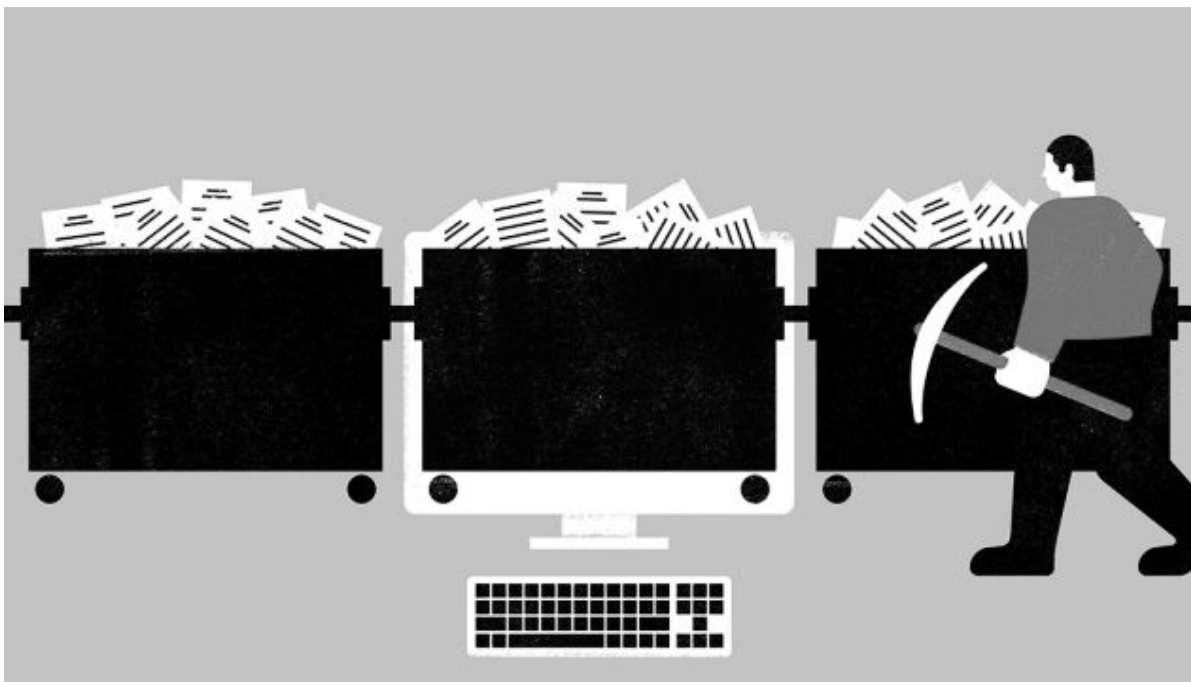
这样的判断对于教科书作者来说是受欢迎的消息。加州大学伯克利分校计算机学家 Stuart Russell 直到《自然》杂志就这篇文章采访时才意识到这一点,他和共同作者 Peter Norvig 在1995年编写的《人工智能》一书是计算机科学领域被指定阅读最多的文献。“我非常出乎意料。”他说。

除了点燃职业自豪感之外,这样的信息有助巩固身职位及促进晋升。编写一本教科书,无论它多么有用或是信息含量多么高,但在学术文章中的引用率总体不高,因此其学术影响力很可能并不高。OSP 则有助改变平衡。“我正处这样一个时期,我认为教员应该对其学时、能力和影响力的整体记录有更多知情权。”麻省理工学院出版社主任 Amy Brand 说。她表示,教学大纲使用的硬数据能够让教职员工“就他们的工作在世界上究竟有什么作用讲述自己的故事”。

目前,开放课程资源管理器可查询到上至2000年的100多万门课程,相互参照2000多万篇文章,从而制作出一篇文章被教授频率的数据。使用者可以通过作者、题目、机构和学科等目录搜索这些数据。这个工作还报告了哪本教科书使用率有多普遍,并对每篇文章的教学频率进行了排序。

一个升级版的工具将于2017年1月21日在该资源管理器的首个周年纪念日开放,它将会包括300万个教学大纲,相互参考文献达1.5亿篇文章。它们将包括来自预印本服务器 arXiv、跨库检索 CrossRef 以及虚拟国际权威文件(VIAF,将相同的书目记录与不同国家图书馆目录相连)等处的标题。此次升级将包括新的搜索选择,如通过日期或单位进行搜索的能力,OSP 项目主任 Joe Karaganis 说。新的版本将融合加拿大和英国更好的数据和信息,如果作者最终同意复制它们,将会

“现在所用的技术还很粗糙,但它们都是可以改善的,数据科学正在变得越来越好。”



The Project Twins 制图

形成全文本的教学大纲。 “我们很有雄心。”Karaganis 说,“现在所用的技术还很粗糙,但它们都是可以改善的,数据科学正在变得越来越好。”

“约”取引用率

OSP 位于哥伦比亚大学公共政策研究所的“美国集会”,由艾尔弗·斯隆基金和阿卡迪亚基金资助。它受到一个叫做教学大纲发现者的搜索引擎的启发,该引擎对2002年到2009年间的教学大纲公共网络进行了分析。该工具由弗吉尼亚州乔治·梅森大学历史学家 Dan Cohen 开创,现在 Cohen 已成为美国数字公共图书馆执行主任。该引擎聚集了 Cohen 所说的当时最多的教学大纲,包含约100万份资料。他在2011年将其作为数据库公布了网站链接。

与 OSP 不同,Cohen 的工具提供了与每个大纲全文本相连的工具,但它仅包括到2009年的课程,在因改为谷歌程序界面而使该工具下线之后,Cohen 的很多同事以及作为幼儿教育专家的妻子都为此感到烦恼。“我仍然会收到邮件让我继续维持教学大纲发现者。”他说。

当 OSP 在2014年启动时,该团队建立了挖掘公共互联网的工具,包括 Cohen 使用的链接(Cohen 因编码错误丢失了一部分数据)。但像 Cohen 一样,Karaganis 推算,该团队也受限可公开获取的教学大纲:粗略估计仅有美国8000万到1.2亿教学大纲中的600万份。教学大纲被封在私人课堂管理软件的墙壁背后而难以接触到。“例如,哥伦比亚大学在过去12-13年间的教学大纲数量为8万份左右。”Karaganis 说,“但一所大型州立学校的教学大纲则是它的两到3倍。”

OSP 团队接下来需要建造工具提取那些大纲的内容。例如,引用率并没有一致的结构,该

项目技术主任 David McClure 说。该工具通过在一个拥有2000万个标题(1100万来自哈佛大学图书馆云以及900万来自JSTOR)的数据库中交叉引用每个教学大纲来搜寻所需要的标题。将文章标题和作者匹配可算作一次引用。“我们对允许模糊性设计了不同的技术,比如把‘由’一词放在作者和题目中间。”McClure 说。

一个新标准

OSP 提取出的数据可通过一种叫作教学分数的单一指标进行排序,它可以表明一篇文章在教学大纲中被指定阅读的次数的多寡。其数值可以从1(很少用于教学)到100(经常用于教学)。

据 Suber 介绍,教学分数为常规学术影响力标准提供了一种替代指标。它们反映了日益兴起的“替代指标”思潮,它旨在对一个人的研究产出进行量化。“我认为教学分数可以参与新的替代指标行动中,在文本影响力方面提供更加敏锐的衡量方法。”他说。

现在,一些研究人员和高校已经在利用这些数据。莱克星顿肯塔基大学发现该校教师 Edward Morris 一篇文章的引用率在13225篇社科类文章中排名第46位之后,举办了新闻发布会。现在,这篇文章在53177篇文章中排名第371位,Morris 计划利用这些数据支撑自己晋升正教授。

美国高校并非唯一引起注意的群体。目前 OSP 每天约1000名访客中大部分来自美国,Karaganis 说,但来自乌克兰、俄罗斯和埃及的访客也占相当数量。

其他研究人员也在利用这些数据编纂一些广泛用于教学的文章的清单,例如通过漫画小说和连环画形式,或者对由女性写作的经常被用于教学的社会学文章进行量化。康涅狄格州纽黑文耶鲁大学博士后 Melanie Martin 利用教

学大纲发现者了解她所在的进化人类学领域教学频率最高的文章,但因为没办法搜索次一级学科的相关数据,例如在次一级学科神经科学或基因组学方面的生物学搜索结果的局限性,她不得不手动浏览1.6万篇人类学题目。“没有更好的过滤器,我觉得这是它的局限。”她说。

夯实专业技能

OSP 数据的另一个潜在用途是课程设计。通过让教职员尤其是青年教学人员在同行的知识基础上开展工作,OSP 有助于使他们更具创造性地投入教学,如发现呈现教学材料的新方法。“它对提高教学质量大有裨益。”Russell 说。它还有助于提高效率,使教职员有更多时间做科研和指导学生。

然而,重要的是不能过度解释这些数据,伊利诺伊大学香槟分校信息文献专家 Lisa Janicke Hinchliffe 说。该项目的样本集可能并非所有教学大纲的最佳代表,尤其对一个具体机构而言。例如,根据资源管理器,哈佛大学位居第二位的文章是由马丁·路德·金撰写的《伯明翰狱中来信》。但 OSP 的80%哈佛教学大纲来自肯尼迪政府学院,Karaganis 说(尽管 OSP 并不总是公开列出其详细资料来源)。因此,很难总结这篇文章在哈佛究竟有多受欢迎。

对于 Hinchliffe 来说,OSP 的价值在于揭示教职员使用资源的广度。“我并不需要明确‘这些是教学中使用率最高的前6本书’。”她说,“我希望看到的是它的丰富程度。”

这样的信息对于简化费时费力的课程设计来说仍有相当长的路要走。Suber 已经教授哲学21年,他对此深有体会。“无论什么时候,我得知要开新的课程,那么我至少需要为它准备1年。”他说,“写40个讲座的教学大纲是个巨大的工程,它比写一本书更难。”而 OSP 的数据将有助减轻这一负担。(晋楠编译)

基因“黑客”要种定制庄稼

基因编辑技术发展使其成为可能

当来自全球的作物工程师日前聚集在英国伦敦时,他们的研究目标颇为宏大:培育更高效利用水分的水稻、需要更少肥料的谷物以及由增强光合作用提供动力的超高产木薯。作物工程联盟研讨会的150名与会者带来了各种想法以及分子工具。多亏了合成生物学和自动化技术的发展,若干项目已拥有1000多个经过改造的基因和其他分子工具,并且准备在研究人员选择的作物中接受测试。不过,这正是他们经常碰壁的地方。用于培育拥有定制基因组(这一过程被称为遗传转化)植物的过时方法烦琐、不可靠且耗费时间。

当被问及该领域存在哪些障碍,诺福克约翰伊恩斯中心植物发育生物学家 Giles Oldroyd 有一个现成的答案:“最大的问题是改善植物遗传转化。”

“我们所有人都面临着输送问题。”美国明尼苏达大学植物生物学家 Dan Voytas 表示,“我们拥有强大的试剂,但你如何将它们输送到细胞中?”

引发广泛讨论的是持续了几十年的难题:很难修改植物基因组然后利用一些转化细胞重新产生全新的作物。诸如 CRISPR-Cas9 等基因编辑技术让人们看到了复杂作物工程的希望,而这曾经是不可思议的事情。不过,当研究人员遇到原有的障碍时,现实让他们变得更加沮丧。

美国国家科学基金会(NSF)已意识到这一挫折,并在9月28日宣布将资助针对更好转化方法开展的研究。这是一个新的植物基因组研究项目关注的4个焦点中的一个。整个项目将收到共计1500万美元的资助。



通过利用更好的基因编辑技术,农作物产量和用水效率都能得以改善。图片来源:AFP/Getty

“每个人都会同意,这真的是基因组工程的瓶颈。”去年11月,NSF 关于植物转化的研讨会共同组织者、田纳西大学植物生物学家 Neal Stewart 表示。

利用可将基因添加到植物基因组中的细菌,诸如植物界的“小白鼠”——拟南芥等一些植物很容易被转化。研究人员将想要测试的基

因插入农杆菌中,然后“诱骗”细菌感染植物的生殖细胞。当作物随后产生后代时,其中一些会表达新的基因。

不过,这对很多作物不起作用,而且利用农杆菌会引发诸如美国农业部等政府机构的额外审查,因为它被视为一种植物害虫。作为替代方案,研究人员可利用“基因枪”向植物细胞中发

射被黄金珠包裹的DNA。随后,这些细胞“沐浴”在生长激素中并被“诱骗”重新产生完整的植株。诸如玉米等一些植物很容易屈从于这种方法,小麦、高粱等其他作物则不行。

对于一些“顽固”的作物来说,重新产生全新的植株,需要持续数月艰苦的细胞培养工作(优化生长条件和激素浓度)。成功需要的条件不仅依据每种作物而不同,在相同的植物物种之间也不一样。

来自康奈尔大学的 Joyce van Eck 是该领域的专家。在她看来,植物转化专家凤毛麟角。“我们所做的事情包含很多艺术。”她在伦敦研讨会上表示,“拥有这种训练的专家很难找到。”再加上对新方法的资助匮乏,研究人员不得不依靠了几十年的旧技术。

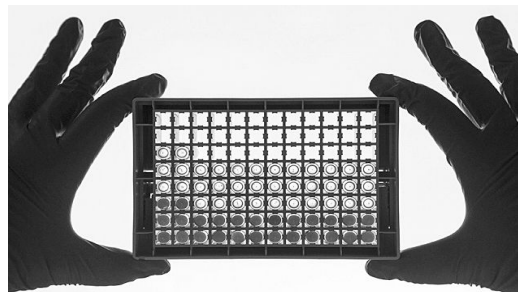
不过,随着对替代方法的寻找正在升温,这一切可能会发生改变。Stewart 和合作者已研制出和用手工相比能更加快速、准确地执行原生质体转化技术的机器人。该方法利用酶消化细胞壁,从而使研究人员更容易引入新基因。不过,重新产生完整植株的问题仍然存在。研究人员利用一种没有机器人的类似方法,在包括生菜和水稻在内的多种植物中进行了 CRISPR-Cas9 基因编辑。

细胞培养步骤仍然困难。Stewart 介绍说,他的实验室中有人用了两年时间转化一种被用于生物燃料研究的秆草,但最终还是以失败告终。不过,酶的成本不断下降使研究人员得以开展更多试验,机器人学则能提高生产能力。Stewart 是如此迷恋他的造物,以至于为其创作了一首歌曲。“现在,它是我们的孩子。”Stewart 说。(宗华编译)

科学线人

全球科技政策新闻与解析

调查发现近一半临床试验未公布结果



关于报告临床试验结果的新规定应当会为整个过程带来更多透明度。图片来源:Dan Kitwood/Getty

一个自动工具搜罗了关于全球重要临床试验数据库的上千条记录,以揭示哪些制药公司和学术机构未将其试验结果公布。

这种失职行为已经有据可查:多项研究不同程度地报告称,25%~50%的临床试验结果在试验完成数年后仍未公开。今年9月,美国健康和人类服务部宣布了更加严厉的法规,迫使其资助的研究人员公开临床试验设计和结果。

一篇日前在线发表于 F1000Research 且描述了上述工具的论文通讯作者、英国牛津大学临床研究访问学者 Ben Goldacre 表示,软件使得更加复杂的搜索成为可能。

过程自动化还意味着结果可被定期更新,而这会对未报告结果的研究发起人带来持续的压力,并迫使其采取措施提高成绩。“如果任何人想提高他们的成绩或改善排名,其要做的所有事情是公布他们的结果。”Goldacre 表示。

Goldacre 和在牛津大学工作的论文共同作者 Anna Powell-Smith 开发了上述工具,目的是搜索 ClinicalTrials.gov 数据库以寻找至少在两年前完成的试验。随后,计算机机会尝试将这些试验同该数据库或研究知识库——PubMed 中已发表的结果进行匹配。

在被评估的近2.6万项试验中,有45.2%未将结果发表。该团队还创建了一个网站,让用户能按照谁在公布结果方面做得最好或最差的顺序查看临床试验发起人。列表包括了来自全球各地的学术界和业界发起者。

美国纽约大学朗格尼医学中心医学伦理学家 Jennifer Miller 表示,对于调查临床试验透明性的研究来说,自动化而非手动分析正日益成为标准做法。

不过,Goldacre 承认,搜索自动化会导致精度上的牺牲。比如,搜索可能会错过已发表的结果,如果它们未被标上 ClinicalTrials.gov 数据库分配的号码,或者发表它们的期刊不在 PubMed 的列表中。

尽管 Goldacre 表示其团队确实发现在单个研究如何被计分上存有一些差异,但他的搜索工具产生的总体趋势和由此前针对小型数据库进行的手动调查发布的结果类似。他同时希望,定期更新结果的能力将刺激试验发起者改善成绩。(宗华)

美国国家心理卫生研究所所长 解读优先关注领域



Joshua Gordon
图片来源:NIMH

当 Joshua Gordon 还是美国加州大学旧金山分校一年级医学博士生并且下决心在癌症生物学领域有一番作为时,他听到了一场描述用电刺激猴子大脑一个关键区域的细胞将如何改变动物对圆点在电脑屏幕上朝哪个方向移动的认知的演讲。这种操作的威力及其显示出的探索大脑的巨大可能性令 Gordon 着迷。他将博士期间的关注点转移到了神经科学。

25年后,49岁的 Gordon 已成为拥有15.5亿美元经费的美国国家心理卫生研究所(NIMH)所长。在哥伦比亚大学呆了19年后,他在不久前就任 NIMH 所长。

在被《科学》杂志问及上任后优先考虑的事情是什么时,Gordon 表示,神经回路和计算精神病学是自己首要关注的两个领域。他认为,将更多的计算上的形式体系带入基础和临床等各个层面的精神病学研究,甚至是实践研究,将有助于人们把工作做得更好。

与此同时,Gordon 表示,自杀是一个巨大的临床问题。该问题涉及各种交叉的诊断方法,并且拥有至少短期的循证治疗和预防。

关于如何以及为何将神经回路研究工作列为 NIMH 优先关注的领域,Gordon 解释说,技术上的进步带来了日益强大的工具,使神经科学家得以测量并影响特定神经回路中的神经活动。比如,利用这些工具,人们可以减少小鼠模型中类似于抑郁症的行为,改变这些工具并将其用于人类是一件非常诱人的事情。

他同时表示,人们尚不确定这些技术能否在治疗中派上用场。不过,除非从现在开始奠定基础,否则人们将无法知晓答案。“我们可以从发明必需的工具和方法开始。”Gordon 表示,“从某种程度上说,‘通过推动创新型神经技术开展大脑研究计划’涉及的便是这个领域。”不过,对 NIMH 感兴趣的领域进行专用性投资也将变得至关重要,比如开发在动物模型中确认并影响同情心和认知回路的方法。(徐徐)