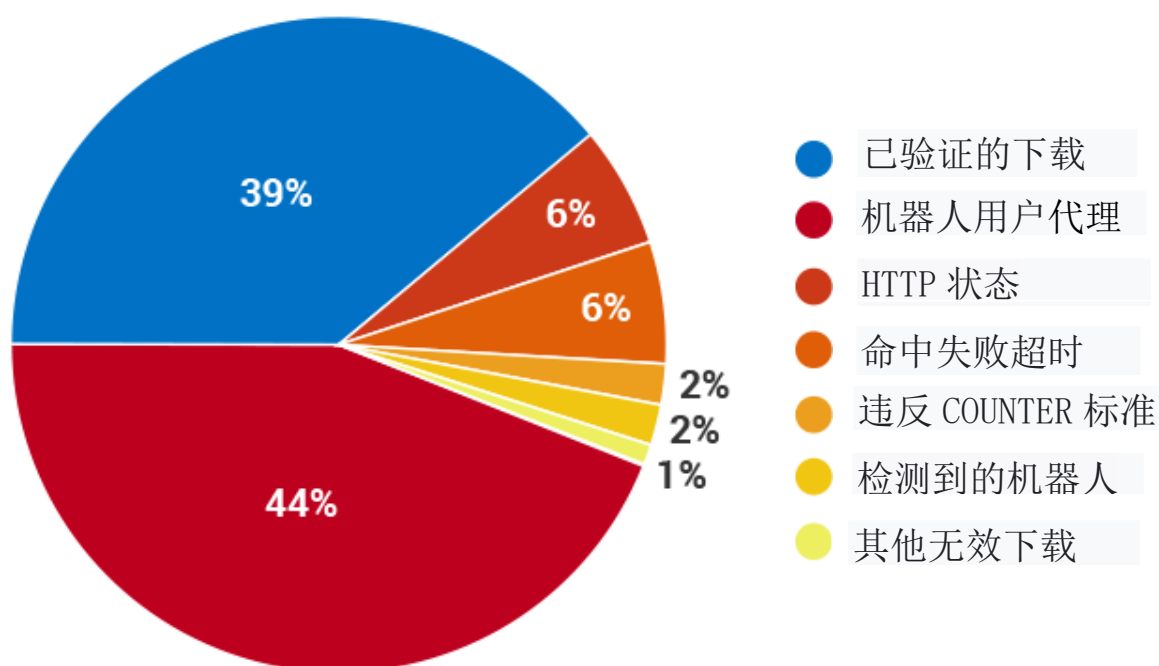


<https://bepress.com/download-counts/>

准确的数据

最新、准确、可证明的数据对于贵机构的学术影响比以往任何时候都相比重要。它也从未像现在这样具有挑战性。

已验证的下载量



经过验证, 只有 39% 的下载是人类活动。

蜘蛛、机器人和（令人毛骨悚然的）爬虫

不是所有的细菌都是“坏的”（例如冷冻酸奶）。相同的道理，一些非人类活动对您的线上内容是有益的。例如，谷歌爬虫有助于通过谷歌的搜索引擎保持高可发现性。但越来越多的日益复杂的代理正在夸大开放获取材料的下载量。夸大到什么程度呢？使用我们在行业中最严格的筛选方法，我们最终得到每 100 次未经过滤或“原始”下载只有大约 39 次人工刻意下载。

大规模的实时过滤

我们与 500 多家机构合作，并且倾向构建由实时数据提供的可视化工具——因此我们没有时间仔细研究上个月的下载数据并一一判定文章下载量是否真实，例如很多人不可能阅读一篇关于意大利即兴喜剧（意大利语：commedia dell' arte）和游牧生活的文章。因此，我们开发了一种实时过滤方法，让我们继续随着我们（以及那些讨厌的机器人）的学习和成长过程来改进它。

分清良莠

我们如何区分人类读者和机器？当然，我们可以要求登录名和密码，但我们认为这会使开放获取变得不那么开放。因此，我们追踪人类和非人类活动的模式，以建立一个不断发展的过程。我们以现有的 [COUNTER](#) 标准为基础，这些标准旨在消除错误的人类使用模式，同时也删除机器阅读器。

我们：

- 审查来自指定 IP（不包括主机 IP）超过阈值百分比读者的任何下载
- 根据 COUNTER 标准删除下载
- 从已知机器人和程序中删除下载
- 从 bepress 员工（本地和远程）删除下载

下一步是什么？

我们最近加入了 COUNTER 机器人工作组，很高兴能够共同建立一些共同的建议和最佳实践。如果您想更深入地了解我们的工作，您可以观看我们在 Open Repositories 2016 上介绍下载过滤的网络研讨会，以了解有关流程。（如果您在观看视频时遇到问题，请尝试刷新浏览器或清除缓存，然后重试。）