

doi:10.3969/j.issn.1672-5565.2015.02.08

基于微信公众平台的文献定制服务

单光宇, 卢一鸣*, 屈武斌, 张成岗*

(军事医学科学院放射与辐射医学研究所, 蛋白质组学国家重点实验室, 全军军事认知与心理卫生研究中心, 北京 100850)

摘要:文献学习是科研人员跟踪领域进展, 思考课题发展的必要途径。为了解决移动端跟踪文献的繁琐问题, 本文借助移动互联网即时性、便捷性、个性化等特性, 开发了基于微信公众平台的文献定制服务。该服务通过解析用户输入, 动态匹配自构建的期刊名表, 索引 PubMed 数据库, 实现期刊个性化订阅、文献查询、影响因子查询等便捷功能。对于提高科研人员的文献学习效率, 降低追踪文献的时间成本具有较大的价值。

关键词:微信公众平台; 影响因子查询; 文本挖掘; 期刊定制; 文献查询

中图分类号: TH133; TP183 **文献标志码:** A **文章编号:** 1672-5565(2015)-02-120-05

A customized literature service based on WeChat public platform

SHAN Guangyu, LU Yiming*, QU Wubin, ZHANG Chenggang*

(Institute of Radiation Medicine, Academy of Military Medical Sciences; State Key Laboratory of Proteomics;
Key Laboratory of Military Cognition and Mental Hygiene, PLA; Beijing 100850, China)

Abstract: Literature study is a necessary way for tracking progress in a certain research field and conceiving the project development. In order to solve the onerous literature-tracking problem in a mobile terminal, we developed a PubMed Customized Retrieving system (PCR) on WeChat Public Platform with the features of immediacy, convenience and personalization of the mobile internet. The service provides several convenient features such as literature customization, literature querying and Impact Factor (IF) Querying by parsing input, matching built-in journals name table dynamically and retrieving PubMed database. This work helps to improve the literature learning efficiency and reduce the time cost for researchers.

Keywords: WeChat Public Platform; IF Search; Text Mining; Journal Customization; Literature Querying

文献学习是贯穿科学研究工作的重要过程, 对本学科领域的各类文献进行全面、系统的追踪对于培养科研嗅觉, 了解领域进展起到了重大的作用。文献追踪的效率直接影响到科研效率, 据 National Science Foundation (NSF) 统计, 一个科研人员花费在查找和消化科技资料的时间需占全部科研时间的 51%^[1]。因此, 提高文献追踪的效率便成为一项重要工作。

随着科技的发展, 文献追踪从过去的图书馆查阅到今日的网络检索, 科研人员的科研效率大幅提高^[2]。

近几年, 随着移动互联网浪潮的到来, 智能手机已经成为了人们获取信息的主要渠道^[3], 截至 2015 年 2 月, 国内移动互联网总数已达 8.83 亿户^[4], 而微信在中国大陆的市场渗透率已达 93%^[5]。毫不夸张地说, 微信已经走进每个人的生活。工欲善其事, 必先利其器, 微信的普及为再次提高科研效率提供了良好契机。

微信公众平台是基于微信的服务平台, 它的可二次开发性使得更加高效的文献追踪成为可能。随着期刊数量的增加, 如何高效便捷的获取文献便成为了重要的课题。当前科研人员使用的文献服务例

收稿日期: 2015-04-04; 修回日期: 2015-04-20.

基金项目: 国家自然科学基金项目 (31401141, 31371345); 蛋白质组学国家重点实验室优秀青年课题 (SKLP-Y201303)。

作者简介: 单光宇, 男, 硕士研究生, 研究方向: 生物信息学; E-mail: sci@shanguangyu.com.

* 通信作者: 卢一鸣, 男, 军事医学科学院助理研究员, 研究方向: 生物信息学; E-mail: luyim@outlook.com.

张成岗, 男, 研究员, 博士生导师, 研究方向: 生物信息学; E-mail: zhangcg@bmi.ac.cn.

如 PubMed^[6]、MEDLINE^[7]、Google Scholar^[8] 虽然能够在服务类目上提供一定程度的适用性,但存在操作过程繁琐、条目过多、移动端体验不友好等特点,难以满足高节奏的科研需要。因此,本研究着重于文献追踪的友好性与高效性,开发了基于微信公众平台的文献定制服务 PubMed Customized Retrieving (PCR) System。该服务能够自动识别用户输入、利用文本挖掘技术匹配自构建的关键词和期刊名表进行 PubMed 数据库索引,并且定时推送新增文章到用户微信号中。除此之外,考虑到科研人员即时查询的需要以及对期刊影响因子的关注,我们继续开发了文献实时查询以及期刊影响因子即时查询等模

块,较好的提高了本服务平台的立体性。

1 平台功能介绍

本平台基于 Python 语言,引入了广泛使用的 PubMed 数据库以及文本挖掘技术为研究人员提供便捷的文献定制服务。该服务主要包括三个主要功能模块:(1) 期刊文献订阅(支持关键词+杂志);(2) 文献实时查询;(3) 期刊影响因子实时查询。

基本工作流见图 1,用户可以通过如下方法关注使用:(1) 使用手机扫描二维码(见图 2);(2) 进入微信,添加公众平台,输入 PCR,进行关注。

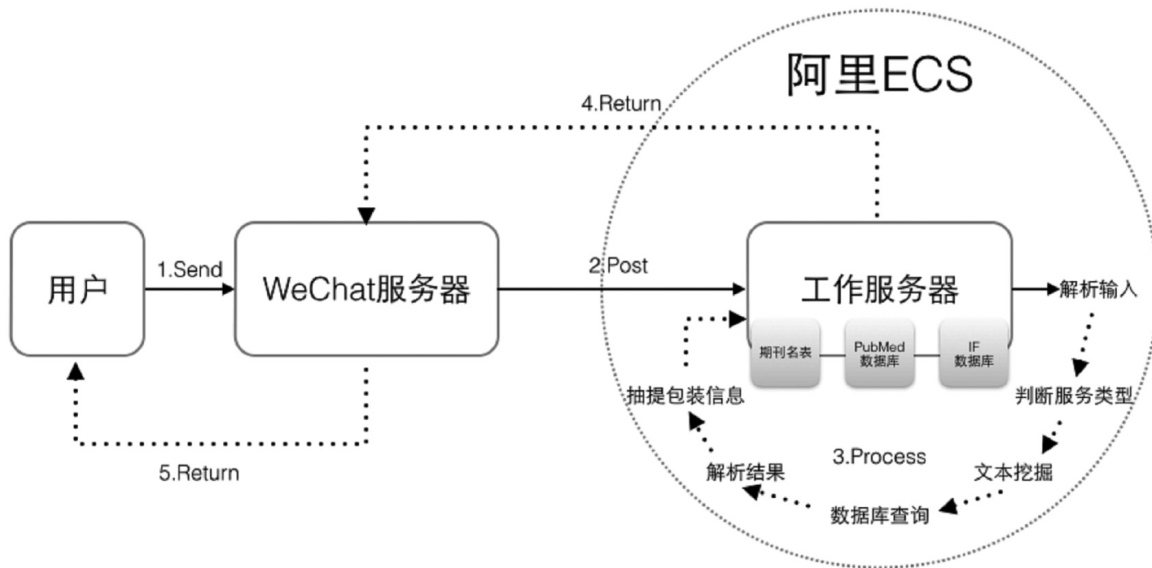


图 1 公众平台基本工作流程图

Fig.1 Basic flowchart for public platform



图 2 平台二维码

Fig.2 QRcode for public platform

1.1 PubMed 文献个性定制

对于持续追踪领域内文献,关键词显得尤为重要。此外,每个领域都有特定的优秀期刊。为提高个性化定制的友好性,本平台极大简化了用户的输入负担。用户只需在对话框中键入关键词及杂志名称,后台能够利用文本挖掘技术自动匹配关键词与期刊名并索引 PubMed 数据库,将文章数量、文章题目、摘要等关键信息存入目前广泛采用的非结构化数据库 MongoDB 中。一旦新增文章,便能够定时为用户推送,保证用户能够及时得到关注领域的最新文章,较好地提高了科研人员的文献追踪效率(见图 3)。

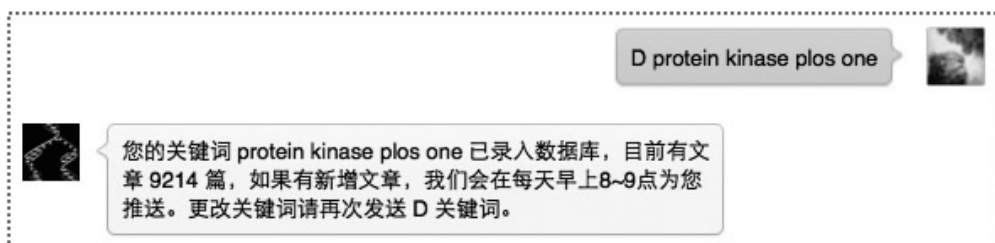


图 3 使用关键词与期刊名的订阅功能

Fig.3 Subscribe articles by keywords and journal name

1.2 文献实时查询

定制好用户的搜索模式后,针对科研人员的需求特性,我们继续开发了文献实时查询模块,可以直接通过本微信公众平台实现文献查询。用户只需要

简单输入想要追踪的关键词及期刊名称,平台通过索引利用文本挖掘技术建立的关键词-杂志名称映射表自动对其识别并调用 PubMed 数据库查询(见图 4)。

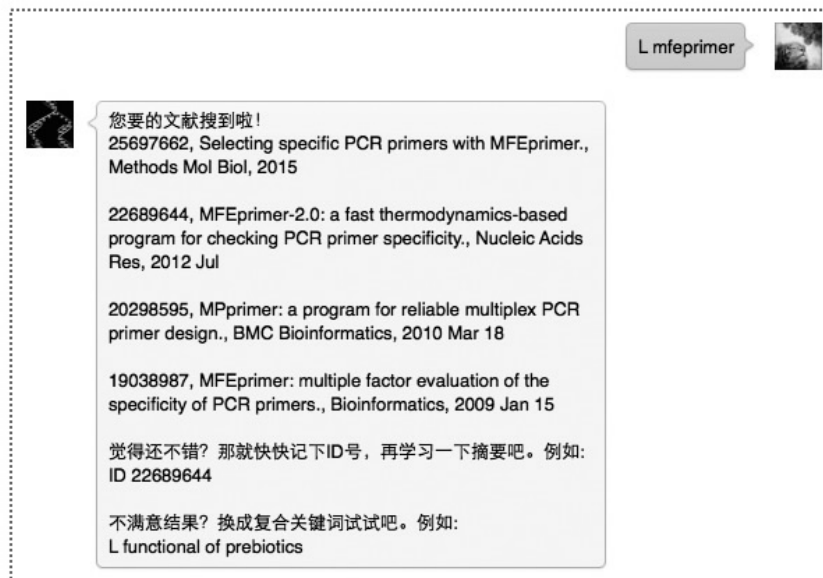


图 4 目的关键词的文献查询

Fig.4 Retrieve target articles by keywords

为了进一步增强平台的友好性,方便用户操作,本平台还对部分展示系统进行优化,包括:(1)建立人性化引导机制,简化用户操作难度;(2)使用正则表达式,对输出结果进行部分限定和过滤;(3)提供

二次追踪服务,用户可以针对感兴趣的文章查看摘要内容(见图 5)。

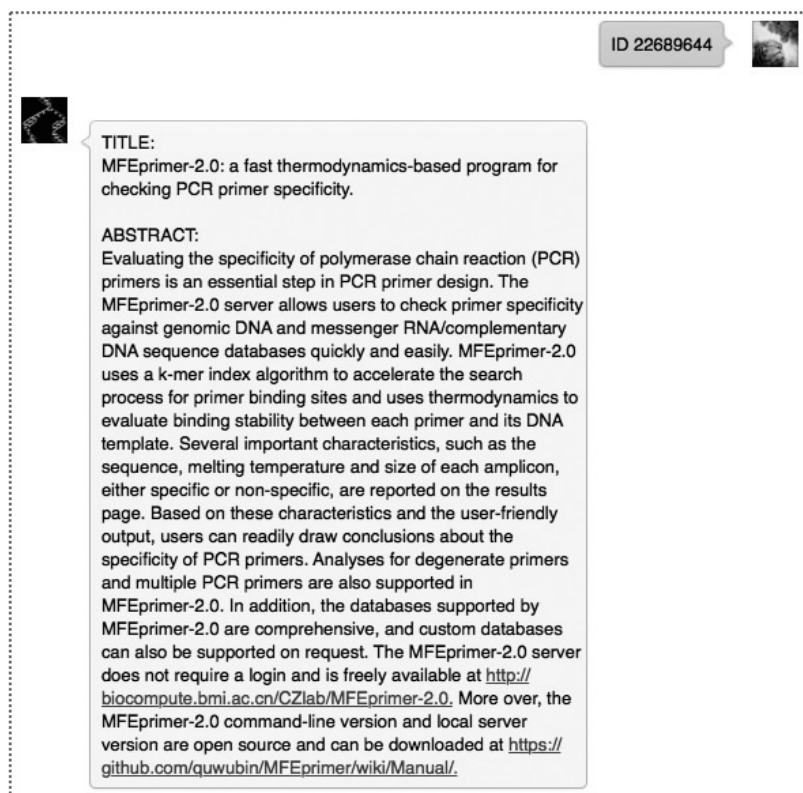


图 5 目的文章的摘要信息查询

Fig.5 Retrieve abstract information for target article

1.3 期刊影响因子查询

在建立起文献个性定制以及实时查询两个功能模块后,我们了解到研究人员对便捷查询期刊影响因子的需求,于是开发了期刊影响因子查询模块。期刊影响因子是研究人员选择投稿期刊的一项主要参考指标,了解目的期刊历年来的影响因子变化趋势也有助于辅助研究人员进行决策^[9]。为此,我们收集了 2009~2014 年 Journal of Citation Reports (JCR)^[10]的所有影响因子数据,构建起了影响因子数据库。基于该数据库,我们针对用户的查询期刊,

绘制六年影响因子趋势变化图即时返回给用户,较好的提升了用户体验。

另外,为了增强移动端的友好性,我们还在后台进行相似度计算,尤其是在用户不明确知道期刊名时,利用全局数据库比对,动态推荐杂志(见图 6)。为了方便用户查看,对检索结果建立二次链接,只需点击图片便可以得到该杂志的六年影响因子趋势变化图,十分便捷(见图 7),对于帮助科研人员了解领域热点,理解领域的前景与深度具有重要意义^[11]。

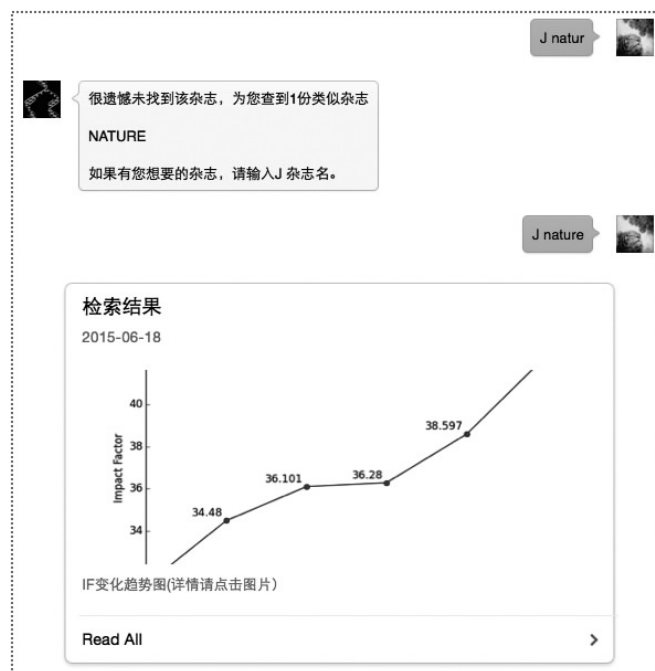


图 6 影响因子实时查询

Fig.6 Search Impact factor in real-time querying

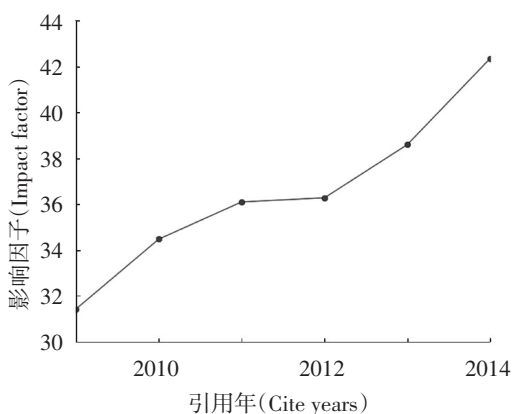


图 7 Nature 六年影响因子趋势变化图

Fig.7 Impact factor history chart of Nature Journal in six years

为估计推荐算法的用户满意度,我们检测了 20 位用户的二次操作查找率(即按照推荐杂志进行查找的操作)。评估原则为:(1)不进行任何干预,输

错杂志名的用户自动入实验组,记录前 20 个用户信息;(2)结果表明,17/20 的用户在二次操作中查找推荐杂志,我们认为推荐算法具有较好的可靠性,即推荐的杂志基本满足用户的心理预期。

2 讨论

随着文献数量的急剧增长、科研节奏也在逐渐加快,如何高效率的获取文献,追踪领域进展成为科研人员亟待解决的问题。为此,本文设计开发了本公众平台,较好的完善了科研人员移动端追踪文献的需求。

值得一提的是,除了开发功能模块,我们还对服务器进行了较为完备的压力测试,保障其可以承受高并发的用户需求,并且对计算出的杂志影响因子趋势图进行了 CDN (Content Delivery Network) 加速,在用户查询时,输出图片可以动态切换到离用户

速度最快的网络节点进行传输,保障了查询的稳定性、时效性,平均反应时可达0.2秒,基本达到即输即现,极大地提升了用户体验。

然而,正反共存,优缺点同在。作为依托微信的服务,既享受到其提供的便捷服务,还受到诸多限制,例如发送消息的长度、发送图片的规格、消息展示的风格等等。因此,下一步将继续针对微信公众平台的优势与不足,找到一个平衡点,以期在接受限制的情况下开发更多便捷高效的服务,例如出具统计报告。对用户操作进行适当的群体行为分析,出具统计报告有益于用户了解查询群体的兴趣点所在,对于规划研究方向,了解研究热点能够起到一定的辅助作用,这也是下一步的工作重点所在。

参考文献(References)

- [1] 郭栋. ISI Web of Knowledge 在科研中的作用与价值 [EB/OL]. <http://lib.utsz.edu.cn/ekcms/UploadFile/2008-10-25/20081025045124433.ppt>, 2011, 8.
- GUO Dong. The value and effects of ISI Web of Knowledge inresearch[EB/OL]. <http://lib.utsz.edu.cn/ekcms/UploadFile/2008-10-25/20081025045124433.ppt>, 2011, 8.
- [2] LAGEMANN E C. An elusive science: The troubling history of education research [M]. USA: University of Chicago Press, 2002.
- [3] ZHOU Y, ZHANG X, JIANG X, et al. Taming information-stealing smartphone applications (on android) [M]//Trust and Trustworthy Computing. Springer Berlin Heidelberg, 2011. 93-107.
- [4] 工信部. 2015年2月份通信业经济运行情况[EB/OL]. <http://www.miit.gov.cn/n11293472/n11293832/n11294132/n12858447/16505685.html>, 2015, 2.
- China's ministry of industry and information technology. Economic operation status of telecommunications in February, 2015 [EB/OL]. <http://www.miit.gov.cn/n11293472/n11293832/n11294132/n12858447/16505685.html>, 2015, 2.
- [5] MEEKER M. Internet trends 2014-code conference [J]. Retrieved May, 2014, 28: 10-23.
- [6] FALAGAS M E, PITSOUNI E I, MALIETZIS G A, et al. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses [J]. The FASEB Journal, 2008, 22(2): 338-342.
- [7] DICKERSIN K, HEWITT P, MUTCH L, et al. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database [J]. Controlled Clinical Trials, 1985, 6(4): 306-317.
- [8] JACSÓ P. Google scholar: the pros and the cons [J]. Online Information Review, 2005, 29(2): 208-214.
- [9] THOMPSON P J. How to choose the right journal for your manuscript [J]. CHEST Journal, 2007, 132(3): 1073-1076.
- [10] 闵波, 屈武斌, 卢一鸣, 等. 影响因子 (IF) 即输即现快速查询的实现 [J]. 生物信息学, 2013, 11(1): 29-32.
- MIN Bo, QU Wubin, LU Yiming, et al. A web application for fast impact factor search in an input-as-a-show way [J]. Chinese Journal of Bioinformatics, 2013, 11(1): 29-32.
- [11] KARAGEORGOPOULOS D E, LAMNATOU V, SARDI T A, et al. Temporal trends in the impact factor of european versus USA biomedical journals [J]. PloS One, 2011, 6(2): 1-4.