

用正确的搜索策略 加快药物研发

药物的研发上市是一个漫长的过程——药物的临床前研究、临床试验等过程都极其耗时，除此之外，还有大量的时间被浪费在了搜索临床试验、实验室报告等数据上。本文作者 Jeff Everham 认为，通过采用正确的搜索策略，可以节省大量的研发时间，从而加快药物的上市。

文 / Jeff Everham

PROCESS-Tip

* 更多信息请登录：
pharm.vogel.com.cn
* 反馈编码：220924

据 Deloitte 公司称，现在一种药物的平均研发成本为 22 亿美元，同时近 10 年来首轮临床试验的成功率一直低于 10%，也就是说药物研发的投资回报早已不再稳定，所以制药企业都想尽早完成药物的上市，迅速回收昂贵的研发成本。

药物研发意味着：来自不同地方的众多专家必须访问分散的信息，对其进行组合、处理并找出其中的关联，这也是研发过程进展缓慢的原因所在。软件制造商 M-Files 在其 2019 年出品的智能信息管理报告中，向全球 1500 家公司询问了他们在数据管理方面遇到的问题，并汇总如下：

• 信息迷宫：几乎近半数受访者表示很难找到正确的信息；

本文作者系 Sinequa 公司的产品战略副总裁。



Jeff Everham 认为，通过采用正确的搜索策略，可以节省大量的研发时间，从而加快药物的上市

• 版本升级：超过 2/3 的受访者表示难以找到文档的最新版本；

• 文档重复：超过 80% 的受访者表示他们需要重新创建已有的文档，因为他们无法在自己的企业网络中找到它。

太多的研究论文

信息过量是使事情变得如此

困难的主要原因，这些信息分布在各种各样的地方，而且据 Gartner 公司称，它们当中的 80% 不够系统化，这意味着：没有易于分析的表格或类似的东西可供直接参考，只有医生的笔记、实验室报告、研究观察和图片等。每年都有超过 300 万篇研究论文发表在全球 33 000 多种专业期刊上，一个人仅完成阅读摘要便需要 85 年。自 COVID-19 为人所知以来，关于该病毒的论文便已发表了 50 万篇。

用智能搜索取代关键字匹配

研究人员无法通过标准商业智能和分析工具找到并使用隐藏在非系统化数据中的信息，需要比关键字匹配更优质的新智能搜索技术来提供有关主题的上下文信息（就像普通搜索引擎所做的

那样)。企业搜索软件产品(如阿斯利康和 UCB 使用的 Sinequa)采用了自然语言和深度学习方面的知识,即人工智能技术(AI)来优化搜索。这些产品能够了解搜索者的真正意图,能够发现非系统化文本中的关系和模式,提供相关且有意义的结果。Gartner 公司便在其推出的数据集成工具魔力象限中谈到了洞察引擎。

实验室记录本中的批评性意见

如果某些信息隐藏在难以访问的系统中,研究人员因无法全面了解药物的安全性会付出怎样的代价?如果研究人员因为临床试验结果分散在多个数据集中而无法全面了解药物信息,又会造成怎样的后果?如果实验室记录本或医生笔记中的批评性意见被忽视,研究人员会错过哪些机会?通过采用智能搜索可大幅降低这些问题背后潜藏的风险。

例如,阿斯利康公司的内部和外部专家每天都会创建大量技术性很强的文档:研究文档、医学数据库中的条目、测试报告、专利申请等,此外还有不同专业领域的研究人员之间的交流信息。在阿斯利康的研发过程中,每天都有人在集团中寻找有关特定主题的专家或信息。例如,谁特别熟悉“动脉硬化”?谁了解药物中的活性成分和活性分子?有哪些关于副作用和专利的文件?那么如何高效找到这些信息呢?

使用内容分析技术

阿斯利康首先尝试了使用简单的搜索引擎,根据其在文件上留下的痕迹来定位专业人员,但他们发现最终只有使用企业搜索软件才有可能从来源各异的大量数据中过滤出正确的信息。内容分析技术基于

语义和语言分析,可用于搜索非系统化和系统化的数据,通过这种方式,用户可以找到那些根本没有出现实际搜索词,而只是有同义词或内容有相似概念的文章。

在最初的过程中,阿斯利康使用企业搜索软件分析了大约 2 亿份来自研发领域的内部和外部文件,还确定了术语(同义词和语义相关的术语)之间的特定主题关系。该索引产生了一个数据库,在搜索掩码中输入术语后,软件便能从数据库所有关于该主题的内容中找出最好的专家或研究小组,在最短时间内组建跨学科的专家团队,就一个共同主题开展跨地区的工作,这使得阿斯利康等制药企业在将创新研究成果尽快转化为成熟产品方面具有了明显的竞争优势。

数以百万计的数据记录

国际制药企业 UCB 也以类似的方式使用了企业搜索软件,这家总部位于布鲁塞尔的生物制药企业在全球拥有 8600 多名员工,致力于研究和开发中枢神经系统和免疫疾病领域的创新治疗方案。

在药物疗法的开发过程中,UCB 开展了广泛的临床研究,开发药物时研究人员必须不断访问这一过程中形成的大量数据内容,以便找到相关信息进行深入的统计分析。在过去 15 年里,UCB 积累了大约 1000 万个文件,相当于 20TB 的海量数据,目前这一文件量仍以每年 30% 的速度增长,这包括超过 100 万条、具有数十亿行数据的 SAS 记录、SAS 程序代码、ASCII、Microsoft Office 以及 PDF 文件。

借助人工智能技术支持的搜索和分析,制药企业可以更轻松地内部和外部资源中获取这些庞大和多样的数据,加快研究速度,从而缩短药物的上市时间。



上海意发玛制药设备有限公司

E-PHARMA (SHANGHAI) PHARMACEUTICAL EQUIPMENT CO., LTD



We are **E-Pharma**
sanitary heat exchanger
supplier

我们是 **意发玛**
卫生级换热器提供商

Your good **Partner**
良好的 **合作伙伴**

Specialized, Focus, Try for the Best
专业、专注、精益求精

Web: www.epharmatech.com

E-mail: info@epharmatech.com

Tel: +86-21-32503721 32503722

No.39, Xinjin Road Qingpu District, Shanghai, PRC

中国·上海市青浦区新金路39号



广告