

Thomson Pharma 数据库使用详解及实例解析

李明辉

(中科院上海药物研究所图情室 上海 201203)

〔摘要〕 简述 Thomson pharma 数据库在药物研发各阶段的重要作用、资源体系结构以及主要功能, 通过实例分析重点介绍其强大的检索和分析功能, 由此引导用户更便捷、更有针对性地使用该数据库。

〔关键词〕 Thomson pharma; 信息检索; 药物研发; 分析报告

Thomson Pharma Using Explanation and Example Analysis LI Ming-hui, the Department of Literature and Information, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

〔Abstract〕 The paper briefly introduces the importante function of Thomson Pharma database in each phase of the drug research and development, describes the main system architecture and its functions, highlights its retrieving and analyzing functions through example analysis, which will guide users utilizing the database much easier and more targeted.

〔Keywords〕 Thomson pharma; Information retrieval; Drug research and development; Analysis reports

1 数据库简介

Thomson pharma 数据库是 Thomson 公司 2005 年推出的一个强大的综合医药信息平台。其信息来源于各种科技文献、专利、新闻及科技会议论文等。目前提供了 31 000 种药物、280 万种化合物、7 500 家公司、330 万项专利、13 000 多个药物靶点、500 万个基因和蛋白序列、超过 6 000 种的专业期刊以及 13 000 多项交易信息^[1]。这些信息可以满足药物研发不同阶段以及不同专业人员的需求。

中科院上海药物所自 2006 年购买 Thomson pharma 数据库以来, 不断探索、使用此数据库, 充分利用其强大的搜索、分析、导出等各项功能, 对

很多项目和课题的实施起到完善和推动作用。之前已有关于此数据库的报道^[2-4], 本文将详细介绍 Thomson pharma 数据库的使用方法, 简介若干新增功能和内容变化, 并就使用过程中遇到的问题及实例进行解析。

2 数据库使用详解和实例解析

2.1 信息分类

2.1.1 按资源类型分 作为动态、集成的药物和生物技术信息平台, Thomson pharma 将各类信息按照资源类型分为 10 类, 分别为药物 (Drugs)、知识产权 (Intellectual property)、文献/新闻 (Literature and news)、公司 (Company)、靶标 (Targets)、化学 (Chemistry)、序列 (Sequences)、临床方案 (Protocols)、临床结果 (Outcome) 和交易 (Deals)。此分类名称相比前几年有所变化^[2]。

2.1.2 按用户群分 Thomson pharma 数据库还提供了

〔修回日期〕 2010-05-10

〔作者简介〕 李明辉, 硕士, 馆员, 发表论文 5 篇。

不同的专题页面,如按角色 (Speciality) 分类,分为生物学 (Biology)、化学 (Chemistry)、临床阶段 (Clinical pipeline) 和临床试验 (Clinical trials) 4 种;按主题 (Subject) 分类,分为注册和商业 (Licensing and business)、竞争性情报 (Competitive intelligence)、新闻会议 (News and meetings) 和知识产权 (Intellectual property) 4 种。用户可以根据自己的角色和感兴趣的专题进入相应的专题页面,快速获取相应的信息。如化学工作者,点击“Chemistry”可进入“新化合物浏览 (New compound viewer)”,快速浏览最近新出现的化合物。“合成方法 (Synthetic methods)”来自于化学合成方面的主要期刊,定期将新出现的合成方法汇总列举,“结构式检索 (Structure search)”,支持 MDL Isis draw 画图软件,可实现精确结构和亚结构检索。

2.1.3 信息过滤 可以根据 12 种治疗领域过滤 (Filter by therapy area) 信息,当用户选定领域时,专题页面的内容会重新刷新。如进入“Biology”主页,标签为浅黄色的部分 (如: Drugs 和 Focused journals),内容均可以按照治疗领域过滤,标签为蓝色的部分 (如 Feature 和 Current Opinion Journals) 则不可以过滤信息。

2.2 检索功能

2.2.1 快速检索 (Quick search) 所有页面上部都提供一个快速检索入口,有 10 种检索途径,分别为药物名称 (Drug name)、化合物名称 (Compound name)、专利号 (Patent number)、公司 (Company)、文献/新闻标题 (Lit&News title)、作者/发明人 (Author/Inventor)、靶点 (Target)、治疗领域 (Therapy area)、作用机制 (Action) 和技术类型 (Technology)。2009 年 8 月 Thomson pharma 升级为 9.1 版本后,新增部分功能^[5]。(1) 增加了通配符,以前仅可以使用%,现在 * 和 % 均可以作为通配符,另增加? 代替一个字符。(2) 根据用户输入的信息,给出检索词建议,可以保证用户输入正确的检索词,防止输入错误而无检索结果,同时节省时间,用户无需打开树状图标浏览选定检索词。此外可以显示同义词,让用户检索更全面。(3) 支持布尔逻辑 And 和 Or 的运算。输入一个检索词后,按空格键,布尔逻辑算符会出现在下拉菜单中,从而实现限制性检索。

2.2.2 表格检索 (Form search) 表格检索分为 7 种,包括普通检索 (General search)、临床试验检索 (Clinical trials)、药物检索 (Drug search)、知识产权检索 (Intellectual property)、文献和新闻检索 (Literature and news)、序列检索 (Sequences) 和交易检索 (Deals)。用户可以根据

不同的需求,从不同入口进入,达到高级检索目的。如从 Drug search 进入,系统提供了 8 种途径供用户限制性高级检索,包括主题词、公司、药名、作用机制、治疗领域、技术类型、国家和结构式检索。其中结构式检索可实现精确结构检索、亚结构检索、模糊结构检索。除了主题词检索框,其它 6 种检索框后面都提供一个索引图标,可以点击图标进入标准词表,用户可以搜索或查看树状索引,选定标准检索词加入检索式。此外,页面右侧的复选框可实现对日期的限定以及对药物研发阶段的设定。总之,用户可以根据特定的需求,构建复杂的检索式,满足检索目的。

2.2.3 向导检索 (Guided search) 便于用户快速高效地一步一步构建检索式提供了 5 种入口,包括生物学 - ADMET (Biology - ADMET),治疗领域竞争性情报 (Competitive intelligence - therapy area),药物竞争性情报 (Competitive intelligence - drugs),注册和商业 (Licensing and business - deal finder),文献引用搜索 (Cited reference search)。例如通过“生物学-ADMET”入口,查找药物吸收、分布、代谢、排泄和毒性的信息,向导检索提示输入药物的名称或结构,可直接命中目的药物的相关 ADMET 数据,并提供相关文献链接。再如查找药物竞争性情报信息,向导提示 3 个步骤,按销售额选择一组药物;选择具体的要分析的药物,可单选或多选;结果呈现,以表格报告形式直观显示药物发展状态、作用机制、专利组以及销售情况。

2.2.4 自由式检索 (Write your own) 目前,此功能只支持“Drug”检索。用户可以自由编写检索式,实现利用其它检索方式无法完成的检索。例如想查找 H1N1 方面的上市药物情况,利用“Form search”检索,会把已经上市的药物和疫苗全部搜出,无法排除其中的大部分疫苗信息。利用“Write your own”,编写检索式 ("influenza virus infection" [Therapy area] and launched [Dev status linked]) not (vaccine [Action]),使用逻辑非 (Not) 将上市疫苗排除。

2.3 分析和导出功能

Thomson pharma 数据库提供动态的、多方位的分析功能以及多种结果输出方式。例如查找 II 型糖尿病领域上市药物状况,利用表格检索中的药物检索入口,治疗领域中输入标准检索词“non - insulin dependent diabetes”,右侧研发状态复选框中勾选“Linked to TA/Company/Country”和“Launched”,进入搜索结果界面。

2.3.1 结果显示 查询结果,目前有 40 种 II 型糖尿病

上市药物，有 4 种结果显示方式，分别是只显示概要 (Summary)、含摘要的概要 (Summary with abstract)、表格概要 (Summary table) 和含详细研发状态的表格概要 (Summary with dev. Status)，后两种显示方式更直观和全面，将药物的研发状态、原研公司、治疗领域、作用机制、技术类型和靶标都一一注释，对化学工作者最有利的是给出了化合物结构。此外，还有 4 种排序方式，分别是按照日期、公司、标题和最高研发状态排序，用户可以自由选择是升序或降序。

2.3.2 分析功能 粗过滤：结果列表的上方出现两个彩色饼图，见图 1，可以实现粗过滤。根据左侧饼图的注释，点击右方饼图的相应部位，会根据左侧相应的过滤条件进行过滤。

Rank by 分析：根据左侧的“Rank by”实现更详细的结

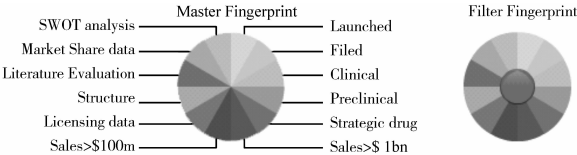


图 1 饼图注释

果分析。目前系统提供 9 种分析方式，如“Rank by target”，可以将 II 型糖尿病的 40 种上市药物按照靶点分析，根据选择，输出柱状图 (Bar chart)、饼图 (Pie chart) 或矩阵图 (Matrix chart)，如果选择的是矩阵分析，则可以实现二维分析，即可以同时选择靶点和公司 (或其它) 两种元素进行结果分析。点击图表中的相应部分，如点击某个靶点的柱形图，会显示出相应靶点下的药物结果列表。然后选择不同的输出方式输出。

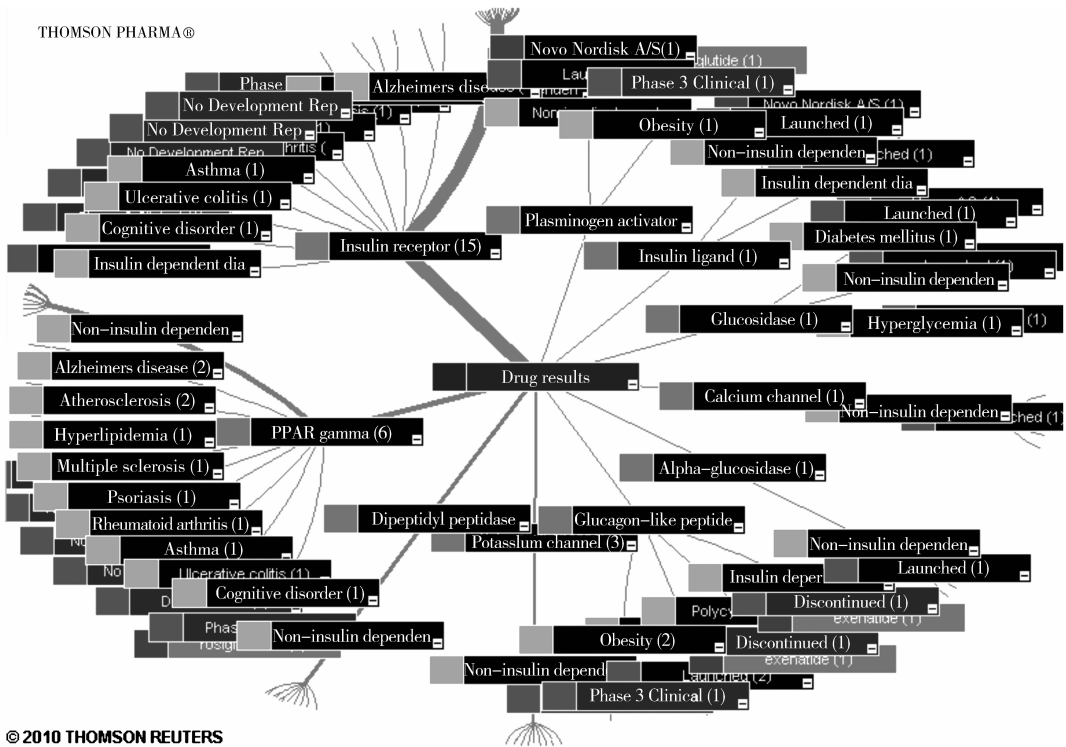


图 2 树状动态可视化分析

Star tree 分析：点击右上角的“Star tree”可以进行树状的动态可视化分析，根据需求创建树状分级。仍以上述检索为例，对已上市的 40 种 II 型糖尿病药物进行树状图分析，分级选择 Target -> Therapy area -> Highest development phase -> Parent company，结果见图 2，40 种药物首先按照靶点分析，树枝越粗代表针对此靶点的药物最多，从图中看出，针对胰岛素受体开发的药物最多，其次是 PPARgamma 类药物，逐级可以按照治疗领域、最高研发状态、原研公司

分析。双击每个树枝分枝的节点，可以打开相应的药物结果记录。

2.3.3 结果输出 检索结果输出可以采用打印、导出 (Export)、发送电子邮件、保存在数据库中的文件夹中 4 种方式，对用户来说，Export 是最常用的方式，可以将结果下载到本地，实现离线查看，还可以自由编辑自己需要的内容。Export 可以选择不同的格式，如 Ms word 或 Ms excel，如选择 Ms word 格式导出，又可以选择不同表现形式的内

容。对于单个药物的报告，还可以实现 PDF 和 HTML 格式的全内容下载。

2.4 分析报告的类型以及报告间的转换

2.4.1 报告类型 Thomson pharma 数据库以报告的形式汇总药物数据信息，并用一系列标签将信息进行分类管理。目前，可以提供 10 种类型的分析报告，分别是药物、

知识产权、文献和新闻、公司、临床过程、临床结果、化学、序列、靶标、和交易报告。每种报告都有其独特的内容标签，下面仅主要介绍药物报告^[6]。药物报告的缩写为 DR，标题即是药物名称。药物报告在药物研发的早期创建，随着药物研发继续进行，不断增加研发信息、临床试验和药物上市的销售信息等，某些报告还提供交易信息。表 1 列举了各标签目前的最新名称及主要内容，Thomson pharma 9.1 版中标签的名称和内容都有少量的改变。

表 1 药物报告内容详解	
标签名称	主要内容
概述 (Overview)	Summary, Overview, Drug names, News, Licensing summary, Deals
研发状态 (Development status)	Development status, Development status history, Therapeutic use
文献评价 (Literature evaluation)	Literature evaluation, Biology, Metabolism, Clinical
化学 (Chemistry)	Chemistry, CAS RN, Scientific data/Chemistry
药物索引评价 (Drugdex evaluation)	Overview, Dosing information, Pharmacokinetics, Cautions, Clinical applications, References, Author information
战略性文件 (Strategic profile)	SWOT analysis, Sales & Market share, US Rx profiles
专利 (Patents)	Patent position, Orange book
临床试验 (Clinical trials)	Clinical trials by sponsor
参考文献 (Bibliography)	Bibliography

2.4.2 报告间转换 目前，提供结果列表转换功能，用户可以在浏览结果时链接到其它相关报告，优点是用户可以快速找到与当前检索结果相关的报告，而无需另外执行任何检索。用户还可以发现通过检索无法发现的信息，如要查找所有 II 型糖尿病上市药物的公司，用户可以在 Drug form search 进行药物检索，然后从药物报告结果列表直接转换至公司报告结果列表。目前系统支持的转换有：药物报告对公司报告、药物报告对化学报告、药物报告对专利报告、药物报告对临床试验报告、药物报告对交易报告、公司报告对交易报告，目前系统最多只可执行两个层级的转换。

2.5 索引功能与个性化功能

2.5.1 索引 (Browse) 功能 首先，Browse 可引导用户快速便捷地选定标准检索词，快速进入信息检索。此外，可以引导用户通过索引系统快速进入各种报告。索引系统包括 7 个方面：作用机制、公司、期刊、会议、靶标、技术类型和治疗领域。如作用机制、技术类型和治疗领域索引均为树状分级索引，引导用户进入相应的药物报告、专利报告、交易报告和临床方案报告（仅治疗领域中有此项）。

2.5.2 个性化功能 用户可以个性化定制主页，存储检索策略，设置个性化报告。通过个性化定制主页，用户可以及时获得各个领域最新的信息及相关领域的最新会议信息。定制 Smart alerts，可以监控第 1 次出现的信息。此外，用户还可以在个人文件夹下保存自己感兴趣的各種报告，通

过项目管理工具，和同事们共享自己的报告。Email 警报还可以对研发人员所关注的对象进行实时跟踪。

Thomson pharma 数据库内容丰富，功能强大，在药物研发的各个阶段均有作用，应得到充分利用。

参考文献

1 Thomson Pharma – world leading pharmaceutical competitor intelligence solution [EB/OL]. [2010 – 01 – 05]. <https://login.thomson-pharma.com/tp/media/pdf/thomson-pharma-fs-en.pdf>.

2 陈湘萍. 强大的 Thomson Pharma 综合医药信息平台 [J]. 生物技术世界, 2006, 10 (5): 59 – 60.

3 马建玲. 新一代制药和生物技术企业动态信息服务系统——Thomson Pharma [J]. 中华医学图书情报学杂志, 2007, 16 (1): 58 – 60.

4 马淑芹, 张玢. 新药研发数据库 ThomsonPharma 分析 [J]. 医学信息学杂志, 2009, 30 (6): 24 – 26.

5 Thomson pharma 客户支持 [EB/OL]. [2010 – 01 – 05]. <http://www.thomsonscientific.com.cn/training/ThomsonPharma.htm>.

6 Introducing Thomson Pharma [EB/OL]. [2010 – 01 – 05]. https://www.thomson-pharma.com/tp/docs/help/tp_help_intro.html.