

医学文献检索的查全查准策略研究 ——以 PubMed 为例

李 盈

(济宁医学院图书馆, 山东 济宁 272067)

摘 要 本文以 PubMed 为例, 介绍提高文献检索查全查准的方法步骤, 从而使广大医务研究人员掌握外文数据库文献检索的必要技巧, 促进医学信息素养的提高。

关键词 PubMed; 检索; 查全; 查准

中图分类号: G252.7 **文献标识码**: A **文章编号**: 1000-9760(2012)06-202-05

Recall and Precision Strategies Research of Medical Literature Retrieval ——Taking the PubMed as Example

LI Ying

(Library of Jining Medical University, Jining 272067, China)

Abstract: In order to achieve the necessary skills in retrieving foreign data pool for medical researchers and enhance their medical information literacy, the paper introduces the methods and steps to retrieve comprehensively and precisely in PubMed.

Key words: PubMed; Retrieval; Recall; Precision

文献信息数字化与计算机网络化、全球化, 推动了知识的广泛普及和充分利用, 知识和信息已成为推动科技进步和社会发展的主要因素。医学工作人员必须能敏锐、准确地捕捉各种有价值的信息, 从而达到开拓思路、开展有效研究和创新工作的要求^[1]。利用各种数据库资源查阅文献是获取医学信息最常用最基础的手段之一, 文献检索的查全查准关乎一个人的学术水平与工作能力的提高。

PubMed 是生物医学文献信息的核心检索系统, 提供世界上最权威的免费文献题录摘要信息, 甚至部分免费全文。因其具有信息量大、使用方便、更新速度快的优点, 赢得了我国医学研究者们的赞誉。另一方面, 语种及检索习惯的差异, 增加了我国学者掌握国外数据库检索技巧的难度。本文介绍 PubMed 查全查准的检索方法, 希望能对提高我国学者检索国外数据库的水平起到积极的作用。

1 PubMed 简介

PubMed 是美国国家医学图书馆(NLM)所属的国家生物技术信息中心(NCBI)开发的因特网生物医学信息检索系统, 位于美国国立卫生研究院(NIH)的平台上。该系统通过网络途径免费提供自 1950 年以来全世界 70 多个国家 4 300 多种主要生物医学文献的书目索引和摘要, 并提供部分免费和付费全文链接服务。

1.1 PubMed 的主要信息数据源

MEDLINE 与 PREMEDLINE 是 PubMed 检索系统的主要信息数据源。MEDLINE 是当今世界上最具权威的生物医学文献数据库, PREMEDLINE 是 1996 年 8 月开始增加的一个新的数据库, 该数据库收录进入 MEDLINE 数据库以前未经标引的数据, 数据库每天进行更新, 每个星期又把经过标引后完整的记录转入 MEDLINE 数据库, 转入后立即从 PREMEDLINE 中删除相应记录。进入 PubMed 以后, 可以直接输入作者姓名、

主题词、自由词和期刊名称等进行检索,在 PubMed 上可以获取最新的文献摘要,有的参考文献提供免费的链接,也可以得到全文^[2]。

1.2 PubMed 中文献的 5 种状态及其检索途径

按照文献在 PubMed 中的流程,可分为 5 种状态(见表 1):

1)“as supplied by publisher”是文献刚交给 PubMed 出版社时的状态,此时没有标引主题词;

2)“in process”阶段,文献进入了 PubMed 数据库,并进入了主题词标引阶段,但仍然没有主题词;

3)“index for medline”阶段,文献被标引了主题词,具有了 Medline 格式,即带有详细的作者标引和详细的主题词和副主题词组配的标引;

4)“premedline”状态表示的是非生物医学类的文献,这类文献不进入 MEDLINE 数据库;

5)“oldmedline for pre1966”是 1966 年以前的文献,这类文献一般无文摘,也没有相应的主题词。

表 1 PubMed 中文献的 5 种状态

文献的类型	文献的状态	主题词
as supplied by publisher	文献进入 PubMed,在 PREMEDLINE 库	无主题词
in process	主题词标引阶段,在 PREMEDLINE 库	无主题词
index for medline	在 MEDLINE 库,具有 Medline 格式	有主题词
premedline	在 PREMEDLINE 库	无主题词
oldmedline for pre1966	在 MEDLINE 库	无主题词

在 PubMed 中,根据文献是否被标引了主题词,确定不同类型文献的检索途径:

1)“as supplied by publisher”和“in process”两类文献时效性最强,缺点是没有标引主题词,所以在检索时无法使用主题词来检索文献,可以使用自由词对文献的题名、摘要进行检索;

2)“index for medline”状态的文献都做了主题词及副主题词的标引,因此,利用主题词检索就一定能找到相关文献,若文章存在却找不到,说明主题词与副主题词的搭配选择有错误;

3)一般的医学文献检索不涉及非生物医学类和 1966 年以前的文献。在检索 PubMed 数据库时,把前 3 类中符合条件的文献检索出来,即可达到查全的目的。

2 PubMed 的查全查准策略

数据库检索的查全查准除了数据库本身资源的限制外,主要受检索途径、检索策略及检索结果的检验与调整等因素的影响。查全建立在查准基础上,是若干个查准的集合。查准的条件就是收集正确的检索式。常用的检索式由作者、作者单位、主题词和副主题词的组配构成。

2.1 利用 PubMed 中的 MeSH 数据库确定检索词

2.1.1 选择正确的主题词 主题词是经过规范化处理并记录在词表中,用于表达文献主题的词^[3]。PubMed 按照医学主题词表(MeSH)对文献进行标引,相比自由词检索,主题词检索文献的结果更具有针对性。进入 MeSH 数据库,在检索框内输入主题词或自由词,在检索结果中可通过两种方式确定所需要的主题词:1)通过字面意思选择(见图 1);2)根据树状结构的上下位隶属关系,确定主题词的学科范围,来判断是否是检索所需要的主题词。图 2 标出的是主题词“Hepatitis”在 MeSH 表中的树状结构及其上下位主题词。

2.1.2 正确选择副主题词 在 MeSH 中检索出主题词后,下面不仅列出了它的词语解释,而且列出了能够与该主题词搭配的副主题词。PubMed 中收录了 83 个副主题词,但在检索结果中,只列出了能够与所检主题词组配的副主题词。副主题词是为了扩大文献的主题词表述范围。例如,图 2 中列出了能够与“Hepatitis”这个主题词组配的副主题词。

2.1.3 选择是否要限定于主要主题词 主要主题词指被检主题词或其与副主题词的组配是文章重点中的重点。选择了该选项,则文献必须满足检索词是文献的主要主题词才能被检索到。这样能够提高检准率,缺点是容易漏检。若强调查全率,该选项不选(见图 2)。

2.1.4 确定是否要扩展主题词 每一个主题词都在 MeSH 库的树状目录上有相应位置及上下位的隶属关系,若选择扩展该主题词选项,则其下位主题词也是检索词,含有下位主题词的文献也可以被检索到。该选项扩大了检索范围,利于提高查全率(见图 2)。



图 1 在 MeSH Database 中检索“Hepatitis”

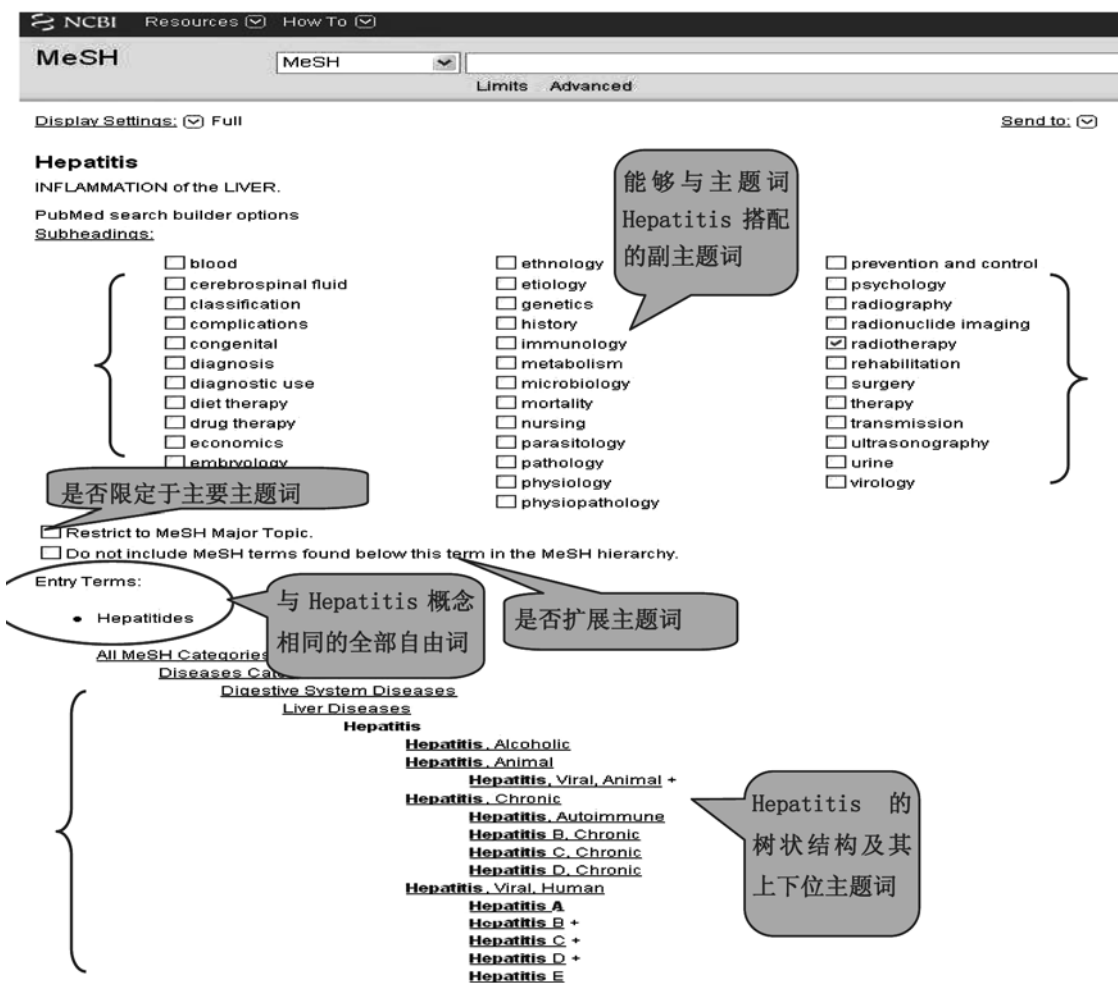


图 2 在 MeSH Database 中检索“Hepatitis”的结果

2.2 利用 MeSH 数据库收集自由词

PREMEDLINE 数据库中的文献,没有标引主题词,这部分文献要使用自由词检索。对于一个概念,研究人员难以掌握其全部的自由词,导致了文献的漏检。

在 MeSH 检索结果的 Entry Terms 项目下, PubMed 列出了与已知检索词概念相同的全部自由词,筛选出有用的自由词来检索文献,可以大大提高查全率。图 2 中“Hepatitides”是与“Hepatitis”概念相似的自由词。

2.3 利用 History 构建检索式

在 PubMed 检索中,常用的检索途径包括主题词和副主题词的组配、自由词、作者、作者单位等。利用 History 可以将单一的检索式进一步整合,构建满足多条件的复杂检索式。

2.3.1 主题词和副主题词的组配 利用 MeSH Database 检索主题词,选择与其匹配的副主题词后,在 PubMed Search Builder 中提交检索式(见图 3)。点击 Advanced,在 PubMed Advanced Search 界面中,Search History 即显示出该检索式(见图 4)。为了保证检索质量,建议每次只用一个主题词与副主题词的组配检索。

2.3.2 利用 Preview 选项检索作者、作者单位、自由词 在 PubMed Advanced Search 界面下的 Search Builder 中选择检索域并输入检索词,点击

Preview 选项,使检索式出现在 Search History 选项中。

2.3.3 使用布尔逻辑(and、or、not)整合 Search History 中的检索式到 Search Box 中(见图 4),点击 Search 得到 PubMed 的检索结果。

2.4 利用 PubMed 检索文献的步骤

PubMed 中,可按照以下步骤进行文献检索,达到查准查全的目的:

1)通过已有的参考文献收集检索所需的作者名、作者单位、主题词及副主题词的组配、期刊名等信息;

2)在 MeSH Database 中,对收集的主题词进行检索,确定其上下位隶属关系,及是否是树状结构中的末位主题词,是否需要向上位或下位主题词进行扩展,同时收集自由词;

3)通过对作者、机构、主题词等的收集,建立正确的检索式,每次只提交一个检索式。在 History 检索框内用布尔逻辑组合已建立的检索式,达到查准的目的;

4)保存每次通过布尔逻辑检索的文献,观察这些文献的作者、机构、主题词等检索项,反复之前的检索过程,达到更多查准;

5)把所有保存的已达到查准要求的文献导入 Endnote 等文献管理软件中进行合并,去除相同结果,达到查全。

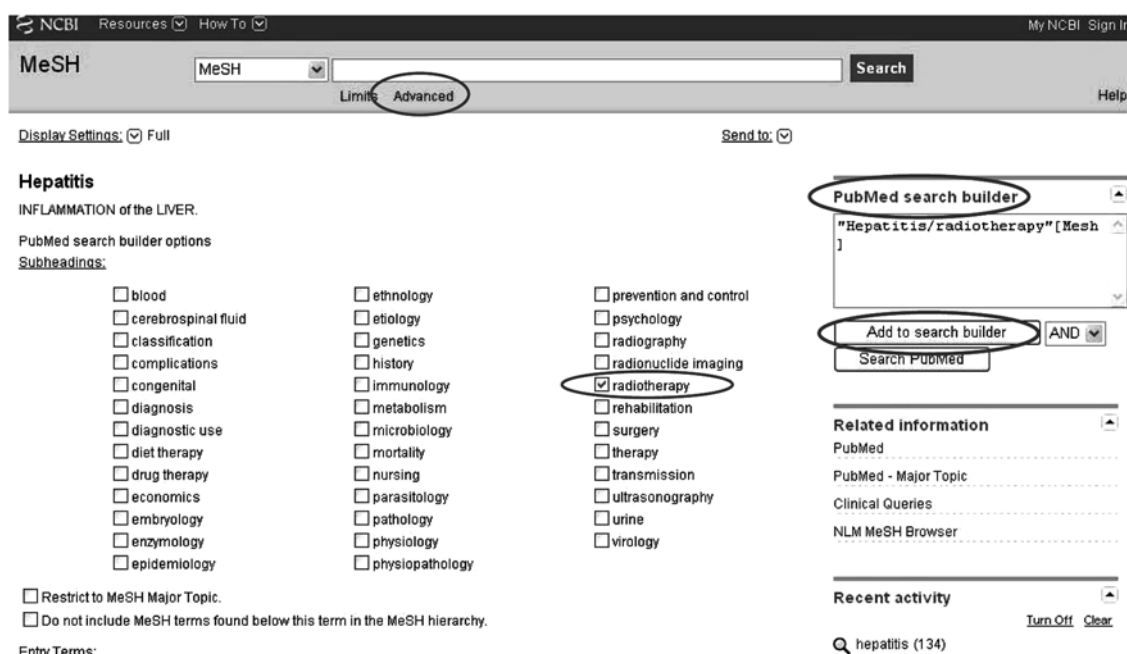


图 3 在 PubMed Search Builder 中提交主题词与副主题词的组配

(下转第 209 页)

- [4] 蒋迎春. 产后抑郁的预防护理体会[J]. 河北医药, 2001, 7(12):1113-1114.
- [5] Murray L, Fiori-Cowley A, Hooper R, et al. The impact of postnatal depression and associated adversity on early mother-infant interactions and later infant outcome[J]. Child Dev, 1996, 67(11):2522-2526.
- [6] 仇剑基, 王祖承, 罗来敏, 等. 爱丁堡产后抑郁量表的应用[J]. 上海精神医学, 2001, 13(4):219-221.
- [7] 林邦杰. 田纳西自我概念量表的修订[J]. 中国测验年刊(台湾), 1980, 27:71-78.
- [8] 李惠萍, 孔祥军. 中学生自我概念的调查研究[J]. 中国健康心理学杂志, 2005, 13(5):324-326.
- [9] 樊富珉, 付吉元. 大学生自我概念与心理健康的相关研究

- [J]. 中国心理卫生杂志, 2001, 15(2):67-68.
- [10] 吕探云. 健康评估[M]. 北京:人民卫生出版社, 2001:139.
- [11] 龙瑞芳, 李连启. 认知行为疗法在心理护理工作中的应用[J]. 山东精神医学, 2001, 14(4):261-262.
- [12] 陈秀梅, 张惠阳. 抑郁症病人的心理干预[J]. 中国健康心理学杂志, 2005, 13(1):77-78.
- [13] 周周桓. 音乐疗法在行为康复中的应用[J]. 中国康复理论与实践, 2002, 8(6):354-356.
- [14] 何明翔. 最新精神病防治技术规范与康复、服务实用全书[M]. 北京:复旦大学出版社, 2005:156-162.

(收稿日期 2012-05-16)

(上接第 205 页)

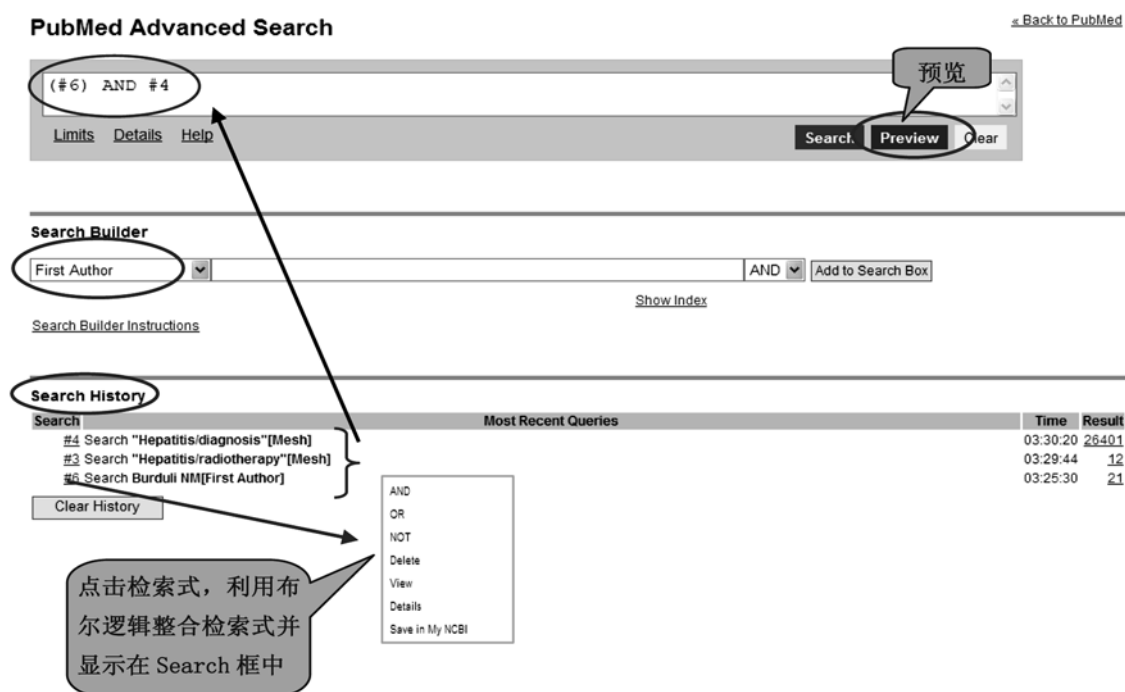


图 4 Preview 和 Search History 的应用

3 总结

本文以生物医学文献检索查准查全为目的,从 PubMed 的高级检索功能入手,重点介绍了 MeSH Database 库中主题词、副主题词、自由词的选择技巧;同时介绍了检索式的构建,以及在 Search History 中使用布尔逻辑组合建立检索式的方法。文章对 PubMed 检索技巧的介绍,蕴含了检索外文数据库的基本思路,希望能对我国学者使用其他类

型国外数据库提供积极的帮助。

参考文献:

- [1] 曾煦欣,王芳,吴剑峰. 药文献检索教学方法初探[J]. 人力资源管理, 2010, 5(2):113.
- [2] 周宇葵,陈先来. 生物医学信息资源文献检索与数据描述[J]. 图书馆理论与实践, 2007, 19(5):80-82.
- [3] 李弘. 医学主题词的选取[J]. 中国病理生理杂志, 2000, 16(5):478-480.

(收稿日期 2012-03-11)