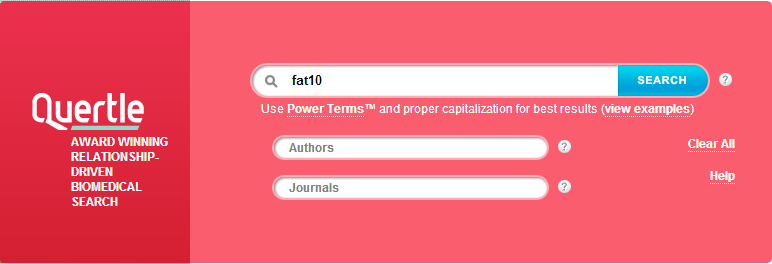
文献搜索工具资料整理

http://www.360doc.com/content/12/0112/10/176942\_178897648.shtml

1. Quertle：这是一个文献搜索的网页工具，包含Medicine’s MEDLINE 和TOXLINE的文献库。

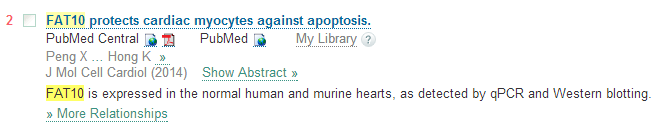
（[http://www.quertle.info](http://www.quertle.info/)）



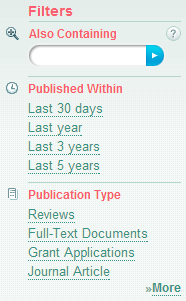
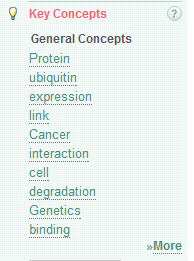
上图是它的搜索界面，第一行可以输入任意关键词，而下面两行则可以根据作者以及杂志名来搜。



点击Search后，会根据结果文献与关键词之间的关联紧密性分成Focused与Broader两类，



显示的文献信息如上图，包括了题目、摘要、网址、PDF信息，并且还将文中与关键词相关的地方在Relationship中显示，搜集信息时可以根据它提供的进行初步判断此文是否符合我们的要求。

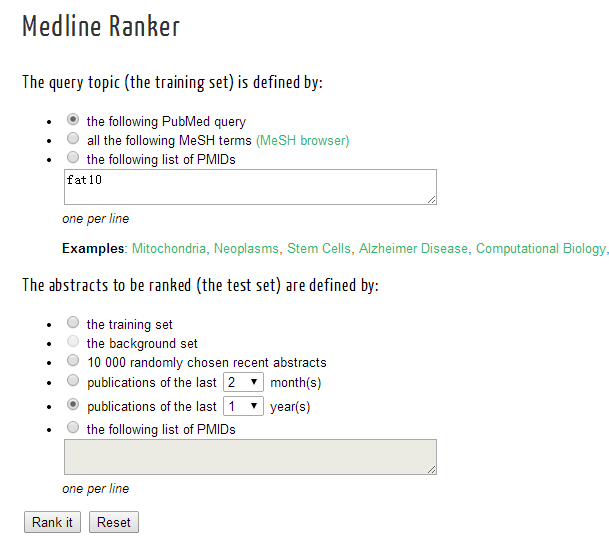
它还提供了类似Advanced Search之类的过滤器，可以加入更多搜索条件，并且根据提供的关键词提供了文献的方向（如蛋白、泛素），还有一般与关键词紧密相关的词汇（conjugation等）。

根据搜索结果以及上手难易程度，这个工具可以算是比较理想的。

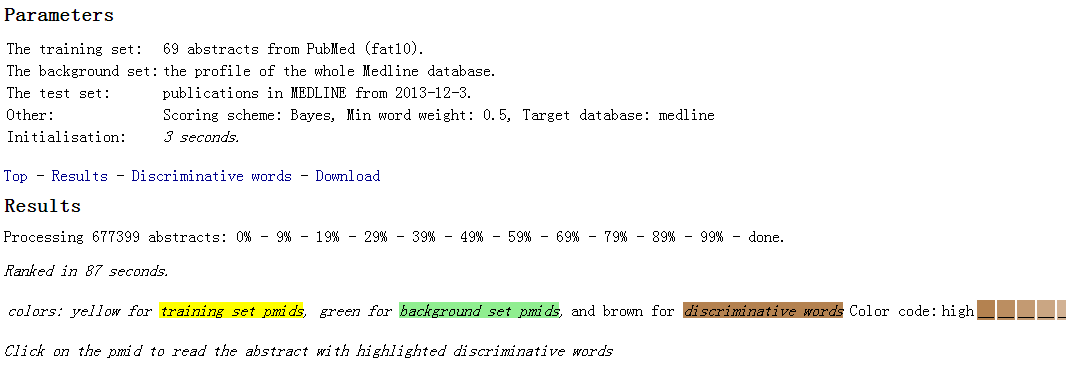
1. MedlineRanker：这也是一个网页版工具，但感觉不怎么适合我们。

MedlineRanker使用贝叶斯模型对某一主题进行阐释，当一些关键词被提出来之后，某一主题的大致样貌便浮出水面，然后利用贝叶斯模型来进行修改。这样更有利于当新文献出现后可以对某一主题重新审视。

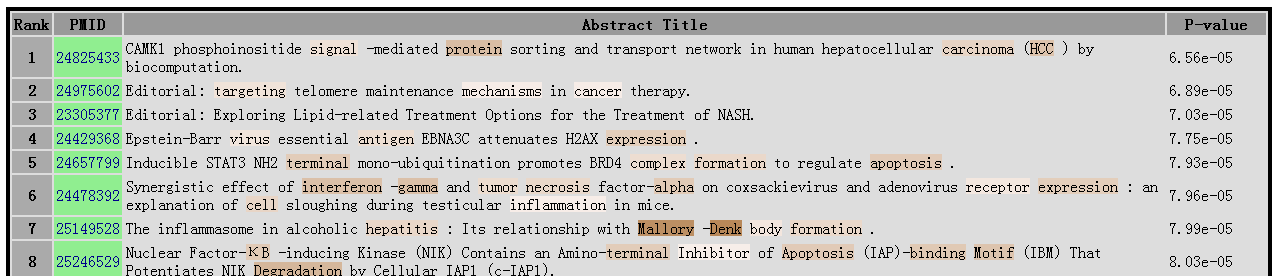
它有Medline Ranker Search以及Medline Ranker Advanced Search两个版块，下面先说一下前者的用户体验。



搜索时可以选择它的训练集即搜索的文献库范围吧，下面则是对搜索的限制条件设置，通过点击Rank it进行搜索。



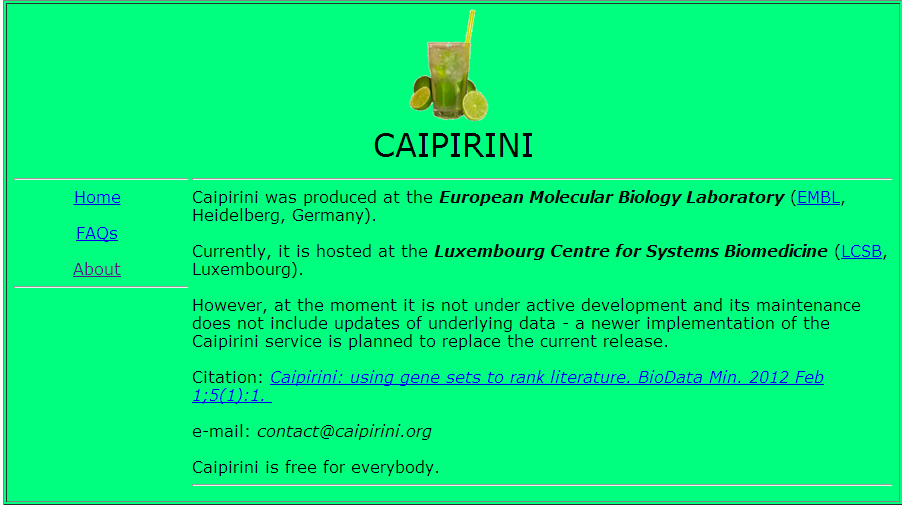
上图是结果界面的一部分，首先是参数，应该就是搜索出来的文献集，并将其作为一个数据集进行了相关的训练，然后将文献的Title与Abstract中的与搜索关键词相关的词汇根据与主题关联程度进行了一个分级，并用深浅不一的棕色标出。



但从文献来看，虽然是能将关键词的基本背景以及相关的词汇都标出，但若是只想获得相关研究文献的话，这并不是一个好的选择，的确只是有利于对一个主题的重新审视，或许对于想从另外一个角度来进行研究的科学家来说比较有用。

我再试了一下Advanced Search，没有任何区别，这个比较奇怪。

1. Caipirini：





这一个工具是能够让用户通过提交感兴趣的GeneID List或者PubmedID List以及相关关键词，过滤出更精细的文献列表。通过点击More options还能提交不感兴趣的列表，以及测试列表来进行更好的training。因为这个实在不知道要提交什么列表，所以没有结果显示，但基本是对文献的一个过滤和分析工具。

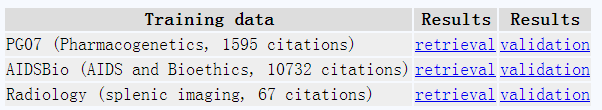
1. MiSearch

Misearch也是种基于用户反馈的排序系统，与RefMed不同的是它可以记录用户的检索词，然后根据这些有反馈的检索词以及用户点击文章的顺序来获得用户的实际需求。

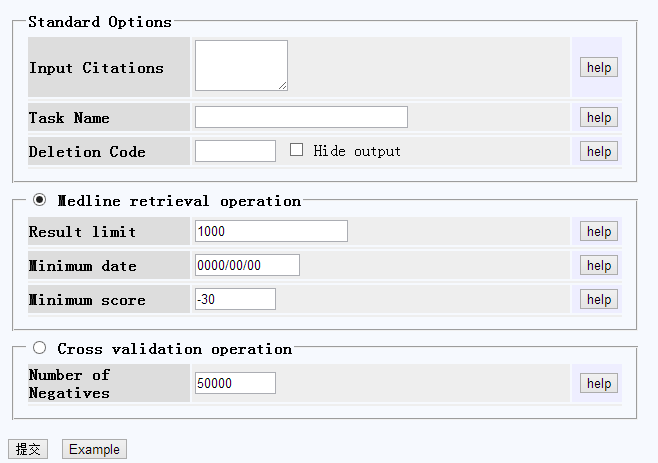
但它最近在系统维护，所以未能进行使用体验。

1. MScanner

MScanner与MedlineRanker最为相似，它不使用检索词，而使用期刊缩写和Mesh。



上图是它的训练数据



从上图可知，它的搜索和分类可以分成三个步骤，首先需要上传一系列的PubMedID，它主要是从上传的文献题目通过算法估计用户想要搜索的文章题目的模式，然后设置参数，获取结果。

我觉得这主要还是用于从PubMed中搜到文献量较多时，需要对它们进行过滤分类的时候，可以用到这个工具。

1. eTBLAST

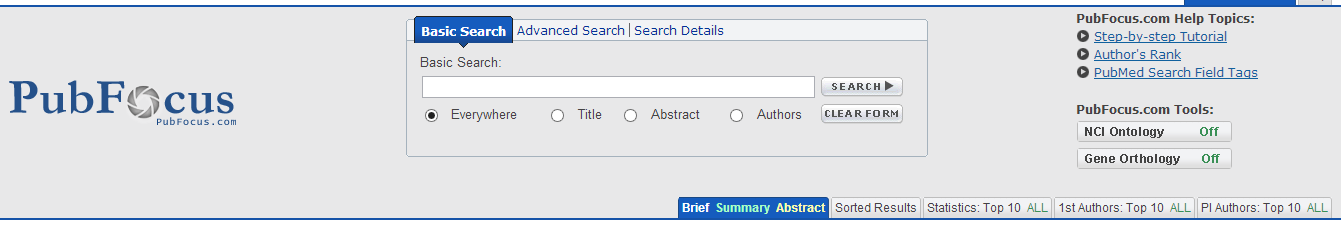
下面是它所发表的文章的摘要：

作为对提交到MEDLINE/PubMed的一次查询的反应，获得没有关联的文章是常见的。当提交一个多-单词查询(这是大多数提交的查询)时，所有查询的单词在每篇文章中的出现可能是检索相关文章的必要条件，但是还不够。理想地，文章中查询的单词间的一个关系也是需要的。我们提议如果两个单词出现在一篇文章中，当单词出现在临近的句子中时，它们之间的一个关系的可能性被解释为比出现在远程句子中更高。因此，句子水平的同时出现能够用作单词之间关系存在的一种替代。为了避免不相关的文章，一种解决方案是增加搜索特异性。另一种解决方案是估计一个相关性得分来排序检索到的文章。但是面向MEDLINE可用的30多个检索服务中，只有几个估计一个相关性得分，并且没有检测和合并查询单词间的关系作为相关性得分的一部分。结果：我们开发了"Relemed"，面向MEDLINE的一个搜索引擎。Relemed通过搜索句子中的查询单词而非整篇文章提高了检索的特异性和准确性。它使用句子水平的同时出现作为单词间关系存在的一种统计替代。它也估计了一个相关性得分，并基于此排序结果，从而移动不相关的文章到列表的尾部。在两个案例研究中，我们证明了最相关的文章出现在Relemed结果的顶部，而这不一定是PubMed搜索中的情况。我们也显示Relemed搜索不仅包括由PubMed检索到的所有文章，而且包括潜在额外的相关文章，归因于在Relemed中实现的扩展的"自动术语映射"和文本-单词搜索特征。结论：通过使用句子水平的匹配，Relemed能够提供更高的特异性，因此排除了更高假阳性的文章。通过引入一个合适的相关性矩阵，用户期望关注的最相关的文章被首先列出。Relemed也减少了显示文本，并因此缩短了花费在扫描文章上的时间。

根据它的摘要，我在其网站上以两个单词为关键词进行搜索。它的确能够搜到较为精确的文献，当在PubMed上搜到文献量过多，结果中包含太多无关文献时可以用这个来搜一下，不过由于用到的是两个或以上的单词，需要对搜索的背景有一定的了解。

1. PubFocus

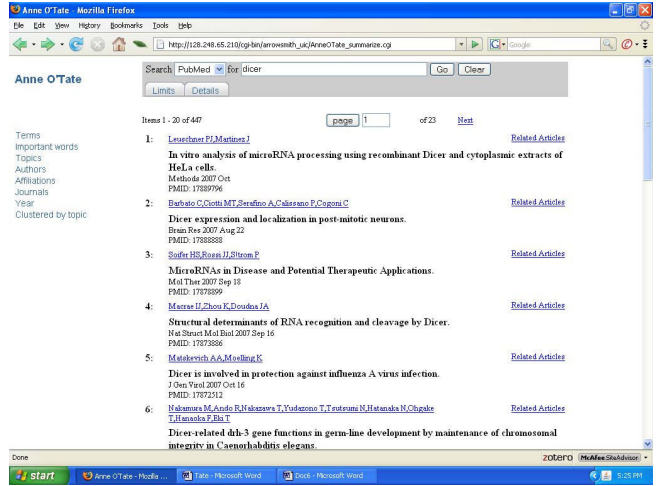
PubFocus则利用一些特定的参数来控制文献排序，例如影响因子、每一作者的贡献、文献历史、文献动态（reference dynamics）。



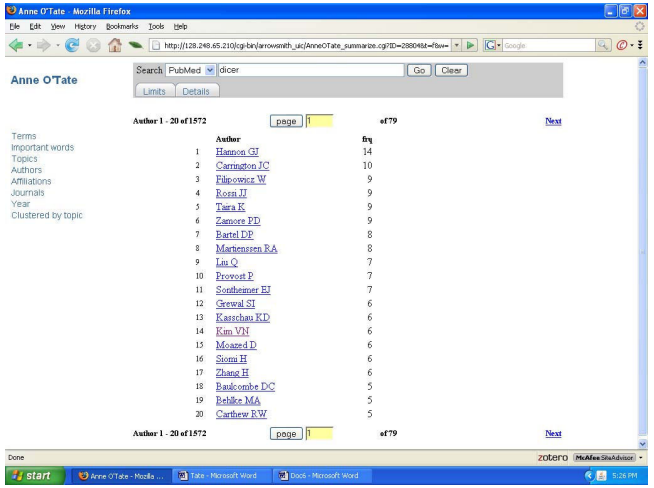
它的用法与PubMed差不多，就是通过关键词进行搜索文献，但是精确度要高一些，要是用Advanced Search锁定的文献量会更少更准确，不过它的sorted Result一直未能弄出来，应该能够将文献结果进行一个分类。

1. Anne O’Tate

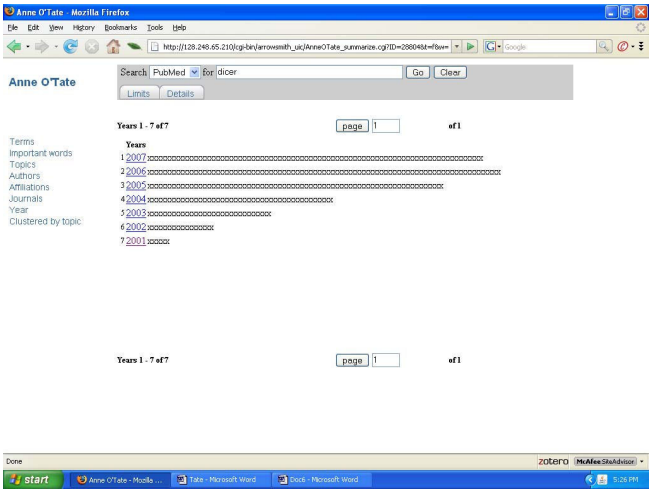
它的网址打不开，就将它文章中的相关图片看了一下。



搜索的界面和PubMed的差不多，应该也很好上手。



上图是查看结果文献中作者的统计信息。

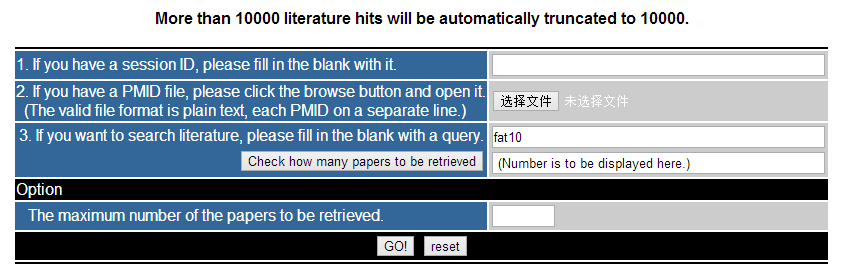


年份的统计

因为未能亲自使用，只能从文章来看，它的特色估计是能从结果中进行信息的统计。

1. McSyBi

McSyBi最重要的一点是可以确定参考文献间的相关性,并允许用户对其进行重排。



这是它的搜索界面。

1. PuReD—MCL

PuReD—MCL （PubMedRelatedDocuments—MCL)是一个基于图的 PubMed的文档聚类方法.它不直接使用自然语言处理技术,而是充分利用PubMed现有资源,然后使 用基于流动模拟图的MCL图聚类算法.这一过程允许用户利用重要线索分析结果,最终使用交互式图形布局算法,可视化群簇和所有相关信息。

这个实在未能找到它的网址。

11、GoPubMed

Gopubmed是一种Pubmed检索工具，是对于Pubmed资源的另一种组合检索方式，它比Pubmed的优点集中在于对于某一方面的研究我们能够直观得知研究时间、研究者、地域、研究过程、研究动态和各种信息之间的相关性？摘自上海交通大学-生命医药学科博客：http://blog.lib.sjtu.edu.cn/lam/article.asp?id=269

GoPubMed其数据源跟美国国立医学图书馆的PubMed完全一样，其本身并没有数据库，GoPubMed：用基因本体和医学主题词表对PubMed检索结果进行探索。 其原理是将读者检索提问词提交给PubMed，接收PubMed的检索结果，利用GO（gene ontology,GO-基因本体）和MESH（医学主题词表）对检索结果进行提炼，利用算法从中提取GO术语和MESH主题词，自动生成临时基因本体和医学主题词表，从而对检索检索进行分类，读者可以根据这些分类快速找到自己需要的文献，而不需要将检索到所有文献进行阅读，节省时间！

12、XPlorMed

<http://www.doc88.com/p-2999819498317.html>这网址是它比较详细的介绍。

它也是对从PubMed上搜到的结果进行过滤。

13、FACTA: a text search engine for finding associated biomedical concepts

网址：<http://text0.mib.man.ac.uk/software/facta/main.html>

基于MEDLINE摘要的文本挖掘工具，帮助用户浏览生命医学概念，如genes/proteins，diseases，enzymes和化合物。通过文本格式展示给用户，基于出现频率来排序。

搜索词汇：free keywords or Boolean combinations of keywords/concepts

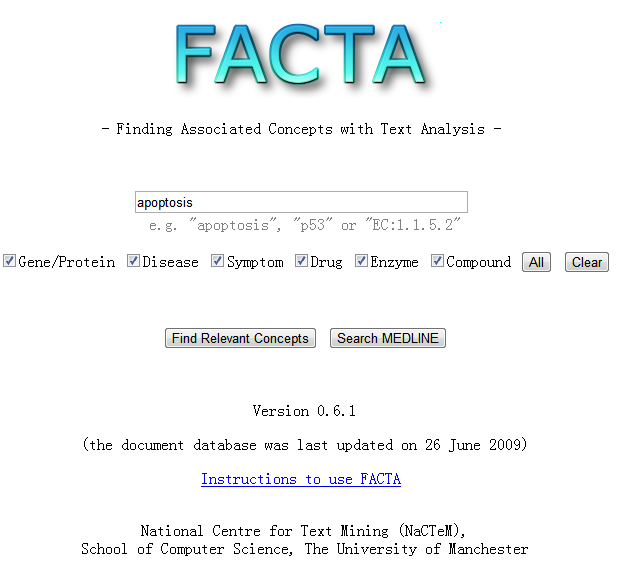
针对搜索内容能够快速展示出结果；借助了Uniprot (the concept IDs for genes/proteins)，BioThesaurus(collected their names and synonyms)，UMLS (diseases and symptoms)，HMDB (The concept IDs and names for drugs)，KEGG(The concept IDs and names for enzymes)，DrugBank (The concept IDs and names for chemical compounds)数据库

优势： a flexible query andprovide real-time responses

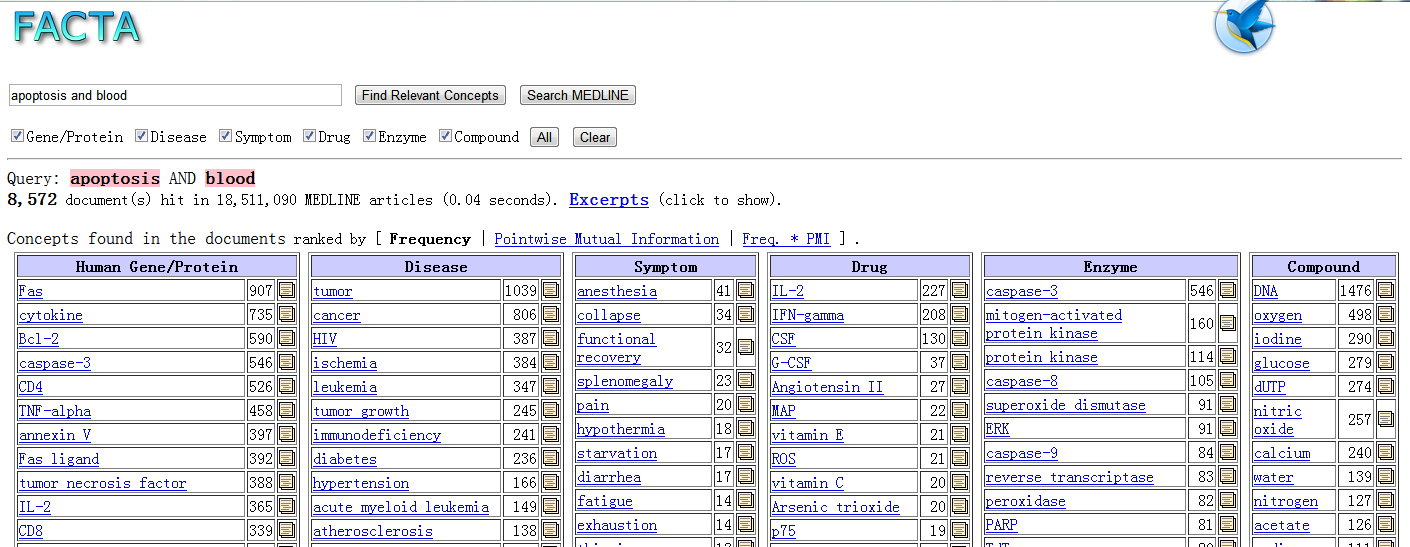
( online computation of association statistics—FACTA analyzes the documents retrieved by the query dynamically, using pre-indexed words and concepts)

缺点：index的时候会出现一些Ambiguity， intra-category ambiguity

软件界面：



搜索结果：



The system then retrieves all the documents that match the query from MEDLINE using word/concept indexes. The concepts contained in the documents are then counted and ranked according to their relevance to the query. The results are presented to the user in a tabular format.

索引的方法：indexing scheme and implementation of the analysis engines in C++

 two indexes built offline—one for the words and the other for the concepts

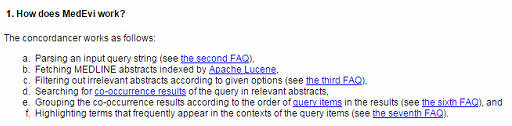
排序的方法： pointwise mutual information, which is defined as log p(*x*, *y*)/(*p*(*x*)*p*(*y*)), where *p*(*x*) is the proportion of the documents that match the query, *p*(*y*) is the proportion of the documents that contain the concept, and *p*(*x*, *y*) is the proportion of the documents that match the query and contain the concept. Pointwise mutual information gives an indication of how much more the query and concept co-occur than we expect by chance. For example, if their occurrences are completely independent (i.e. *p*(*x*, *y*)=*p*(*x*)*p*(*y*)), the measure gives a value of zero.

改进：

Discovering and visualizing indirect associations between biomedical concepts

FACTA+：<http://www.nactem.ac.uk/facta/>

(i) detecting biomolecular events in text using a machine learning model, (ii) discovering hidden associations using co-occurrence statistics between concepts, and (iii) visualizing associations to improve the interpretability of the output.

14 MedEvi http://www.ebi.ac.uk/Rebholz-srv/ebimed/

软件界面：

