



链滴

DocFetcher 使用教程——高效的文档文本检索工具

作者: [wangzhijun97](#)

原文链接: <https://ld246.com/article/1568960776746>

来源网站: 链滴

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

<p></p>

<h2 id="DocFetcher简介">DocFetcher 简介</h2>

<p>DocFetcher 是一个使用 Java 语言开发的开源桌面搜索应用程序：它允许您搜索计算机上的文内容，能够在 Windows, Linux 和 OS X 上运行。</p>

<p>本人在工作中已经使用 DocFetcher 超过半年时间，主要用来在代码库中查找包含关键字或关键的组合的源码文件，基本能够满足我的工作中检索源码的需要。而 Docfetcher 奇快的查找速度着实了我不种草都不行的理由，当然，Docfetcher 也并非完美无瑕，在日常使用中也会发现其存在一些 bu，但不影响主要功能的使用。</p>

<p></p>

<h2 id="下载安装">下载安装</h2>

<p>运行环境：Java Runtime Environment (JRE) 1.7 +</p>

<p>下载地址：http://docfetcher.sourceforge.net/en/download.html</p>

<p>安装方法：windows 平台请使用 .exe 安装包进行安装，其他平台请参照官方文档</p>

<p></p>

<h2 id="使用方法">使用方法</h2>

<h3 id="part1-建立索引">part1 建立索引</h3>

<p>为了快速的进行文本检索，DocFetcher 在检索文档文本前需要建立索引，用户可以根据需要为文件夹或者整个磁盘建立索引，当然，仅仅在第一次检索时需要为所要检索的文件夹或磁盘建立索引，立索引后，当文件内容发生了变化 DocFetcher 会自动更新索引。以下是建立索引的详细方法：</p>

<h4 id="step-1">step 1</h4>

<p>如图所示，右击[搜索范围]框的空白处，选择弹出菜单中的[建立索引 - 资料夹]。</p>

<p></p>

<h4 id="step-2">step 2</h4>

<p>在弹出的选择窗口中选择需要检索文档的位置，点击[确认]。</p>

<p></p>

<h4 id="step-3">step 3</h4>

<p>在弹出的窗口中可以对索引的细节进行设置，可以选择所要检索的文件格式，在排除列表中可以用正则表达式过滤不需要建立索引的文件和文件格式，其他选项没有特殊要求使用默认值即可。最后择[执行]，将开始建立索引。</p>

<p></p>

<h4 id="step-4">step 4</h4>

<p>建立索引的过程耗时较长，具体时间取决于所选建立索引的文件夹中包含的文件数量及文件大小索引建立完成后在首页的[搜索范围]框内便有了对应的目录。</p>

<p></p>

<h3 id="part2-检索数据">part2 检索数据</h3>

<p>对于建立好索引的目录，可以通过勾选检索目录、勾选文件类型、设定文件大小、输入关键字来进行文本检索。</p>

<h4 id="1--检索语法">1. 检索语法</h4>

<p>这里仅对 DocFetcher 的检索语法做简单介绍，详细内容请参照官方文档。</p>

<p>短语搜索：需要将所要检索的短语放入英文双引号内。如：“可转债”，返回包含词语“可转债”的件；</p>

<p>布尔运算：支持复杂条件的检索，常用运算符包含 AND、OR、NOT 等，短语间没有运算符时默认使用 OR。如 "可转债" AND "可交换债"，返回既包含短语"可转债"又包含短语"可交换债"的文
; </p>

<p>通配符：问号（<code>'?'</code>）和星号（<code>'*</code>）可用于表示某些字符未知如："可*", 返回包含"可以"、"可转债"、"可交换债"等包含以"可"开头的语句的文件；</p>

<p>模糊搜索：可以搜索与给定单词类似的单词。如，roam~，返回包含 foam 和的词的文档 roams 的文件。</p>

<h4 id="2-检索示例">2.检索示例</h4>

<p>检索包含短语"Bond Issue"的文件，点击检索结果里的文件可以在下方的预览框预览文件的内容预览中命中的检索关键字将被标红，如下图所示。

 </p>

<p>检索即包含"Bond Issue"和短语"WBPJRQ"的文件，如下图所示。

 </p>

<h3 id="part3-设置">part3 设置</h3>

<p>通过点击右上角的齿轮状按钮可以进入[设置]。在设置中可以对默认布尔运算符进行修改、自动览一个人符合的项目、使用及时搜索等，还可以对预览的字体，命中文本的颜色，快捷键等进行设置</p>

<p> </p>

<h3 id="part4-注意事项">part4 注意事项</h3>

<h4 id="1--内存大小限制">1. 内存大小限制</h4>

<p>DocFetcher 的默认内存限制为 256 MB，因此 DocFetcher 运行时占用的内存大小和资源都是当小的，但是当检索速度过慢或者索引建立时存在单个文件过大而导致索引建立失败则需要提高内存小。</p>

<p>Windows 平台下，DocFetcher 内置了多个不同内存大小的启动器，其位于 DocFetcher 安装录下的 misc 目录下。</p>

<p>在 misc 目录下，文件名"DocFetcher-2048 64-bit-Java.exe"中的"-2048"即是表示使用 2048 B 的内存启动，"64-bit-Java"表示为 64 位的启动程序，适用于 64 位系统。可根据自己的实际情况选择合适的启动器。</p>

<p>很重要的一点是，这些备用启动器在使用时需要将其从 misc 目录复制到 DocFetcher 安装目录，复制完成后，双击该启动器即可启动对应内存大小的 DocFetcher。其他平台的增加内存方法请参照 DocFetcher-设置-说明。</p>

<h4 id="2-检索关键字格式">2.检索关键字格式</h4>

<p>检索中文，因为 DocFetcher 中文分词器的原因，如果想搜索整个词语请务必使用英文双引号括，否则将导致 DocFetcher 将词语分词后使用 OR 运算进行检索，最终使检索结果出乎意料。</p>

<h4 id="3-搜寻范围越界的问题">3.搜寻范围越界的问题</h4>

<p>使用 DocFetcher 检索时，可能只是勾选了一个搜寻目录，但搜索结果可能包含了未勾选的目录内文件，这可能是 DocFetcher 的一个 BUG，目前暂不清楚原因。解决方法是尝试勾选其他目录，取消勾选，再点击搜索。</p>

<hr>

<h2 id="P-S">P.S</h2>

<p>欢迎反馈与交流。</p>

<p>转载或引用本文请注明来源网址与作者。 </p>

<p>版权所有，转载或引用本网站内容不得进行如下活动： </p>
<blockquote>
<p>a. 损害本人或他人利益； </p>
<p>b. 任何违法行为； </p>
<p>c. 任何可能破坏公秩良俗的行为； </p>
<p>d. 擅自同意他人继续转载、引用本网内容。 </p>
</blockquote>

