Boost搜索引擎

正排索引：根据文档id找到文档内容。

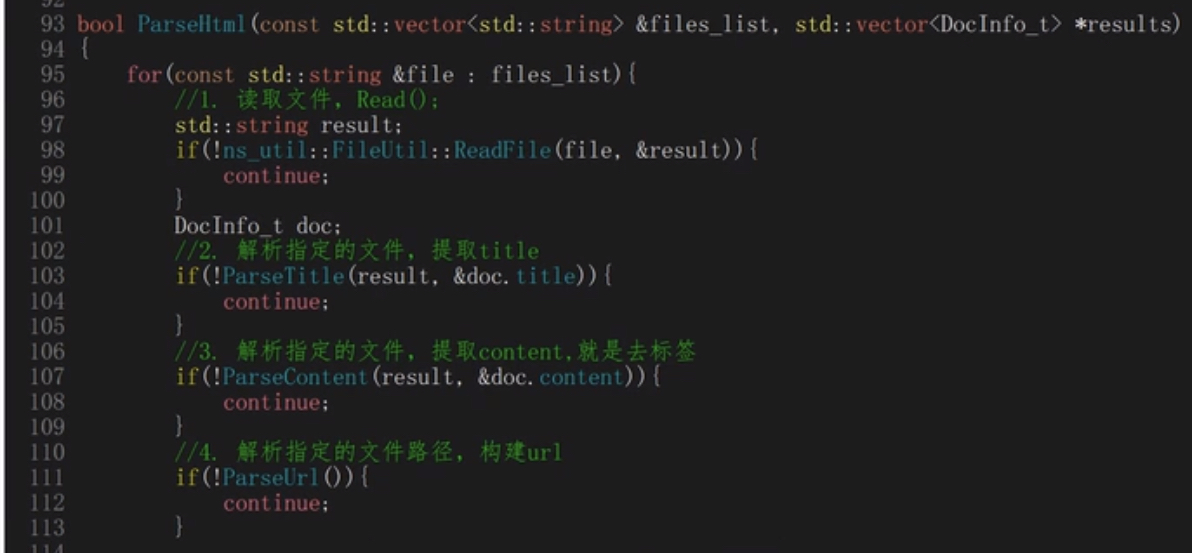
倒排索引：根据关键字找到相应的文件id。

查找过程：倒排索引 -》 文档id -》 正排索引 -》文档内容摘要。

使用boost进行相关的搜索内容。

需要去除相应的文件之中的<>之中的文件信息，使用程序进行去除相关的信息。命名为pares.cc。 需要使用boost\_devel库里面的东西进行访问。

1. 首先让文件储存到路径之中，方便读取。
2. 按照读取每一个文件的内容。（进行去标签化等操作！）



1. 解析文件以/3为分隔符。

URL = url\_head + url\_tail;两部分组成相关的情况。获取路径名字。