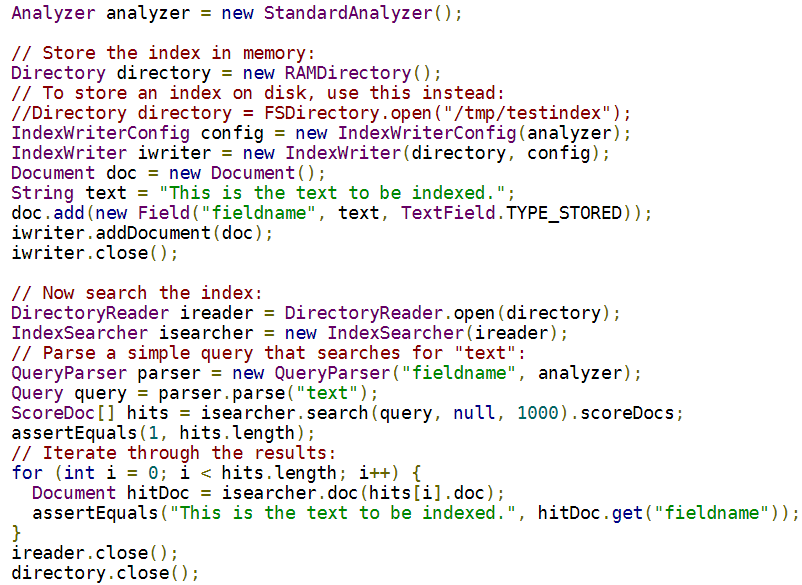
Lucene, an indexing and search library, accepts only plain text input.



To use Lucene, an application should:

1. Create Documents by adding Fields;
2. Create an IndexWriter and add documents to it with addDocument();
3. Call QueryParser.parse() to build a query from a string;
4. Create an IndexSearcher and pass the query to its search() method.

Parsing转换

 Lucene does not care about the Parsing of these and other document formats, and it is the responsibility of the application using Lucene to use an appropriate Parser to convert the original format into plain text before passing that plain text to Lucene.

Tokenization分词

Tokenization is the process of breaking input text into small indexing elements – tokens. The way input text is broken into tokens heavily influences how people will then be able to search for that text.

Core Analysis分析

The analysis package provides the mechanism to convert Strings and Readers into tokens that can be indexed by Lucene.

文档A Document is a collection of IndexableFields. A IndexableField is a logical representation of a user's content that needs to be indexed or stored.IndexableFields have a number of properties that tell Lucene how to treat the content (like indexed, tokenized, stored, etc.)

Collectors are primarily meant to be used to gather raw results from a search, and implement sorting or custom result filtering, collation, etc.搜索结果的过滤器