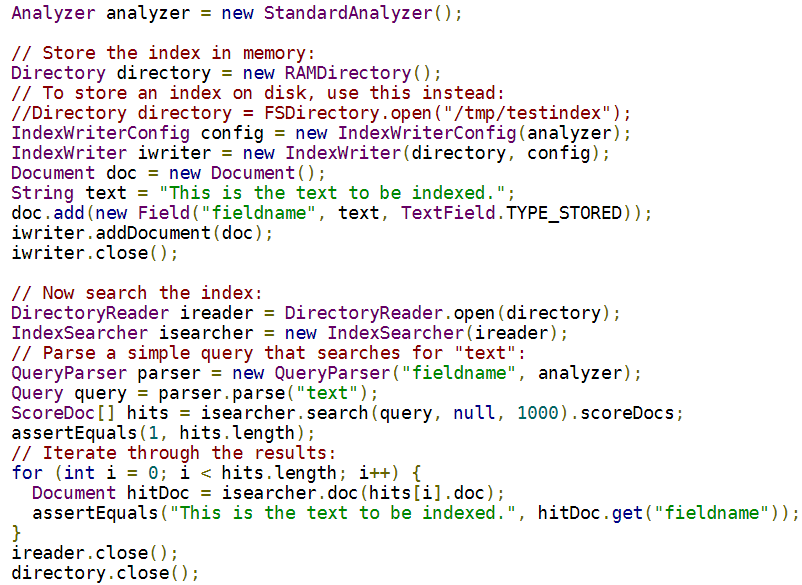
Lucene, an indexing and search library, accepts only plain text input.



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