工作

Lucene

**Lucene** **是一个高效的，基于Java** **的全文检索库。**

**1 创建索引，通过IndexWriter对不同的文件进行索引的创建，并将其保存在索引相关文件存储的位置中。**

**2 通过索引查寻关键字相关文档。**

如何创建索引

### 将原文档传给分词器(Tokenizer)。

### 通过分词结果建立倒排索引：上面经过解析后的对应关系是：“文章号”对“文章中所有关键词”。 这种对应比较符合我们正常的思维习惯，反过来对应的话，就要颠倒思维，所以这种算法叫做倒排！ 倒排索引把这个关系倒过来，变成：“关键词”对“拥有该关键词的所有文章号”。记录关键词在文章中出现次数和出现的位置

输入搜索语句，在倒排索引中找文件

**分词器(Tokenizer)会做以下几件事情(** **此过程称为Tokenize)** **：**

**1.** **将文档分成一个一个单独的单词。**

**2.** **去除标点符号。**

**3.** **去除停词(Stop word)** **。**

gitHub

Http在测试中很常用。HttpClient的原始用法有些不容易使用，因为如果不是很熟悉就会不知道各种函数以及其用法和调用顺序。在读effective java时，看到builder模式，所以就想到可以用builder模式简化httpClient的开发。比如Headers，有时会忘掉怎么加以及加什么，如果用build模式就可以通过eclipse的提醒功能直接一个一个写，把Header的属性放到一个类中，用static final标识。而且还用了工厂函数优于构造函数的原则，即newInstance

Header[] headers=HttpHeader.newInstance().userAgent("Mozilla/5.0").connection(HttpConstants.KEEP\_ALIVE).contentType(HttpConstants.APPLICATION\_JSON).build();

如果不用这个就得

Headers = new BasicHeader(userAgent, "Mozilla/5.0");

类似：

HttpClient client = EasyHttpClientBuilder.newInstance().configSSL().timeout(timeout).pool(maxTotal, maxPerRoute).build();

这里把比较复杂的ssl也通过构造模式写出来，因为ssl需要

\* 1.重写X509TrustManager和HostnameVerifier方法，使客户端信任服务器证书

\* 2.通过SSLContext设置SSL的算法，证书等上下文

\* 3.通过SSLConnectionSocketFactory（基于SSLContext和TrustManager）产生ssl socket

比较麻烦，这里直接用configSSL函数，直接在RegistryBuilder里注册了Socket，还用了PoolingHttpClientConnectionManager配置maxTotal最大连接数和maxPerRoute路由连接数

Builder模式可以隐藏复杂的方法调用关系，而且可以动态地增减，挑选需要的配置

EasyHttpClient还用了配置类，HttpConfig，它有httpClient，headers，url等配置，这样可以把所有配置都集中到一块来，通过eclipse提示简化操作

response = EasyHttpClient.get(config.headers(headers).url(baseUrl));

还有httpmethod类用了enum，就像effective java里说的enum更优

总的来说分三块，一个是Builder，EasyHttpClientBuilder，负责通过工厂模式创建httpClient；Header等通过build模式创建头。

第二是HttpConfig，存储需要的所有参数和httpClient对象。

最后是EasyHttpClient，负责执行http方法

其中，一共用了effective java里的三个原则，工厂函数优于构造函数；builder模式；enum更优（单例，可读性更好）

感想：