链滴

Lucene 中 Analyzer 语句分析

作者：Sysecho
原文链接：https://ld246.com/article/1498787825766
来源网站：链滴

Lucene中Analyzer语句分析,利用lucene中自带的词法分析具Analyzer，进行对句子的分析。

源码如下：

```
package com.test;

import java.io.IOException;
import java.io.StringReader;
import java.util.List;

import org.apache.lucene.analysis.Analyzer;
import org.apache.lucene.analysis.SimpleAnalyzer;
import org.apache.lucene.analysis.StopAnalyzer;
import org.apache.lucene.analysis.Token;
import org.apache.lucene.analysis.TokenStream;
import org.apache.lucene.analysis.WhitespaceAnalyzer;
import org.apache.lucene.analysis.standard.StandardAnalyzer;
import org.apache.lucene.analysis.tokenattributes.TermToBytesRefAttribute;
import org.apache.lucene.util.Version;

import com.bean.mashupDerscriptionTest;
import com.daoImpl.MashupDaoImpl;
import com.gargoylesoftware.htmlunit.javascript.host.Comment;

public class KeyWordsTest {

    /**
     * @param args
     */
    public static void main(String[] args) {
        MashupDaoImpl mashupDao = new MashupDaoImpl();
        List<mashupDerscriptionTest> list = mashupDao
                .findAllmashupDescripteonTest();
        int i = 1;
        String comment = null;
        for (mashupDerscriptionTest mashup : list) {
            // 描述为空去名字作为描述
            if (mashup.getComments().equals("")) {
                comment = mashup.getName();
            } else {
                comment = mashup.getComments();
            }
//              System.out.println(comment);
            //对读取的描述利用Lucene中的Analyzer进行句子分析产生
            //空格及各种符号分割,去掉停止词，停止词包括 is,are,in,on,the等无实际意义的词
            StringReader reader = new StringReader(comment);
            Analyzer analyzer = new StopAnalyzer();
            TokenStream tStream = analyzer.tokenStream("", reader);
            Token t;
            try {
                while ((t = tStream.next()) != null) {
```

```
                //对每个单词采用
                System.out.print(t.termText()+" ");
            }
            System.out.println((i++)+"条描述分词结束！");
        } catch (IOException e) {
            e.printStackTrace();
        }
    }
  }
}
```

注:数据来源于数据库中.....