牛人计划-高级项目课（13）

牛客网
NOWCODER

- 全文搜索
- solr安装
- solr中文分词
- solr数据库导入
- solr数据查询
- solrj接口调用

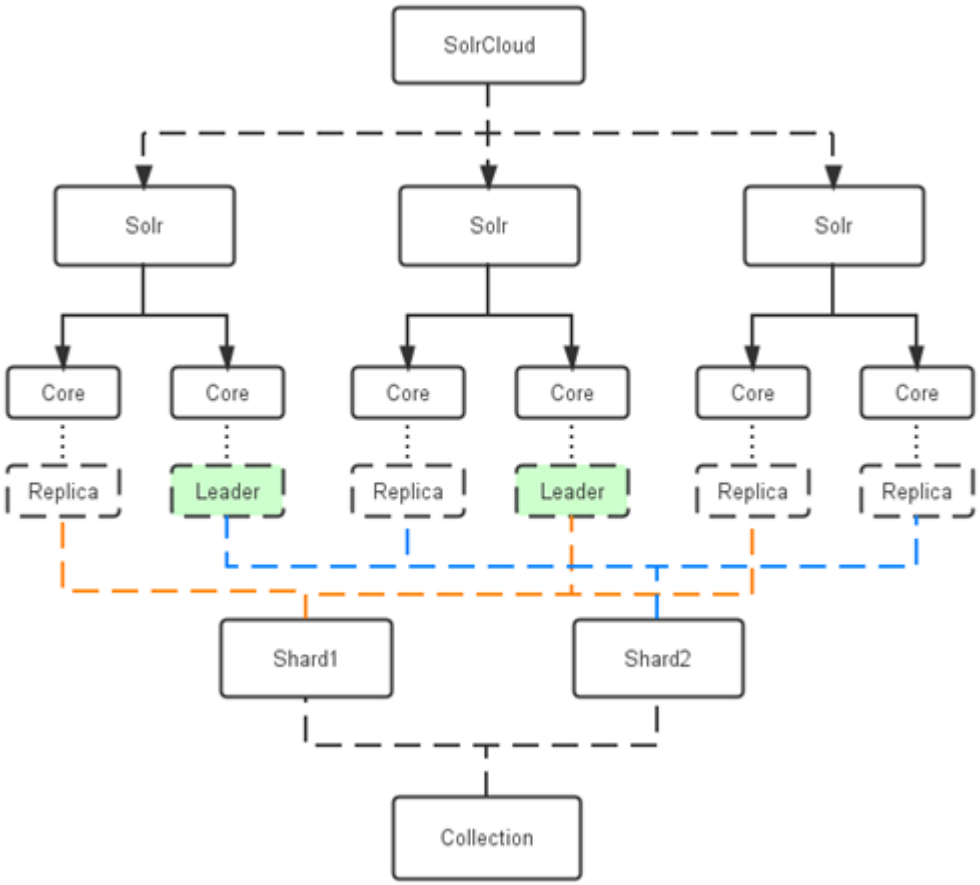
Solr简介

官方网站：http://lucene.apache.org/solr/

搜索语法：

1. +包含关键词
2. -不需要某关键词

Solr简介



云模式 solr start -e cloud –noprompt
单机 solr start
solr create_collection wenda

核心概念：
Document：每个被索引的文档
Field：文档里的各个属性值

导入数据
java -Dauto -Dc=gettingstarted -Drecursive=yes -jar example\exampledocs\post.jar docs/

中文分词(IK-Analyzer)

源码地址：
https://code.google.com/archive/p/ik-analyzer/downloads
http://git.oschina.net/wltea/IK-Analyzer-2012FF

采用了特有的正向迭代最细粒度切分算法
支持细粒度和智能分词两种切分模式

分词效果	useSmart（true）	useSmart（false）
你好清华	你好|清华	你好|你|好清|清华

IKAnanlyzer分词配置（managed-schema）

```
<field name="question_title" type="text_ik" indexed="true" stored="true" multiValued="true"/>
<field name="question_content" type="text_ik" indexed="true" stored="true" multiValued="true"/>
```

```
<fieldType name="text_ik" class="solr.TextField">
  <!--索引时候的分词器-->
  <analyzer type="index">
    <tokenizer class="org.wltea.analyzer.util.IKTokenizerFactory" useSmart=“false"/>
    <filter class="solr.LowerCaseFilterFactory"/>
  </analyzer>
  <!--查询时候的分词器-->
  <analyzer type="query">
    <tokenizer class="org.wltea.analyzer.util.IKTokenizerFactory" useSmart=“true"/>
  </analyzer>
</fieldType>
```

数据库数据导入

solrconfig.xml

```
<requestHandler name="/dataimport" class="org.apache.solr.handler.dataimport.DataImportHandler">
    <lst name="defaults">
      <str name="config">data-config.xml</str>
    </lst>
</requestHandler>
```

data-config.xml

```
<dataConfig>
  <dataSource type="JdbcDataSource"
          driver="com.mysql.jdbc.Driver"
          url="jdbc:mysql://localhost/wenda"
          user="root"
          password="nowcoder"/>
  <document>
    <entity name="question"  query="select id,title,content from question">
      <field column="content" name="question_content"/>
      <field column="title" name="question_title"/>
    </entity>
  </document>
</dataConfig>
```

数据库相关jar包导入，参考资料：http://wiki.apache.org/solr/DIHQuickStart

IK-Analyzer（自己编译）

IKAnalyzer.java

```
public final class IKAnalyzer extends Analyzer{
    /**
     * 重载Analyzer接口，构造分词组件
     */
    @Override
    protected TokenStreamComponents createComponents(String fieldName) {
        Tokenizer _IKTokenizer = new IKTokenizer(this.useSmart());
        return new TokenStreamComponents(_IKTokenizer);
    }

}
```

IKTokenizerFactory.java

pom打包带dic文件

```
public

    @Override
    public Tokenizer create(AttributeFactory attributeFactory) {
        Tokenizer tokenizer = new IKTokenizer(useSmart);
        return tokenizer;
    }
}
```

搜索结果页（代码演示）

课后练习

1. SQL like实现标题搜索
2. 中文分词配置
3. 数据库增量更新配置
4. 索引评论表

Thanks