项目介绍

项目名称：Jcseg
码云地址：http://git.oschina.net/lionsoul/jcseg
开发作者：狮子的魂，冬芽
项目简介：Jcseg是一个轻量级中文分词器，同时集成了关键字提取，关键短语提取，关键句子提取和文章自动摘要等功能，并且提供了一个基于Jetty的web服务器，方便各大语言直接http调用，同时提供了最新版本的lucene, solr, elasticsearch的分词接口！

狮子的魂 / jcseg

Java

捐赠 0 Watch 13

代码 Issues 50 Pull Requests 0 附件 0 Wiki 1 统计 服务 管理

Jcseg是基于mmseg算法的一个轻量级中文分词器，同时集成了关键字提取，关键短语提取，关键句子提取和文章自动摘要等功能，并且提供了一个基于Jetty的web服务器，方便各大语言直接http调用，同时提供了最新版本的lucene，solr和elasticsearch的分词接口！
https://github.com/lionsoul2014/jcseg -- 编辑

407 Commits 2 Branches 10 Tags 0 Releases

Pull Request Issue 插件 克隆/下载

lionsoul 最后提交于 8小时前 . fix the lexicon part-of-speech bug reported at ...

jcseg-analyzer	lionsoul	Optimize the JcsegTokenizer implementation for Attributes clear	25天前
jcseg-core	lionsoul	Fix the bug reported at http://git.oschina.net/lionsoul/jcseg/issues/49	10天前
jcseg-elasticsearch	lionsoul	Disable the doc strict check for java8	2月前
jcseg-server	lionsoul	Disable the doc strict check for java8	2月前
vendors	lionsoul	fix the lexicon part-of-speech bug reported at https://github.com/lion...	8小时前
.gitignore	lionsoul	bug fixed for src ignore	4月前
CHANGES.md	lionsoul	Optimize the datetime entity demo	10天前
LICENSE.md	lionsoul	add Free statement under the license	3月前
README.md	lionsoul	Update the Isegment create doc	2月前
build.xml	lionsoul	Prepare the release of version 2.1.0	2月前
jcseg-server.properties	lionsoul	some optimization work for the server module	4月前
jcseg.properties	lionsoul	NLPSeg forward with mixed word recognition optimazation impelment...	4月前
pom.xml	lionsoul	Disable the doc strict check for java8	2月前

码云项目地址：

http://git.oschina.net/lionsoul/jcseg

中文分词

(1).简易模式：FMM算法。
(2).复杂模式：MMSEG四种过滤算法，具有较高的歧义去除，分词准确率达到了98.41%（算法作者的数据）。
(3).检测模式：只返回词库中已有的词条。
(4).检索模式：细粒度切分，专为检索而生。
(5).分隔符模式：按照给定的字符切分词条，默认是空格。
(6). NLP模式：继承自复杂模式，增加电子邮件，大陆手机号码，网址，人名，地名，datetime日期，货币等以及无限种自定义实体的识别与返回。

其他核心功能

2，关键字提取：基于textRank算法。
3，关键句子提取：基于textRank算法。
4，文章自动摘要：基于BM25+textRank算法。
5，自动词性标注：目前只是基于词库，效果不是很理想。
6，Restful api：嵌入jetty提供了一个server模块，包含全部功能的http接口，标准化json输出格式，方便各种语言客户端直接调用。

发展计划

长期目标：打造成一个NLP处理框架。
1，完善功能：分词，词性标注，关键字提取和自动摘要采用统计模型优化。
2，新增功能：文本分类，情感分析。