Atitit nlp用到的技术与功能自然语言处理 v3 t99.docx

Atitit nlp用到的技术与常见类库

目录

[1. 常用的技术 1](#_Toc11696)

[1.1. 语言处理基础技术 分词 相似度等 1](#_Toc31070)

[1.2. 新闻摘要 2](#_Toc26727)

[1.3. 情感倾向分析 2](#_Toc7233)

[1.4. 文章标签 2](#_Toc22599)

[1.5. 文章分类 2](#_Toc20007)

[1.6. 文本审核 2](#_Toc7325)

[1.7. 机器翻译 2](#_Toc22309)

[1.8. 智能写作平台 3](#_Toc21317)

[1.9. other 3](#_Toc17265)

[2. Ati常用技术 3](#_Toc20020)

[2.1. 文档读取处理类 3](#_Toc16709)

[2.2. 文本信息抽取 3](#_Toc32375)

[2.3. 关键词提取 4](#_Toc21092)

[2.4. 摘要 关键词提取 与目录提取 4](#_Toc2047)

[2.5. 分词，ik分词器 4](#_Toc5941)

[2.6. 索引与检索类 4](#_Toc19697)

[2.7. 情感分析（歌词） 4](#_Toc26453)

[3. Ref 4](#_Toc29780)

# 常用的技术

## [语言处理基础技术](https://ai.baidu.com/tech/nlp_basic" \t "https://ai.baidu.com/tech/_blank) 分词 相似度等

[词法分析](https://ai.baidu.com/tech/nlp_basic/lexical" \t "https://ai.baidu.com/tech/_blank)

[依存句法分析](https://ai.baidu.com/tech/nlp_basic/dependency_parsing" \t "https://ai.baidu.com/tech/_blank)

[词向量表示](https://ai.baidu.com/tech/nlp_basic/word_embedding" \t "https://ai.baidu.com/tech/_blank)

[DNN语言模型](https://ai.baidu.com/tech/nlp_basic/dnnlm_cn" \t "https://ai.baidu.com/tech/_blank)

[词义相似度](https://ai.baidu.com/tech/nlp_basic/word_emb_sim" \t "https://ai.baidu.com/tech/_blank)

[短文本相似度](https://ai.baidu.com/tech/nlp_basic/simnet" \t "https://ai.baidu.com/tech/_blank)

[语言处理应用技术](https://ai.baidu.com/tech/nlp_apply" \t "https://ai.baidu.com/tech/_blank)

[文本纠错](https://ai.baidu.com/tech/nlp_apply/text_corrector" \t "https://ai.baidu.com/tech/_blank)

[对话情绪识别](https://ai.baidu.com/tech/nlp_apply/emotion_detection" \t "https://ai.baidu.com/tech/_blank)

## [新闻摘要](https://ai.baidu.com/tech/nlp_apply/news_summary" \t "https://ai.baidu.com/tech/_blank)

## [情感倾向分析](https://ai.baidu.com/tech/nlp_apply/sentiment_classify" \t "https://ai.baidu.com/tech/_blank)

## [文章标签](https://ai.baidu.com/tech/nlp_apply/doctagger" \t "https://ai.baidu.com/tech/_blank)

[评论观点抽取](https://ai.baidu.com/tech/nlp_apply/comment_tag" \t "https://ai.baidu.com/tech/_blank)

## [文章分类](https://ai.baidu.com/tech/nlp_apply/topictagger" \t "https://ai.baidu.com/tech/_blank)

## [文本审核](https://ai.baidu.com/tech/textcensoring" \t "https://ai.baidu.com/tech/_blank)

## [机器翻译](http://fanyi-api.baidu.com/api/trans/product/index" \t "https://ai.baidu.com/tech/_blank)

[通用翻译API](http://fanyi-api.baidu.com/api/trans/product/prodinfo" \t "https://ai.baidu.com/tech/_blank)

[拍照翻译SDK](http://fanyi-api.baidu.com/api/trans/product/prodinfo" \l "5" \t "https://ai.baidu.com/tech/_blank)

[定制化翻译API](http://fanyi-api.baidu.com/api/trans/product/prodinfo" \l "2" \t "https://ai.baidu.com/tech/_blank)

[垂直领域API](http://fanyi-api.baidu.com/api/trans/product/prodinfo" \l "1" \t "https://ai.baidu.com/tech/_blank)

[语音翻译SDK](http://fanyi-api.baidu.com/api/trans/product/prodinfo" \l "4" \t "https://ai.baidu.com/tech/_blank)

[语种识别API](http://fanyi-api.baidu.com/api/trans/product/prodinfo" \l "3" \t "https://ai.baidu.com/tech/_blank)

## [智能写作平台](https://ai.baidu.com/tech/intelligentwriting" \t "https://ai.baidu.com/tech/_blank)

## other

* ****词性标注（Part-of Speech Tagging）:****
* ****拼写更正（Spelling Correction）:****基于"你要找的是"风格的检查引擎
* ****数据库文本挖掘（Database Text Mining）****
* ****字符串比较(String Comparison) ：****基于距离与相似度测量，包括权重距离，TF/IDF距离，Jaccard distance, Jaro-Winkler distance,等
* ****语言辨别（Language Identification）****
* 推荐系统
* 【NLP】十分钟快览自然语言处理学习总结 - 伏草惟存 - 博客园.mhtml
* 这个什么有详细的推荐算法

# Ati常用技术

## 文档读取处理类

mail类库，poi类库（处理officex系列），jsoup html类库。

org.apache.tools.zip\_1.9.6 处理zip库

java-unrar 处理rar文档

pdfbox-1.8.16.jar

## 文本信息抽取

Zip压缩文档抽取技术

Office pdf系列文档读取技术

Eml html系列温暖的读取技术

分词

## 关键词提取

****LingPipe的优势是: 貌似不包含关键词提取 内容提取****

FudanNLP进行新闻关键词提取

## 摘要 关键词提取 与目录提取

Html文档摘要提取

索引与 检索

## 分词，ik分词器

## 索引与检索类

Lucence imap类库mail。Jar zip库读取库

## 情感分析（歌词）

# Ref

百度ai

Atitit nlp 自然语言处理的艺术 attilax著作 v2 t55.docx

Atitit nlp 自然语言处理attilax总结 dep

java 语义分析 开源 - fengbin2005 - ITeye博客.mhtml

(17条消息)利用FudanNLP进行新闻关键词提取 - qq\_43685315的博客 - CSDN博客.mhtml

Atitit java的nlp自然语言处理类库v2 t99.docx