Atitit nlp常用的类库和技术

目录

[1. Atitit nlp用到的技术与常见类库 1](#_Toc7157)

[2. Atitit python nlp类库 1](#_Toc23661)

[2.1. Atitit java的nlp自然语言处理类库 2](#_Toc16367)

[3. Other 2](#_Toc13466)

[3.1. Mmseg分词，ik分词器 2](#_Toc3897)

[3.2. 文档读取处理类 2](#_Toc21225)

[3.3. 索引与检索类 2](#_Toc21726)

# Atitit nlp用到的技术与常见类库

目录

1. 常用的技术 1

1.1. 语言处理基础技术 分词 相似度等 1

1.2. 新闻摘要 2

1.3. 情感倾向分析 2

1.4. 文章标签 2

1.5. 文章分类 2

1.6. 文本审核 2

1.7. 机器翻译 2

1.8. 智能写作平台 3

2. Ati常用技术 3

2.1. 分词 关键词提取 3

2.2. 摘要 关键词提取 与目录提取 3

2.3. 抽取 3

2.4. 索引与 检索 3

2.5. Zip压缩文档抽取技术 3

2.6. Office pdf系列文档读取技术 3

2.7. Eml html系列温暖的读取技术 3

# Atitit python nlp类库

目录

1.1. snownlp 1

1.2. ·Spacy是NLTK的主要竞争对手。这两个库可用于相同的任务。 1

1.3. ·Scikit-learn为机器学习提供了一个大型库。此外还提供了用于文本预处理的工具 1

2. Ati 使用记录 2

3. ref 2

OpenNLP

## Atitit java的nlp自然语言处理类库

目录

1.1. OpenNLP是一个基于Java机器学习工具包， 1

1.2.  Java自然语言处理 LingPipe 1

1.3. 2.中文自然语言处理工具包 FudanNLP 1

1.3.1.  5、Stanford CoreNLP 斯坦福大学NLP 2

1.4. 1.IKAnalyzer 2

1.5. 3.FudanNLP 2

1.6. jcseg 3

# Other

## Mmseg分词，ik分词器

## 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

## 索引与检索类

Lucence imap类库mail。Jar zip库读取库

Atitit python nlp类库