Atitit nlp用到的技术与常见类库

目录

[1. 常用的技术 1](#_Toc4293)

[1.1. 语言处理基础技术 分词 相似度等 1](#_Toc30283)

[1.2. 新闻摘要 2](#_Toc26806)

[1.3. 情感倾向分析 2](#_Toc31024)

[1.4. 文章标签 2](#_Toc24193)

[1.5. 文章分类 2](#_Toc16510)

[1.6. 文本审核 2](#_Toc8741)

[1.7. 机器翻译 2](#_Toc28166)

[1.8. 智能写作平台 3](#_Toc24787)

[2. Ati常用技术 3](#_Toc19022)

[2.1. 分词 关键词提取 3](#_Toc14506)

[2.2. 摘要 关键词提取 与目录提取 3](#_Toc23521)

[2.3. 抽取 3](#_Toc6983)

[2.4. 索引与 检索 3](#_Toc17957)

[2.5. Zip压缩文档抽取技术 3](#_Toc7632)

[2.6. Office pdf系列文档读取技术 3](#_Toc5037)

[2.7. Eml html系列温暖的读取技术 3](#_Toc27241)

[3. Nlp常用类库 3](#_Toc28780)

[3.1. Mmseg分词， 3](#_Toc26159)

[3.2. 文档读取处理类 3](#_Toc14899)

[3.3. 索引与检索类 4](#_Toc30765)

[4. Ref 4](#_Toc16987)

# 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)

# Ati常用技术

## 分词 关键词提取

## 摘要 关键词提取 与目录提取

Html文档摘要提取

## 抽取

## 索引与 检索

## 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

# Nlp常用类库

## 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

# 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