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

目录

[1.1. 主要范畴 1](#_Toc18265)

[1.2. 研究难点 2](#_Toc31399)

[2. Ati涉及的领域(文档 tts 分词 抽取 摘要 检索） 3](#_Toc5880)

[3. 中文分词原理与实现 11 3](#_Toc8580)

[4. 第6章　信息提取 271 4](#_Toc7378)

[5. 第7章　自动摘要 304 4](#_Toc28101)

[6. 第8章　文本分类 319 4](#_Toc14150)

[7. 第9章　文本倾向性分析 364 4](#_Toc22245)

[8. 机器翻译 4](#_Toc12311)

[9. 第11章　语音识别 4](#_Toc14084)

[10. 信息检索 4](#_Toc31394)

[11. 应用 4](#_Toc14021)

[11.1. 问答系统 自动翻译 4](#_Toc31957)

[12. Atitit NLP---信息检索概论方面的书籍目录 4](#_Toc19692)

[12.1. Atitit 现代信息检索 4](#_Toc196)

[12.2. 《信息检索导论》(（美）曼宁...)【简介\_书评\_在线阅读】 - dangdang.html 5](#_Toc6163)

[12.3. Atitit 自然语言处理原理与实现 attilax总结 6](#_Toc7188)

[13. 参考 6](#_Toc19188)

[13.1. Atitit.数据检索与网络爬虫与数据采集的原理概论 6](#_Toc918)

[13.2. 推荐新书《自然语言处理原理与技术实现》 - 自然语言处理-炼数成金-Dataguru专业数据分析社区.html 7](#_Toc26651)

## 主要范畴

[编辑](https://baike.baidu.com/item/nlp/javascript:;)

文本朗读（Text to speech）/语音合成（Speech synthesis）

语音识别（Speech recognition）

中文自动分词（Chinese word segmentation）

词性标注（Part-of-speech tagging）

句法分析（Parsing）

自然语言生成（Natural language generation）

文本分类（Text categorization）

信息检索（Information retrieval）

信息抽取（Information extraction） 摘要

文字校对（Text-proofing）

问答系统（Question answering）

机器翻译（Machine translation）

自动摘要（Automatic summarization）

文字蕴涵（Textual entailment）

## 研究难点

[编辑](https://baike.baidu.com/item/nlp/javascript:;)

**单词的边界界定**

在口语中，词与词之间通常是连贯的，而界定字词边界通常使用的办法是取用能让给定的上下文最为通顺且在文法上无误的一种最佳组合。在书写上，汉语也没有词与词之间的边界。

**词义的消歧**

许多字词不单只有一个意思，因而我们必须选出使句意最为通顺的解释。

**句法的模糊性**

自然语言的文法通常是模棱两可的，针对一个句子通常可能会剖析（Parse）出多棵剖析树（Parse Tree），而我们必须要仰赖语意及前后文的资讯才能在其中选择一棵最为适合的剖析树。

**有瑕疵的或不规范的输入**

例如语音处理时遇到外国口音或地方口音，或者在文本的处理中处理拼写，语法或者光学字符识别（OCR）的错误。

**语言行为与计划**

句子常常并不只是字面上的意思；例如，“你能把盐递过来吗”，一个好的回答应当是把盐递过去；在大多数上下文环境中，“能”将是糟糕的回答，虽说回答“不”或者“太远了我拿不到”也是可以接受的。再者，如果一门课程去年没开设，对于提问“这门课程去年有多少学生没通过？”回答“去年没开这门课”要比回答“没人没通过”好。

# Ati涉及的领域(文档 tts 分词 抽取 摘要 检索）

文本朗读（Text to speech）/语音合成（Speech synthesis）

语音识别（Speech recognition）

中文自动分词（Chinese word segmentation）

自然语言生成（Natural language generation）

文本分类（Text categorization）

问答系统（Question answering）

信息检索（Information retrieval）

信息抽取（Information extraction） 摘要

自动摘要（Automatic summarization）

# 中文分词原理与实现 11

2.2 查找词典算法 13  
2.2.1 标准Trie树 14  
2.2.2 三叉Trie树 18  
2.2.3 词典格式 26

# 第6章　信息提取 271

# 第7章　自动摘要 304

# 第8章　文本分类 319

# 第9章　文本倾向性分析 364

# 机器翻译

# 第11章　语音识别

# 信息检索

# 应用

## 问答系统 自动翻译

# Atitit NLP---信息检索概论方面的书籍目录

## Atitit 现代信息检索

Line 1: 第1章　引言

Line 22: 第2章　用户搜索界面

Line 44: 第3章　信息检索建模

Line 79: 第4章　检索评价

Line 104: 第5章　相关反馈与查询扩展

Line 123: 第6章　文档：语言及属性

Line 160: 第7章　查询：语言及属性

Line 175: 第8章　文本分类

Line 210: 第9章　索引和搜索

Line 238: 第10章　并行与分布式信息检索

Line 260: 第11章　Web检索

Line 307: 第12章　Web爬取

Line 333: 第13章　结构化文本检索

Line 360: 第14章　多媒体信息检索

Line 409: 第15章　企业搜索

Line 444: 第16章　图书馆系统

Line 460: 第17章　数字图书馆

## 《信息检索导论》(（美）曼宁...)【简介\_书评\_在线阅读】 - dangdang.html

第1章　布尔检索  
第2章　词项词典及倒排记录表  
第3章　词典及容错式检索  
第4章　索引构建  
第5章　索引压缩  
第6章　文档评分、词项权重计算及向量空间模型  
第7章　一个完整搜索系统中的评分计算  
第8章　信息检索的评价  
第9章　相关反馈及查询扩展  
第10章　XML检索  
第11章　概率检索模型  
第12章　基于语言建模的信息检索模型  
第13章　文本分类及朴素贝叶斯方法  
第14章　基于向量空间模型的文本分类  
第15章　支持向量机及文档机器学习方法

## Atitit 自然语言处理原理与实现 attilax总结

1.1. 中文分词原理与实现 11 1

1.2. 英文分析 194 2

1.3. 第6章　信息提取 271 2

1.4. 第7章　自动摘要 304 2

1.5. 第8章　文本分类 319 2

1.6. 第9章　文本倾向性分析 364 2

1.7. 第10章　问答系统 374 2

1.8. 第11章　语音识别 413 2

1.9. 对自然语言处理的两个学派（语言学派和统计学派 2

1.10. 《自然语言处理综论》 3

1.11. 《统计自然语言处理基础》（Foundations of Statistical Natural Language Processing） 3

1.12. 机器翻译 3

1.13. Atitit 自然语言处理原理与实现 attilax总结 4

# 参考

## Atitit.数据检索与网络爬虫与数据采集的原理概论

几本自然语言处理入门书 \_ 我爱自然语言处理.html

自然语言处理与计算语言学书籍汇总之五：机器翻译 \_ 我爱自然语言处理.html

参考资料

## 推荐新书《自然语言处理原理与技术实现》 - 自然语言处理-炼数成金-Dataguru专业数据分析社区.html

ref

(9+条消息)java的TTS（Text to Speech）的实现 - gudujohn的博客 - CSDN博客

用FreeTTS实现Java语音程序 - - ITeye博客.html