

*JuTzungKuei*

zzk's homepage

## NLP 基本任务

📅 2019-08-11 | 📅 2019-11-09 | 📁 总结 | 📖 阅读量 124

📖 1.1k 字 | ⌚ 4 分钟

[TOC]

### NLP基本任务

#### 1、词法分析（Lexical Analysis）：对自然语言进行词汇层面的分析，是NLP基础性工作

- 分词（Word Segmentation/Tokenization）：对没有明显边界的文本进行切分，得到词序列
- 新词发现（New Words Identification）：找出文本中具有新形势、新意义或是新用法的词
- 形态分析（Morphological Analysis）：分析单词的形态组成，包括词干（Stems）、词根（Roots）、词缀（Prefixes and Suffixes）等
- 词性标注（Part-of-speech Tagging）：确定文本中每个词的词性。词性包括动词（Verb）、名词（Noun）、代词（pronoun）等
- 拼写校正（Spelling Correction）：找出拼写错误的词并进行纠正

#### 2、句子分析（Sentence Analysis）：对自然语言进行句子层面的分析，包括句法分析和其他句子级别的分析任务

- 组块分析（Chunking）：标出句子中的短语块，例如名词短语（NP），动词短语（VP）等
- 超级标签标注（Super Tagging）：给每个句子中的每个词标注上超级标签，超级标签是句法树中与该词相关的树形结构
- 成分句法分析（Constituency Parsing）：分析句子的成分，给出一棵树由终结符和非终结符构成的句法树
- 依存句法分析（Dependency Parsing）：分析句子中词与词之间的依存关系，给一棵由词语依存关系构成的依存句法树
- 语言模型（Language Modeling）：对给定的一个句子进行打分，该分数代表句子合理性（流畅度）的程度

- 语种识别 (Language Identification) : 给定一段文本, 确定该文本属于哪个语种
- 句子边界检测 (Sentence Boundary Detection) : 给没有明显句子边界的文本加边界

### 3、语义分析 (Semantic Analysis) : 对给定文本进行分析和理解, 形成能够表达语义的形式化表示或分布式表示

- 词义消歧 (Word Sense Disambiguation) : 对有歧义的词, 确定其准确的词义
- 语义角色标注 (Semantic Role Labeling) : 标注句子中的语义角色类标, 语义角色, 语义角色包括施事、受事、影响等
- 抽象语义表示分析 (Abstract Meaning Representation Parsing) : AMR是一种抽象语义表示形式, AMR parser把句子解析成AMR结构
- 一阶谓词逻辑演算 (First Order Predicate Calculus) : 使用一阶谓词逻辑系统表达语义
- 框架语义分析 (Frame Semantic Parsing) : 根据框架语义学的观点, 对句子进行语义分析
- 词汇/句子/段落的向量化表示 (Word/Sentence/Paragraph Vector) : 研究词汇、句子、段落的向量化方法, 向量的性质和应用

### 4、信息抽取 (Information Extraction) : 从无结构文本中抽取结构化的信息

- 命名实体识别 (Named Entity Recognition) : 从文本中识别出命名实体, 实体一般包括人名、地名、机构名、时间、日期、货币、百分比等
- 实体消歧 (Entity Disambiguation) : 确定实体指代的现实世界中的对象
- 术语抽取 (Terminology/Glossary Extraction) : 从文本中确定术语
- 共指消解 (Coreference Resolution) : 确定不同实体的等价描述, 包括代词消解和名词消解
- 关系抽取 (Relationship Extraction) : 确定文本中两个实体之间的关系类型
- 事件抽取 (Event Extraction) : 从无结构的文本中抽取结构化事件
- 情感分析 (Sentiment Analysis) : 对文本的主观性情绪进行提取
- 意图识别 (Intent Detection) : 对话系统中的一个重要模块, 对用户给定的对话内容进行分析, 识别用户意图
- 槽位填充 (Slot Filling) : 对话系统中的一个重要模块, 从对话内容中分析出用户意图相关的有效信息

### 5、顶层任务 (High-level Tasks) : 直接面向普通用户, 提供自然语言处理产品服务的系统级任务, 会用到多个层面的自然语言处理技术

- 机器翻译 (Machine Translation) : 通过计算机自动化的把一种语言翻译成另外一种语言
- 文本摘要 (Text summarization/Simplification) : 对较长文本进行内容梗概的提取
- 问答系统 (Question-Answering System) : 针对用户提出的问题, 系统给出相应的答案
- 对话系统 (Dialogue System) : 能够与用户进行聊天对话, 从对话中捕获用户的意图, 并分析执行

- 阅读理解（Reading Comprehension）：机器阅读完一篇文章后，给定一些文章相关问题，机器能够回答
- 自动文章分级（Automatic Essay Grading）：给定一篇文章，对文章的质量进行打分或分级

坚持原创技术分享，您的支持将鼓励我继续创作！

赏

📖 NLP

◀ 解压缩命令-使用教程

Vim 基本配置 ▶

© 2019 ❤️ Zongkui Zhu

由 [Hexo](#) 强力驱动 | 主题 - [NexT.Pisces](#)

👤 访问人数 5809 人 | 👁 总访问量 10616 次