# 术语表

|  |  |
| --- | --- |
| 英文 | 中文 |
| Tokenization | 断词处理 |
| Token | 分词 |
| Stemming | 词干提取 |
| stop word | 停用词 |
| Rare word | 罕见词 |
| Chunking | 分词处理 |
| Spell correction | 拼写纠错 |
| Sentence splitter | 句法分割 |
| Lemmatization | 词形还原 |
| part-of-speech tagging / POS tagging | 词性标注 |
| Parsing | 语法解析 |
| Natural Language Processing / NLP | 自然语言处理 |
| Named-entity recognition / NER | 命名实体识别 |
| Speech recognition | 语音识别 |
| Text to speech | 语音朗读 |
| Speech synthesis | 语音合成 |
| Machine translation | 机器翻译 |
| Maximum Entropy Classifier (MEC) | 最大熵分类器（MEC） |
| Chinese word segmentation | 中文自动分词 |
| Text categorization | 文本分类 |
| Information retrieval | 信息检索 |
| Information extraction | 信息抽取 |
| Text-proofing | 文本校对 |
| Automatic summarization | 自动摘要 |
| Textual entailment | 文字蕴涵 |
| Question answering / QA | 问答系统 |
| Hidden Markov Model (HMM) | 隐马尔可夫模型 |
| Conditional Random Field (CRF) | 条件随机场 |