

### 1.分词 Word Segmentation

[chqiwang/convseg](#) , 基于 CNN 做中文分词, 提供数据和代码。

### 2.词预测 Word Prediction

[Kyubyong/word\\_prediction](#) , 基于 CNN 做词预测, 提供数据和代码。

### 3. 文本蕴涵 Textual Entailment

[Steven-Hewitt/Entailment-with-Tensorflow](#), 基于 Tensorflow 做文本蕴涵, 提供数据和代码。

### 4. 语音识别 Automatic Speech Recognition

[buriburisuri/speech-to-text-wavenet](#), 基于 DeepMind WaveNet 和 Tensorflow 做句子级语音识别。

### 5. 自动摘要 Automatic Summarisation

[PKULCWM/PKUSUMSUM](#), 北大万小军老师团队的自动摘要方法汇总, 包含了他们大量 paper 的实现, 支持单文档摘要、多文档摘要、topic-focused 多文档摘要。

### 6. 文本纠错 Text Correct

[atpaino/deep-text-corrector](#), 基于深度学习做文本纠错, 提供数据和代码。

### 7.字音转换 Grapheme to Phoneme

[cmusphinx/g2p-seq2seq](#), 基于网红 transformer 做, 提供数据和代码。

## 8. 复述检测 Paraphrase Detection 和 问答 Question Answering

[Paraphrase-Driven Learning for Open Question Answering](#), 基于复述驱动学习的开放域问答。

## 9. 音汉互译 Pinyin-To-Chinese

[Kyubyong/neural\\_chinese\\_transliterator](#), 基于 CNN 做音汉互译。

## 10. 情感分析 Sentiment Analysis

情感分析保护的内容太多了, 目前没发现比较全的。推荐两个适合练手的吧:

[Deeply Moving: Deep Learning for Sentiment Analysis](#), <http://sentic.net/about/>。

## 11. 手语识别 Sign Language Recognition

[Home - SignAll](#), 该项目在手语识别做的非常成熟。

## 12. 词性标注 (POS)、命名实体识别 (NER)、句法分析 (parser)、语义角色标注 (SRL) 等。

[HIT-SCIR/ltpl](#), 保护代码、模型、数据, 还有详细的文档, 而且效果还很好。

## 13. 词干 Word Stemming

[snowballstem/snowball](#), 实现的词干效果还不错。

## 14. 语言识别 Language Identification

<https://github.com/saffsd/langid.py>, 语言识别比较好的开源工具。

## 15. 机器翻译 Machine Translation

[OpenNMT/OpenNMT-py](#), 基于 PyTorch 的神经机器翻译, 很适合练手。

## 16. 复述生成 Paraphrase Generation

[vsuthichai/paraphraser](#)，基于 Tensorflow 的句子级复述生成，适合练手。

## 17. 关系抽取 Relationship Extraction

[ankitp94/relationship-extraction](#)，基于核方法的关系抽取。

## 18. 句子边界消歧 Sentence Boundary Disambiguation

<https://github.com/Orekhov/SentenceBreaking>，很有意思。

## 19. 事件抽取 Event Extraction

[liuhuanyong/ComplexEventExtraction](#)，中文复合事件抽取，包括条件事件、因果事件、顺承事件、反转事件等事件抽取，并形成事理图谱。

## 20. 词义消歧 Word Sense Disambiguation

[alvations/pywsd](#)，代码不多，方法简单，适合练手。

## 21. 命名实体消歧 Named Entity Disambiguation

[dice-group/AGDISTIS](#)，实体消歧是很重要的，尤其对于实体融合（比如知识图谱中多源数据融合）、实体链接。

## 22. 幽默检测 Humor Detection

[pln-fing-udelar/pghumor](#)

## 23. 讽刺检测 Sarcasm Detection

[AniSkywalker/SarcasmDetection](#)，基于神经网络的讽刺检测。

#### 24. 实体链接 Entity Linking

[hasibi/EntityLinkingRetrieval-ELR](#), 实体链接用途非常广, 非常适合练手。

#### 25. 指代消歧 Coreference Resolution

[huggingface/neuralcoref](#), 基于神经网络的指代消歧。

#### 26. 关键词/短语抽取和社会标签推荐 Keyphrase Extraction and Social Tag Suggestion

[thunlp/THUTag](#), 用多种方法 实现了多种关键词/短语抽取和社会标签推荐。