

1.入门资料：(统计自然语言处理基础)

全局看：(大概看看就好)

<http://bit.baidu.com/Course/detail/id/56.html>

系统看：(细分到每个小章节) 可以用这个来入门统计自然语言处理以及基本概念

链接：https://pan.baidu.com/s/1126z3_j5ZjpyGqGRQdFy2w

提取码：8st2 (推荐这个)

<https://pan.baidu.com/s/1c1QmoIO#list/path=%2FNLP>

2.统计自然语言处理->深度自然语言处理 (因为现在自然语言处理都需要和深度学习联系在一起下面这些就是)

深度自然语言处理，项目什么的很实用

参考教案(英)：<http://web.stanford.edu/class/cs224n/syllabus.html>

3.进阶参考通过论文来学习自然语言处理的目前主要研究领域：(大部分是经典论文可以研读)

主要领域如下：

Distributed Representations

Entity Recognition (Sequence Tagging)

Language Model

Machine Translation

Question Answering

Relation Extraction

Sentences Matching

Text Classification

可参考如下研读

<https://github.com/mhagiwara/100-nlp-papers>

<https://github.com/llhthinker/NLP-Papers>

4.实际项目入门：

统计自然语言处理的项目入门：（python 版本）

链接：<https://pan.baidu.com/s/1O4P1JJobky1DjHKJ5Vrt9Q>

提取码：pnf4

比赛入门：（通过查看内部的 discussion 以及 kernel 来进行学习）

1) kaggle Spooky Author Identification

可参考：

<https://www.kaggle.com/c/spooky-author-identification/kernels>

2) kaggle quora-question-duplicate:

<https://github.com/udacity/cn-machine-learning/tree/master/quora-question-duplicate>

也可以参考我们队伍解决方案：

该比赛的一个解决方案：<https://github.com/qqgeogor/kaggle-quora-solution-8th>

同时给出我们组方案的 PPT: 链接：https://pan.baidu.com/s/1ITD_erJi8a0XmWjISfx3fw

提取码：bma8

3) Kaggle toxic Comment Classification Challenge

部分设计 deeplearning 代码的实现参考：

<https://github.com/udacity/cn-machine-learning/tree/master/toxic-comment-classification>

<https://www.kaggle.com/qqgeogor/keras-lstm-attention-glove840b-lb-0-043>