NLP目的是让计算机了解人类的语言。发展历程可分为三部分：基于理论的理性主义、偏实际应用的经验主义和深度学习阶段。理性主义主要是设计文本中的语言规则，通过特定的公式来解决问题，典型方法是专家系统，然而该阶段缺乏从数据中学习和处理推理中不确定的能力。经验主义时期的NLP的特点是通过数据语料库和（浅）机器学习、统计或者其他方法使用数据样本。此阶段NLP也叫“浅层机器学习”，即缺乏多层或者“深层”来提取数据的抽象信息。

自然语言处理的任务：

1. 标计划：将文本语料库分为 原子（单词）单位
2. 词义消歧WSD：理解文本中每个单词的正确含义，如这张图真好看与你图啥两句话中“图”的意思并不一样
3. 命名实体识别NER：从给定的文本找出实体
4. 词性标注PoS：
5. 句子/概要分类
6. 文本生成
7. 问答(QA)系统
8. 机器翻译