# 自然语言处理课程大纲

目录

[一、 自然语言处理相关概述](#_Toc21518_WPSOffice_Level1) [3](#_Toc21518_WPSOffice_Level1)

[1.1 自然语言处理概念](#_Toc1456_WPSOffice_Level2) [3](#_Toc1456_WPSOffice_Level2)

[1.2 自然语言处理发展历程与现状](#_Toc3894_WPSOffice_Level2) [3](#_Toc3894_WPSOffice_Level2)

[1.3 自然语言处理目前业界发展状况](#_Toc3617_WPSOffice_Level2) [3](#_Toc3617_WPSOffice_Level2)

[二、 自然语言处理与数学](#_Toc1456_WPSOffice_Level1) [3](#_Toc1456_WPSOffice_Level1)

[2.1 隐马尔科夫模型 (HMM)](#_Toc29404_WPSOffice_Level2) [3](#_Toc29404_WPSOffice_Level2)

[2.2 最大熵马尔科夫模型（MEMM）](#_Toc17512_WPSOffice_Level2) [3](#_Toc17512_WPSOffice_Level2)

[2.3 条件随机场 (CRF)](#_Toc9209_WPSOffice_Level2) [3](#_Toc9209_WPSOffice_Level2)

[三、自然语言处理与深度学习](#_Toc3894_WPSOffice_Level1) [3](#_Toc3894_WPSOffice_Level1)

[3.1 CNN](#_Toc20054_WPSOffice_Level2) [3](#_Toc20054_WPSOffice_Level2)

[3.2 RNN及其变种网络](#_Toc25346_WPSOffice_Level2) [3](#_Toc25346_WPSOffice_Level2)

[3.3 lstm + crf模型架构介绍](#_Toc23782_WPSOffice_Level2) [3](#_Toc23782_WPSOffice_Level2)

[四、自然语言处理基础技术说明](#_Toc3617_WPSOffice_Level1) [3](#_Toc3617_WPSOffice_Level1)

[4.1 语料库与语言知识库](#_Toc11766_WPSOffice_Level2) [3](#_Toc11766_WPSOffice_Level2)

[4.2 词法分析](#_Toc2883_WPSOffice_Level2) [3](#_Toc2883_WPSOffice_Level2)

[4.2.1 中文分词](#_Toc1456_WPSOffice_Level3) [3](#_Toc1456_WPSOffice_Level3)

[4.2.2 词性标注](#_Toc3894_WPSOffice_Level3) [4](#_Toc3894_WPSOffice_Level3)

[4.3 句法分析](#_Toc7122_WPSOffice_Level2) [4](#_Toc7122_WPSOffice_Level2)

[4.4 语义分析](#_Toc25328_WPSOffice_Level2) [4](#_Toc25328_WPSOffice_Level2)

[五、语言模型&神经语言模型](#_Toc29404_WPSOffice_Level1) [4](#_Toc29404_WPSOffice_Level1)

[六、知识图谱](#_Toc17512_WPSOffice_Level1) [4](#_Toc17512_WPSOffice_Level1)

[6.1 知识图谱概述](#_Toc3235_WPSOffice_Level2) [4](#_Toc3235_WPSOffice_Level2)

[6.2 知识建模](#_Toc5651_WPSOffice_Level2) [4](#_Toc5651_WPSOffice_Level2)

[6.3 知识抽取](#_Toc6937_WPSOffice_Level2) [4](#_Toc6937_WPSOffice_Level2)

[6.4 知识挖掘](#_Toc18741_WPSOffice_Level2) [4](#_Toc18741_WPSOffice_Level2)

[6.5 知识融合](#_Toc25547_WPSOffice_Level2) [4](#_Toc25547_WPSOffice_Level2)

[6.6 知识推理](#_Toc20474_WPSOffice_Level2) [4](#_Toc20474_WPSOffice_Level2)

[6.7 知识存储](#_Toc6030_WPSOffice_Level2) [4](#_Toc6030_WPSOffice_Level2)

[6.7.1 图数据库](#_Toc3617_WPSOffice_Level3) [4](#_Toc3617_WPSOffice_Level3)

[6.8 语义搜索](#_Toc7321_WPSOffice_Level2) [4](#_Toc7321_WPSOffice_Level2)

[6.9 知识图谱应用](#_Toc14734_WPSOffice_Level2) [4](#_Toc14734_WPSOffice_Level2)

[七、项目](#_Toc9209_WPSOffice_Level1) [4](#_Toc9209_WPSOffice_Level1)

[7.1 多标签文本分类](#_Toc24611_WPSOffice_Level2) [5](#_Toc24611_WPSOffice_Level2)

[7.2 写诗机器人](#_Toc12030_WPSOffice_Level2) [5](#_Toc12030_WPSOffice_Level2)

[7.3 信息抽取](#_Toc2510_WPSOffice_Level2) [5](#_Toc2510_WPSOffice_Level2)

[7.3.1 实体抽取](#_Toc29404_WPSOffice_Level3) [5](#_Toc29404_WPSOffice_Level3)

[7.3.2 事件抽取](#_Toc17512_WPSOffice_Level3) [5](#_Toc17512_WPSOffice_Level3)

[7.4 基于深度学的的人机对话](#_Toc13737_WPSOffice_Level2) [5](#_Toc13737_WPSOffice_Level2)

### 自然语言处理相关概述

### 1.1 自然语言处理概念

### 1.2 自然语言处理发展历程与现状

### 1.3 自然语言处理目前业界发展状况

## 自然语言处理与数学

### 2.1 隐马尔科夫模型 (HMM)

### 2.2 最大熵马尔科夫模型（MEMM）

### 2.3 条件随机场 (CRF)

## 三、自然语言处理与深度学习

### 3.1 CNN

### 3.2 RNN及其变种网络

### 3.3 lstm + crf模型架构介绍

## 四、自然语言处理基础技术说明

### 4.1 语料库与语言知识库

### 4.2 词法分析

#### 4.2.1 中文分词

#### 4.2.2 词性标注

### 4.3 句法分析

### 4.4 语义分析

## 五、语言模型&神经语言模型

## 六、知识图谱

### 6.1 知识图谱概述

### 6.2 知识建模

### 6.3 知识抽取

### 6.4 知识挖掘

### 6.5 知识融合

### 6.6 知识推理

### 6.7 知识存储

#### 6.7.1 图数据库

### 6.8 语义搜索

### 6.9 知识图谱应用

## 七、项目

### 7.1 多标签文本分类

### 7.2 写诗机器人

### 7.3 信息抽取

#### 7.3.1 项目描述

信息抽取是自然语言处理中一个非常重要模块，目前在工业界已经有很广泛的应用。

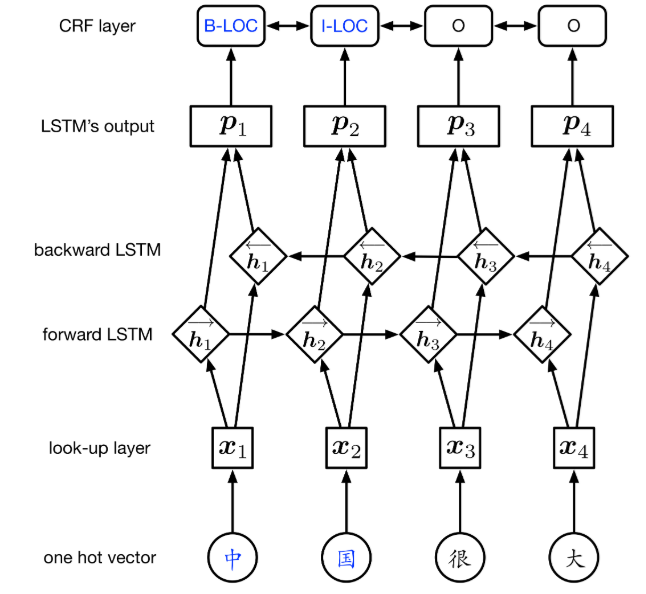
信息抽取是指从文本数据（非结构化）中抽取出特定的事实信息，比如从新闻中抽取出机构名称（ORG）、人名（PER）、地址（LOC）、时间（TIM）、事件（Event）、人物关系（Relation）等等。这些被抽取出来的信息通常以结构化的形式直接存入数据库（或者es等），可以提供用户查询及进一步分析使用，为之后构建知识库、只能问答等提供数据支撑。[这是一个将非结构化数据转化为结构化数据的过程。]

信息抽取对于构建大规模的知识库有着重要的意义，但是目前由于自然语言本身的复杂性、歧义性等特征，而且信息抽取目标只是规模巨大、复杂多样等问题使得信息抽取技术还不是很完善，但是在目前深度学习和只是图谱及语言学的深入研究下，信息抽取技术将会进一步完善和更广阔的前景

#### 7.3.2 子任务

#### 7.3.3 技术架构

信息抽取涉及的主要技术包括实体识别、句法分析、篇章分析及知识库等。目前业内使用的主流技术是深度学习 + 统计规则模式。对于部分规律性比较强的数据则可以利用规则的方法来提取，无规律性的数据则用深度学习模型的方法来提取。



### 7.4 基于深度学的的人机对话