

核心模块-自然语言理解

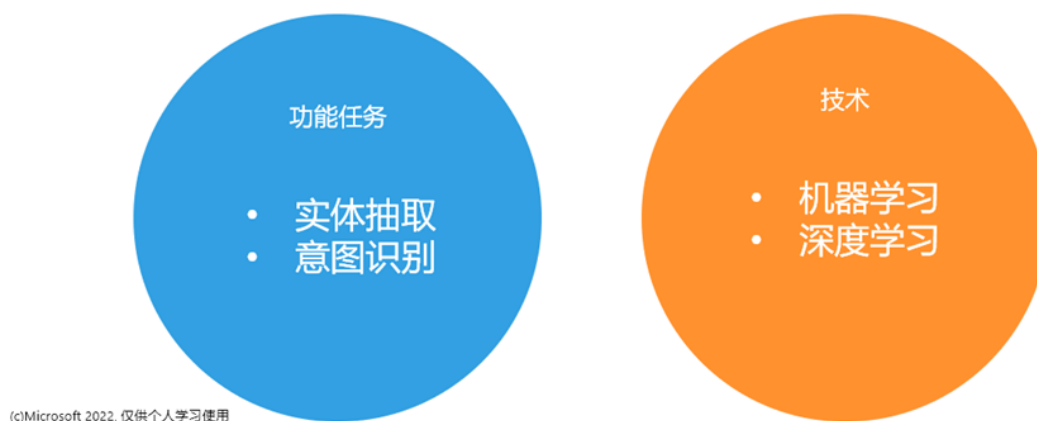
该课程主要为大家讲授如下的内容：

- 自然语言理解
- 文本分词
- 用户意图识别
- 文本相似度

1. 自然语言理解

自然语言理解（NLU）是我们常说的自然语言处理(NLP) 中的一个分支。要完成的功能是对一个词、一句话和一段文字的语义上理解，将理解的结果映射成计算机可以识别的下一步行为、动作或响应。

自然语言理解（Natural Language Understanding）



2. 文本分词

中文的一个特点是需要把句子进行分词处理，分词算法是指将一个汉

字序列切分成一个一个单独的词。

文本分词

实体抽取

Microsoft AI Talent Program

字典-匹配分词

正向最大匹配法
Trie 索引树
Hash 索引
.....

序列预测 (seq2seq)

机器学习: CRF
深度学习: 神经网络

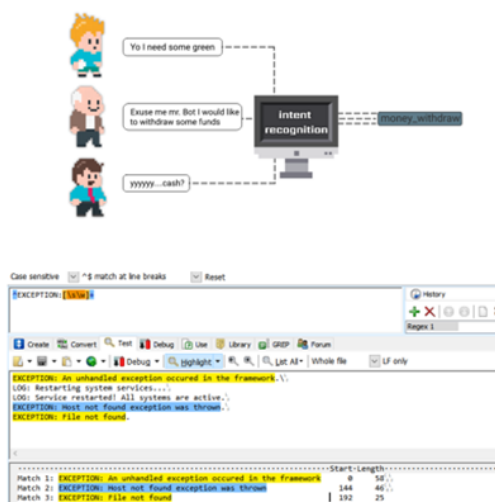
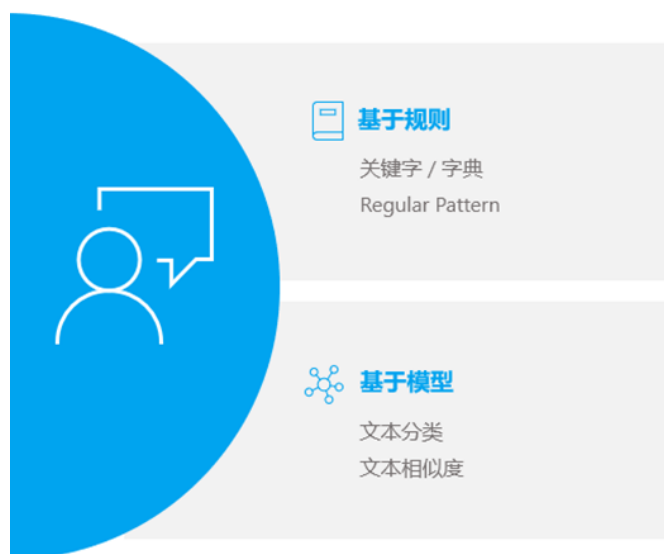
基于模型分词

序列预测模型
神经网络

BiLSTM-CRF + Dictionary

3. 用户意图识别

简单的用户意图识别，可以基于规则，复杂的意图识别，就涉及到机器学习和深度学习等复杂的计算模型了。



4. 文本相似度

计算文本相似度是一个如何把文字组合转变成数学计算的复杂问题，简单来说，就是将自然语言的词句通过词嵌入或句嵌入的方式转化为高维空间的向量，使得不同句子在该空间中的相对位置的远近可以体

现其相似程度的大小。



(c) Microsoft 2022. 仅供个人学习使用

