1. 简答题
2. **分词有哪两种算法，并请分别简述之**。

答：基于词典分词算法。也称字符串匹配分词算法。该算法是按照一定的策略将待匹配的字符串和一个已建立好的“充分大的”词典中的词进行匹配，若找到某个词条，则说明匹配成功，识别了该词。

基于统计机器学习算法。基本思路是对汉字进行标注训练，不仅考虑了词语出现的频率，还考虑上下文，具备较好的学习能力，因此其对歧义词和未登录词的识别都具有良好的效果。

1. **请简述在自然语言处理中怎样理解文档**？

答：对一篇文档如果提取出主题、核心词、关键词，识别出文档的情感倾向就认为是理解了文档。

1. **什么是关键词？从文档中抽取关键词一般有哪些方法？**

答：关键词是指能反映文本主题或者意思的词语。

抽取关键词一般有：1）可以采用词频的方式，认为文章中出现频率大的就是关键词，但是这种方法并不太准确。2）采用主题模型从文档中挖掘出来的各隐含主题下的词汇。

1. **什么是情感分析？它主要分哪两类**？

答：情感分析又称倾向性分析，意见抽取，情感挖掘，主观分析等，是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。

主要分两类：无交互的情感分析；有交互的情感分析。

**5、Google在2018年5月召开的年度开发者大会（Google I/O 2018）上，展现了基于人工智能的自然语言理解最新科技：Google Duplex。其可以自主完成给美发店、餐厅等打电话，并保持自然的交谈，完成帮用户预约时间等真实世界应用。请描述其三大主要模块，并分别说明功能。**

答：Google Duplex三大主要模块为**自动语言识别系统(ASR)**、**循环神经网络(RNN)**、**文本到语音系统(TTS)**。

自动语言识别系统：该模块功能为**将听到的声音转换为文本信息**。

循环神经网络(RNN)：该模块功能为**理解输入的文本信息，并产生对话内容(回答)**。

文本到语音系统(TTS)：将**文本信息转换为语音，并决定语音在语调、语气以及一些语言习惯上的特征**，使之更加自然。

**6、自然语言处理的基本研究领域包括：语音分析、词法分析、句法分析、语义分析、语用分析等。请分别阐述这几个领域的目的。**

答：语音分析：根据音位规则，从语音流中区分出一个个独立的音素，再根据音位形态规则找出一个个音节及其对应的词素或词。

词法分析：分析目的是找出词汇的各个词素，从中获得语言学的信息。

句法分析：对句子和短语的结构进行分析，目的是找出词、短语等的相互关系以及各自在句中的作用。

语义分析：分析目的是找出词义、结构意义及其结合意义，从而确定语言所表达的真正含义或概念。

语用分析：研究语言所存在的外界环境对语言使用者所产生的影响。

1. 论述题
2. **请论述自然语言处理中为什么有交互的情感分析难度大**？

答：评分标准需要包含下列内容：

1. 交互是一个持续的过程而不是短时固定的，这从本质上改变了情感判断的评价方式。
2. 交互的存在将大部分的状态信息都隐藏了起来。
3. 交互涉及到信息的跳跃。
4. **请论述什么是知识图谱，它的应用场景？并举出一个使用知识图谱的实例**。

答：知识图谱本质上是一种叫做语义网络（semantic network）的知识库，即具有有向图结构的一个知识库，其中图的结点代表实体（entity）或者概念（concept），而图的边代表实体/概念之间的各种语义关系，比如说两个实体之间的相似关系。

它可以应用到智能问答、情报分析、反欺诈、机器翻译和自然语言理解等。

实例略。