一面：面试官较严肃

Question1：自我介绍，跳槽原因

Question2：介绍项目

Question3：分词原理，项目应用上分词工具有什么改进？

Question4：关键词和摘要怎么抽取，算法原理

Question5：事件抽取任务描述，模型方法介绍

Question6：阐述序列标注的方法总结（我顺便讲了Lattice Network）

Question7：分本分类预处理、模型

Question8：分本分类模型知道哪些，阐述FastText原理

Question9：n-gram在分类任务中的作用

Question10：有没有尝试过基于ML的分类任务，模型，模型原理，和DL分类的区别

Question11：SVM目标函数、推导（我说到对偶问题）

Question12：手撸前缀字典树，包括数据结构定义、插入和查找（认真的一比，一个很小的bug被挑出来了）

二面：面试官没有问得很宽泛，对我了解的部分，问得巨细

Question1：自我介绍，跳槽原因

Question2：排序数组前后对调的二分查找

Question3：两个项目简介

Question4：事件抽取任务描述，模型方法介绍

Question5：我们舆情预测中，新词发现

Question6：挑一个熟悉的ML模型（我说LR、SVM、CRF）

Question7：写LR带L2正则化的目标函数，LR损失函数，

Question8：sigmoid特性，其他激活函数，sigmoid偏导推导，

Question9：NER中端到端模型详解（CNN-BiLSTM-CRF），CNN输入、LSTM层数选择和LSTM怎么接CRF层的，CRF使用的特征，CRF后怎么求解

Question10：HMM模型参数哪些，含义，HMM解码过程，怎么记录和输出最优序列

Question11：HMM和CRF模型差异性、各自优缺点

Question12：CRF和端到端模型（CNN-LSTM-CRF）各自的优缺点

三面：度秘技术负责人，高T

Question1：自我介绍，跳槽原因

Question2：过了一下简历，核对了一个会哪些NLP任务

Question3：问了分类任务，我们是怎么从问题到解决的

Question4：熟悉哪些模型，各模型结果分析，有没有过一些改进

Question5：对输入有没有做一些尝试（我说不同方法的embedding，他偏向n-gram特征这种）

Question6：讲DPCNN模型

Question7：讲CNN和RNN在分类上的差异

Question8：方案设计

在intent识别问题上，目前有5种情况发生：

1、经常有之前没有定义的intent加入

2、一句话对应多个intent

3、intent在不同设备上，结果差得很大

4、不同intent的样本数量差距很大

5、保证很高的系统吞吐和扩展性

Question9：给我介绍度秘部门结构，各个组的人员配备和职责，技术人员在百度成长路线

1.hmm和crf的区别，为什么引入最大熵hmm？lstm+crf和不加crf有什么不同？crf有哪些应用？

2.ner和分词任务有什么不一样的地方？

3.cnn和rnn在nlp应用中有什么不同？

4.讲讲多任务学习的损失函数

5.用tensorflow实现一下lr算法

6.关系分类用什么模型？涉及到有方向的关系时有什么处理办法？

7.讲讲inception模型的结构，这样的结构有什么好处？

8.lstm的qps比较高，有什么解决办法？

9.自动摘要时采用了什么算法？如何保证摘要的流畅可读性？textrank本身的缺点？

10.情感分析任务有没有对无用词做处理？做了哪些处理？如何构造情感词典？

11.知识图谱中不同数据源实体如何做知识融合？

12.q-learning和sarsa的区别

13.讲讲attention机制，在模型中是怎么用的？可能涉及推导公式

14.word2vec和glove的区别，最近的语言模型接触过哪些？能否讲讲模型结构？elmo和bert，transformer结构

15.能否从数学角度来理解一下什么样的数据会适合什么样的模型？