腾讯面试 2018/3/28 两个小时
问：介绍一下自己
简历上写了几个在公司做的相关的项目，命名实体识别和机器翻译检测等。
第一个面试官问的是机器翻译检测项目，很细，因为是用 cnn 做的聊了 cnn 的一些东西。
问：数据是哪来的，我说。。。
问：CNN 的结构是什么样的　答：...
问：池化哪些　答：最大，平均。
问：哪个好　答：最大吧，现在很多是最大
问：知道 k-maxpooling 么 答：取最大的 k 个
问：数据正负样例的比重怎么选的　答：1 比 3 左右
问：参加的评测里面的模型是什么　答：lstm+crf（简述，crf 的作用说了说）
问：现在效果好的有哪些　答：github 上开源的普遍是这个框架。（注：那个 team 和我们组都参加了一个评测）
问：为什么没用 rnn　答：项目还在进行中，这是个 baseline
问：分词方法有哪些？HMM,CRF 怎么加入到基本的分词算法中？
问：做个题吧，给个词表和句子，怎么分词。　答：字符串匹配的一些巴拉巴拉。（给的白纸手写）
问：还能优化吗？　答：词表先处理一下，然后再匹配
问：还能优化吗？　答：想不起来了
问：试一下前缀树。　答：哦，然后聊了下前缀树的复杂度。
第一个面试官走了，一会第二个就来了。始终是一个人在面我。
第二个面试官问的是命名实体识别的项目。我说了一遍目前是怎么做的。
问：数据哪来的，怎么处理的　答：亚马逊爬的。。（很细的说了一遍）
问：语料怎么构造的　答：字符串匹配然后打标签。
问 :有没有想过加一些负样例,现在你构造的模型也就比使用字符串匹配稍微有点泛化能力。
问：会主题模型吗？　答 :。
问：会无监督学习吗？　答 :。
问 :无监督情况下，怎么对相似文本进行聚类？一个具体的场景是不给用户推荐内容一样的新闻。回答的不怎么好。
问：attention 的理解，memory network 的理解？ 答：我说一种 attention 是由 memory network 演变过来的，他表示有点惊奇。
问：用什么框架？　答：keras 多点
问 :有个场景，现在让你去识别特定领域的词，怎么做？数据怎么做？想没想过负样例？怎么处理和普通词之间的矛盾
（应该是有冲突的时候，怎么标注）？怎么提高模型泛化能力？有没有考虑使用维基百科这样的知识库？词表里如果有阿里你就标注为正样例了，那要是拳王阿里呢？
问：用过 MapReduce 吗？
问：命名实体识别模型的创新点有啥？
问 :再写个题，有个数组，非递减，从中间分开把后面的放到前面去，怎么查找给定元素,log 复杂度
问：你们现在研究生都几年？暑假让实习吗？能实习几个月？
我：你们今年还会做那个评测吗？　答：会（ps,去年他们第一）
这是前后两个面试官，人看着都挺好的样子，感觉懂的很多，貌似正在做推荐新闻的一些任务。我也不知道他们是哪个部门的，都是男的，各子高高的。银科大厦 19 楼。

阿里 1：半小时，应该是筛简历，投的自然语言处理。
介绍一下自己，也是项目一顿问，ner 的任务，他们目前好像有在做情感分析，简历信息抽取，法律文书信息方面的项目。编程语言用啥，python 多线程多进程全局锁迭代器，CNN，RNN 各自优缺点。

阿里 2：半小时，花名叫 guqing，不知道那俩字
介绍一下自己，项目，神经网络怎么调参，也是 cnn,rnn,lstm 一套甩过来。聊完就没说啥了，没有算法。目前在等着。

Ps:项目问的多，不知道简历上的头像是不是发挥了一些加分的作用，推荐一下微笑时刻吧，p 的我自己都不认识了，有点贵，99 元，被师兄带过去的。