

## 写在前面

本文通篇流水账，大佬或可直接忽视并按“退出键”即可。言归正传，本文系在去年 10 月底的初稿上调整、修改而来。从 8 月初开始写简历到 10 月中下旬拿到 offer，历经两个多月，时间很漫长，确实很熬人。稳住心态、坚定求职方向至关重要，不要怂，就是干。另外，不断地、客观地认识自身的实力与短板，方能有的放矢。面试过程中，几乎全部的面试官是根据递交的简历展开提问，然后深度挖掘简历上的每个知识点。以下内容均为面试纪实。

## 科大讯飞（~50min）

### 一轮面：（面试官很放松）

自我介绍.....你的优势是什么？

开始聊项目，小米项目的短文本组和知识图谱组；个人练手的文本生成。

短文本组，采用什么模型？

TFIDF 是怎么算的？TFIDF 的实际意义是什么？

主题模型是怎么用的？

主题模型和统计特征怎么结合的？效果怎么样？

知识图谱组（NER）中，为什么这个模型？主要用什么语言？

这个模型 IDCNNs-CRF 是一种框架还是一个具体的算法？

项目过程中是怎么跟小米那边反馈问题的？做完怎样对接的？

讲一下 Attention 吧？（Seq2Seq-Attention 机器翻译为例，边画图边讲解）.....

了解 Self-Attention 吗？两者有何不同？（不了解）

研究生期间学过哪些算法类的课程？

贪心策略需要满足什么条件？（举了个找零钱的例子讲了下）

智力题：现在有两个玻璃球，以最少的次数测出 100 层楼中，玻璃球往下扔，球不碎的极限楼层？说出你的思路？

试着将你的思路形式化？（1min 没写出来，没再纠结这个问题）.....

唠嗑中.....一轮结束

### 二轮面：（面试官有点严肃，看样子有点像总监）~10min

先简单自我介绍 .....

面试官原话“项目没什么好问的，我给你出一道简单的算法题，既然你用python较多，我就不要求你用C/C++写出代码，你只要给我思路就行，公平吧？”

以最快速度打印出自然数0-N内的所有质数？空气安静的2-3 min！

有思路嘞？（用的是改进版的暴力法）吧啦吧啦 .....

话未讲完，被怼！这是暴力法，效率不是最高！你再想想吧，空气再度安静！

（实在不会.....）尝试地说了声是用动态规划？具体怎么做？

（实在不会.....）开始给我演示最高效率的方法.....

开启怼人模式：项目只是给你安排的课题和任务，那些基础的算法要了解，要有自己的心得和体会.....你可以回去了！

## 苏宁（面试官非常清秀，很有礼貌）40+ min

流程性问题，哪里人（顺便聊了下合肥这座城）？目标薪资？有哪些Offer？

自我介绍，全程聊简历，问问题，加上自我介绍时一些感兴趣的问题（你的优势是什么？）

知识图谱具体是怎么做的？

IDCNN 整个工作流程？IDCNN 与 CNN 有什么不同？

IDCNN 的优点有哪些？

CRF 与 IDCNN 是怎么结合的？

Softmax 的用处？

短文本处理流程？

主题模型承担的角色是什么？如果有一批语料，你们怎么用主题模型？

同一个 query 也可能属于多个主题？怎么认定它属于哪个主题？

统计特征是怎么用的？怎么与主题模型结合的？

模型是怎么评价的，用到哪些指标？

见我文本生成（Seq2Seq-Attention）是练手的，就没往下再问！

除了项目中用到的深度学习模型和框架？机器学习的算法有了解吗？

研究生期间上过数学课吗？上过哪些专业课程？（我只说数据挖掘、人工智能.....其他的，没想起来）

人工智能这门课中，主要讲什么的？（列举了爬山法、模拟退火、遗传、蚁

群等)

你认为人工智能这门课主要是在介绍优化算法的吗？除这些，还有什么？

唠嗑中.....

## 金山云（面试官是做图像处理的，对 NLP 了解不多）20+min

你知道 Siri 团队为啥会解散么，是不是做不下去？（解散了吗？没有吧）

NLP 目前遇到了哪些问题？

计算机体系结构了解吗？知道流水线模型吗？（WTF）

RNN 和 CNN，你最深刻的感悟是什么？

池化层有几种操作？哪个操作是最常用的？有什么优劣势吗？

针对 max pooling 会丢失一部分语义信息，有什么样的改进方法？

卷积操作的代码实现？

## 网易互联网（小哥面无表情）50+min

主题模型与 TFIDF 各自怎么用的？如果给你新的 query，主题模型和 TFIDF 使用流程是什么？

TFIDF 具体到 query 中代表什么？

评估模型时，用的是哪些指标？准确率是怎么算的？

词向量是直接训练好的？还是使用了什么工具包？

IDCNNs 与 BiLSTM 在并行处理的效果方面，自己有做过对比实验吗？

RNN 不能很好并行处理的本质原因是什么？

IDCNNs 最后得到的结果是什么？

CRF 原理了解吗？

未登录词是怎么解决的？

LSTM 与 GRU 的区别是什么？

Linux 常用命令有哪些，除了 cd、rm.？

难忘的九月：

九月面试不多，只有四家，笔试七连跪，所以后面也就没什么面试邀约。  
想要说明的是：笔试编程题真的 **hin** 重要！真的很 **hin** 重要！真的 **hin** 重要！

----- 九月、十月分割线 -----

十月面试大小公司二十多家 .....

## 小米 (50+min)

讲讲 LDA 数学原理吧 ..... 它的处理过程是什么？

介绍一下 CRF 的原理，为什么不是 IDCNNs + HMM ？

了解似然函数吗？

机器学习了解多吗？介绍一个你最熟悉的机器学习模型？

思考两道题：（给出思路）

怎么判断一个单链表中是否有环？

盛最多水的容器

## 斗鱼 TV

### 一轮面（就像在讨论问题一样 90+ min）

介绍项目时，一有问题，他就会打断我，开始讨论 .....

短文本任务中，他认为我们这么做是有问题的，怎么能完全依靠模型呢？而且觉得用主题模型和统计特征加权求和的方式也是有问题的，认为我们对 LDA 主题模型依赖程度太高，一种合理的方式，应该只是让 LDA 提供一种参考，打个辅助。然后开始给我介绍一波斗鱼在做关键词抽取时的一些思路和手法（我听得有点懵逼）。

知识图谱任务中，数据是怎么标注的？

手写 Neo4j cypher 语句

了解到我只涉及 NER，觉得我在知识图谱这部分的内容做得太浅

然后开始探讨 Sequence2Sequence + Attention

你觉得 Attention 是要解决什么问题的？（觉得我对 attention 的理解有点问题后，开始给我介绍 attention .....

讲一讲 RNN、LSTM、GRU

为什么 RNN 记不住久远之前的信息？

### 二轮面（面试官全程抖腿 40+ min）

大致介绍知识图谱的整套流程？每一步流程中都用到哪些模型？

IDCNNs+CRF 训练时，用的是什么损失函数？

NLP 都有哪些领域？

你觉得 NLP 现在的瓶颈是什么？

LDA 的数学原理

Attention 是什么？

开始手撕代码。两个字符串序列，找到两个字符串中共有的字符.....

## 出门问问（台湾人，看我进门连忙过来热情握手致意）

看完简历，问为什么不去科大讯飞？为什么不去小米？

LDA 公式会推吗？

介绍一下 IDCNNs，它和 CNN 有何不同？

CRF 在这里的作用是什么？

为什么不考虑 HMM+Viterbi+forward procedure？这样可以对序列做标注啊？

用 Java 多久？Java 中 `synchronized` 定义在函数上有什么作用？Java 有哪些设计模型？

后来面试官给我介绍求学历程，以及对现在 AI 过热的一些看法，也给些中肯的学习建议（唯一一位主动给我提学习建议的面试官！）

## 后记

其余公司面试内容，大同小异，不一而足。10 月底后，知名的互联网公司秋招工作差不多告一段落。相信此时的大佬们都拿到了心仪的 offer，开始享受短暂的、完美的假期。