

23|微软算法岗武功秘籍

1 微软面经汇总资料

第一节
微软面经
汇总资料
(整理: 江大白)
www.jiangdabai.com

- 1.1 面经汇总参考资料
- 1.2 面经涉及招聘岗位
- 1.3 面试流程时间安排
- 1.4 微软面经整理心得

1.1 面经汇总参考资料

① 参考资料:

- (1) 牛客网: 微软面经-29 篇, [网页链接](#)
- (2) 知乎面经: [点击进入查看](#)
- (3) 面试圈: [点击进入查看](#)

② 面经框架及参考答案:

- (1) 面经框架及参考答案: [点击进入查看](#)
- (2) 大厂目录及整理心得: [点击进入查看](#)

1.2 面经涉及招聘岗位

(1) 实习岗位类

【机器学习实习生】、【微软苏州 SWE 实习生】、【NLP 实习生】、

(2) 全职岗位类

【Bing 团队算法工程师】、【苏州算法工程师】、【STCA 算法工程师】

1.3 面试流程时间安排

微软面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第三面	技术Leader面	自我介绍+项目/实习经验 +技术问答+算法编程	主要是对于项目的深度思考
第四面	HR面	基础人力问题	/

PS：以上流程为大白总结归纳所得，以供参考。

其他注意点：

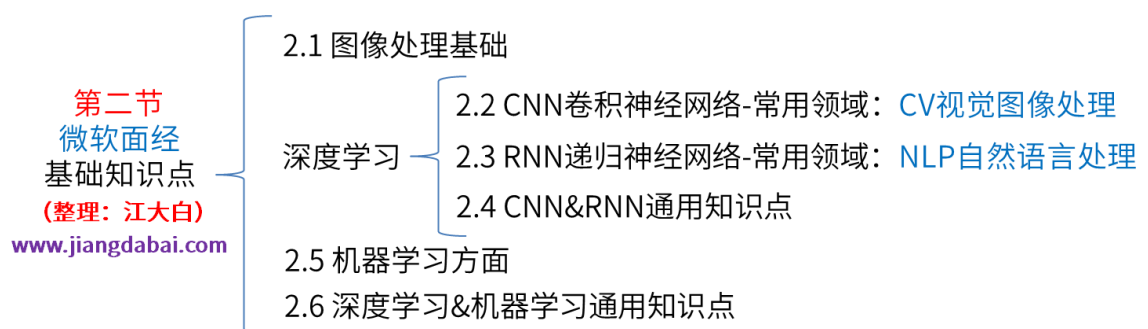
- 实习时，是笔试+3轮远程面试，笔试是4道题，远程视频面，一共三面，每次50分钟，都是先讲15分钟左右项目，然后用skype共享桌面做题。
- 有的面试最后还有Boss面
- 微软的社招面试通常是先进行一轮电话面试，面试通过的话才会邀请进行现场面试
- 电话面试之后会约现场面试，通常会安排5-6轮的面试，每轮一小时，前3轮是基础面，面试结束后面试官商量决定要不要进行后续的面试，当然如果表现比较差，也可能在某一轮直接结束。

1.4 微软面试心得汇总

- ★ 想去微软的话，编程功底比专业知识重要的多，基本leetcode easy, mid都要刷一刷
- ★ onsite面试都是在黑板上写代码，写完告诉面试官逻辑就行。三面在collabedit上做，用自己的IDE调试，需要共享屏幕。

- ★ 做完每道题都会问你怎么测试自己的代码是正确的，考虑哪些边际情况。
- ★ 算法之前一定要冷静思考一下，想一想可能有的坑，要多与面试官沟通
- ★ 有的时候，那一轮面试遇到了英国面试官就变成了英文面试，所以英文介绍最好也要准备一下
- ★ 微软更注重的是编程能力，想面微软的同学建议好好刷题，微软一般每一面都有算法题。
- ★ 面试重点还是在做题上，而且大多都是剑指 offer 和 LeetCode 的原题，所以感觉研究面经不如多刷几个题。另外最好提前准备一个英文自我介绍和项目介绍，虽然我没遇到，但很多人遇到了。
- ★ 微软的面试整体偏向基础，英语能力考察仅限于个人简介和项目描述，如果运气好的话都是中国的面试官，没有英文面试。
- ★ 投递简历之后会有 hr 先和你聊一轮，要求做一个一分钟的英文自我介绍，然后会对英文能力做一个整体评估，告诉你应该怎么准备可能的英文面试。

2 微软面经涉及基础知识点



2.1 图像处理基础

无

2.2 深度学习：CNN 卷积神经网络方面

2.2.1 讲解相关原理

2.2.1.1 卷积方面

- 反卷积具体怎么实现的？
- 为什么 dropout 能减少过拟合？

2.2.1.2 其他方面

- 问 backpropagation 的基本公式，问每一层之间是否能独立传播？

2.3 深度学习：RNN 递归神经网络方面

无

2.4 深度学习：CNN&RNN 通用的问题

无

2.5 传统机器学习方面

2.5.1 讲解相关原理

2.5.1.1 数据准备

无

2.5.1.2 特征工程

无

2.5.1.3 有监督学习-分类和回归方面

- 介绍 bagging 和 boosting？
- GBDT 原理说一下？

- xgb 和 gbd 的区别 (几乎必问的题目, 提前准备一下, 说的要有条理, 有哪些算法优化, 哪些工程实现优化, 可以适当扩展提一下 lgb)
- 深度学习网络 Factorize Machine 相对于线性模型有什么好处? Spark 用过吗?
- 如何构建一个分布式机器学习框架?

2.5.1.4 无监督学习-聚类方面

无

2.6 深度学习&机器学习面经通用知识点

无

3 微软面经涉及项目知识点

第三节
微软面经
项目知识点
(整理: 江大白)
www.jiangdabai.com

- 3.1 深度学习: CNN卷积神经网络方面
- 3.2 深度学习: RNN递归神经网络方面
- 3.3 强化学习方面
- 3.4 机器学习方面

3.1 深度学习: CNN 卷积神经网络方面

3.1.1 目标检测方面

- RCNN, Fast RCNN, Faster RCNN, Yolo, 问了我具体的 yolo 的那个 anchor, 反正好多具体的东西?

3.2 深度学习: RNN 递归神经网络方面

3.2.1 自然语言处理 NLP

① Bert

- BERT 有几种 Embedding 编码, 分词方法?

② Transformer

- Transformer 结构讲一下？

③ Word2vec

- Word2vec 中，负采样相比层次化 softmax，有什么优缺点？层次化 softmax 能保证概率归一化吗？
- word2vec 原理？

④ 其他

- 文本分类的方法有哪些，深度学习和非深度学习的方法都说一下？
- 文本相似度计算方法有什么，当我说完后，面试官说你说的基本都是深度学习方面的，经典的 NLP 方法知道有哪些吗？
- Fasttext 和 textCNN 说一下原理？
- CBOW 和 skip-gram？

3.3 强化学习

无

3.4 机器学习方面

3.4.1 推荐系统

- FM 算法、ALS 矩阵分解、协同过滤算法都说一下，并说下优缺点？

4 数据结构与算法分析相关知识点

第四节
微软面经
数据结构与算法分析
(整理: 江大白)
www.jiangdabai.com

- 4.1 数据结构与算法分析：线性表、属、散列表、图等
- 4.2 算法思想实战及智力题
- 4.3 其他方面：数论、计算几何、矩阵运算等
- 4.4 Leetcode&剑指offer原题

4.1 数据结构与算法分析

4.1.1 线性表

4.1.1.1 数组

- rotate 一次的数组，找 target，例如 [3,4,0,1,2] 找 4 所在的位置，如果不存在返回 -1，要求 $\log n$ 时间 (LeetCode medium 原题，直接二分即可)
- 给一个有重复数字的有序数组和一个数 x ，找出 x 在数组中最左和最右的下标，不存在的返回 [-1, -1]?
- 在排序数组中用二分查找找到某数字的第一个位置?
- 无序数组找第 k 大的数
- 二维数组找递增 target
- 有一个数组元素 $[a_0, a_1 \dots]$ ，从数组中找出连续的数组和为最大。
- 有一个常数 n ，有一个数组元素 $[a_0, a_1 \dots]$ 无重复元素。从数组里面找出所有可能的组合加和是 n ，并且输出。
- 有一个数组（对，全是数组题目），从数组中找出连续数组乘积最大
- 连续子数组的最大和?
- 假设一个数组只有 "a" 和 "b" 两种 string 组成。如何重新安排数组，使得最多有 3 个 a 相邻，3 个 b 相邻。如果不能安排，返回 None
- 求数组最大值时，从前往后遍历，候选值会被更新若干次，求这个次数的数学期望（说思路，我算出来是 $1 + 1/2 + \dots + 1/n$ ）
- 两个有序数组找第 k 大?

4.1.1.2 链表

- BST 转双向链表?
- 单链表找交点?

- 如何判断两个链表是不是有交点？
- 实现两个链表排序？
- 链表栈哈希表的区别？哈希表的原理解决冲突什么的，排序函数以及分别适用的场合？

4.1.1.3 栈

- 给定一个温度的时间序列，判断高于当前温度的那一天在几天后出现？（先写了个 n^2 的，面试官说复杂度太高，短路想不出来，他说用栈，改了个 $O(n)$ 的）
- 用栈模拟队列？
- 最小栈，空间优化？

4.1.1.4 字符串

- 两个字符串的编辑距离（增加、删除、替换一个字符距离为 1） [动态规划]
- 中文字符串，比如一千五百亿八千九百万六百二十这种形式，转换成 long long 的整数？要我考虑很多非法输入。
- 找出字符串中所有连续字母的 subset？
- 字符串中最长连续不重复子序列长度？ [一次遍历，额外数组记录字符最后一次出现的位置]
- 一个字符串 切分成多个回文串，返回所有可能，如 aab 要返回 [[a,a,b],[a,a,b]] 3
- 给定一个字符串，判断是不是合法 IP 地址？
- 假设有 a,b 两个 int, 转成二进制后 $c = a | b$. 假设从 0→1, 1→0 理解为一个 action。最少需要多少个 action 计算 $c = a | b$ ？

4.1.2 树

- 非递归前序遍历二叉树？
- 前序遍历、中序遍历、后序遍历，知道那些可以恢复二叉树？原因？
- sorted linked list 从头到尾翻转一次？二叉树搜索。

- 字典树

4.1.3 排序

- 写归并排序，非递归
- Top K
- 写一个快排，非递归？
- 将一堆大小写字母根据大小写排序？
- 写快排，问时间复杂度和空间复杂度，然后要求输出排序后对应元素的原下标？
- 堆排序以及很多变体？

4.2 算法思想实战及智力题

4.2.1 算法思想实战

- 丑陋数变体。手撕代码，推出了 int 型的丑陋数上限并分析复杂度？
- 给一个 100 万规模的词典，一个长文档，如何快速从里面标注出所有的词，写一下代码？
- 给一句英文，in-place 把每个词顺序翻转；给一个全排列的某个情况，找到这个 permutation 在全排列中的 index？
- 给定一个字符串，判断是否是有效的 IP 地址？（提示，有效的 IP 地址格式是 xxx.xxx.xxx.xxx, xxx 在 0-255 之间）（输入不保证都是这种格式，要自己判断，同时 001 这种是否有效要问面试官）
- 找最深的左括号，follow up: 括号匹配。
- 【二极管能显示的数字】

一组二极管有七个，能表示 0-9 十个数，现在有 n 组二极管，每组二极管中有一些亮，有一些灭，灭可能是因为坏了也可能是因为不需要亮。要求出这 n 组二极管能显示的所有数字的组合

输入：二维数组，每行代表一组二极管，每行七个，0 代表灭，1 代表亮

输出：所有可以显示的数字组合

例：现在两组二极管，显示的是 23，那输出 23，83，88，28，用了个回溯

- 从(0,0)出发，每一步有上下左右和停留五种选择，问 k 步之后回到原点的概率有多大，不能超出边界？（先用了回溯，后面用了 dp，面试官说我最开始不该暴力，应该直接给出最优解）
- 两个长度为 m 的无序数组 A, B，对于任意不相交的区间 ab 和 cd， $val[ab] = \text{sum}(A, a, b) - \text{sum}(B, a, b)$ ， $val[cd] = \text{sum}(B, c, d) - \text{sum}(A, c, d)$
- 求 abcd，使 $val[ab] + val[cd]$ 最大（这题比较难，先写了个暴力解法，然后和面试官逐步讨论优化，没有给出最优解法）
- 汉字数字转数字，如 一百二十转化成 120

4.2.2 智力题

- 丢两个骰子，最可能出现的点数和是多少？3 个骰子呢，不能枚举，面试官让快速估计。
- 给定二维平面一些点，问用一个半径为 1 的圆最多可框住多少个点（说思路：每两个点确定两个对称的圆，当然两点距离不能超过直径 2）
- 给一个 N*M 的棋盘，从(1, 1)移动到(N, M)，只能向右或向下，计算方案数，如果 N 和 M 很大怎么办？

4.3 其他方面

4.3.1 概率分析

- 给一个随机函数 fun，30%的概率产生 1，70%产生 0，如何用 fun 产生等概率的 0 和 1？

4.3.2 矩阵运算

- 顺时针旋转正方形矩阵 90 度；在时间复杂度 $O(\log(m+n))$ 的要求下寻找两个有序数组合并后的中位数？

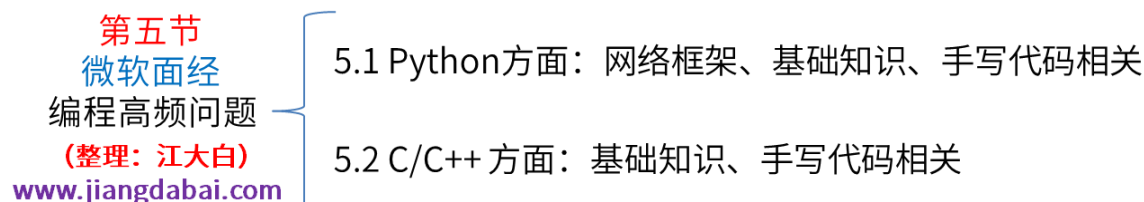
4.3.3 其他

- 给定一堆序列标号（标号用整数）,形式为[a,b]，表示标号为 a，b 的两个物体的体积关系满足： $a > b$ ，用合适的数据结构存储数据，并判断这堆序列是否有效？
- KMP 算法有什么缺点？除了 KMP，还有什么算法可以快速做字符串查找？

4.4 Leetcode&剑指 offer 原题

- Leetcode 72
- Leetcode 137：写通用解法，要求时间复杂度 $O(n)$ ，空间复杂度 $O(1)$ ？
- LeetCode 283：原地移动数组，使得元素对应顺序不变，0 值移动到末尾
- LeetCode 543：二叉树直径
- LeetCode 1497：问复杂度
- Leetcode 原题：股票只能买入卖出一次，和买入卖出多次？
- LeetCode 原题：最大子数组和
- LeetCode 原题：实现 atoi 考虑所有情况，考虑所有异常情况，包括溢出
- 剑指 offer 原题：一个矩阵，每一行从左到右递增，每一列从上到下递增，给一个值，判断这个值是否在矩阵中？面试官后面要求用 binary search 做，问了下时间复杂度。

5 编程高频问题：Python&C/C++方面



5.1 python 方面

无

5.2 C/C++方面

无

6 操作系统高频问题：数据库&线程&常用命令等

第六节
微软面经
操作系统高频问题
(整理：江大白)
www.jiangdabai.com

6.1 数据库方面：基础知识、手写代码相关

6.2 操作系统方面：TCP、线程&进程、常用命令相关

6.1 数据库方面

- 数据库的索引有了解过吗，有哪些优缺点？

6.2 操作系统方面

- 什么是死锁，造成死锁的原因有哪些？

7 技术&产品&开放性问题

7.1 技术方面

- 场景题：如何给问答系统中的新问题推荐答案？
- 场景题：单词纠错怎么做？
- 场景题：如何让对话机器人产生的回答更具情感性，面试官简化了问题：机器人产生回答后，我们给回答加前缀，比如问“今天吃饭了吗？”，回答“【嗯呀】，我吃了”，如何从大规模 QA 数据中统计出要加哪些前缀（如上面的“嗯呀”），然后判断是否需要加前缀，需要加什么前缀？
- 场景题：对话机器人说了一句话后，如何判断该话是否含有***、暴力元素，有标注数据怎么做，无标注数据怎么做？

- n 个准确率为 50% 的分类器，可以通过什么方式提升准确吗？60% 呢？如果可以，提升到 96% 需要多少个？

7.2 产品方面

- 一些推荐算法相关，比如搜搞笑，推荐的视频如何排序，怎么区分推荐的视频是不是重复的等等？