

Yelp HR

1. size of unsigned integer

2. http port no.?

- 20 & 21: File Transfer Protocol (FTP)
- 22: Secure Shell (SSH)
- 23: Telnet remote login service
- 25: Simple Mail Transfer Protocol (SMTP)
- 53: Domain Name System (DNS) service
- 80: Hypertext Transfer Protocol (HTTP) used in the World Wide Web
- 110: Post Office Protocol (POP3)
- 119: Network News Transfer Protocol (NNTP)
- 143: Internet Message Access Protocol (IMAP)
- 161: Simple Network Management Protocol (SNMP)
- 194: Internet Relay Chat (IRC)
- 443: HTTP Secure (HTTPS)
- 465: SMTP Secure (SMTPS)

3. ssl full form?

(Secure Socket Layer, Encrypt in Transportation layer)

4. use of grep and kill

**Grep** is a **command-line** utility for searching plain-text data sets for lines matching a **regular expression**. Grep was originally developed for the **Unix** operating system, but is available today for all **Unix-like** systems. Its name comes from the **ed** command *g/re/p* (*g*lobally *s*earch a *r*egular *e*xpression and *p*rint), which has the same effect: doing a global search with the regular expression and printing all matching lines

In **computing**, **kill** is a **command** that is used in several popular **operating systems** to send signals to running **processes** in order to request the termination of the process.

5. the runtime of adding something to a linked list?

$O(N)$  because it does not support random access

6. SSL 和 TLS (Transport Layer Security) 的区别:

TLS 是 SSL 的升级版 (TLS3.0 = SSL1.0) TLS 先 handshake 再 secure

7. hashmaps, DNS(Domain Name System),

8. python native datatypes

Boolean, Number, String, Byte, List, Tuple, Set, Dictionary

9. What does HTML stand for?

HyperText Markup Language

10. what is the protocol used underneath FTP, what is the size of int

char : 1 byte

short : 2 bytes

int : 4 bytes

long : 4 bytes

float : 4 bytes

double : 8 bytes

File Transfer Protocol (FTP) is a standard Internet protocol for transmitting files between computers on the Internet. underneath it is TCP/IP protocol.

11. would it generally be better to run a binary search on a <<sorted array>  
>, or a linked list

12. immutable python types:

numbers, booleans, strings, tuples, frozensets

int, float, tuple and str

13. Difference between POST and GET:

GET requests a representation of the specified resource.

POST submits data to be processed (e.g., from an HTML form) to the identified resource.

14. What command would you use to retrieve one row of data from a table using SQL?

15. What is the best and worst case time complexities for a hash table?  $O(1)$   
 $O(n)$
16. What is the maximum value of an unsigned integer?  
 $\ll 2^{32} - 1 \gg$
17. What is the command for viewing memory management information in Linux?  
 vmstat, top
18. What command would you use to retrieve one row of data from a table using SQL?  
**SELECT ... FROM ... WHERE ... ORDER BY ... FETCH FIRST n ROWS ONLY**
19. Do INNER JOINS produce the intersection or union between two tables?  $\ll$  Intersection.  $\gg$  Outer join is union.
20. Which operation is more expensive? Multiply,  $\ll$  Divide  $\gg$  or Add
21. What is the average disk access time?  $\ll$  seek time + rotational delay + Transfer time  $\gg$  10-20 ms
22. what's the difference between hashtable and hashmap? 前者 thread safe
23. sed command:  $\ll$  pattern matching  $\gg$   
 Sed is a Stream Editor used for modifying the files in unix
24. How to delete in the SQL table?  
 $\ll$  DELETE \* FROM \* WHERE \*  $\gg$
25. What is Big O notation?
26. How does process communicate with each other?  $\ll$  Socket, Pipe, shared memory  $\gg$

Method	Short Description	Provided by ( <a href="#">operating systems</a> or other environments)
<a href="#">File</a>	A record stored on disk that can be accessed by name by any process	Most operating systems
<a href="#">Signal</a>	A system message sent from one process to another, not usually used to store information but instead	Most operating systems; some systems, such as Win NT subsystem, implement signals in only the C run-time library and provide no support for their use as an IPC method <sup><a href="#">[citation needed]</a></sup> . But other subsystems like the POSIX subsystem provided by default until windows 2000.

	give commands.	Then available with interix in XP/2003 then with « windows services for UNIX » (SFU).
Socket	A data stream sent over a network interface, either to a different process on the same computer or to another computer	Most operating systems
Message queue	An anonymous data stream similar to the pipe, but stores and retrieves information in packets.	Most operating systems
Pipe	A two-way data stream interfaced through standard input and output and is read character by character.	All POSIX systems, Windows
Named pipe	A pipe implemented through a file on the file system instead of standard input and output.	All POSIX systems, Windows
Semaphore	A simple structure that synchronizes threads or processes acting on	All POSIX systems, Windows

	shared resources.	
Shared memory	Multiple processes given access to the same memory, allowing all to change it and read changes made by other processes.	All POSIX systems, Windows
Message passing (shared nothing)	Similar to the message queue.	Used in MPI paradigm, Java RMI, CORBA, DDS, MSMQ, MailSlots, QNX, others
Memory-mapped file	A file mapped to RAM and can be modified by changing memory addresses directly instead of outputting to a stream, shares same benefits as a standard file.	All POSIX systems, Windows

27. explain with the most detail what happens in the background after a URL is put in the browser and pressed enter

28. some linux commands: pid <<

29. What's a static variable?

In computer programming, a **static variable** is a **variable** that has been **allocated statically**—whose **lifetime** or "extent" extends across the entire run of the program. This is in contrast to the more ephemeral **automatic variables** (**local variables** are generally automatic), whose storage is allocated and deallocated on the **call stack**; and in contrast to objects whose storage is **dynamically allocated** in **heap memory**.

In Java, static variable **associated with a class, not objects of that class.**

30. Unix command word count:

<<wc>> Find number of line <<wc -l>>

-c, --bytes

print the byte counts

-m, --chars

print the character counts

-l, --lines

print the newline counts

-L, --max-line-length

print the length of the longest line

-w, --words

print the word counts

31. linux command for checking cpu usage

<<top>>

32. how many bytes are contained in a 32 bit system

<<2^29>>

33. How can you get the ip address of the other machine? <<use ping>>

34. The number in the option of 'kill' command? <<-9>>

Why Yelp?

Why would we hire you?

And then few technical questions:

signal for kill command?

performance of binary tree if implemented thru array or linkedlist?

SQL command to move a row in a table?

protocol used to transfer message in HTTP application?

SSL?

no of bits to represent octal?

Three

max value that can be represented by unsigned int?

unsigned:  $2^{32}-1 = 4 \cdot 1024^3 - 1$ ; signed:  $-2^{31} \dots +2^{31}-1$ , because the sign-bit is the highest bit. Just learn  $2^0=1$

to  $2^{10}=1024$  and combine.  $1024=1k$ ,  $1024^2=1M$ ,  $1024^3=1G$

which one of the following describe inner join? 1. A intersection B, 2. A Union B and ...

Why do you want to work at Yelp? What are three adjectives your professor would use to describe you?

What do you use to search code in Linux? Is an SQL inner join a union, intersection or difference? Which is immutable in Python: tuple, linked list or array? What port does HTTP use?

HTTP/UDP/SSL/TCP

1. Why Yelp?

2. No. of bytes in a 64 bit machine

8

3. What is the value of the unsigned integer in a 32 bit machine?

4 byte

4. How many bits are required to represent an octal digit?

3

5. Where can you apply binary search - sorted array or linked list?

6. Time complexity of binary search.

7. Time complexity of adding a node at the head of the linked list.

8. What is the signal number for kill

9. How would you find a method in a file.

10. How would you find whether a remote computer is online or not

Determine whether the Ping method returns True

11. How does the domain name map to ip address

The Internet *Domain Name Service (DNS)* can translate host names into equivalent IP addresses and *vice versa*, as needed by various Internet programs.

12. How to select one row of a table in SQL

13. Port number for HTTP and protocol used by HTTP

14. Full form of SSL and what does it do?

-number of bit in octal

-some command in linux(grep, kill)

-some sort complexity(binary)

-adding element to first of linked list

-data base question(outerjoin, left joint,...)

-SSL? TCP?

w many bits in octal?, kill -9 signal, how to get all the instances of a class, what is the command in unix to find IP configuration information,

ifconfig

inner join is what out of A union B, A intersect B and A - B, HTTP port number, SSL full form and its purpose, complexity of some search.

3 adjectives coworkers would use to describe you?

- Why would you want to work at Yelp?

- Port number used by HTTP?

- Protocol underneath HTTP?

- Runtime efficiency of searching an element in an array using a binary search algorithm?

- Complexity of adding an element to the head of a linked list?

- Maximum value of 32 bits unsigned int?

- Number of digits needed to code an octet?

- What if we're in base 8?

- Number of bytes to code 64 bits?

- What does SSL stands for and what does it do?

- What is the Linux kill signal?

- Command to highlight the occurrence of a method in a file?

`grep/ack`

- Would you apply a binary search on a linked list or a sorted array?

- Is an inner join an intersection, a union or something else?

Phone interview with engineer:

- Started by presentation of himself, his team, and other teams

- Why computer science?

- Talk about an interesting project you implemented.

- How would you improve it?

- What did you find most interesting in that project?

- If a request takes too long, what could be the reason?

- Describe different steps when a form is posted

- Coding problem to do on a shared pad:

# given a list of words, write a function which

# takes in the list, and groups the words together

# according to which ones are anagrams of eachother

# e.g.

# input = ["art", "rat", "bats", "banana", "stab", "tar"]

# output = [["art", "rat", "tar"], ["bats", "stab"], ["banana"]]

I wasted too much time trying to come up with an iterative solution when the problem is "easily" solvable with a recursive algorithm.



what might be the possible reasons of slow network connection?

C/C++ questions: what is inline function? describe why compiler does not perform inline in some circumstances

coding:

- Write a function to find a hidden guess integer given a helper function that will check if guess is too low, too high, or correct. [Answer Question](#)
- Find the k-lowest elements of an unsorted array. [Answer Question](#)

Given a URL. How do you find a phone no. in it?

braces are balanced and how to make the check faster. Also asked about interface and abstract class, generate random number, some operating system knowledge. Then a coding question about given a lowercase string 'ab', write a program to generate all possible lowercase and uppercase combination {'AB', 'Ab', 'aB' and 'ab'}.

a coding question about given a lowercase string 'ab', write a program to generate all possible lowercase and uppercase combination {'AB', 'Ab', 'aB' and 'ab'}

split shifted linked list into two sorted lists

- Get the k first element of a list. [View Answer](#)
- Compare the output of two textbox and reject if their content is too similar (reordering of paragraph should be rejected) [View Answer](#)

implement this anagram grouping algorithm with a running time of  $O(n)$ .

**Interview Question** – Coding question: given a list of URLs, find out the most frequent top 5 URLs.

[Answer Question](#)

Given a large log file, return or print the ten most visited URLs. [Answer Question](#)

Given a list of words, group all the anagrams. Given a list of words, find the longest common prefix between them.

Implement a cache (had growing requirements, such as a limit on how many things could be stored in it, and should return the most recently placed record). The interviewer didn't seem to care that I was correct, but more how I approached it.

Given a list of strings, write a function to calculate the longest common prefix (LCP) of all those strings.

You're given an array of strings. Sort it so that the result returns an array of an array of strings sorted into its anagrams. (e.g. input: ["aa", "ad", "da"], output: [ ["aa"], ["ad", "da"] ]

So the question was given a array how you have to move the character located in one location and place it at the other location. move(char[] arr, int from, int to). [Answer Question](#)

- Given a list of numbers, write a function that combines those numbers in arithmetic expression to obtain the value T. The allowed operations are +, -, \*, and /. Parenthesis are also allowed.

[View Answer](#)

- Given a list of n words. Print the most frequent 10.

Given an unsorted array of numbers (that may contain repeated numbers), print or produce a data structure that contains all the pairs that have a sum equal to a number k. Do not include pairs that are the same numbers in a different order.

Ex. [3,2,6,2,0,5,3,8,7,10]

[(0,10),(2,8),(3,7)] [View Answers \(2\)](#)

1. What happens when you type yelp.com on your browser
2. Lets say your yelp.com homepage is loading slowly how would you start addressing the problem?
3. inner join vs left outer join

I pretty answered all of them, then was a problem solving question.

Write a function which will determine if a given string is palindrome or not.

bool ispalindrome(string input);

I was able to complete the problem and even the interviewer seemed satisfied with the solution.

Test Cases:

"racecar" => true

"!!a!" => true

"abc" => false

"a" => true

"A Man, A Plan, A Canal – Panama!" => true [Answer Question](#)

题目是找 anagram，Input 是 List<String>，输出 List<List<String>>，anagram 的 string 放到一个 list 里。我就用一个 hashmap，key 是按照字母顺序重新排序的 string,value 是 List<String>扫一遍，就可以了。改了几个小 bug，就直接跑过了。他貌似对结果很满意，一直在那里说 cool,cool

刚面完的 Yelp skype interview.

题目是这样的: I have a very large text file, many gigabytes. I want a uniform random sample of exactly k lines. Write a program to read the file and print the random sample.

We assume we have n lines in the file.

我说假设 memory 可以 handle 的话用一个 hashmap 就解决了, key 是行数, value 就是 string。她说可以, 但是如果 memory 不能 handle 怎么办。我说算一下 string 的 hashCode()。她说可以 但是你怎么能根据 hashCode() 找到 String。我说那就用一个 B+ 数建立 index 吧。然后告诉她怎么 build 这么一个 B+树, 用 bulk loading。然后她说, good, 你把她写出来吧。我顿时就汗颜了, 麻痹的 1000 多行的 B+数 implementation 让我写出来不是扯淡么, 我说我用文字描述行不。她说可以。真心蛋疼。不知道有什么好方法可以不用 B+树的 望指点。

开头 20 分钟问了我的 project, 她是做 search 的, 问我的 project 怎么建的 index, 用的什么方法, 怎么做的 search, 这些我都 high level 的跟她说了说, 她貌似都懂了, 但是问的这个问题实在是蛋疼啊。大家给点意见。

然后是那道经典的系统设计题目: 1 million urls from last hour are stored in the file, find the top K url in terms of the frequency.

直接说了 Hashmap 扫一遍, 然后用 size K 的 Min-Heap 过一遍。然后他就开始追问, 如果 input 改为一个 unlimited data stream, 怎么 online 的 update top K urls for last hour? 我说最笨的方法是你动态的 update HashMap, 然后每次再用 Min-Heap 过一遍。如果只想得到近似解, 可以用一个 size 2K 的 minHeap, 动态的更新 Min-Heap。之后又补充道, 可以对初始的 input 先 sort, 然后根据 HashMap 的变动维护 sorted results, 每一次插入只是  $\lg n$  的开销。他没说什么, 我也不确定最优解是什么。他这时候开始在 collabedit 上敲函数, 想让我 coding 这个问题, 看样子却是个字符串处理。

结果题目还没打完。他说时间不多了, 他又问, 如果想动态的知道 last day, last hour, last minute 的结果, 该怎么改进这个系统, 我说先明确最小时间单位, 如果是分钟, 就按每分钟存 HashMap, 然后根据 query 整合。他说很好的 starting point。英语还是不够好, 他经常不太明白我说什么, 解释一下, 1 个小时的时间就到了。想问一下, 大家觉得这个系统设计问题的最优答案是什么?

<http://ipv6.weiming.info/zhuti/JobHunting/32380269/>

周五的电面

网投简历，一周后 hr 一轮电面，就是 list 一堆题网上都有，udp tcp 区别，ssl，linux 命令，bit 的简单题，match 关键字即可

然后电面，之前大概看了一下 glassdoor 和本版的面经，电面的时候一个都没命中：为什么投 yelp，简历上的 project，多重继承，ood

最后 15 分钟 coding

一个树：food

//\

america, china, mexico, indian

/\

mifan, mianshi

/\

jiaozi baozi

要求先设计树的数据结构，然后任意给一个 string，比如 mianshi(面食) 返回饺子和包子，如果给 china，返回 mifan，jiaozi 和 baozi，就是返回一个 child 下的全部叶子  
当时脑子抽了

第二次 engineer skype 电面，建议事先把视频麦克设置好，我因为这个耽误了 10 分钟

上来让我介绍 research

然后 coding 题就是 cc150 上的 min stack

之后问 unix ls 命令操作系统做了什么

system call 的作用和详细过程

平时用什么版本控制软件，更喜欢哪种，为什么

linux 创建硬链接用什么命令，参数是什么

最后让我问他问题

之后挂了，估计是后面那些操作系统的细节没答好

可能忘掉一些题，单说我记得的吧：

- (1) powerset and permutation
- (2) mapreduce 计算 click-through rate
- (3) 按层 print 树（带 indent，非 binary），输入是只有父节点的 node list
- (4) Top k urls in a string stream
- (5) Mysql 如何解决 deadlock 问题

已悲剧，估计是挂在(5)了，我说我用 mysql 熟，但对 deadlock 解决不熟悉，他还穷追不

舍，要我当场想，我想了说 timeout 吧，他说对，但还有其他方法吗，一定要我想三四个不同方法，然后就 stuck 在那了，也不 move on 到下一个问题，晕死

yelp back-end 的 new grad 面试：

第一轮，screen phone interview。database,unix 和网络方面基础知识，如果都学过，平时也用 unix，基本不用复习。后来听说 glassdoor 上有很全的 screen interview 的面经。

第二轮，skype interview。聊聊之前的 project 和为什么要申请 yelp。然后做题，leetcode 上简单题的难度。如果做过一遍 leetcode oj，应该没有什么问题。当天下午 hr 电话过来问是否 fresh（因为最近半年都在实习），回答是。然后就预约 onsite 时间，没有 code test。

第三轮，onsite interview。因为签了 agreement，所以不露题了。但只要 leetcode 中等以及简单题能轻松搞定，那应该没有什么问题。一共 4 轮面试，都是 manager。问题除了白版 coding 外，主要集中在之前的 projects 以及一些开放性的设计题（不是 amazon 那种 class 设计），怎么来解决一些问题，基本只能看平时的积累以及临场发挥。四轮中有一个直接上机写代码，由于没有怎么用过 mac，非常出丑的请求面试官帮忙 copy 一行前面的代码。。。

最后，hr 随便聊聊，了解比如什么时候可以开始工作，预期工资，对哪个组比较感兴趣之类。个人挺喜欢 yelp 的氛围，感觉挺年轻挺有活力的。

然后又去面了一次 Yelp，Yelp 在三番市区里，我按他们的地址找到了一个连 yelp 标语都没有的大楼，进去后我问 recruiter 为什么找这么隐蔽的地方，他说他们要 low key，避免很多店家缠着他们改搜索算法提高排名。Yelp 比我想象的小很多，所有的 engineer 都在三番的 office，也就 2 层楼 200 来人。中午听了一个关于如何买机票的 presentation，然后开始面试，一共四轮。第一轮让我写一个输入栏自动纠错的 function，题目我忘了，就是一个简单的 recursion。第二面是让我现场读一段 python 代码，这段代码完成的是一个正则表达式的功能，让我找出 bug 并且更正，我读代码特别讨厌旁边有人，那段代码本来就写得很精炼，而且带 recursion，那个哥们还不停的催我，当时就跪了。第三面问一个数组的元素大多出现 2 次，只有一个元素出现 1 次，找出这个出现 1 次的数。我直接跟他说这个题 XOR 一遍就 OK 了。他又问如果不能 XOR 呢，我就用 hashset 写了。第四面问的什么记不清了，后来 recruiter 进来了，问我现在有哪些 offer，我跟他实话实说了，他就问我对 Apple 印象如何，我说 Apple 很好啊，不过 Yelp 也很好啊。我问他们 Yelp 的 mobile app 没有广告如何赚钱，他说我们是一个 community，为人民服务，言下之意就是确实不怎么能挣钱。后来受到了拒信。

<http://www.1point3acres.com/bbs/thread-45048-1-1.html>

第三轮，连着 2 个 45 分钟 skype interview。一个工程师，一个 manager。

工程师问的题：

1. 如何改进 Yelp。
2. 一个矩阵，每个 element 是一个字母。找出矩阵里的所有 valid 的单词，假设有一个函数 `checkDictionary(String)`，返回 true 如果这个单词是 valid 的。

manager 聊我简历的项目聊了 20 分钟，然后做了一个题，我问他几个问题。

题目是：

给一个 Node 数组，乱序，Node 类如下：

```
Node {  
    int id;  
    int parent_id;  
    String label; // 如 NodeA, NodeB  
}
```

要求写一个函数打印如下：

```
NodeA  
  NodeB  
    NodeD  
  NodeC  
  NodeG  
NodeF  
  NodeD
```

二面睡过头了，晚了 15 分钟，打回去，本来是 45 分钟的面试，就面了 30min。

先让我聊了聊炸出翔游戏的实现，然后问我 process 和 thread 的区别，如何 lock thread，如何实现 process 间通信。然后问从旧金山 query 华盛顿的数据库，怎么判断 latency 中 query 时间和传输时间各占多少。如何优化 query 速度，那天晚上刚好考 484，之前做 sample 的时候刚好做到，回答先 select 后 join，correct。

第二天收三面通知。

三面：（30min）

问了问炸出翔，然后让我设计一个排序算法，越慢越好。第一反应是 `shuffle sort`，被告知要优化，想不出。面试官告诉我 `shuffle` 出排序好的数列不要直接 `return`。。再 `shuffle` 一次。。。当场吐血，大呼无赖，面试官随即变态般地大笑并转入 `collabedit`。

先问用户的个人页面为什么要用 `hash` 的 `URL`，大概答了答安全性问题就过了。编程题是假设有一个正确的 22 位 `hash`（区分大小写），现在 `IE` 浏览器不小心把大写全变小写了，求一个算法找出正确的 `hash`，并估测在 `query` 一次用 `1ms` 的网速下大概需要多久得结果。一周后收 `final round interview`。

四面拒绝，被 `HR` 吐槽 "I thought you're really excited to have an opportunity of joining Yelp"..

Reconstruct a binary tree given two sequences of node traversals, one from inorder and one from postorder traversal.

"We have a fairly large log file, about 5GB. Each line of the log file contains an url which a user has visited on our site. We want to figure out what's the most popular 100 urls visited by our users.

3. 网站变慢，可能有什么因素？`load balancing` 有问题，图片太多，数据库操作太频繁，服务器自身的硬件能力有问题 `etc`。。。

4. `min stack` 问题，拓展 `stack` 实现 `min()` 这个函数的功能

3. 求 `onsite` 祝福。。。。

4. 问一个小问题： $T(n) = 4 * T(n/2) + (n^2) * \lg(n)$  这个递归式能否用 `master method` 求解，如果可以，如何求解？

`onsite` 4 个人，三个 `manager`，每个人都是 45 分钟，其中 10 分钟聊 `project/dissertation`:

- 1) 三哥，`coding` 中续后续重建 `bst`
- 2) 白人，给定一组无重复的数字和运算符，看是否能算出目标数
- 3) 白人，给定一组节点以及每个节点该父节点的 `id`，按层打印树
- 4) 白人，给定两个 `doc`，求相似度，写 `code`

感觉 `yelp` 没诚意，除了第二个写 `code` 时间有点紧，其他我都花 10 分钟就写完了。四个 `coding` 题，面试官最后都说了没问题，聊的 `project` 也没什么挑刺的，给的每个 `project` 都

是 top conf 的 paper。

这家办公室看上去很像网吧，一共 120 个 engineer，50 多个 intern。面试种种让人觉得很无奈，比如我问其中一个 interviewer，你加入 yelp 3 年，感觉 skillset 方面有什么不一样？他直接很 rude 回答 I learn how to become a manager.

一边面一边心里在苦笑，看得出这帮人其实都很水，想考我 datalog，自己 relational algebra 和 SQL 什么关系都分不清楚，上来就 piss off 我的 paper。我一边纠正还要一边想怎么给他漂亮的台阶下。这家真是个 joke。

电面（用的 skype）

介绍自己，为什么 yelp

问了点关于 project 的东西

猜一个数在 1-N 之间，怎么猜？还现场玩了一回，呵呵~~

如果不知道 N 怎么猜？

什么是 SQL injection？（上过 Security，大概有点印象）

什么是 XSS？（这我就知道了。。。)

我有个网页，用户反映每次打开加载都很慢，说 3 个原因为什么？

（我也不是那么清楚。。。就想起什么说什么，说了 bandwidth，说了 database 优化不够，数据量大很多 join，还说了网页做的太次，全是 image，可以用 caching 解决）

然后他还问了下，HTML，CSS，Javascript 我们可以 caching 哪几个？

（我说 CSS 吧，因为不常变动）

我当时问了问简历上的一个 Project，然后出了一个题，给一个 int[] 数组，给一个 number，问用数组里面的数字通过加，减，乘，除能否得到那个 number 返回 boolean. 数组里的数每个只能用一次，可以都用，也可以只用部分。用 DFS 解的

<http://www.1point3acres.com/bbs/forum.php?mod=viewthread&tid=75421&highlight=yelp>

第一面过了，然后即使各种刷题也没压上这个题目。而且说实话和那个工程师说了很久，我也没明白

求地里大牛

```
import java.util.HashMap;
```

```
// To execute Java, please define "static void main" on a class named Solution
```



```

class Solution {
    public static void main(String[] args) {
        HashMap<String, Double> choices = new HashMap<String, Double>();
        map.put("foo", 0.5);
        map.put("bar", 0.3);
        map.put("baz", 0.2);
        System.out.println(weightedRandomChoice(choices));
    }

    public static String weightedRandomChoice(HashMap<String, Double> choices) {

    }
}

```

大概意思我记得就是希望你 output 的 choice 保证对应的 entry 和概率。比如你随便取多少次，返回 foo 的概率就是 0.5

我已悲剧，不过还是想明白怎么做这个题目。  
看来刷题，基础是关键啊。

2ndpoet 发表于 2013-11-20 14:19 

随机产生 0~1 的一个 double 数字，如果这个数在 0~0.5 之间输出 foo， 0.5~0.8 之间输出 bar， 0.8~1 之间输出 ba ...

## 1. Dining philosopher

They want food so they wait. And end up eating each other.

1. Given a random number generator that returns a number between  $[0,1]$ , how would you build a RNG that accounted for weighting?

WAT?

1. You have two arrays with  $N$  integers in them. Merge those arrays using a recursive algorithm so that the integers in the final array are sorted.

Why not use merge sort?

1. generators(in python) and about minesweeper.

They cray.

1. given a query and a document, highlight the query in the document and return a snippet of the document itself. Don't use existing large 3rd party libraries.

No.

1. Finding the longest palindrome in a given string.

No.

1. query log, and how to get the top  $K$  most frequent ones
- 2.

Put it into MySQL.

1. Maximum subarray problem

Pobwem?

1. Generate all permutations of an alphanumeric string (lowercase/uppercase only if a letter)

Use library.

1. How can you improve the service of Yelp?

Make it more web scale.

1. Pick a random number from weighted list based on weight

2. `list_of_random_numbers[adjust_weight(random_index)]`
3. given a string length 25, consisting of alphabet (lower and upper case) and numbers. in this string some of the upper case letters are lower cased and sent to me. I have a black box function that returns an integer, something like 13 or 14, that says the string I have got is correct else it throws an exception. find the right string given a corrupted string.
4. Reconstruct a binary tree given two sequences of node traversals, one from inorder and one from postorder traversal.
5. memory management (implementation of a garbage collector), compilation of Python code, what happens when a URL is typed into the browser, what can be done to speed up a select query
6. Difference between POST and GET.
7. How to prevent SQL Injection virus?
8. How to prevent Cross site visiting?
9. a fairly large log file, about 5GB. Each line of the log file contains an url which a user has visited on our site. We want to figure out what's the most popular 100 urls visited by our users.
10. min stack
11. Generic classes in C++ vs Java
12. - operating system, process, multi threads, memory models
13. database, in particular, SQL language
14. scripting language: python (but I think answer in perl may be okay)
15. data structure & algorithm: stack, linked-list, hashmap, tree, heap, etc.
16. security (attacking, encryption, etc.)
17. web standards/protocol: http, dns, tcp/ip, ssl, etc.
18. recursion and/or dynamic programming
19. linux knowledge/commands
20. Operating system questions, explaining some concept like process, thread
21. linked list or array for binary search
22. Given a file that contains chat messages, find out the top 10 chatters in an efficient way. Assume

that the usernames are unique. For example, : Howdy : Some message : Something else : I am good Nick1 is the top chatter, find nine others like this.

23. What command would you use to retrieve one row of data from a table using SQL?
24. What is the command for viewing memory management information in Linux?
25. Implement a Lock.
26. Write a function that returns a random line in a file.
27. Diff POST/GET ? What happens if type a URL in your browser ? How cookies are working ?  
HTTPS ?
28. url shortener script, data structure, basic structure, and what to do if the URL is already in the DB, what to do to get analytics ?
29. Which operation is more expensive? Multiply, Divide or Add
- 30.
31. Best and worst case time for hash table lookup
32. How would you find the top 10 queries Yelp receives?
33. How does memory management work in ?
34. Given a big list of searches, how would you write a function that returns top 10 searches?
35. How would you use all the data at our disposal?
36. Describe the difference between get requests and post requests
37. What is the average disk access time?
38. in memory, what's the difference between a primitive and object
39. what's a way to set up a large database to find things quickly
40. what's the difference between hashtable and hashmap?
41. find equilibrium index
42. Generate all permutations of an alphanumeric string
43. Implement and explain a generator and what it can be used for
44. Find the longest sequence of unrepeated characters in a string.
45. Asked about the search time difference between linked list and array.
46. How to map an IP address to a server?

47. Database questions like inner join, indexing.
48. String processing question
49. Write an function to compute if a string is an anagram. Time and space complexity and how to optimize it.
- 50.
51. Building regular expressions to catch specific patterns.
52. how do would you find and kill a process using a lot of CPU?
53. What is Big O notation?
54. given a huge log file and the way id find the top 10 sites in them.
55. hashtable implementation
56. Describe about process and thread.
- 57.
58. How does process communicate with each other?
59. We have a fairly large log file, about 5GB. Each line of the log file contains an url which a user has visited on our site. We want to figure out what's the most popular 100 urls visited by our users.
60. What are the different synchronization mechanisms? Difference between semaphore and mutex
61. Reverse a singly linked linklist
62. Unix commands -> wordcount , finding out process information
63. Explain process after you click a URL ( on server / client side )
64. Given a Number to guess b/w 1-100 , how to point to the number in minimum time
65. Explain DNS resolution
66. Usual sorting algorithm complexity ... can it be made any quicker
67. How might you keep track of where users were coming from to get to your site?
- 68.
69. If you have less memory storage, what would you use - Hash Table or Tree? Why?
70. Given an Array of integers, find an efficient way to determine if it contains 3 numbers that sum to 0

71. How to merge two arrays
72. How would you find how much system resources a process is consuming on linux.
- 73.
74. Whether to use a list or linked list to implement a search tree
75. What is the average look-up time for a good hash-function
76. What situation arises when you have a parent div and all it contains is a pair of floating divs?  
And how can you remedy the situation.
77. Given a telephone directory which is broken in number of files, how would you find a telephone number in it.
78. u have a table with user information like name,age and sex. Give a smart way to sort the entire based on age.