

20181214 数据结构作业

1800022769

张靖昆

20181214

我承诺诚实作业，没有抄袭他人！

1. 第 1 题：

解：由题得，这 n 个关键码检索成功所花费检索长度分别为 $1, 2, 3 \dots n$ ，因此成功检索的平均检索长度 $ASL(n) = 1 * \frac{1}{2} + 2 * \frac{1}{4} + \dots + n * \frac{1}{2^n}$ 。又 $\frac{1}{2}ASL(n) = 1 * \frac{1}{4} + 2 * \frac{1}{8} + \dots + (n-1) * \frac{1}{2^n} + n * \frac{1}{2^{n+1}}$ ，因此 $ASL(n) - \frac{1}{2}ASL(n) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \dots + \frac{1}{2^n} - n * \frac{1}{2^{n+1}} = 1 - \frac{2}{2^{n+1}} - n * \frac{1}{2^{n+1}}$ ，从而 $ASL(n) = 2 - (n+2) * \frac{1}{2^n}$ ，即 平均检索长度为 $2 - (n+2) * \frac{1}{2^n}$ 。

2. 第 2 题：

解：每一步时的 first、mid 和 last 如下表所示。

步骤序号	first	mid	last	判断
1	1	4	8	$26 < 88, \text{first} = \text{mid} + 1 = 5$
2	5	6	8	$56 < 88, \text{first} = \text{mid} + 1 = 7$

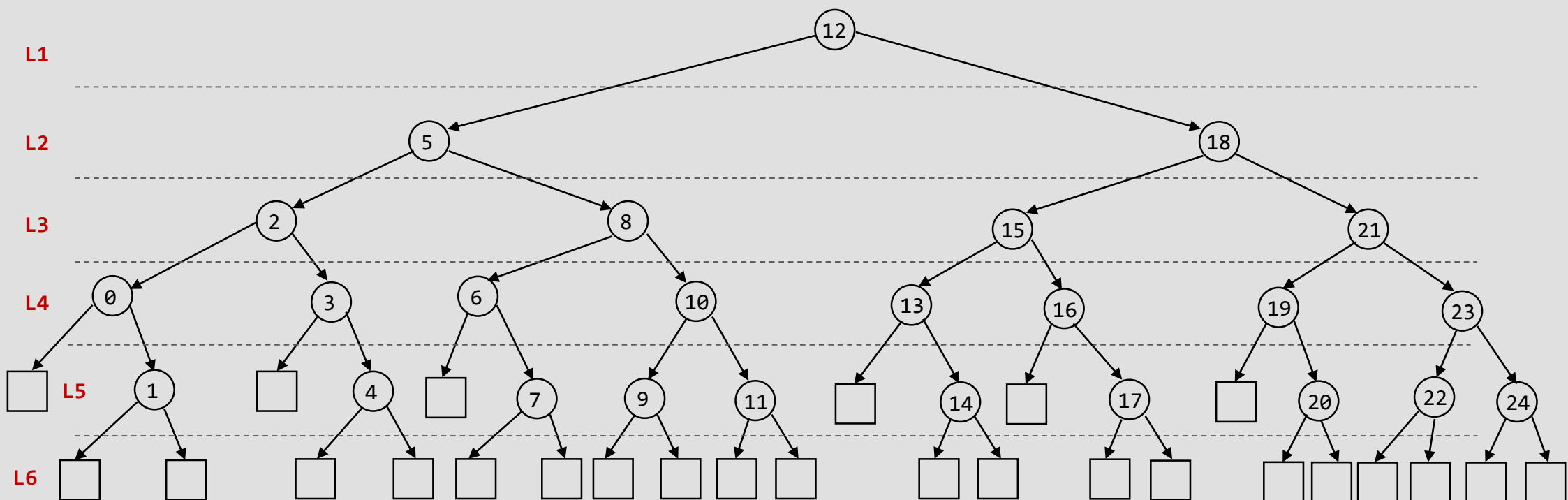
3	7	7	8	找到 88，结束检索
---	---	---	---	------------

3. 第 4 题：

解：以 ASL_s 表示查找成功的平均查找长度，由于题中说查找的是 A[25]中每个元素，因此一定会查找成功。

1) 顺序查找： $ASL_s = (1 + 2 + 3 + \dots + 25) / 25 = 13$ 。

2) 二分检索：以每个元素的下标加圆圈表示内结点，以方块表示外结点，给出对 A[25]进行二分检索的扩充二叉树，如下所示（左子树表示检索值小于当前结点对应的值，右子树表示检索值大于当前结点对应的值）：



从上述扩充二叉树可以看出，查找成功每个值所需要的查找次数如下表所示：

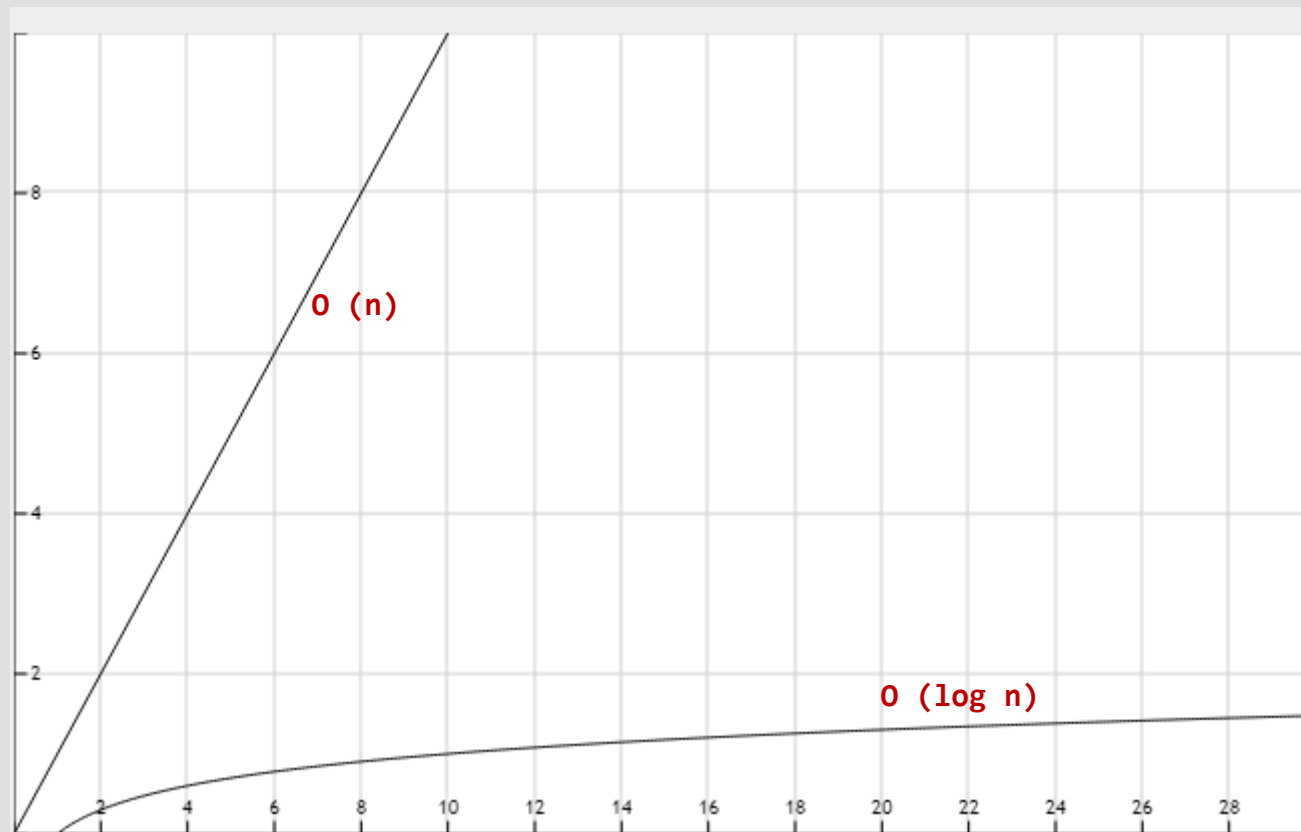
下标	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
次数	4	5	3	4	5	2	4	5	3	5	4	5	1	4	5	3	4	5	2	4	5	3	5	4	5

从而查找成功的平均检索长度(表中次数之和除以 25)为

$$ASL_s = (4+5+3+4+5+\dots+5+4+5)/25 = 99/25 = \boxed{3.96}。$$

3) 分块：由两阶段的顺序检索组成，因此查找成功的平均检索长度为 $ASL_s = (5+1)/2 + (5+1)/2 = \boxed{6}$ 。

4. 第 5 题：请查看 CLion 程序



第 5 小问:

- 1) 如图所示(注意 x 轴的值较大, y 轴值较小), 顺序检索是 $O(n)$ 复杂度, 二分检索是 $O(\log n)$ 复杂度, 上图是算法分析得到的函数图像;

2) 实际运行的情况，先给出几组值：

10K，值域[1-20k]数据

顺序检索进行10k次检索总次数为77542097，成功检索的次数为4111，检索成功百分比为41.11%，每次检索进行的平均比较次数为7754.21。
二分检索进行10k次检索总次数为128252，成功检索的次数为4150，检索成功百分比为41.50%，每次检索进行的平均比较次数为12.83。

20K，值域[1-20k]数据

顺序检索进行20k次检索总次数为246364896，成功检索的次数为12968，检索成功百分比为64.84%，每次检索进行的平均比较次数为12318.24。
二分检索进行20k次检索总次数为266468，成功检索的次数为13052，检索成功百分比为65.26%，每次检索进行的平均比较次数为13.32。

30k，值域[1-30k]数据

顺序检索进行30k次检索总次数为559895140，成功检索的次数为19244，检索成功百分比为64.15%，每次检索进行的平均比较次数为18663.17。
二分检索进行30k次检索总次数为416922，成功检索的次数为19205，检索成功百分比为64.02%，每次检索进行的平均比较次数为13.90。

有如下表：

算法	总检索次数	成功检索次数	成功百分比	平均比较次数
顺序检索	77,542,097	4,111	41.11%	7754.21
顺序检索	246,364,896	12,968	64.84%	12318.24

顺序检索	559,895,140	19,244	64.15%	18663.17
二分检索	128,252	4,150	41.50%	12.83
二分检索	266,468	13,052	65.26%	13.32
二分检索	416,922	19,205	64.02%	13.90

可以看到，二分检索的平均比较次数非常稳定，增长速度很缓慢，这是 $O(\log n)$ 算法的特征，即 n 的规模必须跨越很大倍的提升， $\log n$ 才会有较明显的提升；而顺序检索随着 n 的增大，平均比较次数直线上升，是 $O(n)$ 算法的典型表现。

5. 第 6 题：请查看 CLion 程序

6. 第 11 题：

1) 首先求出每个元素对 23 取余的余数，如下表所示：

元素	32	75	29	63	48	94	25	46	18	70
余数	9	6	6	17	2	2	2	0	18	1

2) 插入元素 32，得到如下散列表：

[illegible]

3) 插入元素 75，得到如下散列表：

[illegible]

4) 插入元素 29，发生冲突，使用线性探测解决冲突得元素 29 插入在下标 7，得如下散列表：

[illegible]

5) 插入元素 63，得如下散列表：

下标	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
----	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----

元素							75	29		32								63						
----	--	--	--	--	--	--	----	----	--	----	--	--	--	--	--	--	--	----	--	--	--	--	--	--

6) 插入元素 48，得如下散列表：

下标	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
元素			48				75	29		32								63						

7) 插入元素 94，发生冲突，使用线性探测解决得元素 94 插入在下标 3，得如下散列表：

下标	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
元素			48	94			75	29		32								63						

8) 插入元素 25，发生冲突，使用线性探测解决得元素 25 插入在下标 4，得如下散列表：

下标	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
元素			48	94	25		75	29		32								63						

9) 插入元素 46，18，70得如下最终散列表：

下标	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
元素	46	70	48	94	25		75	29		32								63	18					

10) 根据上述步骤，我们可以得到每个元素查找成功的次数如下表：

元素	32	75	29	63	48	94	25	46	18	70
查找成功次数	1	1	2	1	1	2	3	1	1	1

**因此平均查找长度为 $(1+1+2+1+1+2+3+1+1+1)/10 = 1.4$ 。

7. 第 13 题：

首先用每个元素对 13 取余，得到如下表：

元素	12	23	45	57	20	03	78	31	15	36
余数	12	10	6	5	7	3	0	5	2	10

1) 利用线性探查法解决冲突，有如下步骤：

a) 插入元素 12,23,45,57,20,03,78 得如下散列表:

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
元素	78			03		57	45	20			23		12

b) 插入元素 31, 发生冲突, 使用线性探查解决得元素 31 插入在下标 6, 得如下散列表:

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
元素	78			45	03	57	<u>31</u>	20			23		12

c) 插入元素 15, 得如下散列表:

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
元素	78		<u>15</u>	45	03	57	31	20			23		12

d) 插入元素 36, 发生冲突, 使用线性探查解决得元素 36 插入在下标 11, 得如下**最终散列表**:

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
元素	78		15	45	03	57	31	20			23	36	12

e) 根据上述过程，得到每个元素的查找成功的检索长度，则得到如下查找次数表：

元素	12	23	45	57	20	03	78	31	15	36
查找成功次数	1	1	1	1	1	2	1	2	1	2

则查找成功的平均检索长度为 $(1+1+1+1+1+2+1+2+1+2)/10 = 1.4$ 。

f) 查找不成功是指对每个散列下标来讲，使用线性探查向后探寻，直至遇到第一个为空的元素停止查找，

因此对每个散列下标查找不成功次数如下表所示：

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
查找失败长度	2	1	3	2	1	5	4	3	2	1	5	4	3

则查找不成功的平均检索长度为 $(2+1+3+2+1+5+4+3+2+1+5+4+3)/13 = 36/13$ 。

2) 使用再散列法寻找下一个下标：

a) 首先插入元素 12,23,45,57,20,03,78 得如下散列表：

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
----	---	---	---	---	---	---	---	---	---	---	----	----	----

元素	78			03		57	45	20			23		12
----	----	--	--	----	--	----	----	----	--	--	----	--	----

b) 插入元素 31, 发生冲突, 使用再散列法计算插入下标: $h_1=5$, $rh(31)=(7*31)\%11+1=9$, $h_2=(5+9)\%13=1$,

发现下标 1 没有元素, 因此可以插入得到如下散列表:

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
元素	78	<u>31</u>		03		57	45	20			23		12

c) 插入元素 15, 得到如下散列表:

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
元素	78	31	<u>15</u>	03		57	45	20			23		12

d) 插入元素 36, 发生冲突, 使用再散列法计算插入下标: $h_1=10$, $rh(36)=(7*36)\%11+1=11$,

$h_2=(10+11)\%13=8$, 发现下标 8 没有元素, 因此可以插入得到如下**最终散列表**:

下标	0	1	2	3	4	5	6	7	8	9	10	11	12
元素	78	31	15	03		57	45	20	36		23		12

g) 根据上述过程，每个元素查找成功的检索长度如下表所示：

元素	12	23	45	57	20	03	78	31	15	36
余数	1	1	1	1	1	1	1	2	1	2

则检索成功的平均检索长度为 $(1*8+2*2)/10 = 1.2$ 。

8. 第 16 题：

1) 解：由于总共有 12 个关键码且负载因子为 0.6，因此 $M=12/0.6=20$ ，则除余法用于取余的因子应当取小于 20 的最大素数即 19，因此散列函数为 $h(K) = K \% 19$ 。

2) 首先 1) 中散列函数求出每个元素对应的余数，如下表：

元素	26	25	20	33	21	24	45	204	42	38	29	31
余数	7	6	1	14	2	5	7	14	4	0	10	12

根据书上所讲，选取小于 19 的最大素数为 17 得到再散列函数 $rh(K) = k \% 17 + 1$ ，从而下一个开放地址的计算公式为 $h_i = (h_{i-1} + rh(k)) \% 19$ 。

****带有□标识的是新插入表中的元素，发生冲突的地方使用深红色加粗标识****

a) 首先，插入元素 26, 25, 20, 33, 21, 24，得如下散列表：

下标	0	1	2	3	4	5	6	7	8	9
元素		20	21			24	25	26		
下标	10	11	12	13	14	15	16	17	18	
元素					33					

b) 插入元素 45，**发生冲突**，根据再散列函数计算得 $rh(45)=45\%17+1=12$ ，则再次计算得到的下标

$h_2=(7+12)\%19=0$ ，发现下标 0 没有元素，插入得到如下表：

下标	0	1	2	3	4	5	6	7	8	9
元素	45	20	21			24	25	26		
下标	10	11	12	13	14	15	16	17	18	
元素					33					

c) 插入元素 204，**发生冲突**，根据再散列函数计算得 $rh(204) = 204 \% 17 + 1 = 1$ ，则再次计算得到的下标

$h_2 = (1+14) \% 19 = 15$ ，发现下标 15 没有元素，插入得到如下表：

下标	0	1	2	3	4	5	6	7	8	9
元素	45	20	21			24	25	26		
下标	10	11	12	13	14	15	16	17	18	
元素					33	204				

d) 插入元素 42，得到如下散列表：

下标	0	1	2	3	4	5	6	7	8	9
元素	45	20	21		42	24	25	26		
下标	10	11	12	13	14	15	16	17	18	
元素					33	204				

e) 插入元素 38，**发生冲突**，根据再散列函数计算得 $rh(38) = 38\%17+1 = 5$ ，则再次计算得到的下标 $h_2 = (0+5)\%19 = 5$ ，**发生冲突**，继续计算 $h_3=(0+5*2)\%19=10$ ，发现下标 10 处没有元素，插入得到如下表：

下标	0	1	2	3	4	5	6	7	8	9
元素	45	20	21		42	24	25	26		
下标	10	11	12	13	14	15	16	17	18	
元素	38				33	204				

f) 插入元素 29，**发生冲突**，根据再散列函数计算得 $rh(29) = 29\%17+1 = 13$ ，则再次计算得到的下标 $h_2 = (10+13)\%19 = 4$ ，**发生冲突**，继续计算 $h_3=(10+13*2)\%19=17$ ，发现下标 17 处没有元素，插入得到如下散列表：

下标	0	1	2	3	4	5	6	7	8	9
元素	45	20	21		42	24	25	26		
下标	10	11	12	13	14	15	16	17	18	

元素	38				33	204		29		
----	----	--	--	--	----	-----	--	----	--	--

g) 插入元素 31，得到如下最终散列表：

下标	0	1	2	3	4	5	6	7	8	9
元素	45	20	21		42	24	25	26		
下标	10	11	12	13	14	15	16	17	18	
元素	38		31		33	204		29		

因此，上表即为最终的经双散列法解决冲突的散列函数表，共发生 6 次冲突。