① 問題定義:

給定 n 个排序好的 key: $k_1, k_2, \dots, k_n$，共有 n+1 个 dummy keys 為: $d_0, d_1, d_2, \dots, d_n$

設其中 k_i 被 search 到的頻率為: p_i, $\forall i=1\dots n$，而 d_i 為 search 到的 frequency 為 q_i

則求 expected search cost (ESC) 最小的 binary search cost，即為 optimal binary search cost

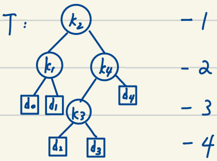
② ESC(T) 之定義

令 depth(k_i) 寫作 $D(k_i)$

$$ESC(T) = \sum_{i=1}^{n}(D(k_i)+1)p_i + \sum_{i=0}^{n}(D(d_i)+1)q_i$$

以右邊的 tree 為 example

T:



i	0	1	2	3	4
p_i		5	2	4	3
q_i	3	2	3	4	2

$ESC(T) = 2\times5 + 2\times1 + 3\times4 + 2\times3 + 3\times3 + 3\times2 + 4\times3 + 4\times4 + 3\times2$

$= 79$

③ characterize an OPT solution

思路: 因為对一个 BST 而言，只會有1个 root 和其左子樹，右子樹

考慮 $k_i, \dots, k_j$ 个 key 的 subproblem，OPT 必定為下面 $j-i+1$ 种

(1). k_i 為 root，$k_{i+1}, \dots, k_j$ 為右子樹，左子樹為空

(2). k_{i+1} 〃 $k_{i+2} \dots k_j$ 〃 〃 k_i

generalized: 設 k_r 為 root，則: $k_i \dots k_{r-1}$ 為左子樹，$k_{r+1} \dots k_j$ 為右子樹

再設 $s[i,j]$ 為 $k_i, \dots, k_j$ 的 OBST 時的 expected search cost

則: 設此 OBST: T^* 的 root 為 k_r，$k_i \dots k_{r-1}$ 為左子樹，$k_{r+1} \dots k_j$ 為右子樹

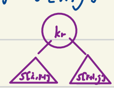
左子樹記作 T_1，$ESC(T_1) = s[i, r-1]$

右子樹 〃 T_2，$ESC(T_2) = s[r+1, j]$

則: $s[i,j] = s[i,r-1] + s[r+1,j] + \underline{\sum_{k=i}^{r-1} p_k + \sum_{k=i-1}^{r-1} q_k + p_r + q_r + \sum_{k=r+1}^{j} p_k + \sum_{k=r+1}^{j} q_k}$

$= \sum_{k=i}^{j} p_k + \sum_{k=i-1}^{j} q_k$

$= W[i,j]$



④ Recursively define the value of an OPT solution

(1). $W[i,j]$: frequency 的和 $= \sum_{k=i}^{j} p_k + \sum_{k=i-1}^{j} q_k$

$$\Rightarrow W[i,j] = \begin{cases} q_{i-1} & \text{if } j=i-1 \\ W[i,j-1] + p_j + q_j & \text{if } i \le j \end{cases}$$

eg $W[1,3]$

i	0	1	2	3	4
p_i		5	2	4	3
q_i	3	2	3	4	2

$W[1,3] = \sum_{j=1}^{3} p_j + \sum_{j=0}^{3} q_j$

$= q_0 + p_1 + q_1 + p_2 + q_2 + p_3 + q_3$

(2). $s[i,j]$ = OBST 的 ESC，即: $ESC(T^*)$

$$\Rightarrow s[i,j] = \begin{cases} q_{i-1} & \text{if } j=i-1 \\ \min\limits_{i \le r \le j} \{ s[i,r-1] + s[r+1,j] + W[i,j] \} & \text{if } i \le j \end{cases}$$

① Recursively define the value of an OPT solution

(1). $W[i,j]$: frequency 的和 $= \sum_{k=i}^{j} p_k + \sum_{k=i-1}^{j} q_k$

$$\Rightarrow W[i,j] = \begin{cases} q_{i-1} & \text{if } j = i-1 \\ W[i,j-1] + p_j + q_j & \text{if } i \le j \end{cases}$$

e.g. $W[1,3]$

i	0	1	2	3	4
p_i		5	2	4	3
q_i	3	2	3	4	2

$$W[1,3] = \sum_{j=1}^{3} p_j + \sum_{j=0}^{3} q_j$$
$$= q_0 + p_1 + q_1 + p_2 + q_2 + p_3 + q_3$$

(2). $S[i,j]$ = OBST 的 ESC，即：$ESC(T^*)$

$$\Rightarrow S[i,j] = \begin{cases} q_{i-1} & \text{if } j = i-1 \\ \min_{i \le r \le j} \{ S[i,r-1] + S[r+1,j] + W[i,j] \} & \text{if } i \le j \end{cases}$$

② Compute value of an OPT solution

例2. 設有4个 key：a_1, a_2, a_3, a_4，會有：$d_0, \ldots, d_4$

而 frequency 為：

i	0	1	2	3	4
p_i	x	5	2	4	3
q_i	3	2	3	4	2

Note：畫表格時，$W[i,j]$ 和 $S[i,j]$ 都是 $i=1, j=0$ 開始

$\because W[i,j]$ 之 definition：$W[i,j] = \sum_{k=i}^{j} p_k + \sum_{k=i-1}^{j} q_k$

$$S[i,j] = \begin{cases} q_{i-1} & \text{if } j = i-1 \\ \min_{i \le r \le j} \{ S[i,r-1] + S[r+1,j] + W[i,j] \} \end{cases}$$

W

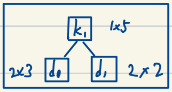
i \ j	0	1	2	3	4
1	3	10	15	23	28
2		2	7	15	20
3			3	11	16
4				4	9
5					2

S

i \ j	0	1	2	3	4
1	3	15	30	56	$\boxed{73}$
2		2	12	31	47
3			3	18	34
4				4	15
5					2

$S[3,3] = S[3,2] + S[4,3] + W[3,3]$

$S[4,4] = S[4,3] + S[5,4] + W[4,4]$



$W[1,1] = W[0,1] + p_1 + q_1$

$W[1,2] = W[1,1] + p_2 + q_2$

$= \sum_{1}^{2} p_i + \sum_{0}^{2} q_i$

$= 5+2+3+2+3 = 15$

$S[1,1] = S[1,0] + S[2,1] + W[1,1] \Rightarrow k_1$ 當root

$S[1,2] = S[1,1] + S[3,2] + W[1,2] = 15 + 3 + 15$

$= S[1,0] + S[2,2] + W[1,2] = 3 + 12 + 15 = 30$

$S[2,2] = S[2,1] + S[3,2] + W[2,2] = 12$

$S[1,3] = S[1,0] + S[2,3] + W[1,3]$

$= S[1,1] + S[3,3] + W[1,3]$

$= S[1,2] + S[4,3] + W[1,3]$

104 交大資工

設有：$k_1 \sim k_5$ 5个 keys

共有：$d_0 \sim d_5$ 6个 dummy keys

令 p_i 為 k_i 被 search 到的机率

q_i 為 d_i 〃

i	0	1	2	3	4	5
p_i	X	0.25	0.2	0.05	0.2	0.3
q_i	0	0	0	0	0	0

☆. 需要 6×6 table

$W[i,j]$

$i \backslash j$	0	1	2	3	4	5
1	0	0.25	0.45	0.5	0.7	1.0
2		0	0.2	0.25	0.45	0.75
3			0	0.05	0.25	0.55
4				0	0.2	0.5
5					0	0.3
6						0

$$W[i,j] = \sum_{k=i}^{j} p_i + \sum_{k=i-1}^{j} q_i$$

$$W[i,j] = \begin{cases} q_{i-1} & \text{if } j = i-1 \\ W[i,j-1] + p_j + q_j \end{cases}$$

$S[1,1] \Rightarrow S[1,0] + S[2,1] + W[1,1]$

$S[2,2] \Rightarrow S[2,1] + S[3,2] + W[2,2]$

$S[i,j]$ $\quad S[1,2] = \min\{(S[1,0] + S[2,2] + W[1,2]),\ (S[1,1] + S[3,2] + W[1,2])\}$

$i \backslash j$	0	1	2	3	4	5
1	0	0.25	0.65	0.8	1.25	2.1
2		0	0.2	0.30	0.75	1.35
3			0	0.05	0.30	0.85
4				0	0.2	0.7
5					0	0.3
6						0

⇒ 可得 OBST 的 expected search cost (optimal value) 為 2.1

$$S[i,j] = \begin{cases} q_{i-1} & \text{if } j = i-1 \\ \max\limits_{i \le r \le j} \{ S[i,r-1] + S[r+1,j] + W[i,j] \} \end{cases}$$