

107 Data Structure Homework 3

一、霍夫曼樹

- Input : read file (input1.txt)
 - 用霍夫曼樹壓縮 input1.txt
 - 使壓縮後的檔案大小比壓縮前的小
 - input1.txt的文字內容由ASCII組成
 - 一個檔案，一個測資
- Output
 - 輸出一個 output.txt，內容為壓縮過後的內容 (50%)
 - 輸出壓縮前跟壓縮後的檔案大小，單位固定為bytes，此結果不要寫在 output.txt內 (10%)

霍夫曼樹

It is a banana.

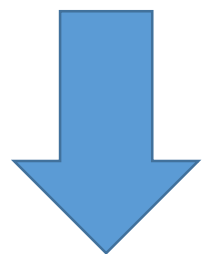
1. 計算頻率

字元	I	t	i	s	a	b	n	.	space
頻率	1	1	1	1	4	1	2	1	3

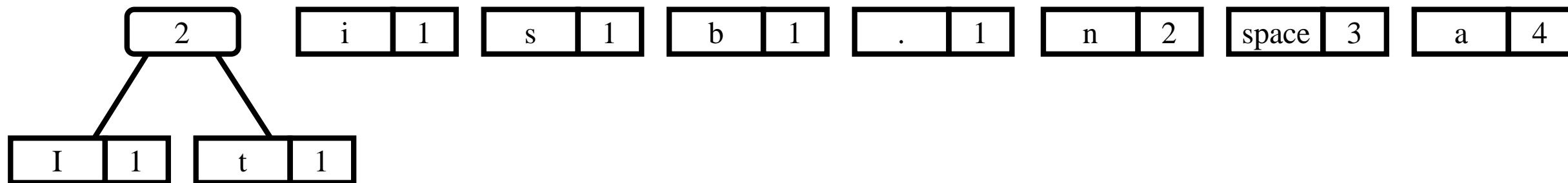
2. 依照頻率由小到大排序

I	1	t	1	i	1	s	1	b	1	.	1	n	2	space	3	a	4
---	---	---	---	---	---	---	---	---	---	---	---	---	---	-------	---	---	---

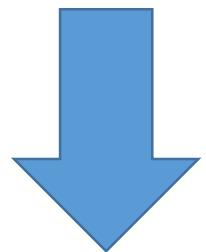
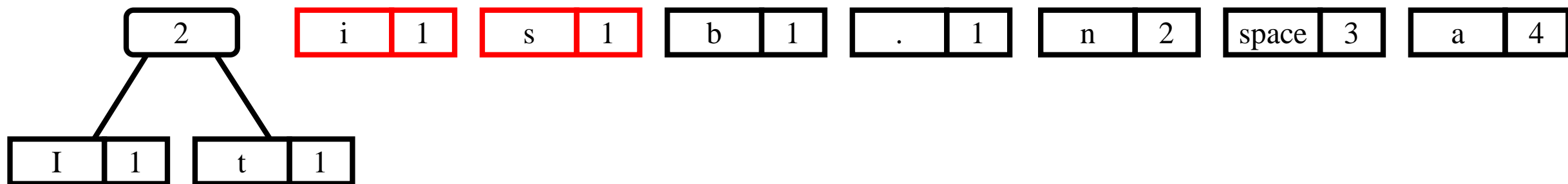
霍夫曼樹



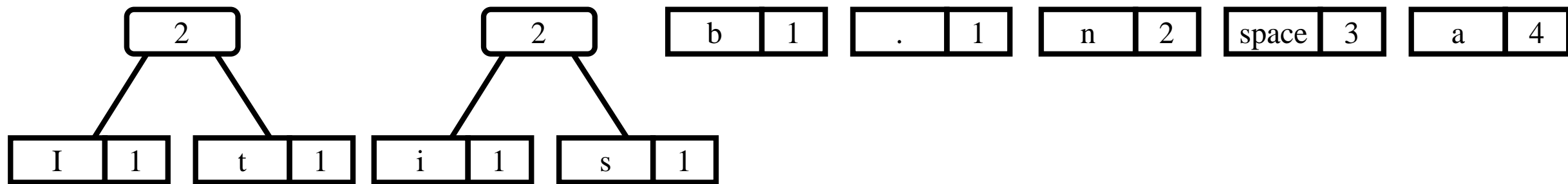
3. 選最小的兩個合併



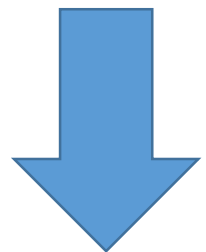
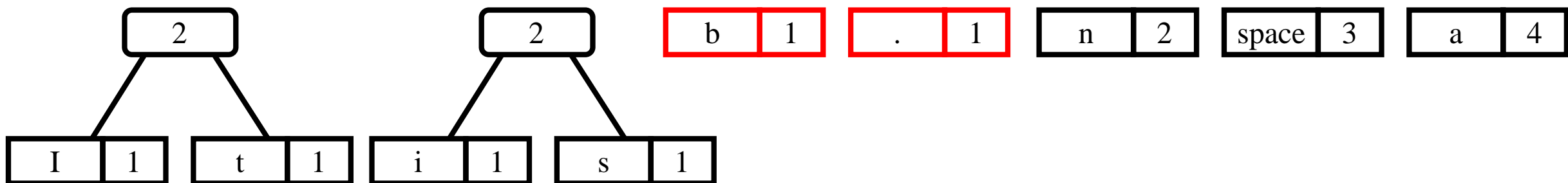
霍夫曼樹



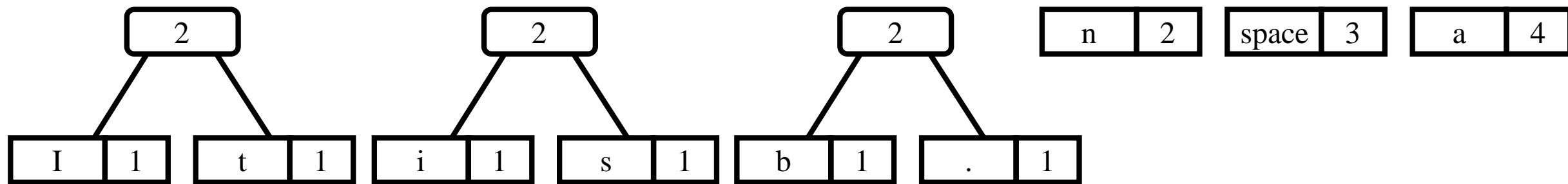
3. 選最小的兩個合併



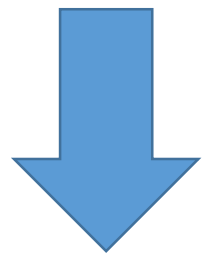
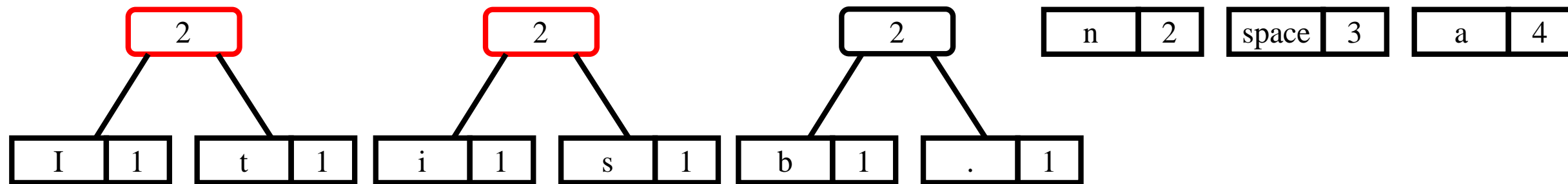
霍夫曼樹



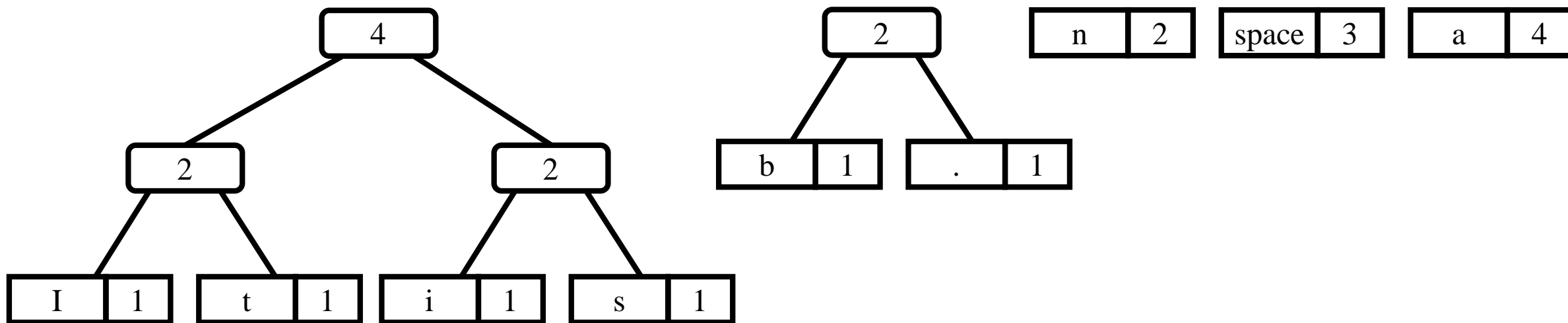
3. 選最小的兩個合併



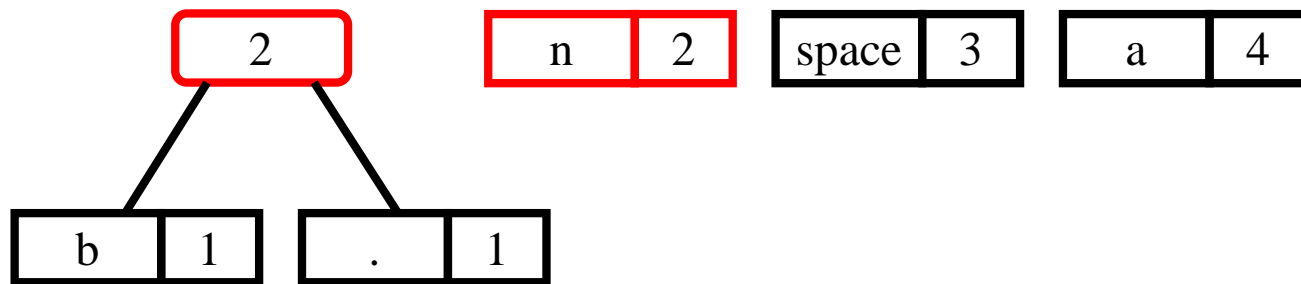
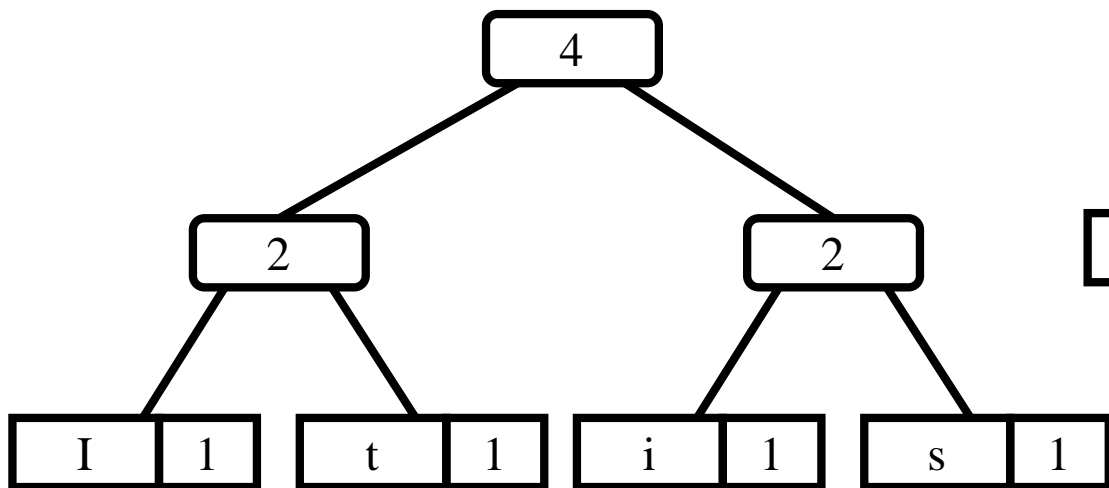
霍夫曼樹



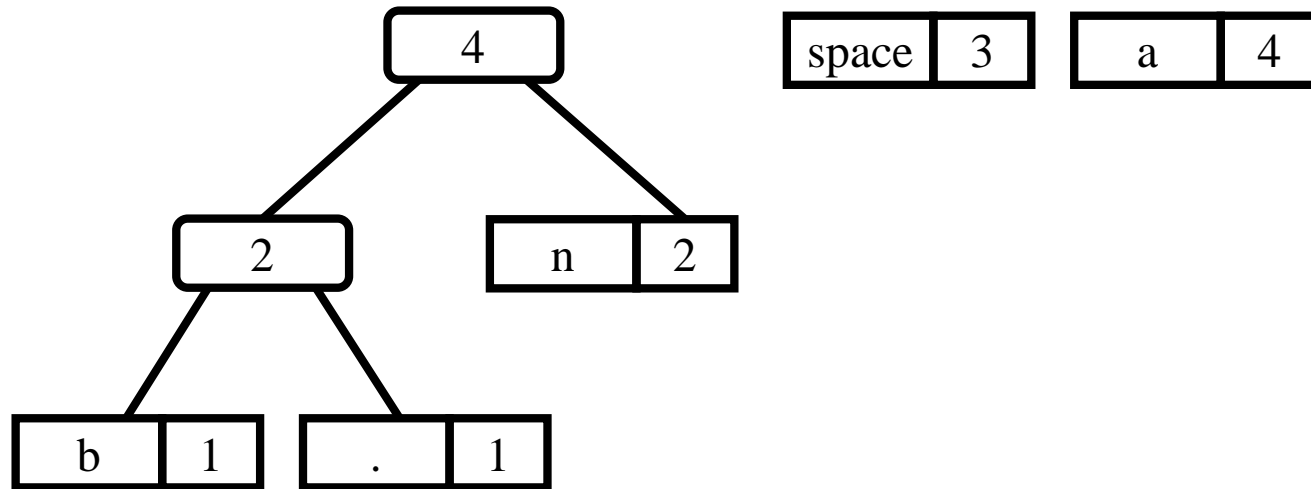
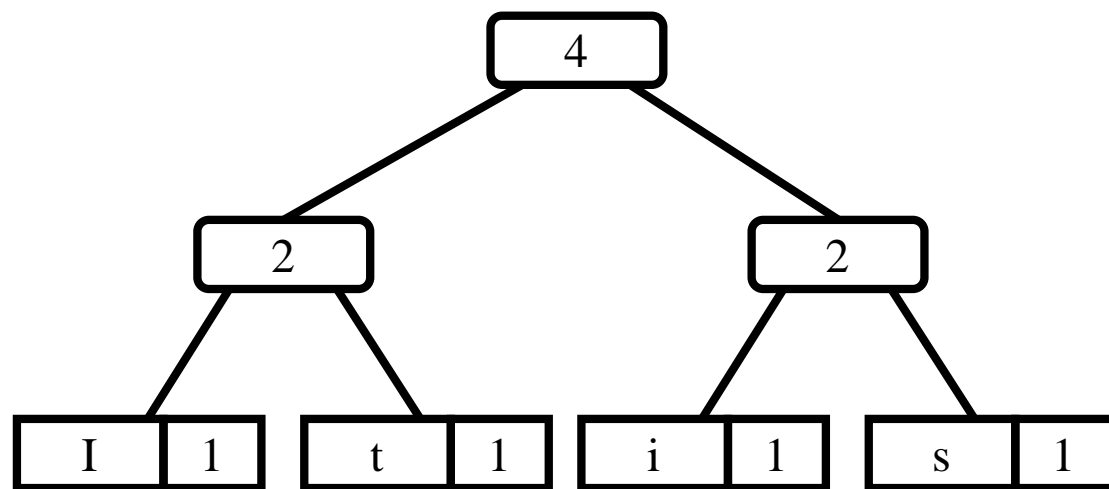
3. 選最小的兩個合併



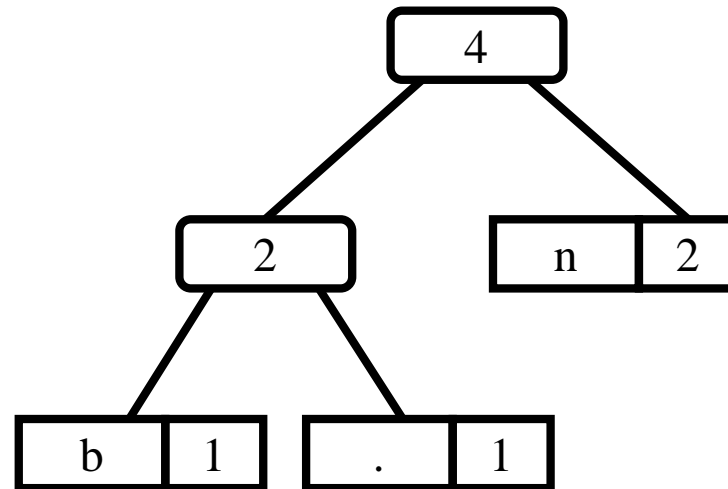
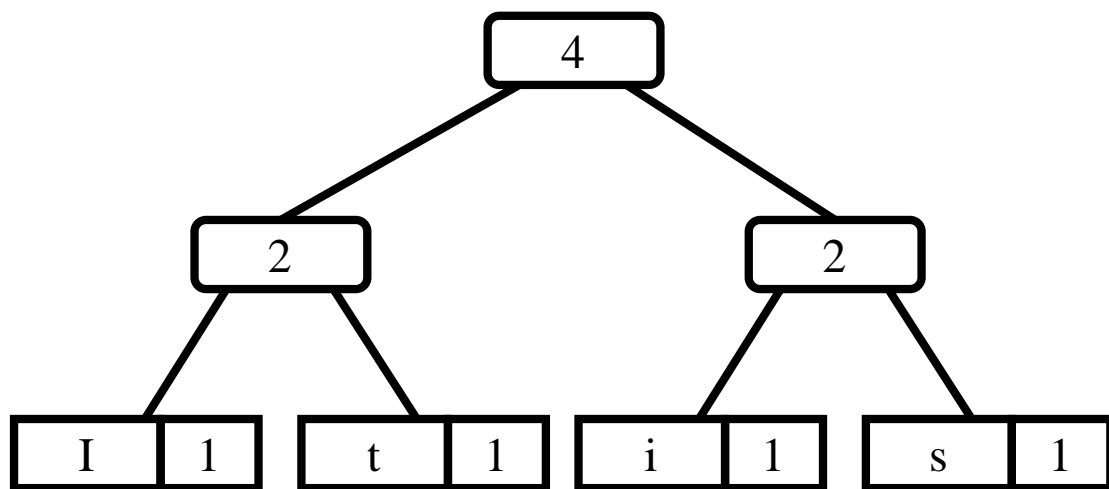
霍夫曼樹



3. 選最小的兩個合併



霍夫曼樹

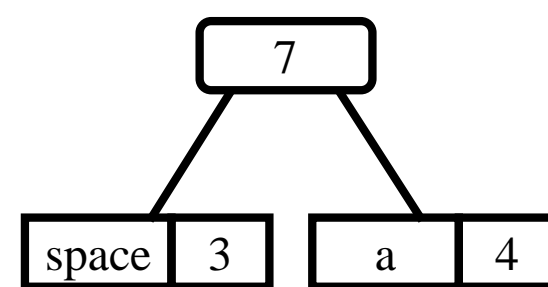
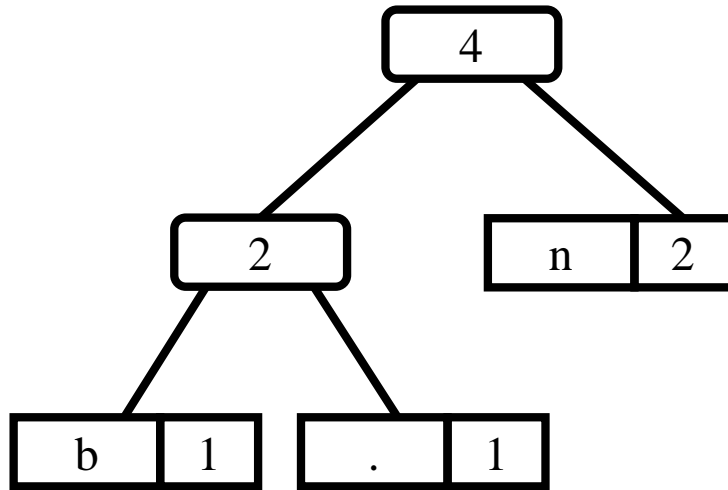
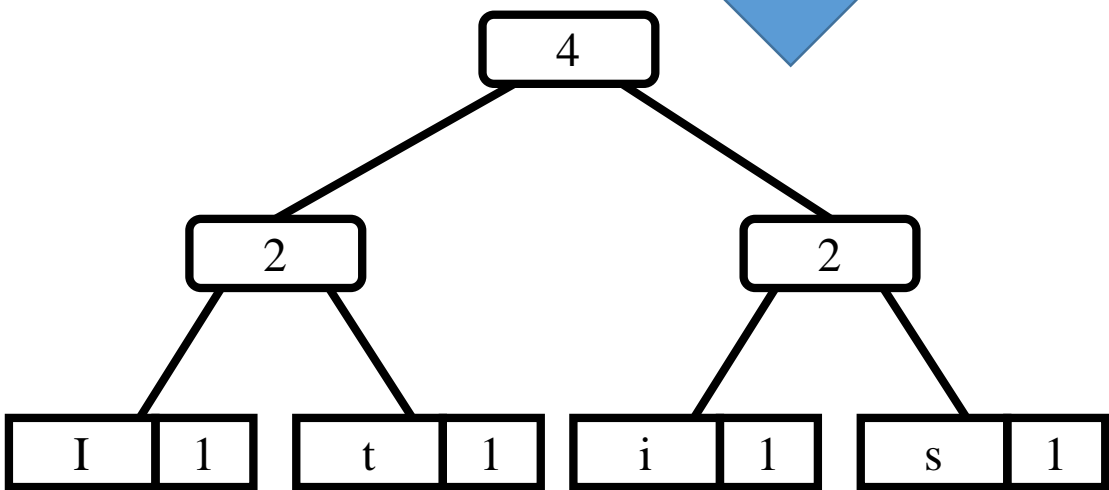


space 3

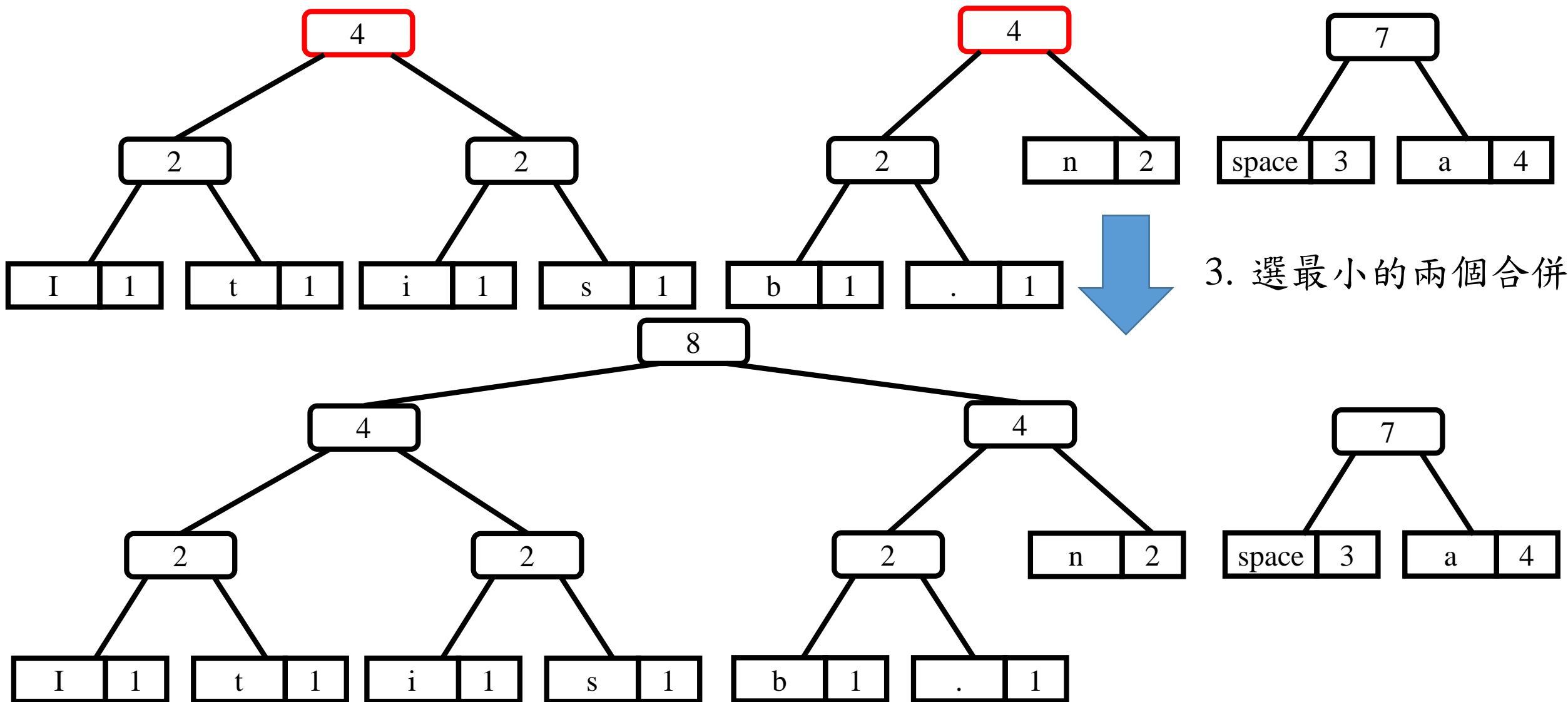
a 4



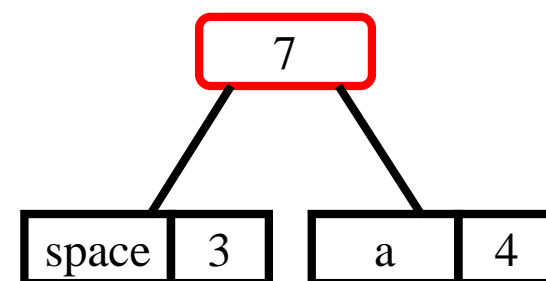
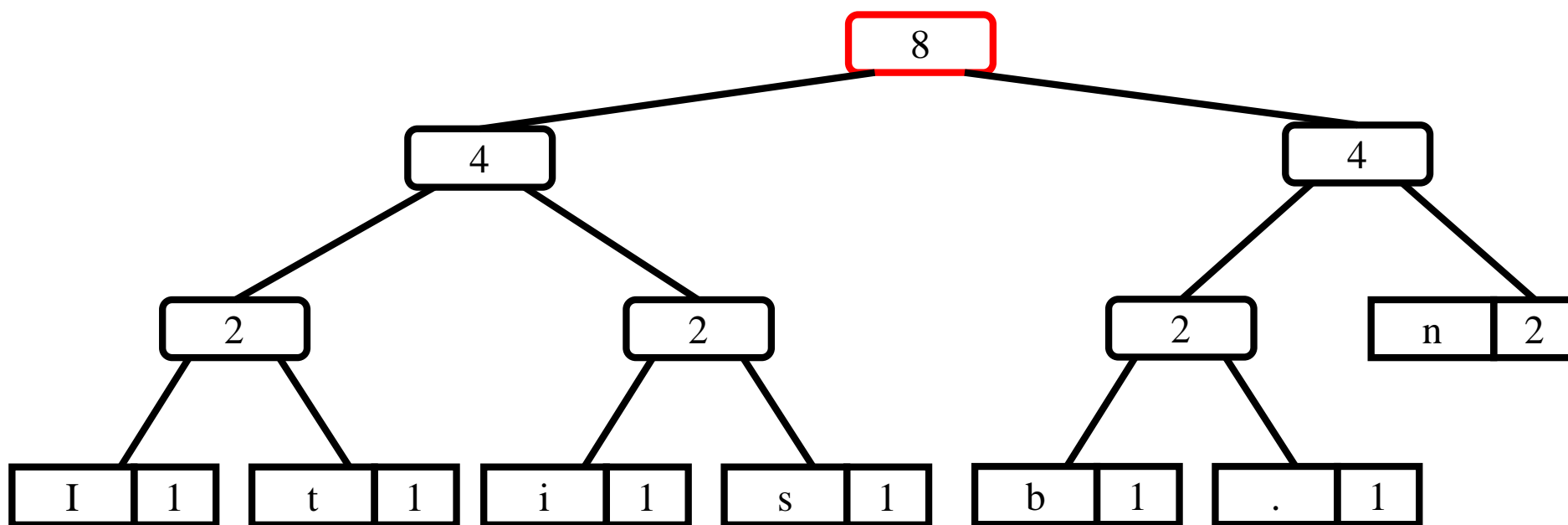
3. 選最小的兩個合併



霍夫曼樹



霍夫曼樹

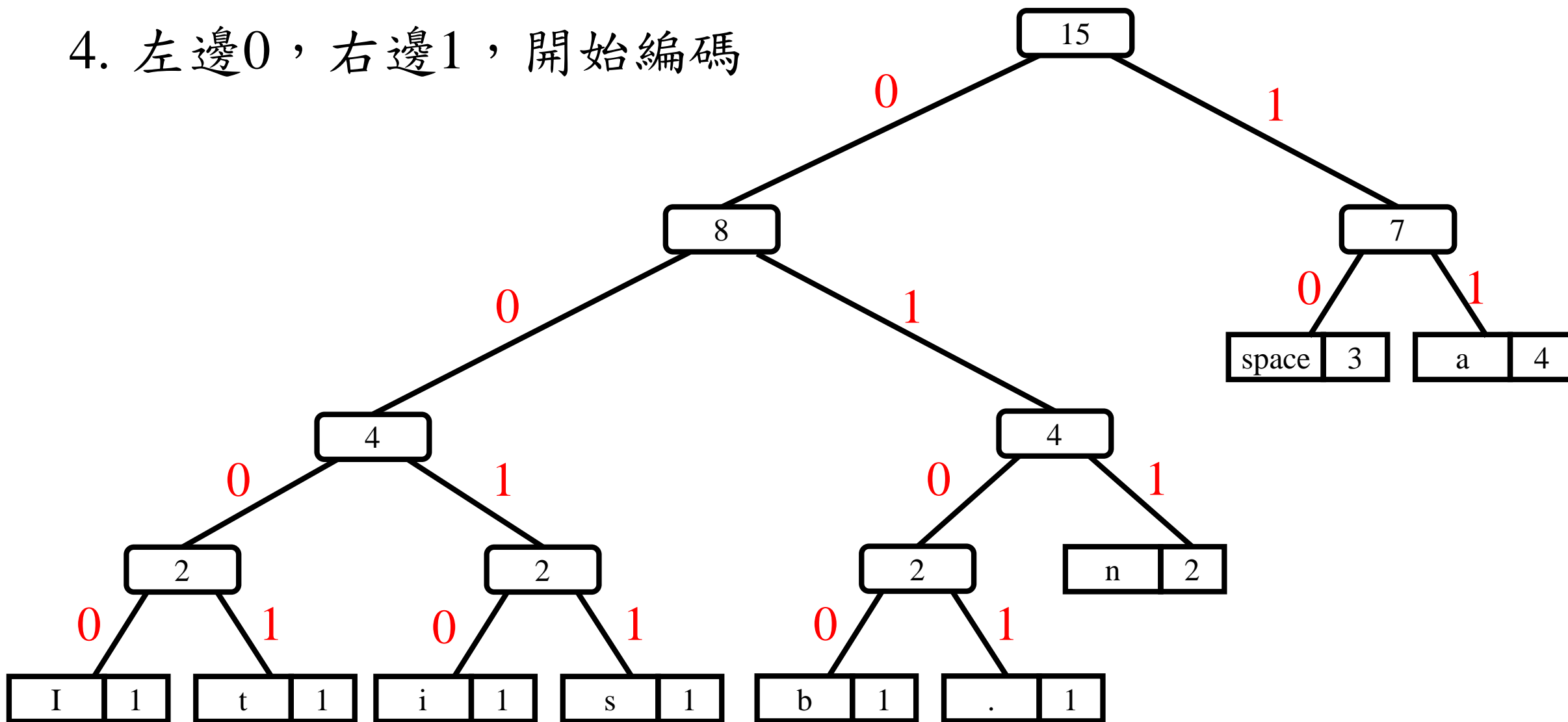


3. 選最小的兩個合併

霍夫曼樹

字元	I	t	i	s	a	b	n	.	space
頻率	1	1	1	1	4	1	2	1	3
編碼	0000	0001	0010	0011	11	0100	011	0101	10

4. 左邊0，右邊1，開始編碼



output.txt

- 編碼結果

字元	I	t	i	s	a	b	n	.	space
頻率	1	1	1	1	4	1	2	1	3
編碼	0000	0001	0010	0011	11	0100	011	0101	10

- It is a banana. 轉換成霍夫曼編碼

I	t	space	i	s	space	a	space	b	a	n	a	n	a	.
0000	0001	10	0010	0011	10	11	10	0100	11	011	11	011	11	0101

- 轉換結果:

00000001100010001110111001001101111011110101

提醒：如果將以上直接寫入檔案會越壓越大，存對的話開出來會是亂碼。

輸出格式

```
file size before compression: 15bytes  
file size after compression: 6bytes
```

```
-----  
Process exited after 0.02046 seconds with return value 0  
請按任意鍵繼續 . . .
```

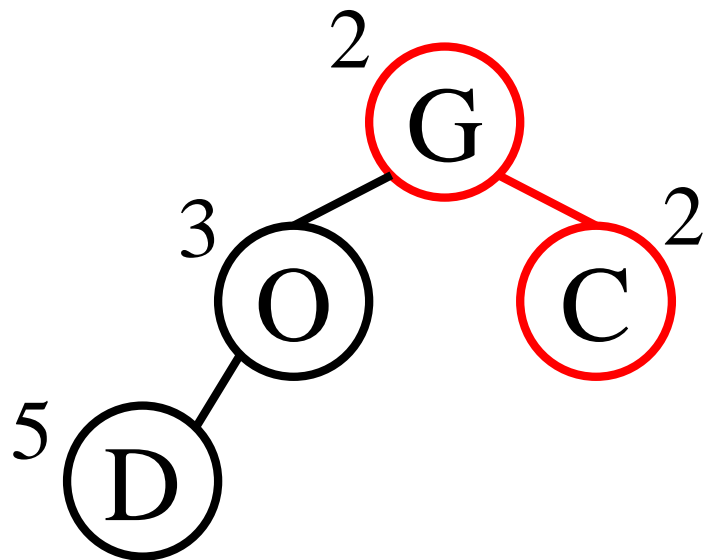
- 判斷標準
 - 要用霍夫曼樹壓縮
 - 檔案大小：壓縮前 > 壓縮後
 - output.txt檔是亂碼

二、完整二元樹

- Input : read file (input2.txt)
 - 輸入多個包含節點名稱及節點權重的節點
 - (節點1名稱,節點1權重), (節點2名稱,節點2權重)...(節點n名稱,節點n權重)
 - 節點名稱：大寫字母表示
 - 節點權重：0 ~ 1000
 - input2.txt 裡有多筆測資，一行代表一筆測資
- Output
 - 由根節點往葉節點的方向輸出最小權重的路徑 (30%)
 - 輸出該路徑的總權重 (10%)

二、完整二元樹

Input : (G,2),(O,3),(C,2),(D,5)

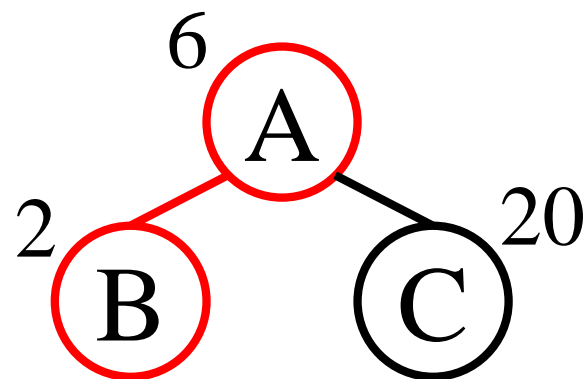


Output :

Path : GC

Weight : 4

Input : (A,6),(B,2),(C,20)



Output :

Path : AB

Weight : 8

Tree Problems

- Send code
- Only accepted C
- Homework upload to **iLearning** platform
- **Deadline : 2018.12.10 23:59**
- 繳交檔名 : main1.c, output.txt, main2.c
- 給PPT上的input和output，有隱藏測資
- 截止日期前可以一直重新上傳
- 作業Deadline後才會改，分數可以直接上**iLearning**看
- **抄襲嚴懲，請打上註解，不接受忘記打**
- **交不合以上規定的作業，將直接不計分(讀檔和編譯問題除外)**