

前置碼 (prefix code)

假設 P 為一個集合，裡面的元素為二元序列，若 P 中任一序列不為其它序列的前置 (prefix)，則稱 P 為前置碼 (prefix code)。

例題：判斷以下集合是否為前置碼。

1. $P_1 = \{0, 1, 00, 01, 10, 11\}$

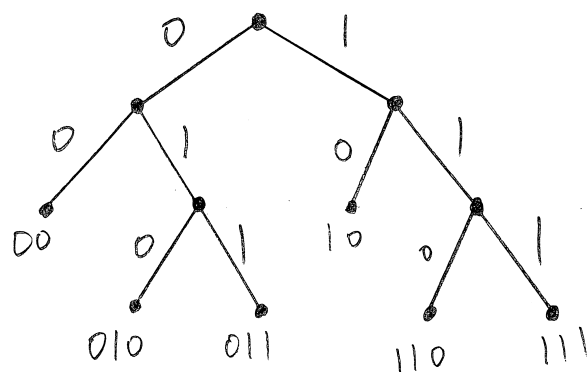
Ans. P_1 不為 prefix code，因為 0 是 00, 01 的 prefix，1 是 10, 11 的 prefix。

2. $P_2 = \{00, 010, 011, 10, 110, 111\}$,

Ans. P_2 為 prefix code，00 和 10 都不為其它序列的 prefix，且有 3 個 bit 的序列都不相同。

Hint: 先檢查短 bit 數的序列，如果 bit 數都相同，且沒有重覆，即此集合為 prefix code。

3. 檢查此滿二元樹是否為 prefix code



Ans. 由左節點路徑為0, 右節點路徑為1的方式標記路徑, 由樹葉記錄路徑, 得出路徑的序列。

將所有樹葉的序列寫成一個集合,

得

$P = \{00, 010, 011, 10, 110, 111\}$, 00和10不為其它序列

的 prefix, 且其它3個 bit 的序列皆不相同,

所以此 full binary tree 為 prefix code.

例題

A prefix code for $\{a, b, c, d, e, f\}$ is given by

$a: 00$, $b: 01$, $c: 101$, $d: x10$, $e: yz1$, $f: 10w1$,

where $x, y, z, w \in \{0, 1\}$. What are the x, y, z , and w ?

Ans.

先檢查 2 個 bit 和 3 個 bit 的 sequence。

$a: 00$, $b: 01$ 對 $c: x10$

則

$x=1$ 才不會有 b 是 c 的 prefix。

再看 $e: yz1$ 對 $a: 00$, $b: 01$, $c: 101$, and $d: 110$

已知 yz 不能是 00 , 01 , and 10 , 那剩下只能是 11

$\therefore y=1, z=1$

最後看 $f: 10w1$ 對 $c: 101$, $d: 110$, and $e: 111$,

可知 c 有可能是 f 的 prefix, 則 w 取 0 才能避免。

$\therefore (x, y, z, w) = (1, 1, 1, 0)$ #

前置碼的應用:

機器語言的世界，只能接受 0, 1 的訊息，若要傳送自然語言 (例如字母)，需將字母編成 0, 1 的二進制編碼。

Hint=

十進位的單位是以 10 為一個單位。

	第3位	第2位	第1位
...	百位	十位	個位
	10^2	10^1	10^0

二進位是以 2 為一個單位。

	第3位	第2位	第1位
...	2^2	2^1	2^0

位數計算：取那進位當底數的 \log 的 ceiling.

例如：

8_{10} , 1013_{10} , and 123_{10} 需幾位十進位，需幾位 2 進位 bit?

Ans.

十進位位數： $\lceil \log_{10} 8 \rceil = 1$, $\lceil \log_{10} 1013 \rceil = 4$, $\lceil \log_{10} 123 \rceil = 3$

二進位位數： $\lceil \log_2 8 \rceil = 3$, $\lceil \log_2 1013 \rceil = 10$, $\lceil \log_2 123 \rceil = 7$

26 個字母可用 $\lceil \log_2 26 \rceil = 5$ 個 bit 來編號 (現在不考慮大小寫)

每個字母固定有 5 bit 的編號，例如給 a 為 00000，

b 為 00001，...，z 為 11010，每次傳輸一個字母

需 5 個 bit 為一組，但是這種傳輸方式，並不有效率，

因為每個字母的使用頻率不同，希望能將使用頻

率高的字母，用少的 bit 來傳輸，這樣便可提高

傳輸效率。

現在問題是每個字母的 bit 數不同，來了一串串聯的

0, 1 序列，機器不知道如何去翻譯，要幾個 bit

為一組。

例如：(此 set 不為 prefix code)

$$P_1 = \{0, 1, 00, 01, 10, 11\}$$

a: 0, b: 1, c: 00, d: 01, e: 10, and f: 11

有事資料為：000111，

機器可能翻譯為 aaabbb, cabf, cabf, ...，

所以有 prefix code 可以保證有唯一的解碼。

ex: $P_2 = \{00, 01, 10, 110, 111\}$, a: 00, b: 01, c: 10, d: 110, e: 111

資料: 10110111 \rightarrow 翻譯為 cde

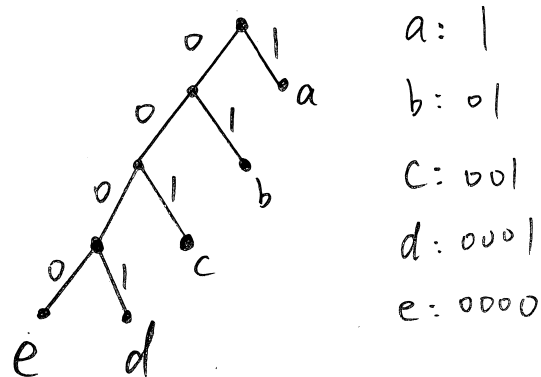
利用建立二元樹來得出前置碼，使用頻率高的離樹根近，使用頻率較低的離樹根就越遠。

現在問題是如何創出二元樹，

給個例題：

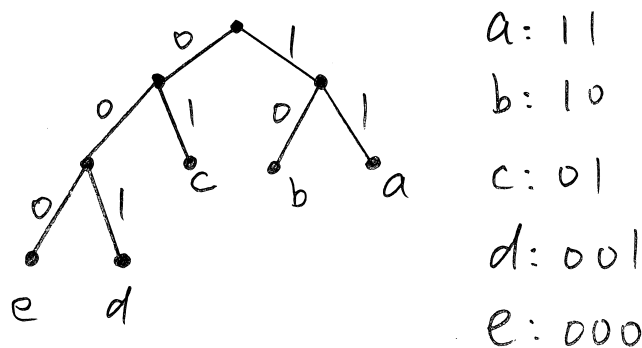
字母	頻率
a	9
b	7
c	5
d	4
e	1

簡單的想法創的 binary tree



每走樹的一個邊都會算是1個成本，且每個邊的權重一樣，這樣走的次數越多，成本就越大，所以需找出走邊數少的方法。

這樣要4個 bit 才能翻譯，而較佳的創建方式為



只需3個 bit，找出這種樹的樹稱為最佳樹(optimal tree)。