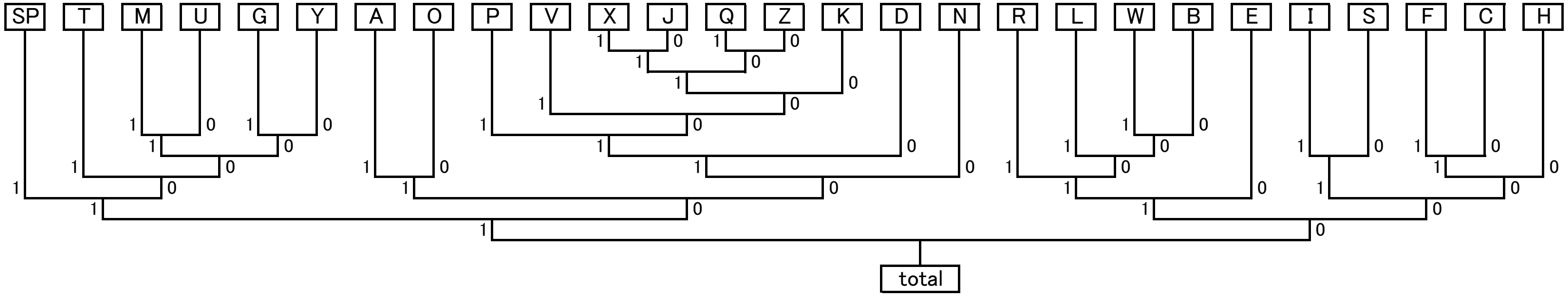
An example of Huffman coding table for alphabetical character by H.Sawami

Table for probability in **http://www.slp.ics.tut.ac.jp/nakagawa/bl/06.pdf** is used.



chr	code	bit	p	−log2(p)
SP	111	3	0.1817	2.4604
A	1011	4	0.0668	3.904
B	011000	6	0.01179	6.4063
C	00010	5	0.0226	5.4675
D	10010	5	0.031	5.0116
E	010	3	0.1073	3.2203
F	00011	5	0.02395	5.3838
G	110001	6	0.01633	5.9363
H	0000	4	0.04305	4.5378
I	0011	4	0.0519	4.2681
J	1001100110	10	0.00108	9.8548
K	10011000	8	0.00344	8.1834
L	01101	5	0.02775	5.1714
M	110011	6	0.02075	5.5907
N	1000	4	0.0581	4.1053
O	1010	4	0.0654	3.9346
P	100111	6	0.01623	5.9452
Q	1001100101	10	0.00099	9.9803
R	0111	4	0.0559	4.161
SP	0010	4	0.0499	4.3248
T	1101	4	0.0856	3.5462
U	110010	6	0.0201	5.6367
V	1001101	7	0.00752	7.0551
W	011001	6	0.0126	6.3104
X	1001100111	10	0.00136	9.5222
Y	110000	6	0.01623	5.9452
Z	1001100100	10	0.00063	10.632

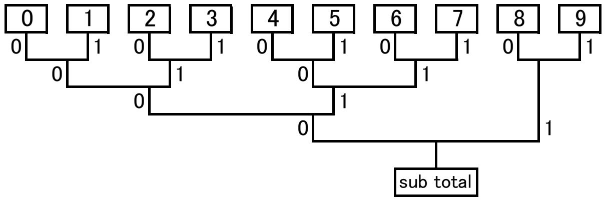
メモ

出現確率の最小となる文字はZ(p=0.0006)そして2番目に小さい文字はQ(p=0.001)であることがわかる。

数字0から9までの確率の和がこの範囲にあるとき，文字Zの確率を数字も含んだものとすると，
文字Zの元の符号1001100100（10ビット）の下に10011001001（11ビット）で表し，符号10011001000（11ビット）の下には，
数字0から9までを4ビット分拡張した符号に分岐させ組み入れることで数字を含む符号化テーブルへと拡張をすることもできる。

例えば、数字0を符号100110010000000（15ビット），．．．，数字7を符号100110010000111（15ビット），
数字8を符号1001100100010（13ビット），数字9を符号1001100100011（13ビット）へと符号化することもできる（下図参照）。

このようにして，多くのビット数（情報量）をより出現確率の低い記号に割り振ることにより符号化をするのがハフマン符号化である。



chr	code	bit	p	length
0	0000	4	1/16	0.25
1	0001	4	1/16	0.25
2	0010	4	1/16	0.25
3	0011	4	1/16	0.25
4	0100	4	1/16	0.25
5	0101	4	1/16	0.25
6	0110	4	1/16	0.25
7	0111	4	1/16	0.25
8	10	2	1/4	0.5
9	11	2	1/4	0.5

<−−−−−−−

平均ビット長　4.10504
エントロピ：　4.06507
（平均ビット長はエントロピよりも長くなる）

−−−−−−−>

平均ビット長　3
エントロピ：　3
（確率が2のマイナス冪乗なので平均ビット長とエントロピは一致する）