

CS 243 — Discrete Math and Functional Programming

Huffman encoding

Nov 14, 2008

From *Discrete Mathematics and Functional Programming* by Thomas VanDrunen; example and code originally adapted from *Structure and Interpretation of Computer Programs* by Abelson and Sussman.

Encoding

ASCII/Unicode (last four bits):

A	0001	F	0110	K	1011
B	0010	G	0111	L	1100
C	0011	H	1000	M	1101
D	0100	I	1001	N	1110
E	0101	J	1010	O	1111

Sample encoding:

0001	1110	1110	1001	1011	0001
A	N	N	I	K	A

Message size: $4 \times 6 = 24$ bits.

Encoding

Variable-length codes (frequent letters are shorter):

A	0	F	100	K	01
B	10	G	101	L	0000
C	001	H	110	M	11
D	010	I	00	N	1
E	011	J	111	O	0001

Sample encoding:

0	1	1	00	01	0
A	N	N	I	K	A

Message size: $1 + 1 + 1 + 2 + 2 + 1 = 8$ bits.

Encoding

A	0	F	100	K	01
B	10	G	101	L	0000
C	001	H	110	M	11
D	010	I	00	N	1
E	011	J	111	O	0001

0	1	1	00	01	0
A	N	N	I	K	A

Or did you mean

011	0001	0
E	O	A

Encoding

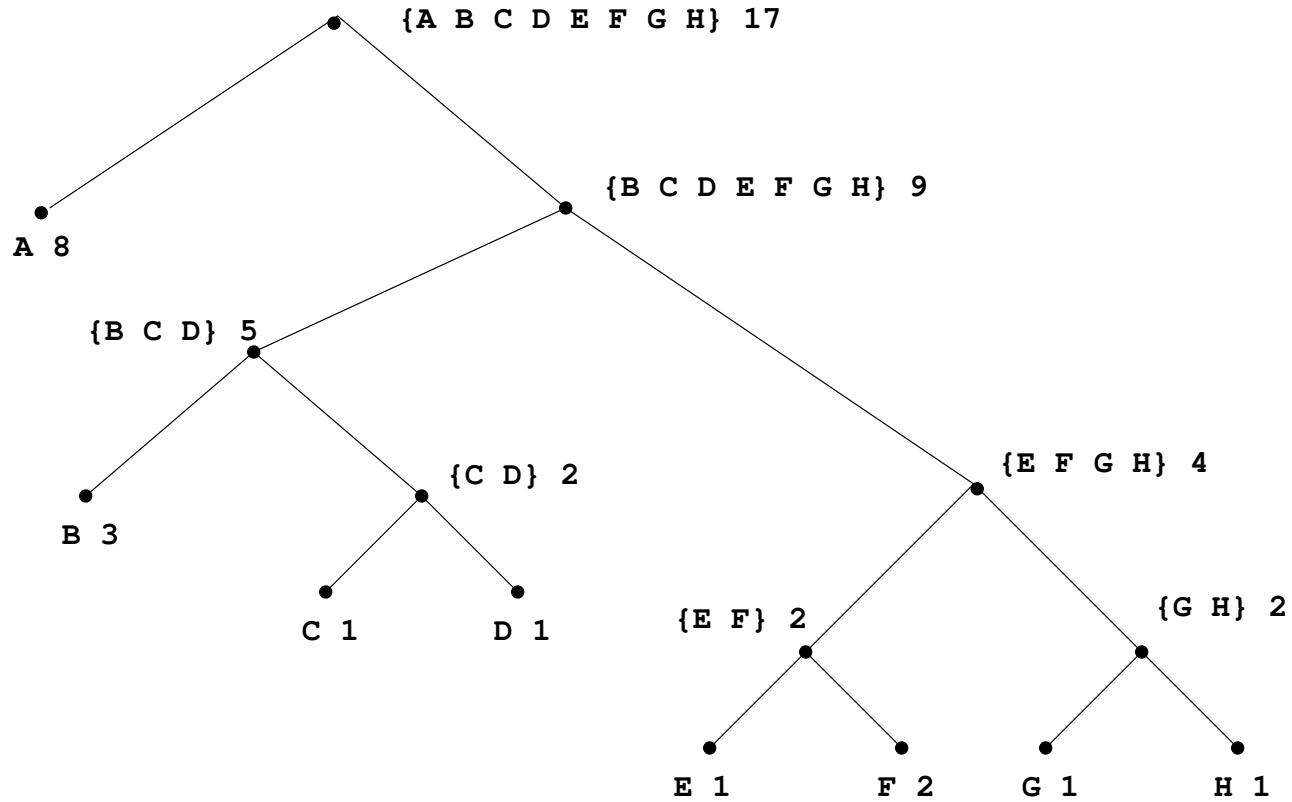
Prefix code

A	0	F	...	K	111
B	...	G	...	L	...
C	...	H	...	M	...
D	...	I	110	N	10
E	...	J	...	O	...

0	10	10	110	111	0
A	N	N	I	K	A

Message size: $1 + 2 + 2 + 3 + 3 + 1 = 12$ bits.

Trees



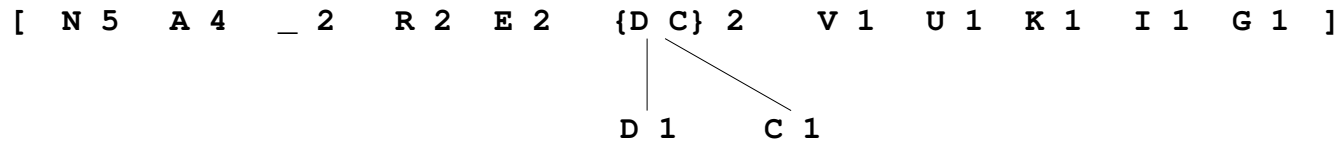
BACADAEAFABBAAGAH

A	0
B	100
C	1010
D	1011
E	1100
F	1101
G	1110
H	1111

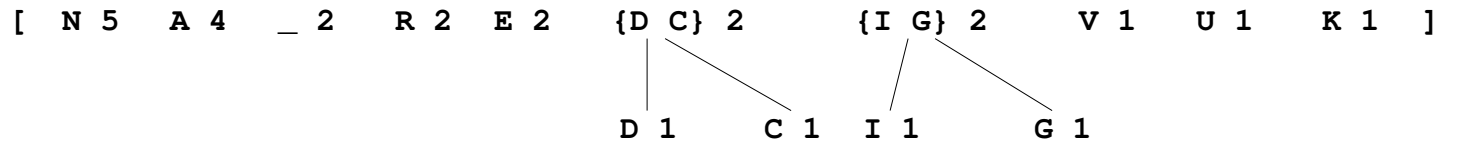
Building the tree

```
[ N 5   A 4   _ 2   R 2   E 2   V 1   U 1   K 1   I 1   G 1   D 1   C 1 ]
```

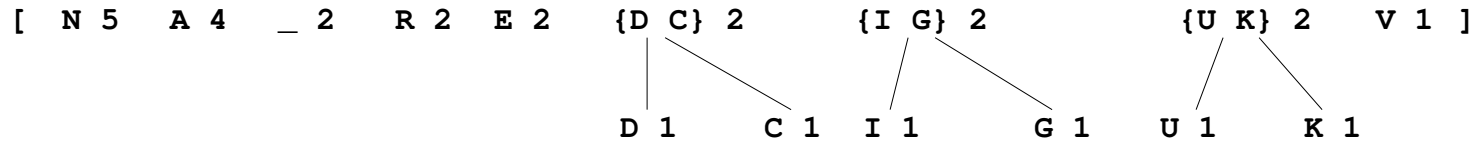
Building the tree



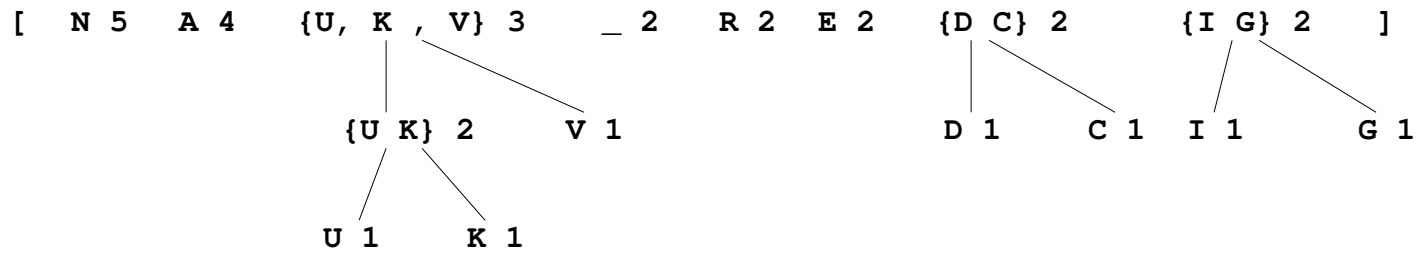
Building the tree



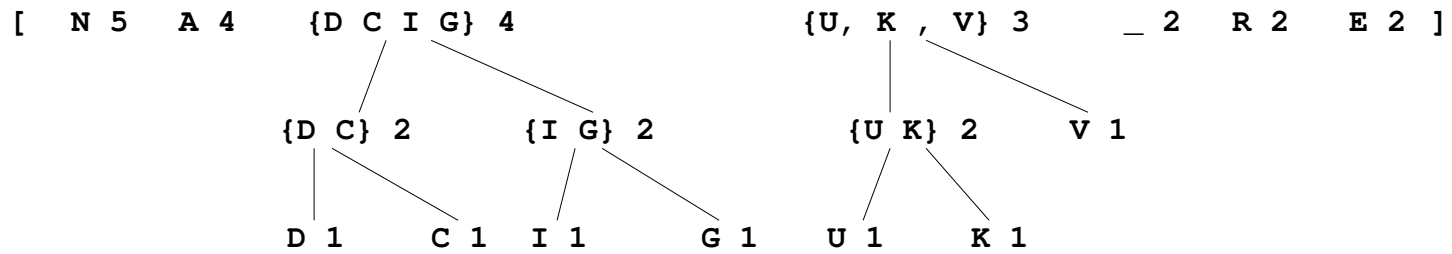
Building the tree



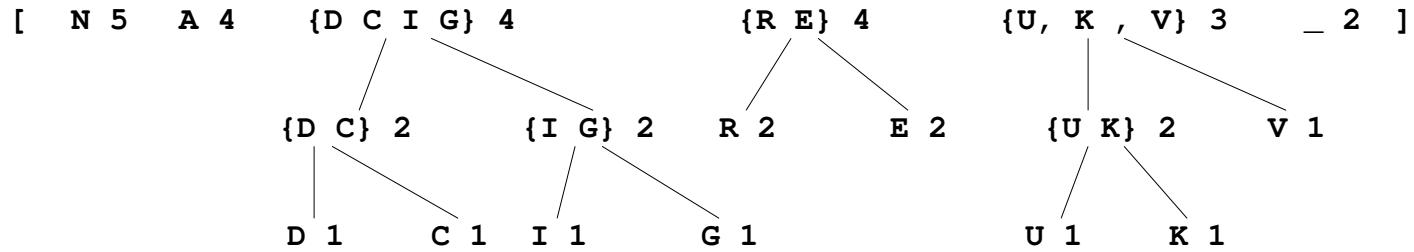
Building the tree



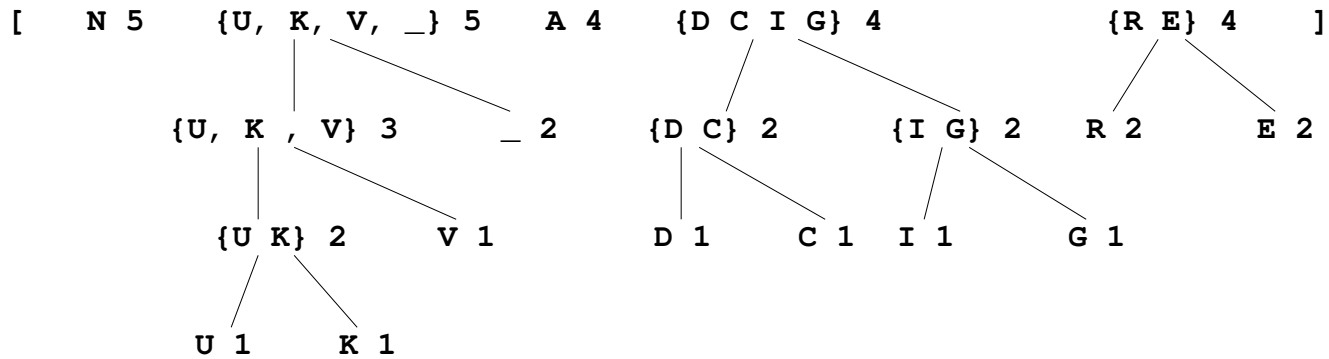
Building the tree



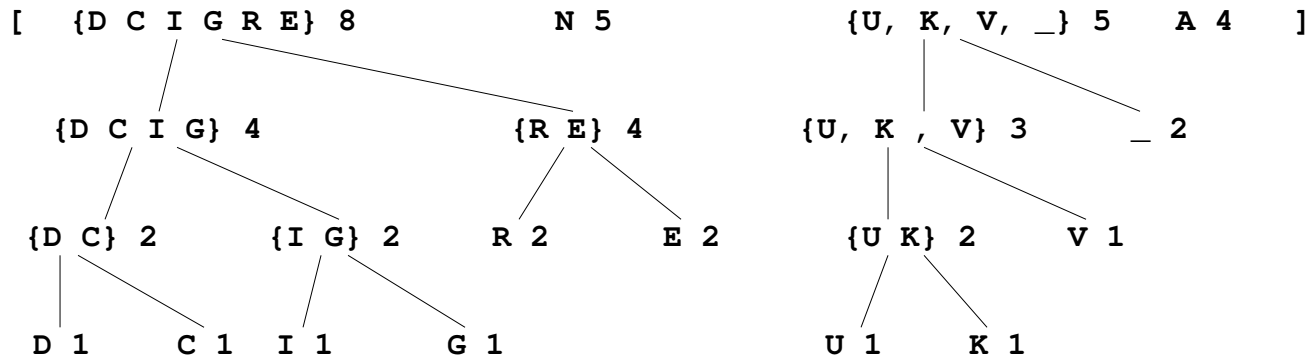
Building the tree



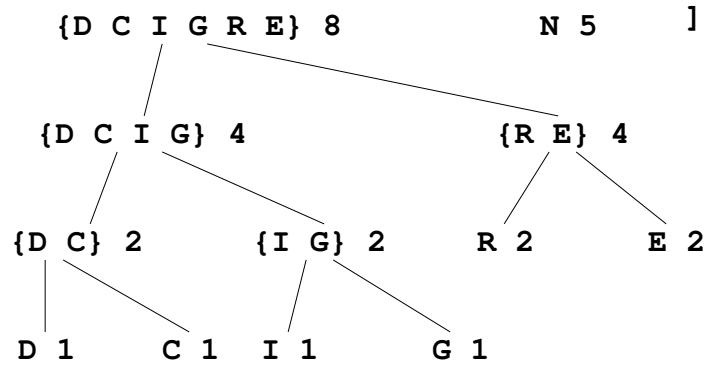
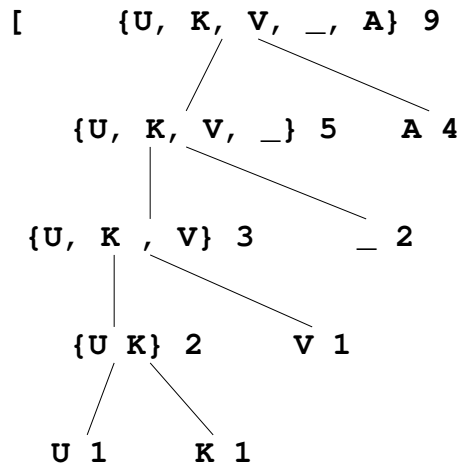
Building the tree



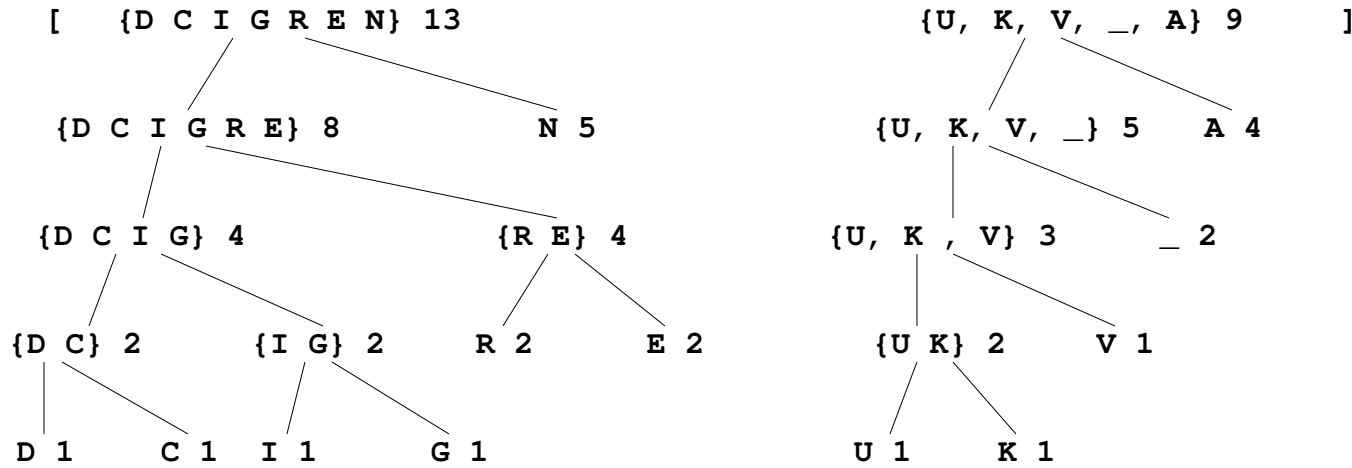
Building the tree



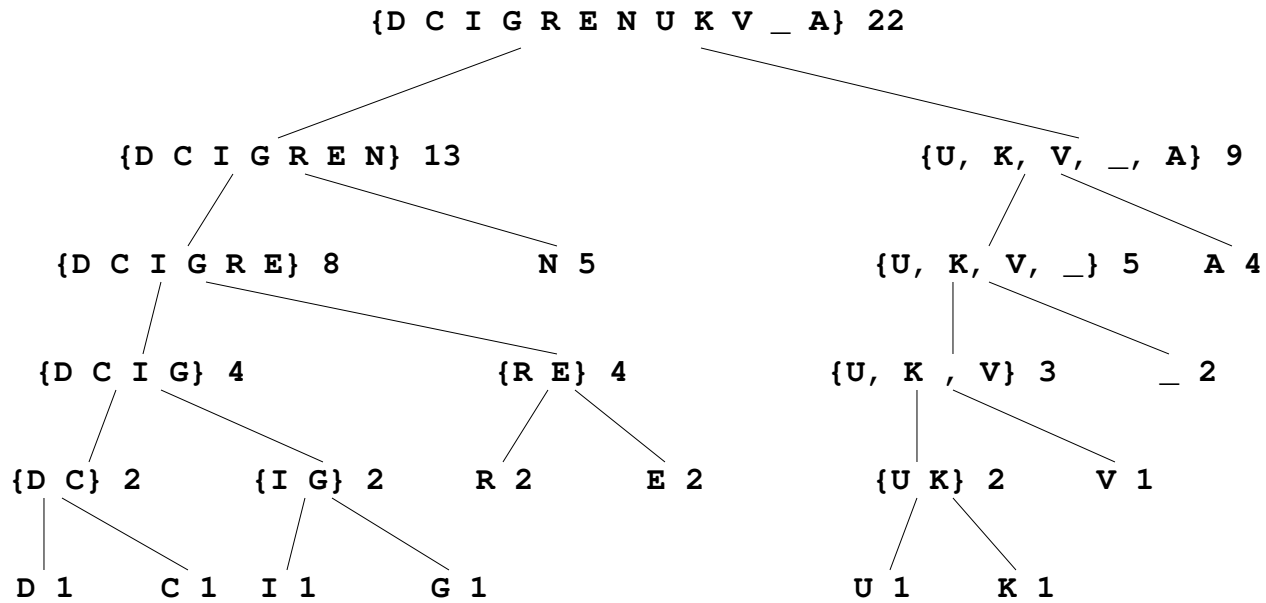
Building the tree



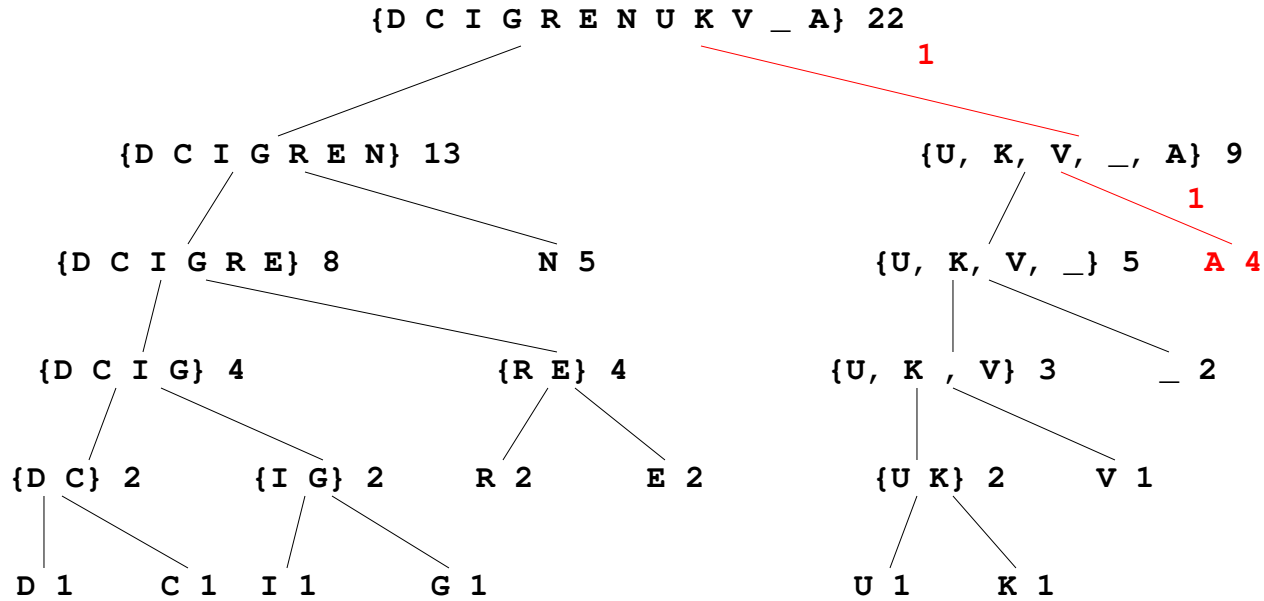
Building the tree



Building the tree

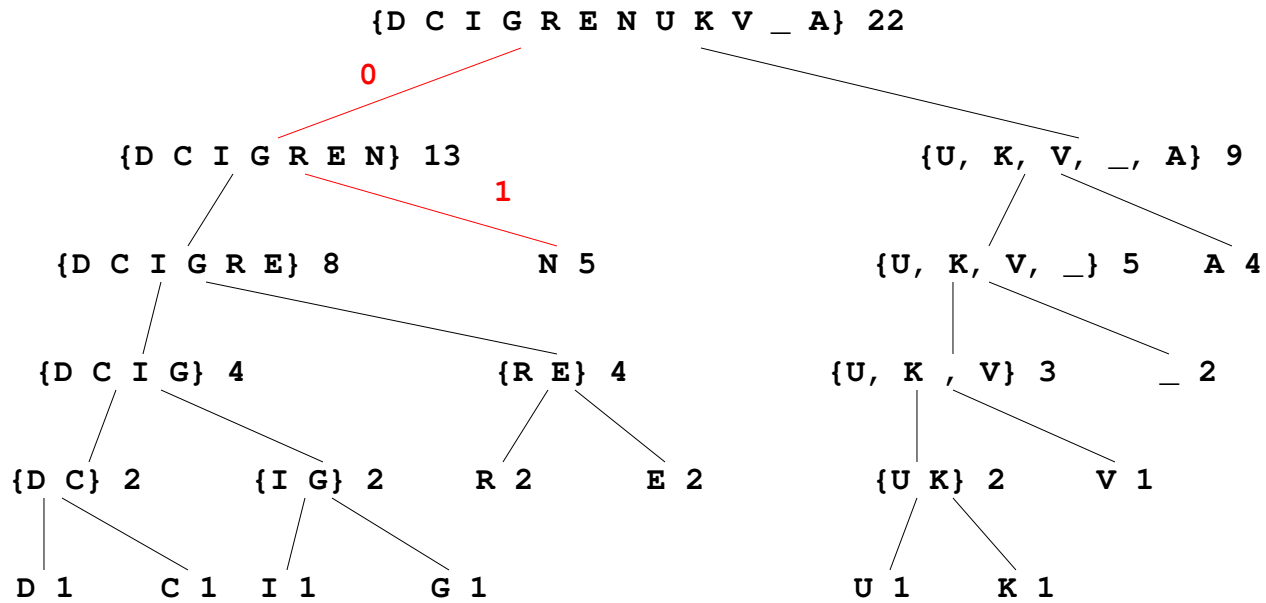


Encoding the message



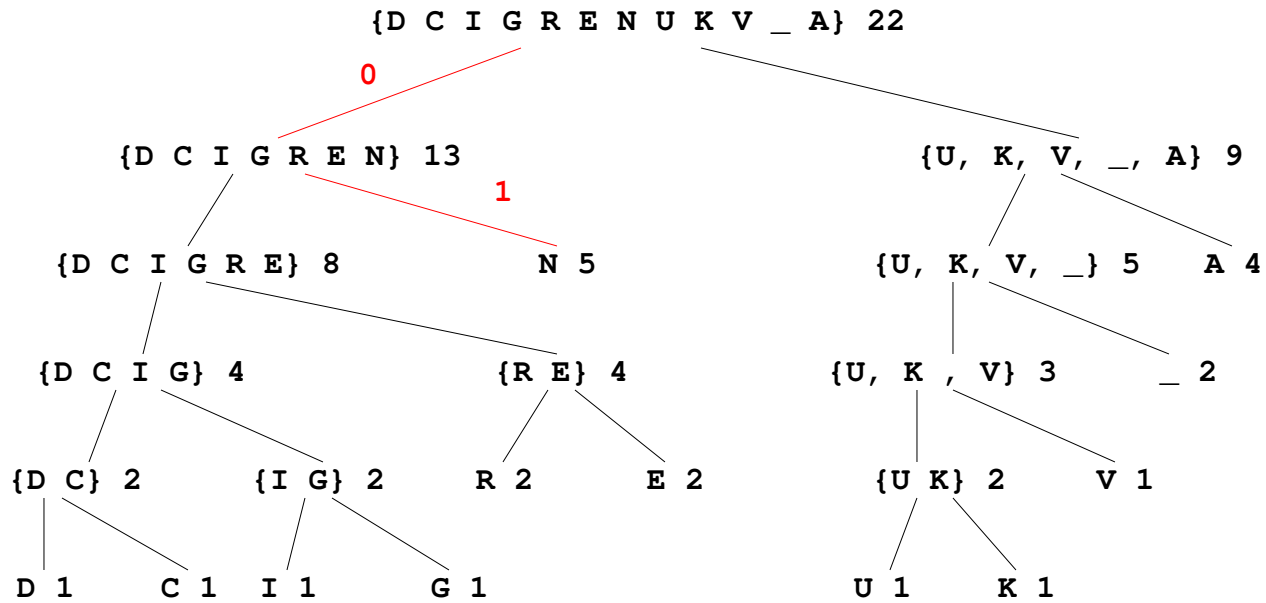
11
A

Encoding the message



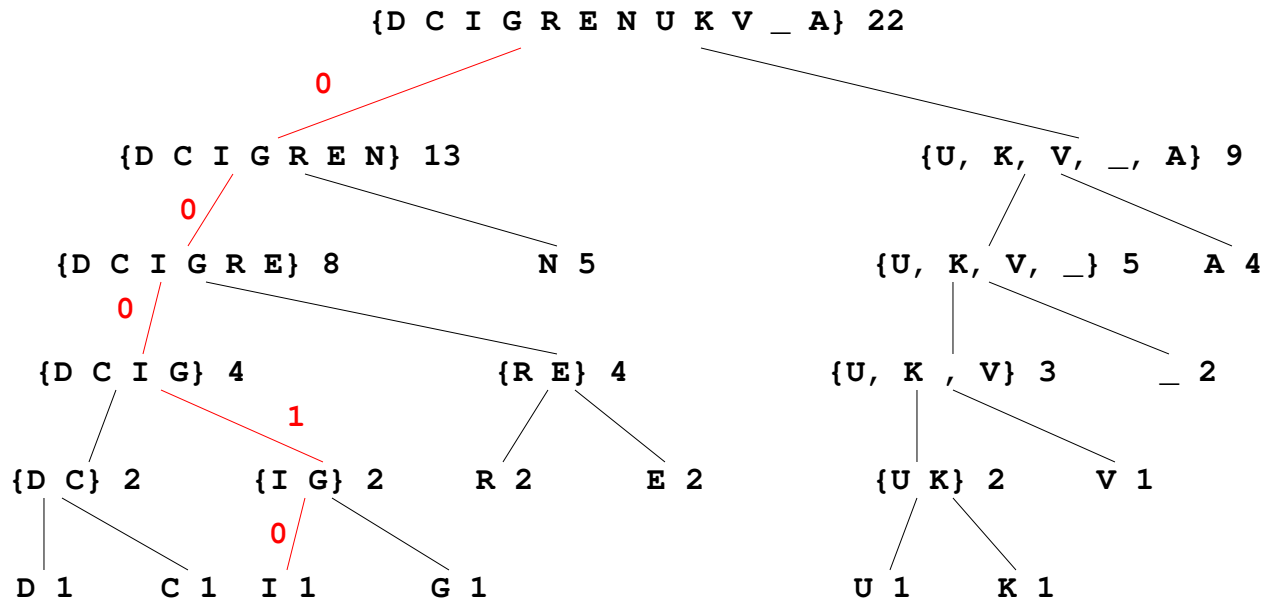
11	01
A	N

Encoding the message



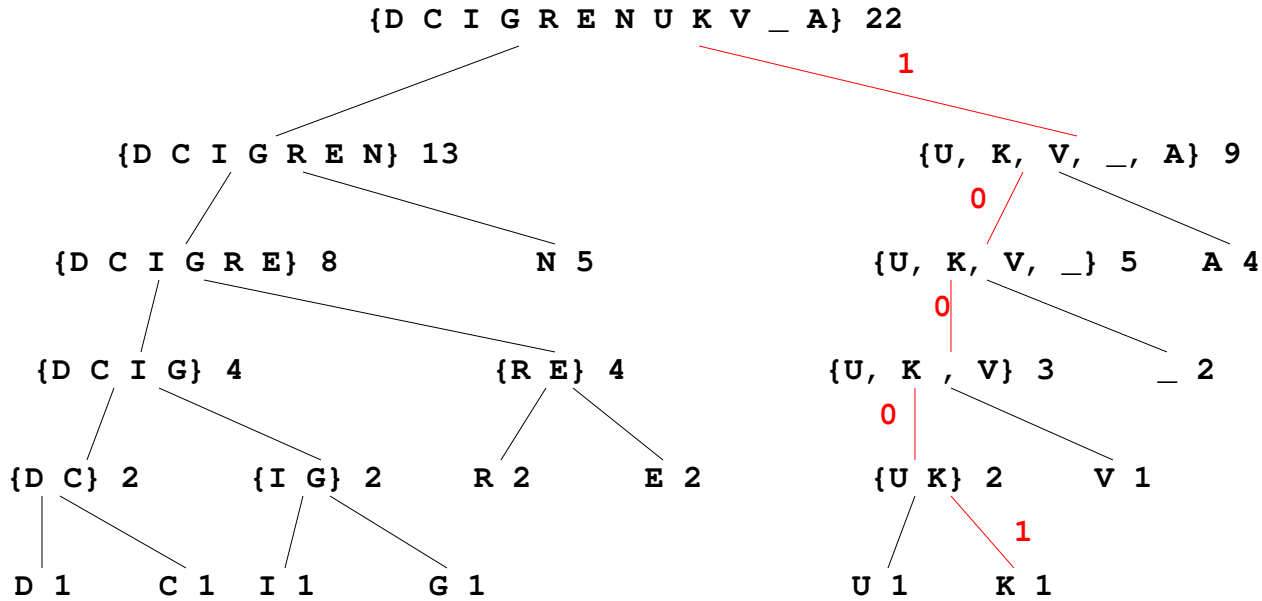
11	01	01
A	N	N

Encoding the message



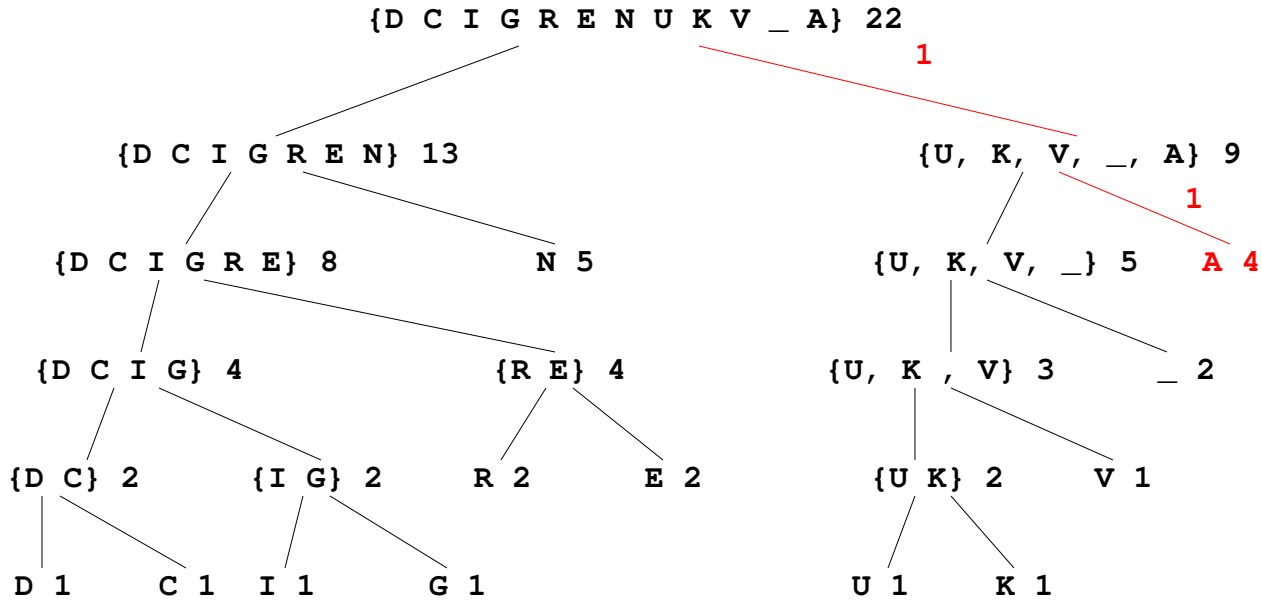
11	01	01	00010
A	N	N	I

Encoding the message



11	01	01	00010	10001
A	N	N	I	K

Encoding the message



11	01	01	00010	10001	11
A	N	N	I	K	A