Lydia Feng
CS323
Writ 4

1. String Sorting
    a. MSD sort on “now is the time for all good men to come to the aid of their party”

now	all	aid	aid	aid	aid	aid	aid
is	aid	all	all	all	all	all	all
the	come	come	come	come	come	come	come
time	for	for	for	for	for	for	for
for	good	good	good	good	good	good	good
all	is	is	is	is	is	is	is
good	now	now	now	now	now	now	now
people	of	of	of	of	of	of	of
to	people	people	party	party	party	party	party
come	party	party	people	people	people	people	people
to	the	the	the	the	the	the	the
the	time	time	time	the	the	the	the
aid	to	to	to	their	their	their	their
of	to	to	to	time	time	time	time
their	the	the	the	to	to	to	to
party	their	their	their	to	to	to	to

    b. Chars in red were not examined
        aid
        all
        come
        for
        good
        is
        now
        of
        party
        people
        the
        the
        their
        time
        to
        to

c. 3-way quicksort

Yellow is key partition

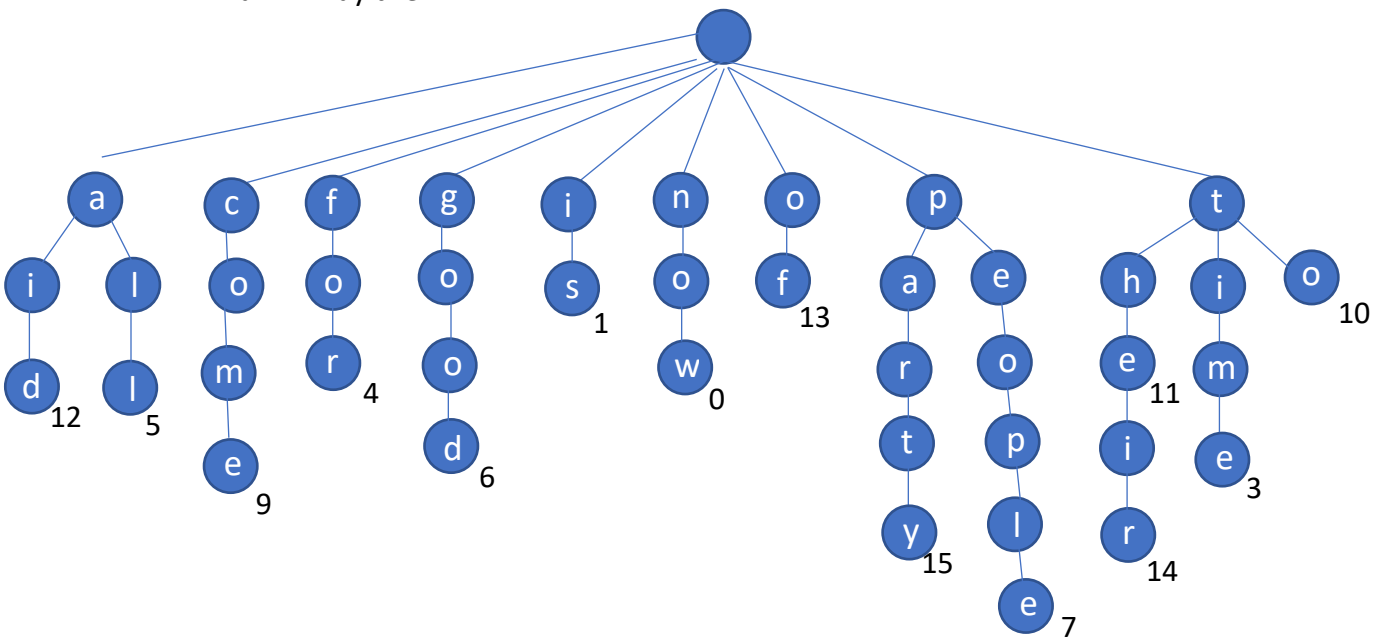
Pink is upper partition

Blue is lower partition

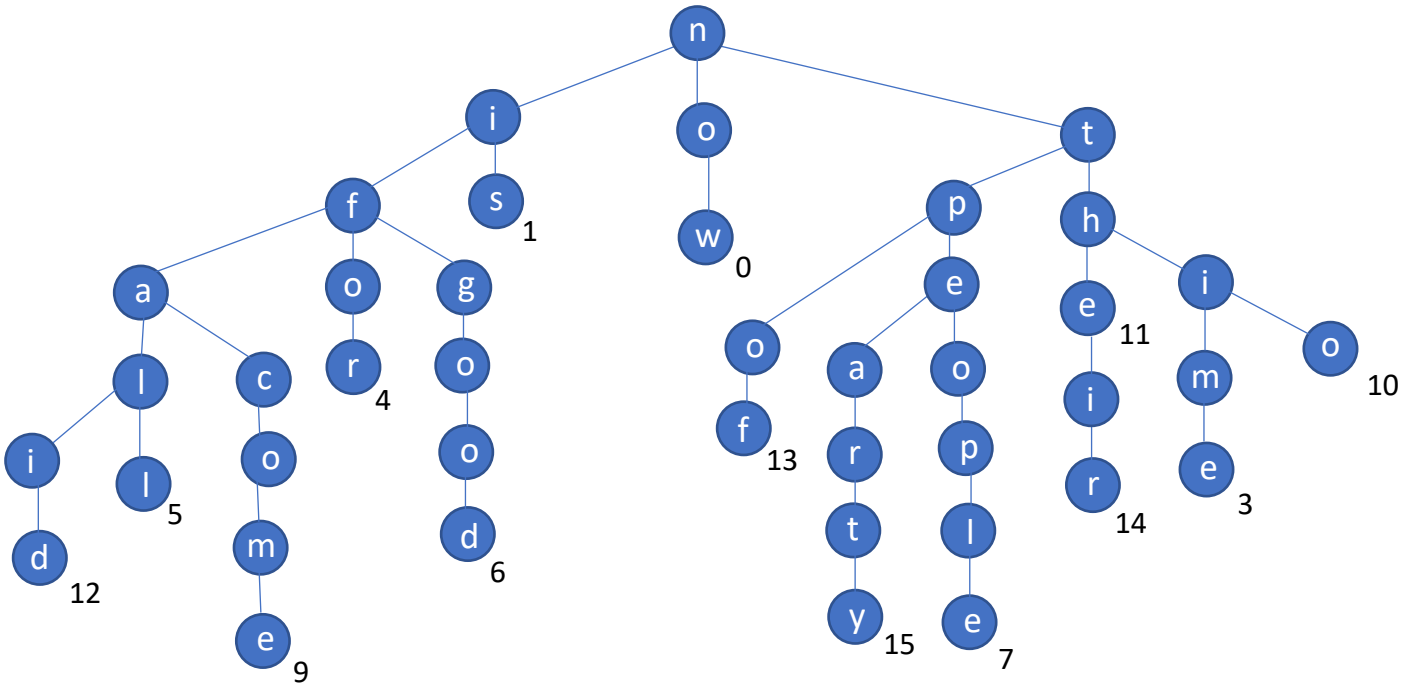
now	is	for	all	all	aid	aid
is	for	all	come	aid	all	all
the	all	good	aid	come	come	come
time	good	come	for	for	for	for
for	come	aid	good	good	good	good
all	aid	is	is	is	is	is
good	now	now	now	now	now	now
people	the	people	of	of	of	of
to	time	of	people	party	party	party
come	people	party	party	people	people	people
to	to	the	the	the	the	the
the	to	time	the	the	the	the
aid	the	to	their	their	their	their
of	of	to	time	time	time	time
their	their	the	to	to	to	to
party	party	their	to	to	to	to

2. Tries

a. R-way trie

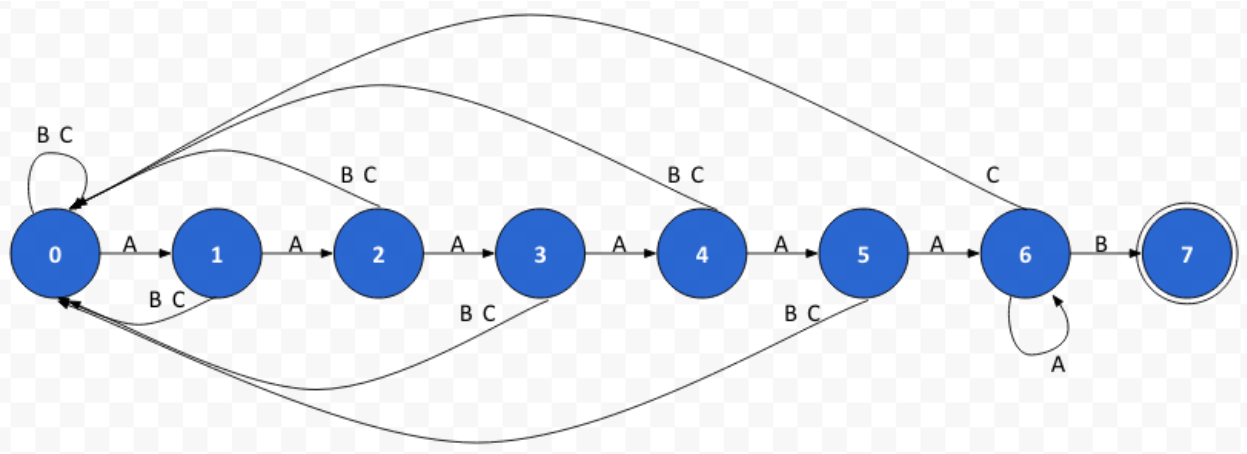


b. Ternary Search Tree

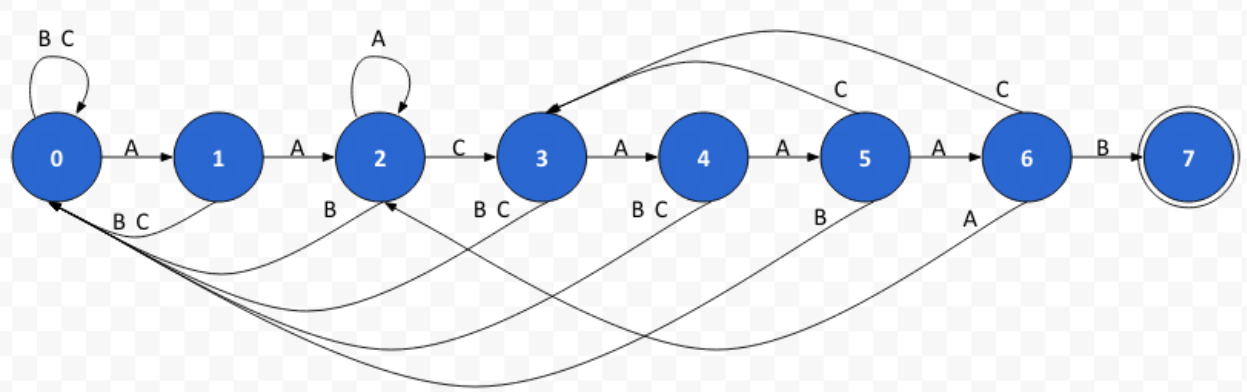


3. KMP DFA

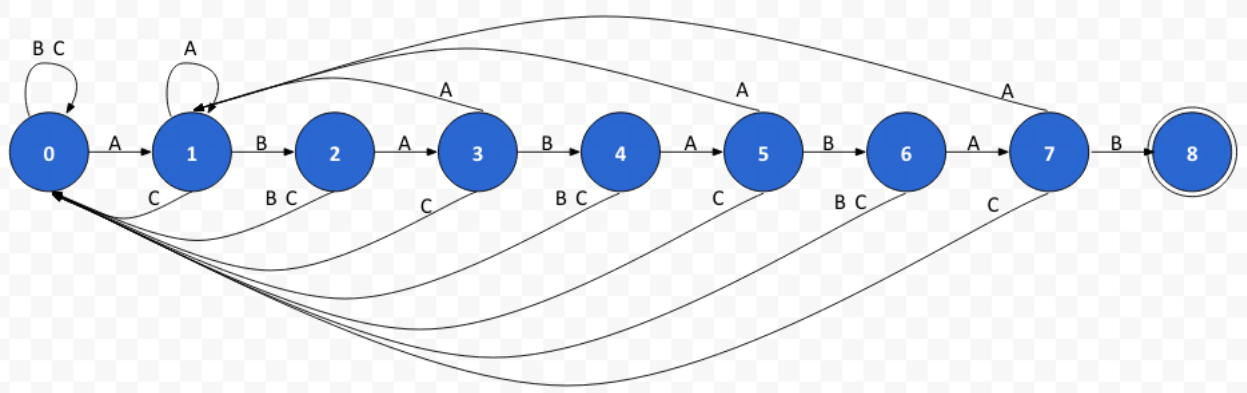
a. AAAAAB



b. AACAAAB

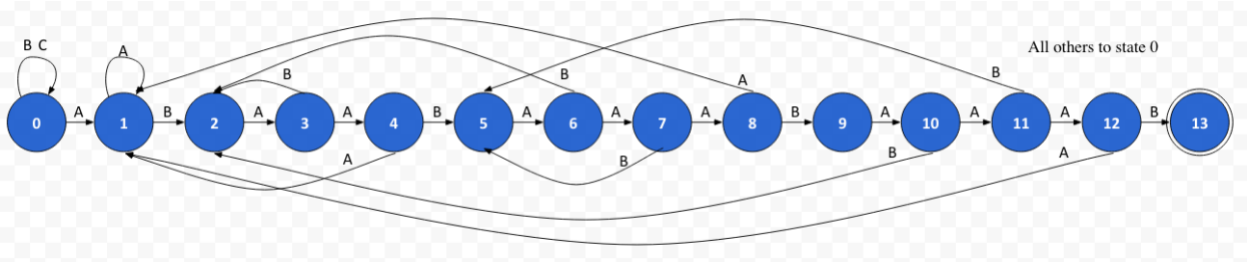


c. ABABABAB



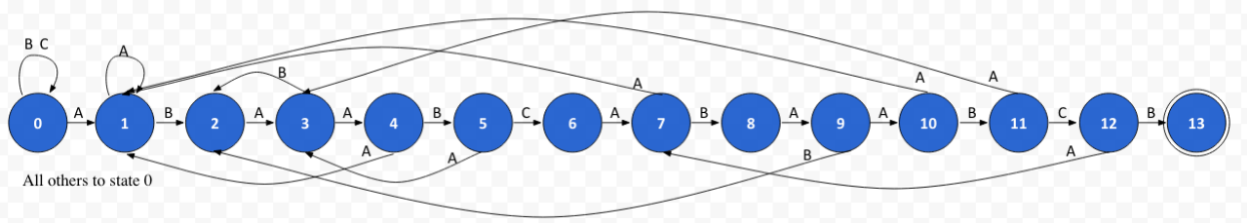
d. ABAABAAABAAAB

(all other arrows go to state 0)

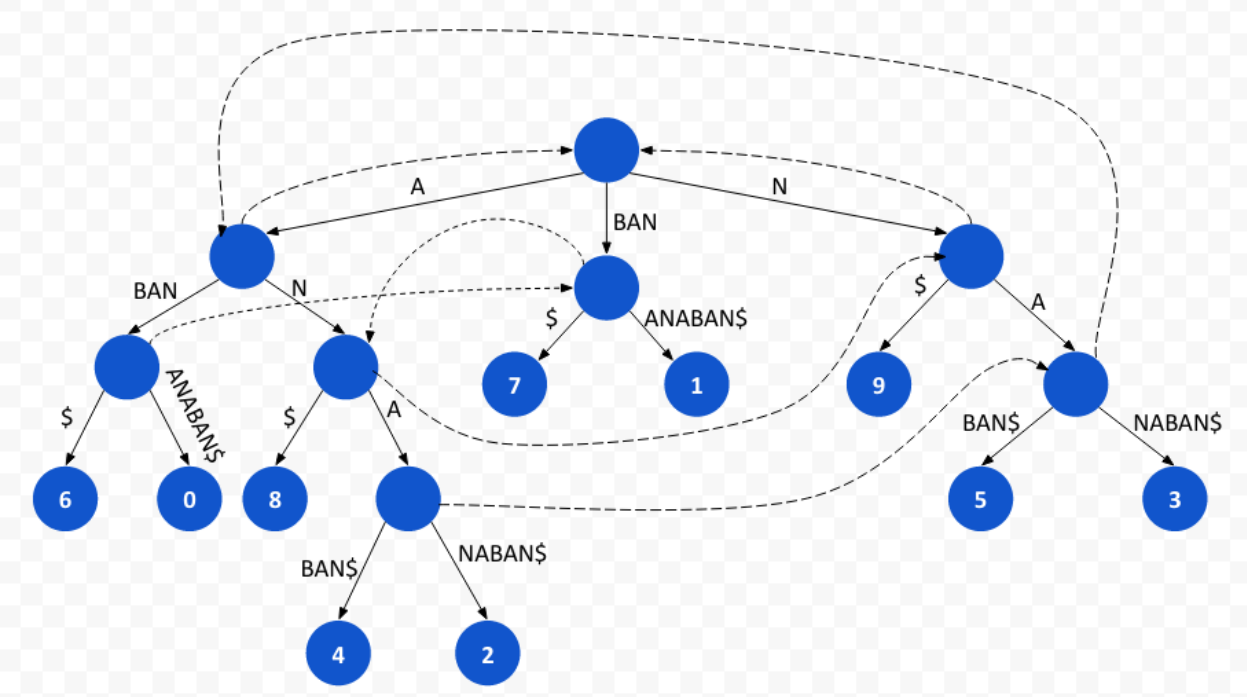


e. ABAABCABAABCB

(all other arrows go to state 0)



4. Suffix Tree



The longest repeated substring of ‘ABANANABAN’ is ‘ABAN’. This is because it is the internal node that is farthest from the root (has the most characters).

5. String Compression
   a. LZW

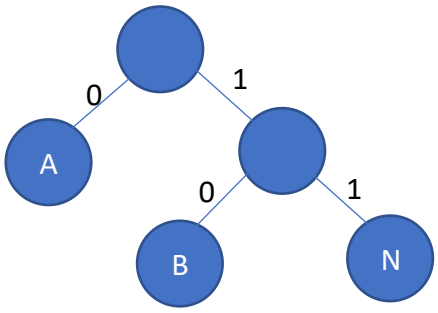
A	B	A	N	A	N	A	B	A	N	
41	42	41	4E	83		81		83		80

Codeword Table:

A	41
B	42
N	4E
AB	81
BA	82
AN	83
NA	84
ANA	85
ABA	86

   b. Huffman Trie

char	frequency	encoding
A	5	0
B	2	10
N	3	11



'ABANANABAN' = 010011011010011 = 15 bits (+1 bit for alignment)

   c. 0 1 01000001 0 1 01000010 1 01001110 = 29 bits (+3 bits for alignment)