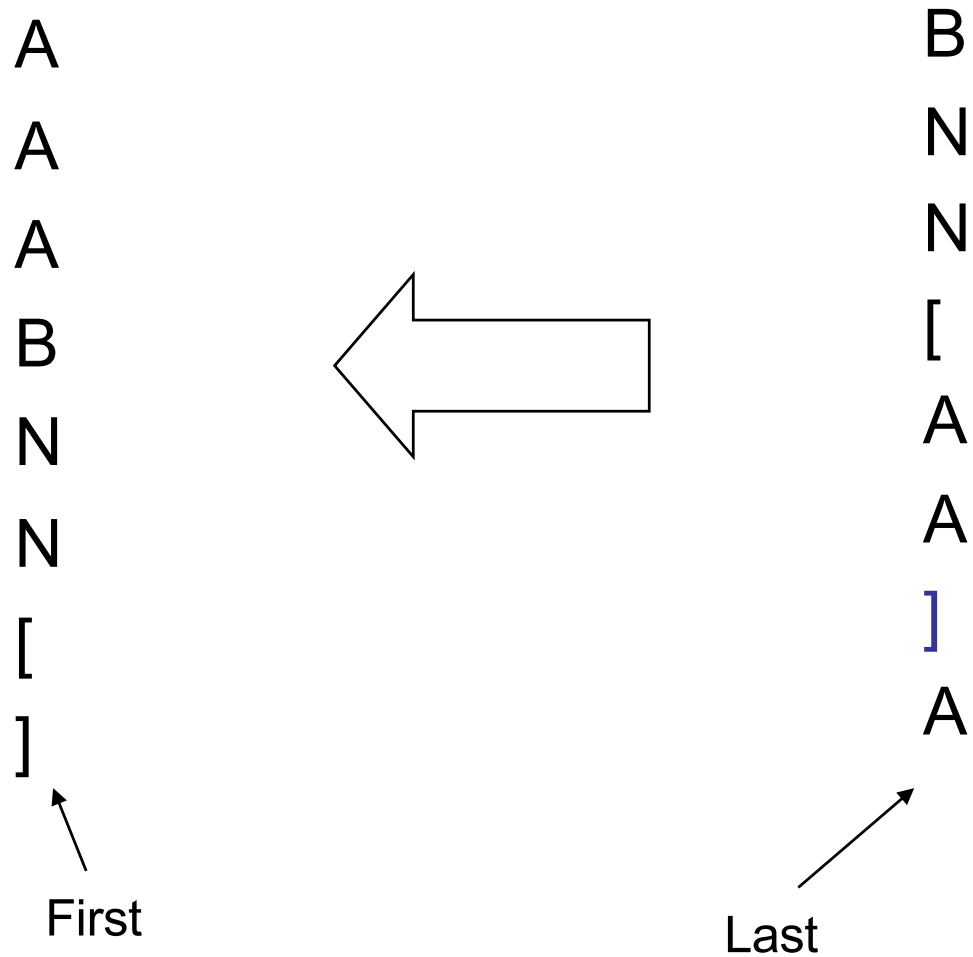

COMP9319 Web Data Compression and Search

BWT revisit

Backward Search overview

Recall: Last column = BWT



A]

A

B

A

N

A

N

B

[

N

A

N

A

[

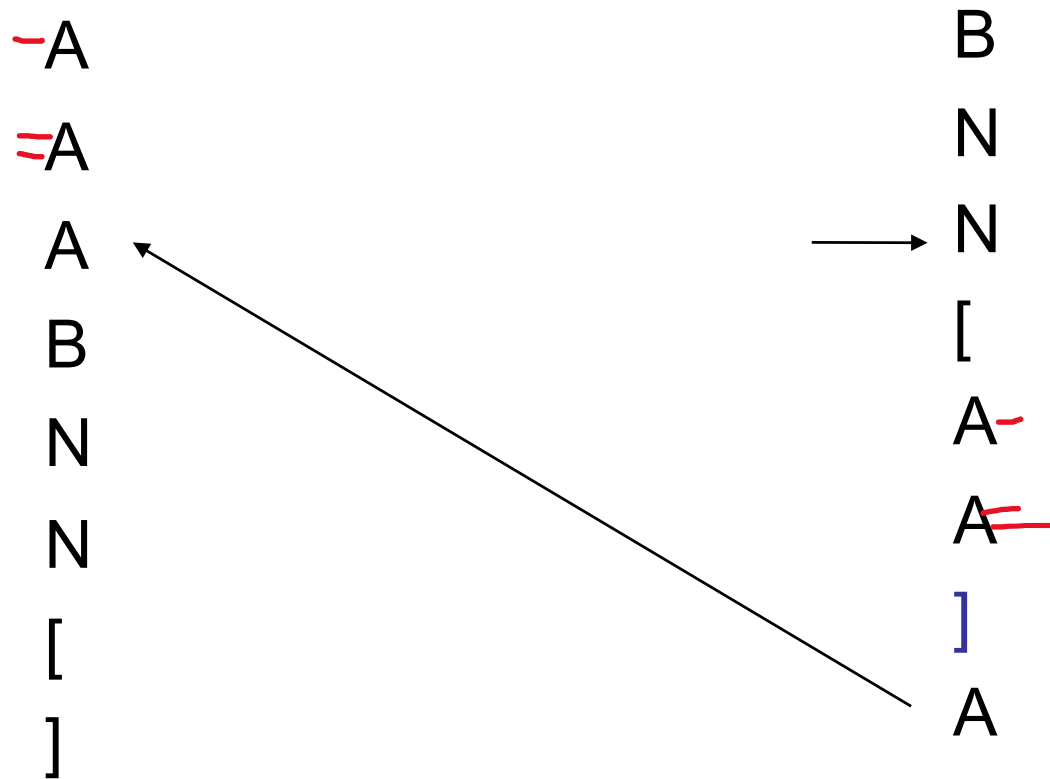
]

]

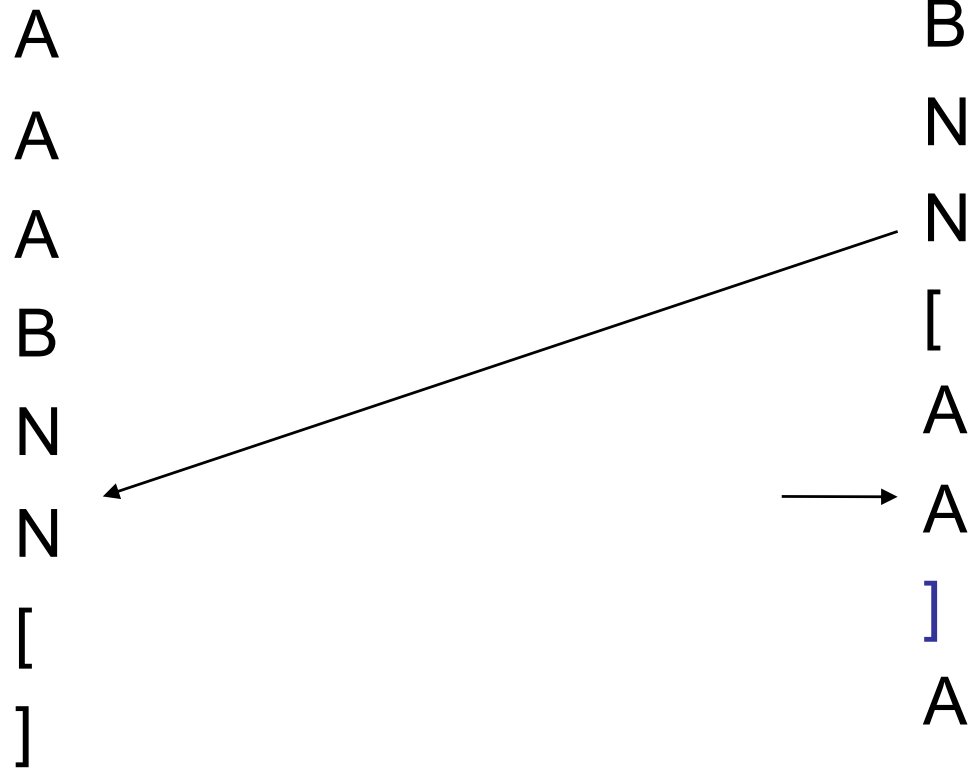
A



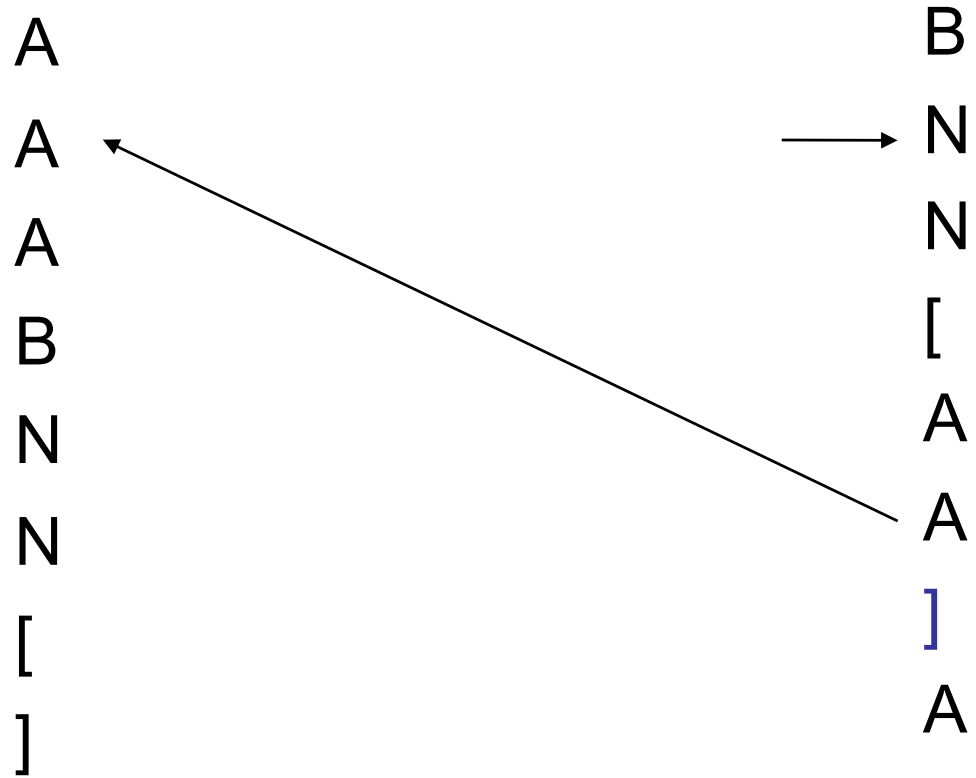
NA]



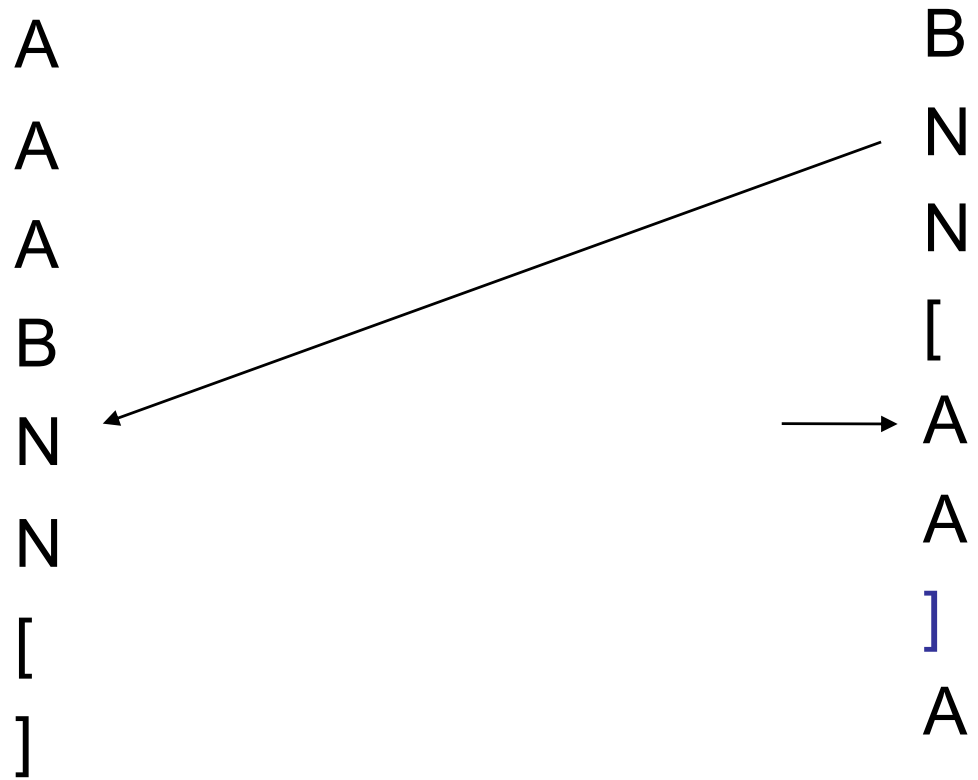
ANA]



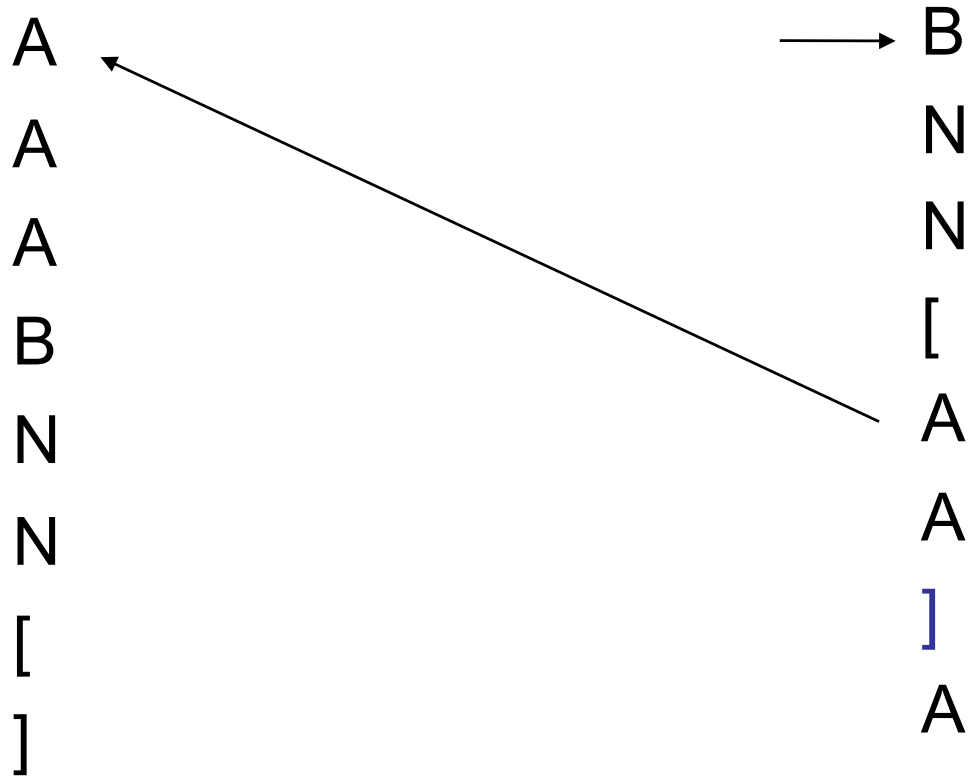
NANA]



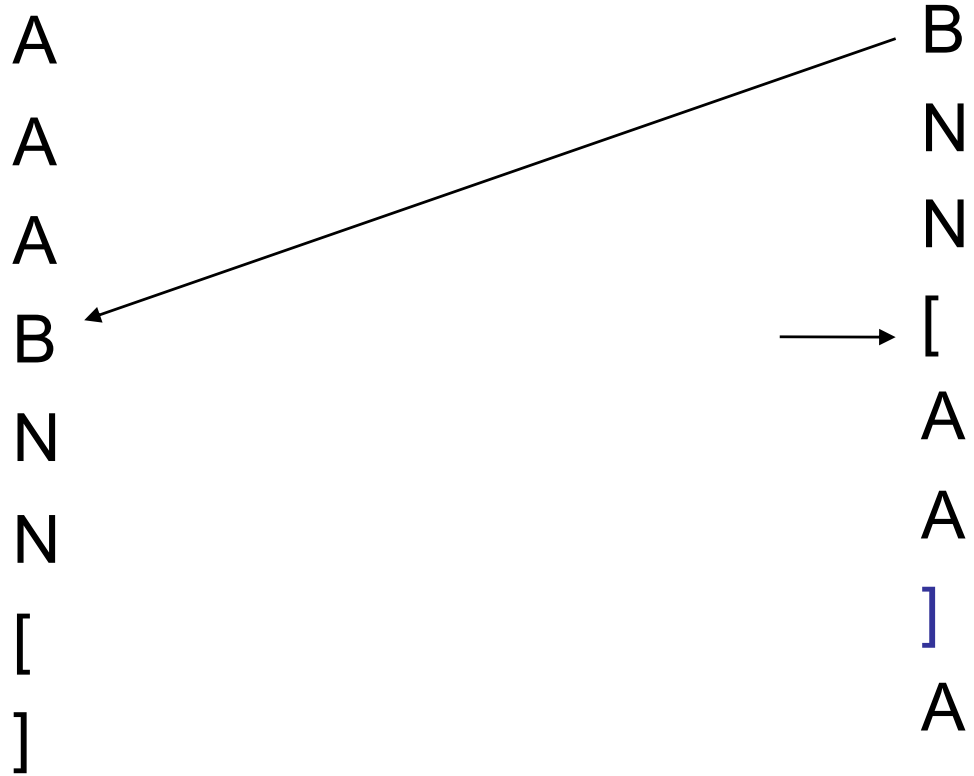
ANANA]



BANANA]



[BANANA]



Example using C[] & Occ[]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

?????]?]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

??????**A**]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

?????NA]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

?????A]NA]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

???NANA]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

??**A**NANA]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

?BANANA]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

[BANANA]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

[BANANA]

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Occ / Rank

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

C[]

C[] & Occ()

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Occ(Symbol, Pos)
=> # Matching

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

C[Symbol] =>
(startPos, endPos)

C[] & Occ()

C[Symbol] =>
(startPos, endPos)

Occ(Symbol, Pos)
=> # Matching

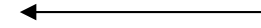
Can these two functions (or tables) be implemented such that they can return the result in constant time ?

Yes, have a precomputed table.

Can they be precomputed efficiently ?

Yes, a single pass.

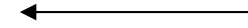
Backward Search for ANA



A
A
A
B
N
N
[
]

B
N
N
[
A
A
]
A

Backward Search for NAN



A
A
A
B
N
N
[
]

B
N
N
[
A
A
]
A

Backward Search for ANA

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Occ(Symbol, Pos)
=> # Matching

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

C[Symbol] =>
(startPos, endPos)

Backward Search for NAN

Position	Symbol	# Matching
0	B	0
1	N	0
2	N	1
3	[0
4	A	0
5	A	1
6]	0
7	A	2

Occ(Symbol, Pos)
=> # Matching

Symbol	# LessThan
A	0
B	3
N	4
[6
]	7

C[Symbol] =>
(startPos, endPos)

Why not Forward Search: ANA →

A

A

A

B

N

N

[

]

B

N

N

[

A

A

]

A

Assignment 2 overview & tips to start...

