CSE 566 Spring 2023

Analysis of Skew Algorithm & Burrows-Wheeler Transform

Instructor: Mingfu Shao

Analysis of Skew Algorithm

- Let T(n) be the running time of skew(S) when |S| = n
- Recurrence: T(n) = T(2n/3) + O(n)
- T(n) = O(n), using master's theorem

$T(n) = T(a \cdot n) + O(n)$

$\Rightarrow \quad T(n) = O(n), \quad \text{if } a < 1$

$T(n) = T(n-1) + O(n) \quad \Rightarrow \quad T(n) = O(n^2)$.

Burrows-Wheeler Transform (BWT)

S = banana$ $|S| = n$, $\$ < a < b < n$

anzb$a2

banana$		$banana	a
anana$b		a$banan	n
nana$ba	**sort** →	ana$ban	n
ana$ban		anana$b	b
na$bana		banana$	$
a$banan		na$bana	a
$banana		nana$ba	a

All cyclic rotations

F L

M(S) = BWT(S)

b2ip2

Constructing BWT

- Trivial algorithm: $O(n^2 \log n)$, by definition
- In fact, can be done in linear time, using suffix array!

S = banana$

banana$	$banana
anana$b	a$banan
nana$ba	ana$ban
ana$ban	anana$b
na$bana	banana$
a$banan	na$bana
$banana	nana$ba

M(S)

- The substring of a cyclic rotation before $ (including $) is the suffix of S!
- **Key**: sorting all suffixes is equivalent to sorting all cyclic rotations!

if there is a unique letter in the middle ?

From Suffix Array To BWT

S = banana$

		SA(S)		BWT[k] = S[SA[k] - 1]
1	banana$	7	$banana	a
2	anana$b	6	a$banan	n
3	nana$ba	4	ana$ban	n
4	ana$ban	2	anana$b	b
5	na$bana	1	banana$	$
6	a$banan	5	na$bana	a
7	$banana	3	nana$ba	a

F L

Li is the predecessor of Fi in S

Recovering S from BWT

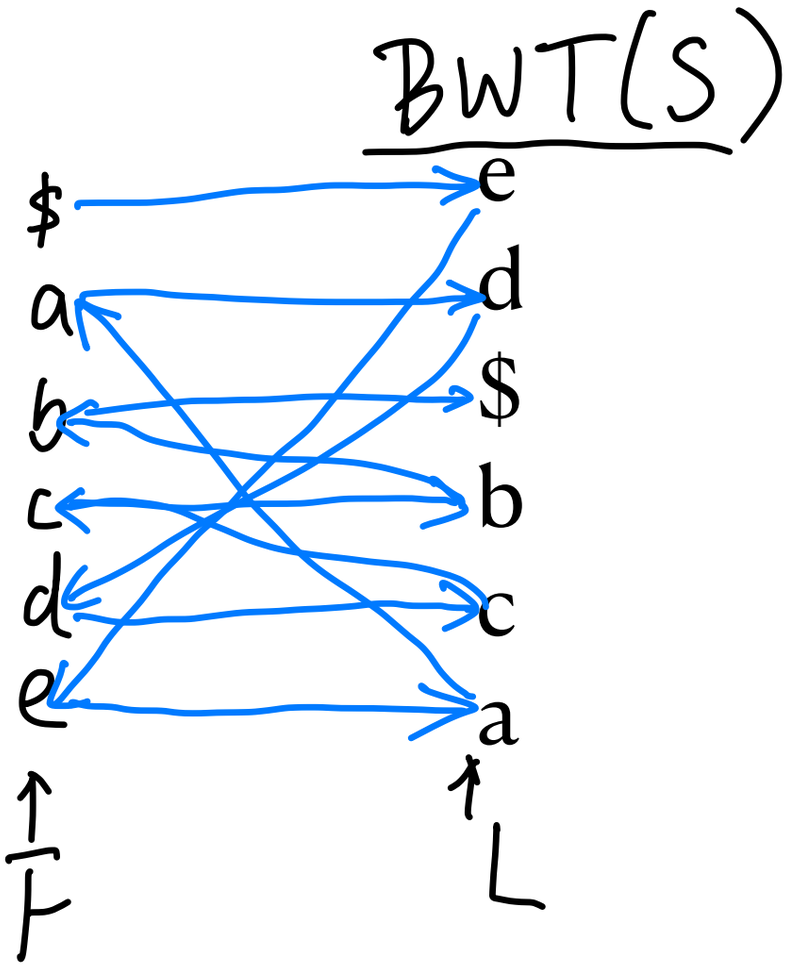
- **An easy case**: all letters in S are distinct

S = bcdae$

M(S)

$bcdae
ae$bcd
bcdae$
cdae$b
dae$bc
e$bcda

L

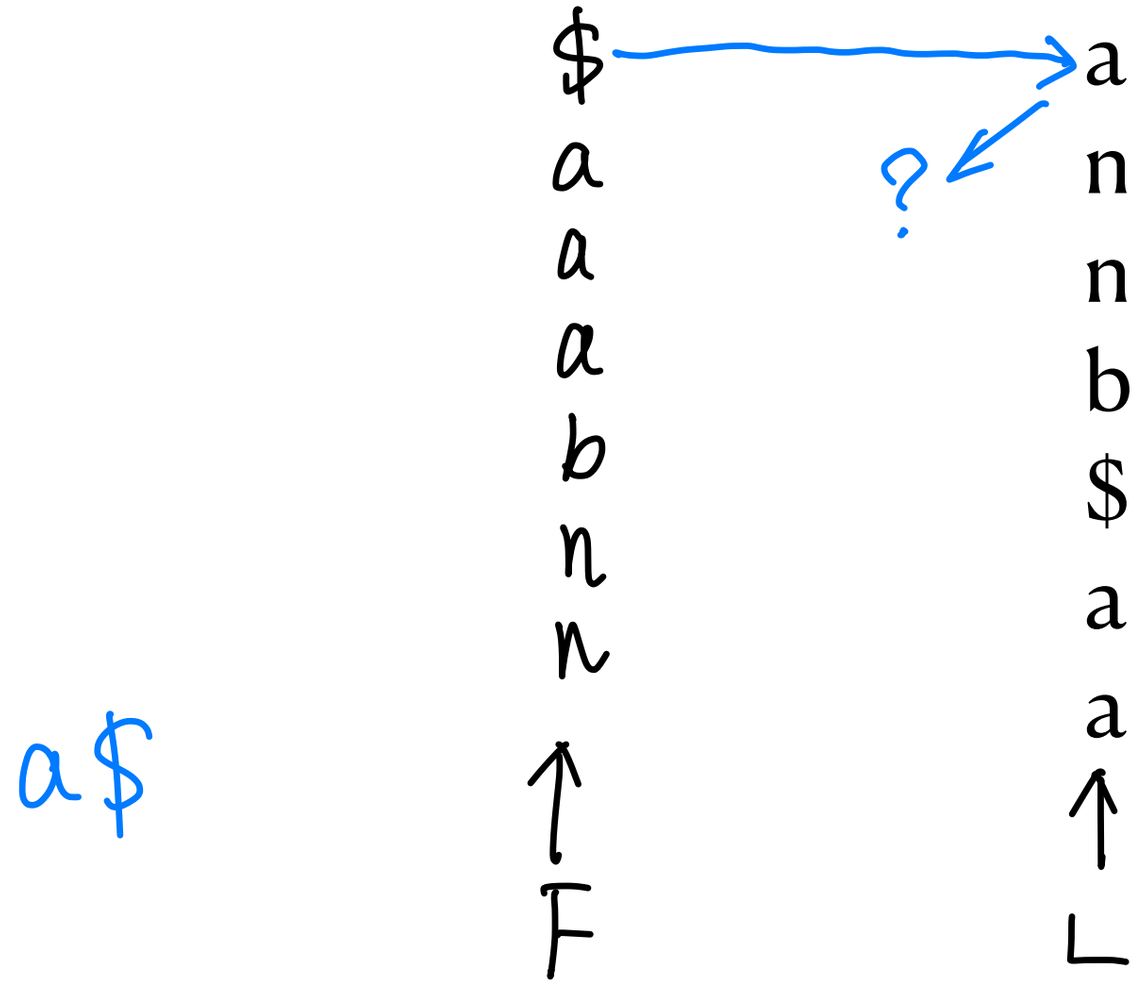


S = bcdae$

But..

S = banana$

BWT(S)



LF Mapping

- **LF Mapping Property**: For every letter c, the j-th occurrence of c in L and the j-th occurrence of c in F correspond to the same letter in S.

S = banana$

F	L
$banana	a
a$banan	n
ana$ban	a
anana$b	$
banana$	b
na$bana	a
nana$ba	a

Proof of LF Mapping Property

S = b$_1$a$_2$n$_3$a$_4$n$_5$a$_6$\$

co(a)=nana\$b

co(a)=na\$ban

co(a)=\$banan

MCS)

\$banana

a\$banan

ana\$ban

anana\$b

banana\$

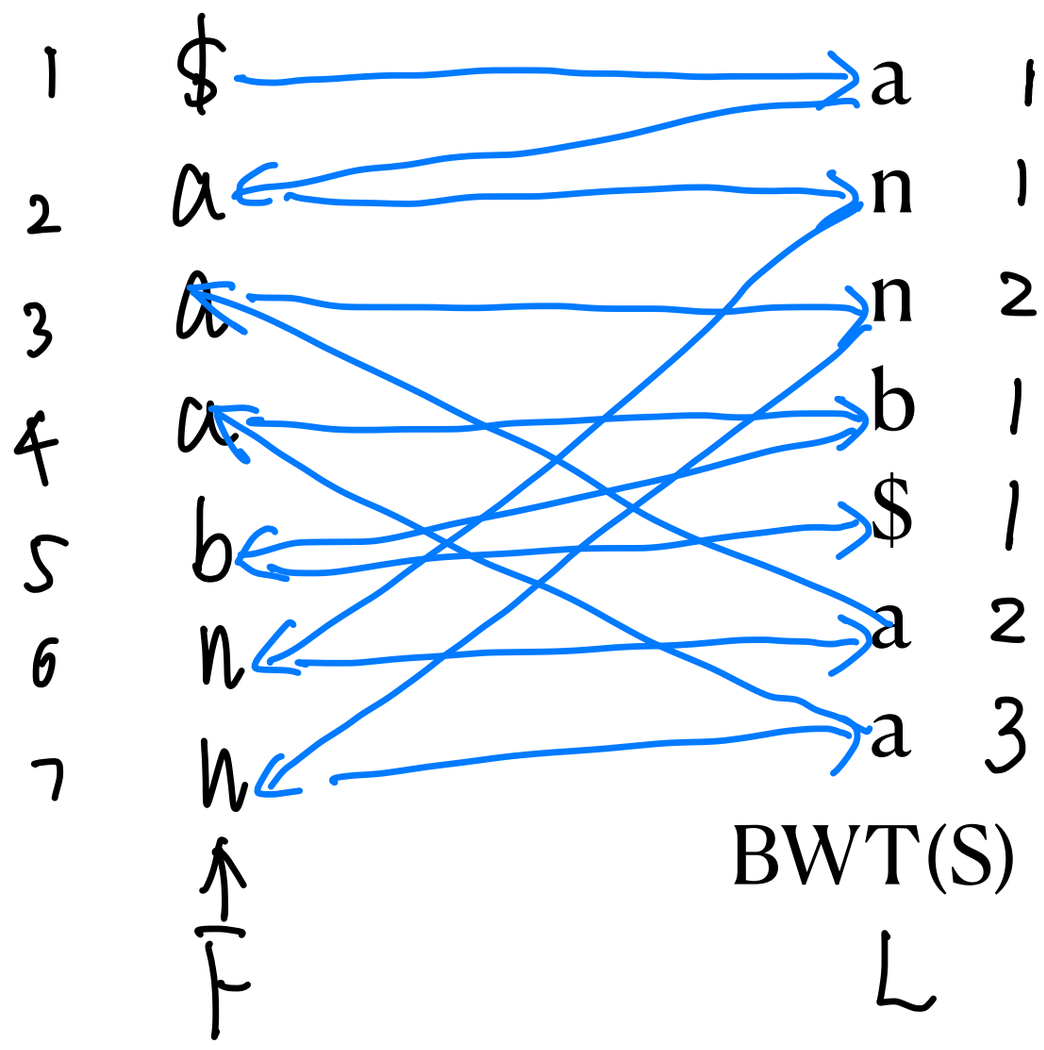
na\$bana

nana\$ba

- **Key**: they are sorted by the same thing!

Algorithm to Recover S from BWT

S = banana$



	$	a	b	n
first	1	2	5	6