Dynamic Burrows-Wheeler Transform

Prague Stringology Conference 2008

Mikaël Salson, Thierry Lecroq, Martine Léonard, Laurent Mouchard

LITIS and University of Rouen, France

1 September 2008

Burrows-Wheeler Transform (1994)

What is it?

- Permutation of a text, that allows better compression.
- Closeness to a widely-used index (suffix array).
- Recent interest in compressed indexing.

Question

- What happens to the transform if the text changes?

Notations

Cyclic shifts

A cyclic shift of a text $T[0\,..\,n]$, of order i is denoted by $T^{[i]} = T[i\,..\,n-1]T[0\,..\,i]$.
The previous cyclic shift of $T^{[i]}$ is $T^{[i-1]}$.

From $T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

From $T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓			sorted $T^{[i]}$			L ↓	
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L: Burrows-Wheeler Transform of T [C G $ T T C C]

From $T = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{C}}\,\overset{3}{\mathrm{T}}\,\overset{4}{\mathrm{G}}\,\overset{5}{\mathrm{C}}\,\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓	sorted $T^{[i]}$					*L* ↓	
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L: Burrows-Wheeler Transform of *T* [C G $ T T C C]

From $T = \overset{0}{C}\overset{1}{T}\overset{2}{C}\overset{3}{T}\overset{4}{G}\overset{5}{C}\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓	sorted $T^{[i]}$					L ↓	SA ↓
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L: Burrows-Wheeler Transform of T [C G $ T T C C]
SA: Suffix Array of T [6 5 0 2 4 1 3]

From $T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓	sorted $T^{[i]}$					L ↓	SA ↓
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L:	Burrows-Wheeler Transform of *T*	[C G $ T T C C]
SA:	Suffix Array of *T*	[6 5 0 2 4 1 3]
ISA:	Inverse Suffix Array of *T*	[2 5 3 6 4 1 0]

$$L[i] = T[(SA[i] - 1) \bmod |T|]$$

From $T = \overset{0}{C}\overset{1}{T}\overset{2}{C}\overset{3}{T}\overset{4}{G}\overset{5}{C}\overset{6}{\$}$ to BWT, SA and ISA

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓	sorted $T^{[i]}$					L ↓	SA ↓
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L:	Burrows-Wheeler Transform of T	[C G $ T T C C]
SA:	Suffix Array of T	[6 5 0 2 4 1 3]
ISA:	Inverse Suffix Array of T	[2 5 3 6 4 1 0]

$$L[i] = T[(SA[i] - 1) \bmod |T|]$$

From $T = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{C}}\,\overset{3}{\mathrm{T}}\,\overset{4}{\mathrm{G}}\,\overset{5}{\mathrm{C}}\,\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓	sorted $T^{[i]}$					L ↓	SA ↓
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L: Burrows-Wheeler Transform of *T* [C G $ T T C C]

SA: Suffix Array of *T* [6 5 0 2 4 1 3]

ISA: Inverse Suffix Array of *T* [2 5 3 6 4 1 0]

$$L[i] = T[(SA[i] - 1) \bmod |T|]$$

From $T = \overset{0}{C}\overset{1}{T}\overset{2}{C}\overset{3}{T}\overset{4}{G}\overset{5}{C}\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓	sorted $T^{[i]}$					L ↓	SA ↓
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L: Burrows-Wheeler Transform of *T* [C G $ T T C C]
SA: Suffix Array of *T* [6 5 0 2 4 1 3]
ISA: Inverse Suffix Array of *T* [2 5 3 6 4 1 0]

$$L[i] = T[(SA[i] - 1) \bmod |T|]$$

From $T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$}$ to *BWT*, *SA* and *ISA*

Burrows-Wheeler Transform and Suffix Array

	unsorted $T^{[i]}$						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	F ↓	sorted $T^{[i]}$					L ↓	SA ↓
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

L:	Burrows-Wheeler Transform of T	[C G $ T T C C]
SA:	Suffix Array of T	[6 5 0 2 4 1 3]
ISA:	Inverse Suffix Array of T	[2 5 3 6 4 1 0]

$$L[i] = T[(SA[i] - 1) \bmod |T|]$$

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓	sorted $T^{[i]}$					L ↓
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

\$

What if I only have access to *BWT*? Can I recover *T*?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓	sorted $T^{[i]}$					L ↓
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

C\$

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓		sorted $T^{[i]}$				L ↓
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

GC\$

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓	sorted $T^{[i]}$					L ↓
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

TGC\$

From *BWT* back to *T*

What if I only have access to *BWT*? Can I recover *T*?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F			sorted $T^{[i]}$			L
	↓						↓
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

CTGC\$

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓	sorted $T^{[i]}$					L ↓
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

TCTGC\$

From *BWT* back to *T*

What if I only have access to *BWT*? Can I recover *T*?

Recovering *T* is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F			sorted $T^{[i]}$			L
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

CTCTGC\$

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓	sorted $T^{[i]}$					L ↓
0	$	C	T	C	T	G	C
1	C	$	C	T	C	T	G
2	C	T	C	T	G	C	$
3	C	T	G	C	$	C	T
4	G	C	$	C	T	C	T
5	T	C	T	G	C	$	C
6	T	G	C	$	C	T	C

CTCTGC$ $= T$

What if I only have access to *BWT*? Can I recover *T*?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F $\downarrow$			sorted $T^{[i]}$			L $\downarrow$
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F $\downarrow$			sorted $T^{[i]}$			L $\downarrow$
0	\$	C	T	C	T	G	C
1	C	\$	C	T	C	T	G
2	C	T	C	T	G	C	\$
3	C	T	G	C	\$	C	T
4	G	C	\$	C	T	C	T
5	T	C	T	G	C	\$	C
6	T	G	C	\$	C	T	C

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F $\downarrow$	sorted $T^{[i]}$					L $\downarrow$	LF
0	\$	C	T	C	T	G	C	1
1	C	\$	C	T	C	T	G	
2	C	T	C	T	G	C	\$	
3	C	T	G	C	\$	C	T	
4	G	C	\$	C	T	C	T	
5	T	C	T	G	C	\$	C	
6	T	G	C	\$	C	T	C	

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	$F \downarrow$			sorted $T^{[i]}$			$L \downarrow$	LF
0	\$	C	T	C	T	G	C	1
1	C	\$	C	T	C	T	G	4
2	C	T	C	T	G	C	\$	
3	C	T	G	C	\$	C	T	
4	G	C	\$	C	T	C	T	
5	T	C	T	G	C	\$	C	
6	T	G	C	\$	C	T	C	

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓	sorted $T^{[i]}$					L ↓	LF
0	$	C	T	C	T	G	C	1
1	C	$	C	T	C	T	G	4
2	C	T	C	T	G	C	$	
3	C	T	G	C	$	C	T	
4	G	C	$	C	T	C	T	6
5	T	C	T	G	C	$	C	
6	T	G	C	$	C	T	C	

CTCTGC$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	$F \downarrow$			sorted $T^{[i]}$			$L \downarrow$	LF
0	\$	C	T	C	T	G	C	1
1	C	\$	C	T	C	T	G	4
2	C	T	C	T	G	C	\$	
3	C	T	G	C	\$	C	T	
4	G	C	\$	C	T	C	T	6
5	T	C	T	G	C	\$	C	
6	T	G	C	\$	C	T	C	3

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	$F \downarrow$	sorted $T^{[i]}$					$L \downarrow$	LF
0	\$	C	T	C	T	G	C	1
1	C	\$	C	T	C	T	G	4
2	C	T	C	T	G	C	\$	
3	C	T	G	C	\$	C	T	5
4	G	C	\$	C	T	C	T	6
5	T	C	T	G	C	\$	C	
6	T	G	C	\$	C	T	C	3

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	F ↓	sorted $T^{[i]}$					L ↓	LF
0	\$	C	T	C	T	G	C	1
1	C	\$	C	T	C	T	G	4
2	C	T	C	T	G	C	\$	
3	C	T	G	C	\$	C	T	5
4	G	C	\$	C	T	C	T	6
5	T	C	T	G	C	\$	C	2
6	T	G	C	\$	C	T	C	3

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	$F \downarrow$			sorted $T^{[i]}$			$L \downarrow$	LF
0	\$	C	T	C	T	G	C	1
1	C	\$	C	T	C	T	G	4
2	C	T	C	T	G	C	\$	0
3	C	T	G	C	\$	C	T	5
4	G	C	\$	C	T	C	T	6
5	T	C	T	G	C	\$	C	2
6	T	G	C	\$	C	T	C	3

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

From BWT back to T

What if I only have access to BWT? Can I recover T?

Recovering T is easy if, given a position in the table, we can find the position of the previous cyclic shift.

Example

	$F \downarrow$	sorted $T^{[i]}$					$L \downarrow$	LF
0	\$	C	T	C	T	G	C	1
1	C	\$	C	T	C	T	G	4
2	C	T	C	T	G	C	\$	0
3	C	T	G	C	\$	C	T	5
4	G	C	\$	C	T	C	T	6
5	T	C	T	G	C	\$	C	2
6	T	G	C	\$	C	T	C	3

CTCTGC\$ $= T$

Property

Since cyclic shifts are sorted, $T^{[i]}[n] = T[i-1]$ appears as many times

- in L from position 0 to the position of $T^{[i]}$ as
- in F from position 0 to the position of $T^{[i-1]}$.

So, L can be used instead of T

L contains all the information that is needed for recovering the original T.

$$T = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{C}}\,\overset{3}{\mathrm{T}}\,\overset{4}{\mathrm{G}}\,\overset{5}{\mathrm{C}}\,\overset{6}{\$} \rightarrow T' = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{G}}\,\overset{3}{\mathrm{C}}\,\overset{4}{\mathrm{T}}\,\overset{5}{\mathrm{G}}\,\overset{6}{\mathrm{C}}\,\overset{7}{\$}$$

What is the impact of a single insertion of G at position $i = 2$?

	unsorted CS of T						
0	C	T	C	T	G	C	$
1	T	C	T	G	C	$	C
2	C	T	G	C	$	C	T
3	T	G	C	$	C	T	C
4	G	C	$	C	T	C	T
5	C	$	C	T	C	T	G
6	$	C	T	C	T	G	C

	unsorted CS of T'							
0	C	T	G	C	T	G	C	$
1	T	G	C	T	G	C	$	C
2	G	C	T	G	C	$	C	T
3	C	T	G	C	$	C	T	G
4	T	G	C	$	C	T	G	C
5	G	C	$	C	T	G	C	T
6	C	$	C	T	G	C	T	G
7	$	C	T	G	C	T	G	C

$$T = \overset{0}{C}\overset{1}{T}\overset{2}{C}\overset{3}{T}\overset{4}{G}\overset{5}{C}\overset{6}{\$} \longrightarrow T' = \overset{0}{C}\overset{1}{T}\overset{2}{\color{red}G}\overset{3}{C}\overset{4}{T}\overset{5}{G}\overset{6}{C}\overset{7}{\$}$$

What is the impact of a single insertion of G at position $i = 2$?

	F ↓	sorted CS of T					L ↓	SA ↓
0	$	C	T	C	T	G	C	6
1	C	$	C	T	C	T	G	5
2	C	T	C	T	G	C	$	0
3	C	T	G	C	$	C	T	2
4	G	C	$	C	T	C	T	4
5	T	C	T	G	C	$	C	1
6	T	G	C	$	C	T	C	3

	F ↓	sorted CS of T'						L ↓	SA ↓
0	$	C	T	G	C	T	G	C	7
1	C	$	C	T	G	C	T	G	6
2	C	T	G	C	$	C	T	G	3
3	C	T	G	C	T	G	C	$	0
4	G	C	$	C	T	G	C	T	5
5	G	C	T	G	C	$	C	T	2
6	T	G	C	$	C	T	G	C	4
7	T	G	C	T	G	C	$	C	1

Stage 1: $T'^{[j]}$ for all $j > i + 1$

Cyclic shifts where the inserted letter G appears after \$ and before L.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTCTGC
1	C	G	C\$CTCTG
2	C	\$	CTCTGC\$
3	C	T	CTGC\$CT
4	G	T	GC\$CTCT
5	T	C	TCTGC\$C
6	T	C	TGC\$CTC

Stage 1: $T'^{[j]}$ for all $j > i + 1$

Cyclic shifts where the inserted letter G appears after \$ and before L.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTCTGC	←
1	C	G	C\$CTCTG	←
2	C	\$	CTCTGC\$	
3	C	T	CTGC\$CT	
4	G	T	GC\$CTCT	←
5	T	C	TCTGC\$C	
6	T	C	TGC\$CTC	←

Stage 1: $T'^{[j]}$ for all $j > i + 1$

Cyclic shifts where the inserted letter G appears after \$ and before L.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	←
1	C	G	C\$CTGCTG	←
2	C	\$	CTCTGC\$	
3	C	T	CTGC\$CT	
4	G	T	GC\$CTGCT	←
5	T	C	TCTGC\$C	
6	T	C	TGC\$CTGC	←

Stage 1: $T'^{[j]}$ for all $j > i + 1$

Cyclic shifts where the inserted letter G appears after \$ and before L.

Impact on M: none

The respective ranking of these cyclic shifts is preserved.
F: no direct modification.
L: no direct modification.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	←
1	C	G	C\$CTGCTG	←
2	C	\$	CTCTGC\$	
3	C	T	CTGC\$CT	
4	G	T	GC\$CTGCT	←
5	T	C	TCTGC\$C	
6	T	C	TGC\$CTGC	←

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

What are we observing?

	F	L	cyclic shifts
0	$	C	$CTGCTGC
1	C	G	C$CTGCTG
2	C	$	CTCTGC$
3	C	T	CTGC$CT
4	G	T	GC$CTGCT
5	T	C	TCTGC$C
6	T	C	TGC$CTGC

$$T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$} \rightarrow T' = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{G}\,\overset{3}{C}\,\overset{4}{T}\,\overset{5}{G}\,\overset{6}{C}\,\overset{7}{\$}$$

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

What are we observing?

	F	L	cyclic shifts
0	$	C	$CTGCTGC
1	C	G	C$CTGCTG
2	C	$	CTCTGC$
3	C	T	CTGC$CT
4	G	T	GC$CTGCT
5	T	C	TCTGC$C
6	T	C	TGC$CTGC

How can we compute the position of the modification?

We are looking for the position of $T'^{[3]}$ (corresponding to $T^{[2]}$).

	0	1	2	3	4	5	6
ISA	2	5	3	6	4	1	0

$$T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$} \rightarrow T' = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{G}\,\overset{3}{C}\,\overset{4}{T}\,\overset{5}{G}\,\overset{6}{C}\,\overset{7}{\$}$$

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTCTGC\$	
3	C	T	CTGC\$CT	$\leftarrow T^{[2]}$
4	G	T	GC\$CTGCT	
5	T	C	TCTGC\$C	
6	T	C	TGC\$CTGC	

How can we compute the position of the modification?

We are looking for the position of $T'^{[3]}$ (corresponding to $T^{[2]}$).

	0	1	**2**	3	4	5	6
ISA	2	5	3	6	4	1	0

$$T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$} \longrightarrow T' = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{G}\,\overset{3}{C}\,\overset{4}{T}\,\overset{5}{G}\,\overset{6}{C}\,\overset{7}{\$}$$

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTCTGC\$	
3	C	T	CTGC\$CT	$\longleftarrow T^{[2]}$
4	G	T	GC\$CTGCT	
5	T	C	TCTGC\$C	
6	T	C	TGC\$CTGC	

Position of the previous cyclic

In what follows, we need the position of the previous cyclic shift $T^{[1]}$ (corresponding to $T'^{[1]}$).

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTCTGC\$	
3	C	T	First T in L	$\longleftarrow T^{[2]}$
4	G	T	GC\$CTGCT	
5	T	C	TCTGC\$C	
6	T	C	TGC\$CTGC	

Position of the previous cyclic

In what follows, we need the position of the previous cyclic shift $T^{[1]}$ (corresponding to $T'^{[1]}$).

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTCTGC\$
3	C	T	First T in L $\leftarrow T^{[2]}$
4	G	T	GC\$CTGCT
5	T	C	First T in F
6	T	C	TGC\$CTGC

Position of the previous cyclic

In what follows, we need the position of the previous cyclic shift $T^{[1]}$ (corresponding to $T'^{[1]}$).

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

What are we observing?

	F	L	cyclic shifts	
0	$	C	$CTGCTGC	
1	C	G	C$CTGCTG	
2	C	$	CTCTGC$	
3	C	T	First T in L	$\longleftarrow T^{[2]}$
4	G	T	GC$CTGCT	
5	T		First T in F C	
6	T	C	TGC$CTGC	

Position of the previous cyclic

In what follows, we need the position of the previous cyclic shift $T^{[1]}$ (corresponding to $T'^{[1]}$).
$\rightarrow LF(3) = 5$, we store 5 in *previous_cs*.

Stage 2: $T'^{[i+1]}$

The cyclic shift where the inserted letter G appears in L.

Impact on M: substitution

F: no direct modification.
L: substitution T (stored) $\longrightarrow$ G.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTCTGC\$	
3	C	G	CTGC\$CTG	$\longleftarrow T'^{[3]}$
4	G	T	GC\$CTGCT	
5	T	C	TCTGC\$C	
6	T	C	TGC\$CTGC	

previous_cs = 5

Stage 3: $T'^{[i]}$

The cyclic shift where the inserted letter G appears in F.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	T	C	TCTGC\$C
6	T	C	TGC\$CTGC

$previous\_cs = 5$

Stage 3: $T'^{[i]}$

The cyclic shift where the inserted letter G appears in F.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	T	C	TCTGC\$C
6	T	C	TGC\$CTGC

Where does the insertion take place?

- We know the position of $T'^{[3]}$ (we have just modified it).
- Now, we need the position of the *new* cyclic shift $T'^{[2]}$ = GCTGC\$CT.
- That's what LF computes: the position of the previous cyclic shift!

previous_cs = 5

Stage 3: $T'^{[i]}$

The cyclic shift where the inserted letter G appears in F.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTCTGC\$
3	C	G	Second G in L
4	G	T	GC\$CTGCT
5	T	C	TCTGC\$C
6	T	C	TGC\$CTGC

Where does the insertion take place?

- We know the position of $T'^{[3]}$ (we have just modified it).
- Now, we need the position of the *new* cyclic shift $T'^{[2]}$ = GCTGC\$CT.
- That's what LF computes: the position of the previous cyclic shift!

previous_cs = 5

$$T = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{C}}\,\overset{3}{\mathrm{T}}\,\overset{4}{\mathrm{G}}\,\overset{5}{\mathrm{C}}\,\overset{6}{\$} \rightarrow T' = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{G}}\,\overset{3}{\mathrm{C}}\,\overset{4}{\mathrm{T}}\,\overset{5}{\mathrm{G}}\,\overset{6}{\mathrm{C}}\,\overset{7}{\$}$$

Stage 3: $T'^{[i]}$

The cyclic shift where the inserted letter G appears in F.

What are we observing?

	F	L	cyclic shifts	
0	$	C	$CTGCTGC	
1	C	G	C$CTGCTG	
2	C	$	CTCTGC$	
3	C	G	CTGC$CTG	
4	G	T	GC$CTGCT	
5	G	T	GCTGC$CT	←
6	T	C	TCTGC$C	
7	T	C	TGC$CTGC	

Where does the insertion take place?

- We know the position of $T'^{[3]}$ (we have just modified it).
- Now, we need the position of the *new* cyclic shift $T'^{[2]}$ = GCTGC$CT.
- That's what LF computes: the position of the previous cyclic shift!

$previous\_cs = \cancel{5}\,6$

Stage 3: $T'^{[i]}$

The cyclic shift where the inserted letter G appears in F.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	G		Second G in F ←
6	T	C	TCTGC\$C
7	T	C	TGC\$CTGC

$previous\_cs = \cancel{5}6$

Stage 3: $T'^{[i]}$

The cyclic shift where the inserted letter G appears in F.

Impact on M: insertion

A new row starting with the inserted letter G and ending with the stored T is inserted.

F: inserted letter G.

L: (stored) T.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTCTGC\$	
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	←
6	T	C	TCTGC\$C	
7	T	C	TGC\$CTGC	

$previous\_cs = \cancel{5}6$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	G	T	GCTGC\$CT
6	T	C	TCTGC\$C
7	T	C	TGC\$CTGC

$previous\_cs = 6$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	←
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGCTGC\$C	←
7	T	C	TGC\$CTGC	

previous_cs = 6

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

How to reorder cyclic shifts?

- Reordering from right to left (from $j = i - 1$ downto 0)
- Comparison between the actual position (value of *previous_cs*) and the position computed with LF.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTGCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	G	T	GCTGC\$CT
6	T	C	TGCTGC\$C
7	T	C	TGC\$CTGC

previous_cs $= 6$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGCTGC\$C	$T'^{[1]}$
7	T	C	TGC\$CTGC	

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position *previous_cs* $= 6$.
Is this the correct position for $T'^{[1]}$? LF can tell us!

previous_cs $= 6$

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	$T'^{[2]}$
6	T	C	TGCTGC\$C	$T'^{[1]}$
7	T	C	TGC\$CTGC	

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position $previous\_cs = 6$.
Is this the correct position for $T'^{[1]}$? LF can tell us!
$T'^{[2]}$ has just been inserted $\rightarrow$ its location is correct.
$T'^{[2]}$ is at position 5, let's compute $LF(5)$.

$previous\_cs = 6$

$$T = \overset{0}{C}\overset{1}{T}\overset{2}{C}\overset{3}{T}\overset{4}{G}\overset{5}{C}\overset{6}{\$} \rightarrow T' = \overset{0}{C}\overset{1}{T}\overset{2}{\color{red}G}\overset{3}{C}\overset{4}{T}\overset{5}{G}\overset{6}{C}\overset{7}{\$}$$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	Second T in L.	[2]
6	T	C	TGCTGC\$C	$T'^{[1]}$
7	T	C	TGC\$CTGC	

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position *previous_cs* $= 6$.
Is this the correct position for $T'^{[1]}$? LF can tell us!
$T'^{[2]}$ has just been inserted $\rightarrow$ its location is correct.
$T'^{[2]}$ is at position 5, let's compute $LF(5)$.

previous_cs $= 6$

$$T = \overset{0}{\mathrm{C}}\overset{1}{\mathrm{T}}\overset{2}{\mathrm{C}}\overset{3}{\mathrm{T}}\overset{4}{\mathrm{G}}\overset{5}{\mathrm{C}}\overset{6}{\$} \longrightarrow T' = \overset{0}{\mathrm{C}}\overset{1}{\mathrm{T}}\overset{2}{\mathrm{G}}\overset{3}{\mathrm{C}}\overset{4}{\mathrm{T}}\overset{5}{\mathrm{G}}\overset{6}{\mathrm{C}}\overset{7}{\$}$$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTGCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	G	T	Second T in L. [2]
6	T	C	TGCTGC\$C $T'^{[1]}$
7	T		Second T in F.

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position *previous_cs* $= 6$.
Is this the correct position for $T'^{[1]}$? LF can tell us!
$T'^{[2]}$ has just been inserted $\rightarrow$ its location is correct.
$T'^{[2]}$ is at position 5, let's compute $LF(5)$.

previous_cs $= 6$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTGCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	G	T	
6	T	C	
7	T	C	

Second T in L. [2]

Second T in F.
→ expected position of $T'^{[1]}$!

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position *previous_cs* $= 6$.
Is this the correct position for $T'^{[1]}$? LF can tell us!
$T'^{[2]}$ has just been inserted → its location is correct.
$T'^{[2]}$ is at position 5, let's compute $LF(5)$.

previous_cs $= 6$

$$T = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{C}}\,\overset{3}{\mathrm{T}}\,\overset{4}{\mathrm{G}}\,\overset{5}{\mathrm{C}}\,\overset{6}{\$} \rightarrow T' = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{G}}\,\overset{3}{\mathrm{C}}\,\overset{4}{\mathrm{T}}\,\overset{5}{\mathrm{G}}\,\overset{6}{\mathrm{C}}\,\overset{7}{\$}$$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTGCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	G	T	
6	T	C	TGCTGC\$C $T'^{[1]}$
7	T		

Second T in L.

Second T in F.

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position 6 but should be at position 7.

$previous\_cs = 6$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGCTGC\$C	$T'^{[1]}$
7	T	C	TGC\$CTGC	

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position 6 but should be at position 7.
Before moving $T'^{[1]}$, we compute the actual position of $T'^{[0]}$ and store it in *previous_cs*.

$previous\_cs = 6$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts
0	\$	C	\$CTGCTGC
1	C	G	C\$CTGCTG
2	C	\$	CTGCTGC\$
3	C	G	CTGC\$CTG
4	G	T	GC\$CTGCT
5	G	T	GCTGC\$CT
6	T	C	Second C in L. [1]
7	T	C	TGC\$CTGC

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position 6 but should be at position 7.
Before moving $T'^{[1]}$, we compute the actual position of $T'^{[0]}$ and store it in *previous_cs*.

previous_cs $= 6$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C			
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGCTGC\$C	$T'^{[1]}$
7	T	C	TGC\$CTGC	

Second C in F.

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position 6 but should be at position 7.
Before moving $T'^{[1]}$, we compute the actual position of $T'^{[0]}$ and store it in *previous_cs*.

$previous\_cs = 6$

$$T = \overset{0}{\mathsf{C}}\,\overset{1}{\mathsf{T}}\,\overset{2}{\mathsf{C}}\,\overset{3}{\mathsf{T}}\,\overset{4}{\mathsf{G}}\,\overset{5}{\mathsf{C}}\,\overset{6}{\$} \rightarrow T' = \overset{0}{\mathsf{C}}\,\overset{1}{\mathsf{T}}\,\overset{2}{\mathsf{G}}\,\overset{3}{\mathsf{C}}\,\overset{4}{\mathsf{T}}\,\overset{5}{\mathsf{G}}\,\overset{6}{\mathsf{C}}\,\overset{7}{\$}$$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	$T'^{[0]}$
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGCTGC\$C	$T'^{[1]}$
7	T	C	TGC\$CTGC	

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position 6 but should be at position 7.
Before moving $T'^{[1]}$, we compute the actual position of $T'^{[0]}$ and store it in *previous_cs*.

previous_cs $= 2$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	$T'^{[0]}$
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGC\$CTGC	
7	T	C	TGCTGC\$C	$T'^{[1]}$

Reordering $T'^{[1]}$

$T'^{[1]}$ is at position 6 but should be at position 7.
Before moving $T'^{[1]}$, we compute the actual position of $T'^{[0]}$ and store it in *previous_cs*.

$previous\_cs = 2$

$$T = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{C}\,\overset{3}{T}\,\overset{4}{G}\,\overset{5}{C}\,\overset{6}{\$} \rightarrow T' = \overset{0}{C}\,\overset{1}{T}\,\overset{2}{\color{red}G}\,\overset{3}{C}\,\overset{4}{T}\,\overset{5}{G}\,\overset{6}{C}\,\overset{7}{\$}$$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	$T'^{[0]}$
3	C	G	CTGC\$CTG	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGC\$CTGC	
7	T	C	TGCTGC\$C	$T'^{[1]}$

Reordering $T'^{[0]}$

Now, let's compute the correct position of $T'^{[0]}$ using $LF(7)$ (7 is the correct position of $T'^{[1]}$).

$previous\_cs = 2$

What are we observing?

	F	L	cyclic shifts	
0	$	C	$CTGCTGC	
1	C	G	C$CTGCTG	
2	C	$	CTGCTGC$	$T'^{[0]}$
3	C	G	CTGC$CTG	
4	G	T	GC$CTGCT	
5	G	T	GCTGC$CT	
6	T	C	TGC$CTGC	
7	T	C	Third C in L.	$T'^{[1]}$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

Reordering $T'^{[0]}$

Now, let's compute the correct position of $T'^{[0]}$ using $LF(7)$ (7 is the correct position of $T'^{[1]}$).

$previous\_cs = 2$

$$T = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{C}}\,\overset{3}{\mathrm{T}}\,\overset{4}{\mathrm{G}}\,\overset{5}{\mathrm{C}}\,\overset{6}{\$} \longrightarrow T' = \overset{0}{\mathrm{C}}\,\overset{1}{\mathrm{T}}\,\overset{2}{\mathrm{G}}\,\overset{3}{\mathrm{C}}\,\overset{4}{\mathrm{T}}\,\overset{5}{\mathrm{G}}\,\overset{6}{\mathrm{C}}\,\overset{7}{\$}$$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	$T'^{[0]}$
3	C		Third C in F.	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGC\$CTGC	
7	T	C	Third C in L.	$T'^{[1]}$

Reordering $T'^{[0]}$

Now, let's compute the correct position of $T'^{[0]}$ using $LF(7)$ (7 is the correct position of $T'^{[1]}$).

previous_cs $= 2$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	\$	CTGCTGC\$	$T'^{[0]}$
3	C	[illegible]	Third C in F.	
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGC\$CTGC	
7	T	C	Third C in L.	$T'^{[1]}$

Reordering $T'^{[0]}$

Now, let's compute the correct position of $T'^{[0]}$ using $LF(7)$ (7 is the correct position of $T'^{[1]}$).
$T'^{[0]}$ should be at position 3.

$previous\_cs = 2$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	G	CTGC\$CTG	
3	C	\$	CTGCTGC\$	$T'^{[0]}$
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGC\$CTGC	
7	T	C	TGCTGC\$C	$T'^{[1]}$

Reordering $T'^{[0]}$

Now, let's compute the correct position of $T'^{[0]}$ using $LF(7)$ (7 is the correct position of $T'^{[1]}$).
$T'^{[0]}$ should be at position 3.

$previous\_cs = 2$

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	G	CTGC\$CTG	
3	C	\$	CTGCTGC\$	$T'^{[0]}$
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGC\$CTGC	
7	T	C	TGCTGC\$C	$T'^{[1]}$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

Reordering $T'^{[0]}$

Now, let's compute the correct position of $T'^{[0]}$ using $LF(7)$ (7 is the correct position of $T'^{[1]}$).
$T'^{[0]}$ should be at position 3.
Position of $T'^{[0]}$ is correct $\rightarrow$ all cyclic shifts are well ordered.

$previous\_cs = 2$

Stage 4: $T'^{[j]}$ for all $j < i$

Cyclic shifts where the inserted letter G appears after F and before \$.

Impact on M: reordering

Depending on the inserted letter, rows might locally rotate.
F: no modification.
L: possible local reorderings.

What are we observing?

	F	L	cyclic shifts	
0	\$	C	\$CTGCTGC	
1	C	G	C\$CTGCTG	
2	C	G	CTGC\$CTG	
3	C	\$	CTGCTGC\$	$T'^{[0]}$
4	G	T	GC\$CTGCT	
5	G	T	GCTGC\$CT	
6	T	C	TGC\$CTGC	
7	T	C	TGCTGC\$C	$T'^{[1]}$

What are we using?

	L	ISA
0	C	2
1	G	
2	$	
3	T	6
4	T	
5	C	
6	C	0

Explanations

1. L and a subsampling of ISA;
2. $\text{rank}_c(L, i)$;
3. F and Count;
4. $LF(i)=\text{rank}_{L[i]}(L,i)+\text{Count}(L[i])-1$;

What are we using?

	L	ISA
0	C	2
1	G	
2	\$	
3	T	6
4	T	
5	C	
6	C	0

$\text{rank}_c(L, i)$

	0	1	2	3	4	5	6
\$	0	0	1	1	1	1	1
C	1	1	1	1	1	2	3
G	0	1	1	1	1	1	1
T	0	0	0	1	2	2	2

Explanations

1. L and a subsampling of ISA;
2. $\text{rank}_c(L, i)$;
3. F and Count;
4. $LF(i) = \text{rank}_{L[i]}(L, i) + \text{Count}(L[i]) - 1$;

$$T = \overset{0}{C}\overset{1}{T}\overset{2}{C}\overset{3}{T}\overset{4}{G}\overset{5}{C}\overset{6}{\$} \rightarrow T' = \overset{0}{C}\overset{1}{T}\overset{2}{\color{red}G}\overset{3}{C}\overset{4}{T}\overset{5}{G}\overset{6}{C}\overset{7}{\$}$$

What are we using?

	F	L	ISA
0	\$	C	2
1	C	G	
2	C	\$	
3	C	T	6
4	G	T	
5	T	C	
6	T	C	0

$\text{rank}_c(L, i)$

	0	1	2	3	4	5	6
\$	0	0	1	1	1	1	1
C	1	1	1	1	1	2	3
G	0	1	1	1	1	1	1
T	0	0	0	1	2	2	2

Count

\$	C	G	T
0	1	4	5

Explanations

1. L and a subsampling of ISA;
2. $\text{rank}_c(L, i)$;
3. F and Count;
4. $LF(i) = \text{rank}_{L[i]}(L, i) + \text{Count}(L[i])$-1;

$$T = \overset{0}{C}\overset{1}{T}\overset{2}{C}\overset{3}{T}\overset{4}{G}\overset{5}{C}\overset{6}{\$} \rightarrow T' = \overset{0}{C}\overset{1}{T}\overset{2}{\color{red}G}\overset{3}{C}\overset{4}{T}\overset{5}{G}\overset{6}{C}\overset{7}{\$}$$

What are we using?

	F	L	ISA
0	\$	C	2
1	C	G	
2	C	\$	
3	C	T	6
4	G	T	
5	T	C	
6	T	C	0

$\text{rank}_c(L, i)$

	0	1	2	3	4	5	6
\$	0	0	1	1	1	1	1
C	1	1	1	1	1	2	3
G	0	1	1	1	1	1	1
T	0	0	0	1	2	2	2

Count

\$	C	G	T
0	1	4	5

Explanations

1. L and a subsampling of ISA;
2. $\text{rank}_c(L, i)$;
3. F and Count;
4. $LF(i) = \text{rank}_{L[i]}(L, i) + \text{Count}(L[i])\text{-}1$;

$\text{rank}_{L[i]}(L, i)$ returns the number of times, t, $L[i]$ appears in L from position 0 to i.

Therefore, $\text{rank}_{L[i]}(L, i) + \text{Count}(L[i])\text{-}1$ returns the position of the t-th $L[i]$ in F.

What are we using?

	F	L	ISA
0	\$	C	2
1	C	G	
2	C	\$	
3	C	T	6
4	G	T	
5	T	C	
6	T	C	0

$\text{rank}_c(L, i)$

	0	1	2	3	4	5	6
\$	0	0	1	1	1	1	1
C	1	1	1	1	1	2	3
G	0	1	1	1	1	1	1
T	0	0	0	1	2	2	2

Count

\$	C	G	T
0	1	4	5

Explanations

1. L and a subsampling of ISA;
2. $\text{rank}_c(L, i)$;
3. F and Count;
4. $LF(i)=\text{rank}_{L[i]}(L, i)+\text{Count}(L[i])\text{-}1$;

Note that $\text{rank}_c(L, i)$ gives L and Count gives F, so storing and maintaining these two functions is normally sufficient...
Note also that $\text{rank}_c(L, i)$ is stored in a more efficient way!

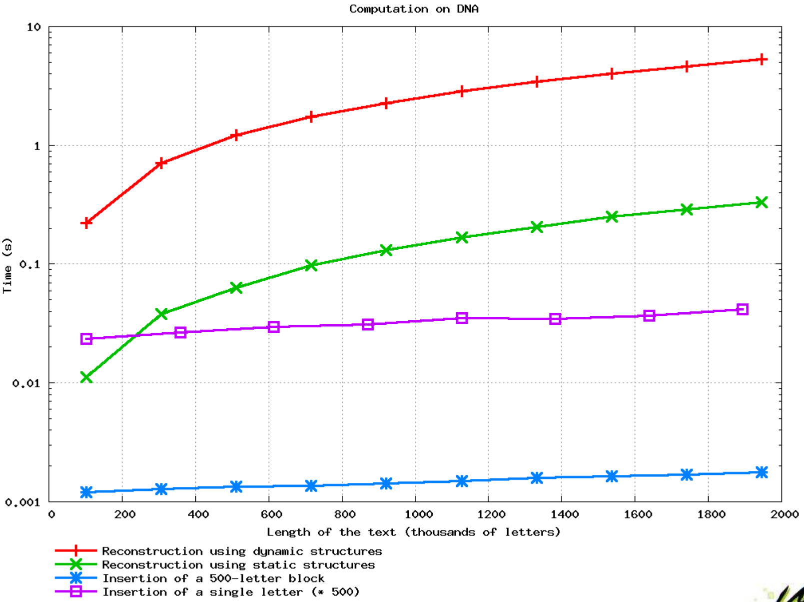
From Theory to Practice

The reordering step of our algorithm requires at most n iterations.

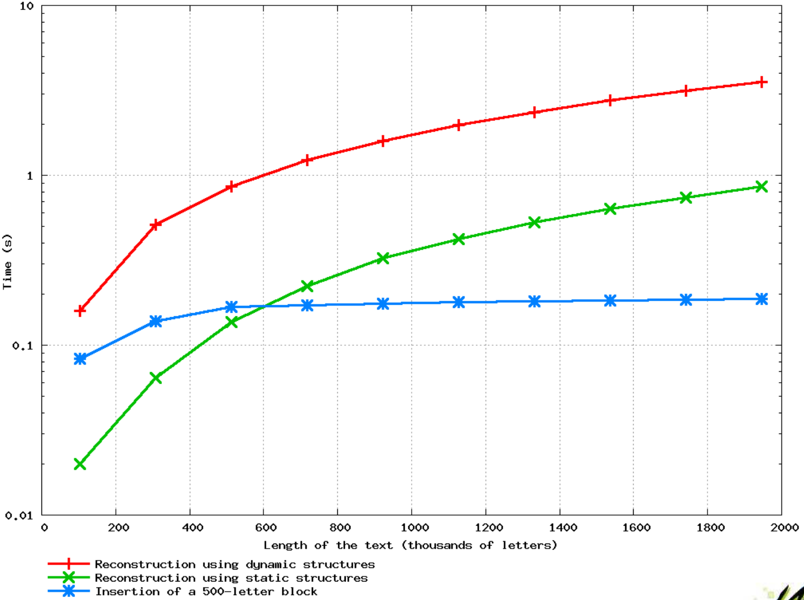
How our Algorithm Behaves in Practice?

- Is the reordering step too time-consuming?
- Is it quicker to update the BWT than recomputing it entirely?
- Is the algorithm slowed down because of the dynamic structures?

Experiments on Human Genome



Experiments on a Fibonacci Word



Conclusion

Generalization

We can handle insertions/deletions/substitutions of a factor as well.

Complexity

$O(n)$ iterations of the algorithm Reorder.
Worst-case scenario ($A^n\$ \rightarrow A^nC\$$).

The operations (*rank*, insertion, deletion) on the dynamic structure storing L are performed in at most $O(\log n(1 + \log \sigma / \log \log n))$.
Overall worst-case complexity: $O(n \log n(1 + \log \sigma / \log \log n))$.

Perspectives

- Dynamic FM-index (using SA, ISA subsamples) *submitted to JDA*
- Dynamic suffix array + LCP *submitted to JDA*
- Dynamic suffix tree *work in progress*

Conclusion

Generalization

We can handle insertions/deletions/substitutions of a factor as well.

Complexity

$O(n)$ iterations of the algorithm Reorder.
Worst-case scenario ($\mathrm{A}^n\$ \rightarrow \mathrm{A}^n\mathrm{C}\$$).

The operations (*rank*, insertion, deletion) on the dynamic structure storing L are performed in at most $O(\log n(1 + \log \sigma / \log \log n))$.
Overall worst-case complexity: $O(n \log n(1 + \log \sigma / \log \log n))$.

Perspectives

- Dynamic FM-index (using SA, ISA subsamples) *submitted to JDA*
- Dynamic suffix array + LCP *submitted to JDA*
- Dynamic suffix tree *work in progress*

Conclusion

Generalization

We can handle insertions/deletions/substitutions of a factor as well.

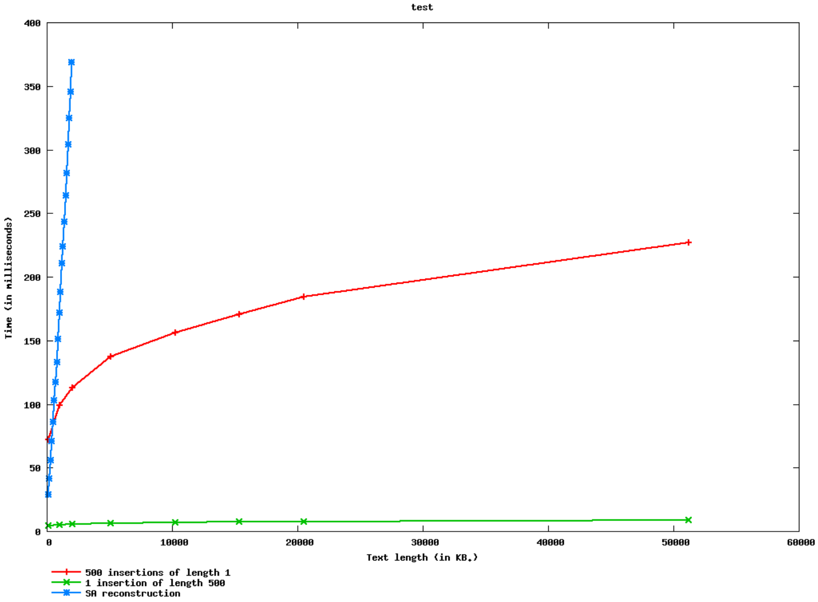
Complexity

$O(n)$ iterations of the algorithm Reorder.
Worst-case scenario ($\mathrm{A}^n\$ \to \mathrm{A}^n\mathrm{C}\$$).

The operations (*rank*, insertion, deletion) on the dynamic structure storing L are performed in at most $O(\log n(1 + \log \sigma / \log\log n))$.
Overall worst-case complexity: $O(n \log n(1 + \log \sigma / \log\log n))$.

Perspectives

- Dynamic FM-index (using SA, ISA subsamples) *submitted to JDA*
- Dynamic suffix array + LCP *submitted to JDA*
- Dynamic suffix tree *work in progress*


test
400
350
300
250
200
150
100
50
0
Time (in milliseconds)
0
10000
20000
30000
40000
50000
60000
Text length (in KB.)
500 insertions of length 1
1 insertion of length 500
SA reconstruction