
COMP9319 Web Data Compression and Search

An Occ Implementation,
RLFM (Compressed FM Index) Revisit

An example Occ implementation

FM Index ($L(x) = c$)

	<u>F</u>	<u>L</u>	<u>c</u>	<u>C</u>	
0	#	i	0	#	0
1	i	p	0	i	1
2	i	s	0	m	5
3	i	s	1	p	6
4	i	m	0	s	8
5	m	#	0		
6	p	p	1		
7	p	i	1		
8	s	s	2		
9	s	s	3		
10	s	i	2		
11	s	i	3		

FM Index (when reversing from L[5])

	<u>F</u>	<u>L</u>	<u>C</u>		<u>C</u>
0	#	i	0		# 0
1	i	p	0		i 1
2	i	s	0		m 5
3	i	s	1		p 6
4	i	m	0	LF[5] = 0+0 = 0, i	s 8
5	m	#	0		
6	p	p	1		
7	p	i	1		
8	s	s	2		
9	s	s	3		
10	s	i	2		
11	s	i	3		

FM Index (when reversing from L[5])

	<u>F</u>	<u>L</u>	<u>C</u>		<u>C</u>	
0	#	i	0		#	0
1	i	p	0		i	1
2	i	s	0		m	5
3	i	s	1		p	6
4	i	m	0	LF[5] = 0+0 = 0, i	s	8
5	m	#	0	LF[0] = 1+0 = 1, p		
6	p	p	1	LF[1] = 6+0 = 6, p		
7	p	i	1	LF[6] = 6+1 = 7, i		
8	s	s	2	LF[7] = 1+1 = 2, s		
9	s	s	3	LF[2] = 8+0 = 8, s		
10	s	i	2	LF[8] = 8+2 = 10, i		
11	s	i	3	LF[10] = 1+2 = 3, s		
				LF[3] = 8+1 = 9, s		
				LF[9] = 8+3 = 11, i		
				LF[11] = 1+3 = 4, m		

FM Index ($L(x) \neq c$)

	<u>F</u>	<u>L</u>	<u>i</u> <u>m</u> <u>p</u> <u>s</u>	<u>C</u>
0	#	i	1 0 0 0	# 0
1	i	p	1 0 1 0	i 1
2	i	s	1 0 1 1	m 5
3	i	s	1 0 1 2	p 6
4	i	m	1 1 1 2	s 8
5	m	#	1 1 1 2	
6	p	p	1 1 2 2	
7	p	i	2 1 2 2	
8	s	s	2 1 2 3	
9	s	s	2 1 2 4	
10	s	i	3 1 2 4	
11	s	i	4 1 2 4	

FM Index (con't)

		<u>F</u>	<u>L</u>	<u>i</u>	<u>m</u>	<u>p</u>	<u>s</u>	<u>C</u>	
pss <u>i</u>	0	#	i	1	0	0	0	#	0
➔	1	i	p	1	0	1	0	i	1
	2	i	s	1	0	1	1	m	5
	3	i	s	1	0	1	2	p	6
➔	4	i	m	1	1	1	2	s	8
	5	m	#	1	1	1	2		
	6	p	p	1	1	2	2		
Fst=1	7	p	i	2	1	2	2		
Lst=4	8	s	s	2	1	2	3		
	9	s	s	2	1	2	4		
	10	s	i	3	1	2	4		
	11	s	i	4	1	2	4		



FM Index (con't)

		<u>F</u>	<u>L</u>	<u>i</u> <u>m</u> <u>p</u> <u>s</u>	<u>C</u>
pssi	0	#	i	1 0 0 0	# 0
→	1	i	p	1 0 1 0	i 1
	2	i	s	1 0 1 1	m 5
	3	i	s	1 0 1 2	p 6
→	4	i	m	1 1 1 2	s 8
	5	m	#	1 1 1 2	
	6	p	p	1 1 2 2	
	7	p	i	2 1 2 2	
	8	s	s	2 1 2 3	
	9	s	s	2 1 2 4	
	10	s	i	3 1 2 4	
	11	s	i	4 1 2 4	

$$Fst = 8 + 0$$

$$Lst = (8 + 2) - 1$$

FM Index (con't)

		<u>F</u>	<u>L</u>	<u>i</u>	<u>m</u>	<u>p</u>	<u>s</u>	<u>C</u>	
pssi	0	#	i	1	0	0	0	#	0
	1	i	p	1	0	1	0	i	1
	2	i	s	1	0	1	1	m	5
	3	i	s	1	0	1	2	p	6
	4	i	m	1	1	1	2	s	8
	5	m	#	1	1	1	2		
	6	p	p	1	1	2	2		
	7	p	i	2	1	2	2		
	8	s	s	2	1	2	3		
	9	s	s	2	1	2	4		
	10	s	i	3	1	2	4		
	11	s	i	4	1	2	4		



$Fst = 8 + 0$
 $Lst = (8 + 2) - 1$

FM Index (con't)

		<u>F</u>	<u>L</u>	<u>i</u> <u>m</u> <u>p</u> <u>s</u>	<u>C</u>
<u>pssi</u>	0	#	i	1 0 0 0	# 0
	1	i	p	1 0 1 0	i 1
	2	i	s	1 0 1 1	m 5
	3	i	s	1 0 1 2	p 6
	4	i	m	1 1 1 2	s 8
	5	m	#	1 1 1 2	
	6	p	p	1 1 2 2	
	7	p	i	2 1 2 2	
➡	8	s	s	2 1 2 3	
➡	9	s	s	2 1 2 4	
	10	s	i	3 1 2 4	
	11	s	i	4 1 2 4	

$Fst = 8 + 2$
 $Lst = (8 + 4) - 1$

FM Index (con't)

		<u>F</u>	<u>L</u>	<u>i</u>	<u>m</u>	<u>p</u>	<u>s</u>	<u>C</u>	
<u>pssi</u>	0	#	i	1	0	0	0	#	0
	1	i	p	1	0	1	0	i	1
	2	i	s	1	0	1	1	m	5
	3	i	s	1	0	1	2	p	6
	4	i	m	1	1	1	2	s	8
	5	m	#	1	1	1	2		
	6	p	p	1	1	2	2		
	7	p	i	2	1	2	2		
	8	s	s	2	1	2	3		
	9	s	s	2	1	2	4		
	10	s	i	3	1	2	4		
	11	s	i	4	1	2	4		

$$Fst = 8 + 2$$

$$Lst = (8 + 4) - 1$$

FM Index (con't)

<u>pss</u> <u>i</u>		<u>F</u>	<u>L</u>	<u>i</u>	<u>m</u>	<u>p</u>	<u>s</u>	<u>C</u>	
	0	#	i	1	0	0	0	#	0
	1	i	p	1	0	1	0	i	1
	2	i	s	1	0	1	1	m	5
	3	i	s	1	0	1	2	p	6
	4	i	m	1	1	1	2	s	8
	5	m	#	1	1	1	2		
	6	p	p	1	1	2	2		
	7	p	i	2	1	2	2		
	8	s	s	2	1	2	3		
	9	s	s	2	1	2	4		
→	10	s	i	3	1	2	4		
→	11	s	i	4	1	2	4		

$Fst = 6 + 2$
 $Lst = (6 + 2) - 1$

FM Index (con't)

		<u>F</u>	<u>L</u>	<u>i</u>	<u>m</u>	<u>p</u>	<u>s</u>	<u>C</u>	
<u>pssi</u>	0	#	i	1	0	0	0	#	0
	1	i	p	1	0	1	0	i	1
	2	i	s	1	0	1	1	m	5
	3	i	s	1	0	1	2	p	6
	4	i	m	1	1	1	2	s	8
	5	m	#	1	1	1	2		
	6	p	p	1	1	2	2		
	7	p	i	2	1	2	2		
	8	s	s	2	1	2	3		
	9	s	s	2	1	2	4		
→	10	s	i	3	1	2	4		
→	11	s	i	4	1	2	4		

$$Fst = 6 + 2$$

$$Lst = (6 + 2) - 1$$

Fst > Lst => No match

FM Index (con't)

<u>pssi</u>		<u>F</u>	<u>L</u>	<u>i m p s</u>	<u>C</u>
	0	#	i		# 0
	1	i	p		i 1
	2	i	s		m 5
	3	i	s	1 0 1 2	p 6
	4	i	m		s 8
	5	m	#		
	6	p	p		
	7	p	i	2 1 2 2	
	8	s	s		
	9	s	s		
	10	s	i		
	11	s	i	4 1 2 4	

To reduce space

FM Index (con't)

		<u>F</u>	<u>L</u>	<u>i m p s</u>	<u>C</u>
<u>psssi</u>	0	#	i		# 0
	1	i	p		i 1
	2	i	s	←→	m 5
	3	i	s	1 0 1 2	p 6
	4	i	m		s 8
	5	m	#		
	6	p	p		
	7	p	i	2 1 2 2	
	8	s	s		
	9	s	s		
	10	s	i		
	11	s	i	4 1 2 4	

To reduce space

Similar when $L(x) = c$

	<u>F</u>	<u>L</u>	<u>i m p s</u>	<u>C</u>
0	#	i		# 0
1	i	p		i 1
2	i	s		m 5
3	i	s	1 0 1 2	p 6
4	i	m		s 8
5	m	#		
6	p	p		
7	p	i	2 1 2 2	
8	s	s		
9	s	s		
10	s	i		
11	s	i	4 1 2 4	

To reduce space

RLFM Index (Revisit)

RLFM Index (Derive B' from LF)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	
1	a	c	a	0	1	1	c	a	a	0
2	a	c	c	3	2	0	a	a	c	2
3	a	c	g	6	3	0	g	c	g	3
4	c	a	t	8	4	1	a	g	t	4
5	c	a			5	0	t	t		
6	c	g			6	1				
7	g	g			7	0				
8	g	a			8	1				
9	t	t			9	1				
10	t	t			10	0				

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	a	c	a	0	1	1	c	a	a	0
2	a	c	c	3	2	0	a	a	c	2
3	a	c	g	6	3	0	g	c	g	3
4	c	a	t	8	4	1	a	g	t	4
5	c	a			5	0	t	t		1
6	c	g			6	1				0
7	g	g			7	0				0
8	g	a			8	1				
9	t	t			9	1				
10	t	t			10	0				

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	a	c	a	0	1	1	c	a	a 0	1
2	a	c	c	3	2	0	a	a	c 2	0
3	a	c	g	6	3	0	g	c	g 3	
4	c	a	t	8	4	1	a	g	t 4	1
5	c	a			5	0	t	t		0
6	c	g			6	1				0
7	g	g			7	0				
8	g	a			8	1				
9	t	t			9	1				
10	t	t			10	0				

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	a	c	a	0	1	1	c	a	a	0	1
2	a	c	c	3	2	0	a	a	c	2	0
3	a	c	g	6	3	0	g	c	g	3	
4	c	a	t	8	4	1	a	g	t	4	1
5	c	a			5	0	t	t			0
6	c	g			6	1					0
7	g	g			7	0					1
8	g	a			8	1					0
9	t	t			9	1					
10	t	t			10	0					

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	a	c	a	0	1	1	c	a	a 0	1
2	a	c	c	3	2	0	a	a	c 2	0
3	a	c	g	6	3	0	g	c	g 3	1
4	c	a	t	8	4	1	a	g	t 4	1
5	c	a			5	0	t	t		0
6	c	g			6	1				0
7	g	g			7	0				1
8	g	a			8	1				0
9	t	t			9	1				
10	t	t			10	0				

RLFM Index (no L & F, nor LF)

	<u>B</u>	<u>S</u>	<u>B'</u>
1	1	c	
2	0	a	
3	0	g	
4	1	a	
5	0	t	
6	1		
7	0		
8	1		
9	1		
10	0		

If only B and S are stored and given... then how ???

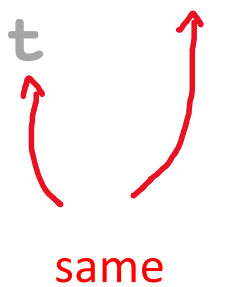
RLFM Index (no L & F, nor LF)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>B'</u>
1	1	c	a	
2	0	a	a	
3	0	g	c	
4	1	a	g	
5	0	t	t	
6	1			
7	0			
8	1			
9	1			
10	0			

If only B and S are stored and given... then how ???

RLFM Index (no L & F, nor LF)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	1	c	a	a	0
2	0	a	a	c	2
3	0	g	c	g	3
4	1	a	g	t	4
5	0	t	t		
6	1				
7	0				
8	1				
9	1				
10	0				



same

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	1	c	a	a	0
2	0	a	a	c	2
3	0	g	c	g	3
4	1	a	g	t	4
5	0	t	t		
6	1				
7	0				
8	1				
9	1				
10	0				

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	1	c	a	a	0
2	0	a	a	c	2
3	0	g	c	g	3
4	1	a	g	t	4
5	0	t	t		
6	1				
7	0				
8	1				
9	1				
10	0				

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	1	c	a	a	0
2	0	a	a	c	2
3	0	g	c	g	3
4	1	a	g	t	4
5	0	t	t		
6	1				
7	0				
8	1				
9	1				
10	0				

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>	<u>B'</u>
1	1	c	a	a	0
2	0	a	a	c	2
3	0	g	c	g	3
4	1	a	g	t	4
5	0	t	t		
6	1				
7	0				
8	1				
9	1				
10	0				

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a 0		1
2	0	a	a	c 2		0
3	0	g	c	g 3		
4	1	a	g	t 4		
5	0	t	t			
6	1					
7	0					
8	1					
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	
4	1	a	g	t	4	
5	0	t	t			
6	1					
7	0					
8	1					
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	
5	0	t	t			
6	1					
7	0					
8	1					
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	
5	0	t	t			
6	1					
7	0					
8	1					
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	
5	0	t	t			
6	1					
7	0					
8	1					
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					
8	1					
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					
8	1					
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					
10	0					

RLFM Index (No LF mapping)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

Now we have B, S, **B'**
Let's **reverse (decode)** using LF mapping

CHANGES TO FORMULAS

- Recall that we need to compute $C_T[c] + \text{rank}_c(L.i)$ in the backward search.
- Theorem:** $C[c] + \text{rank}_c(L.i)$ is equivalent to $\text{select}_1(B', C_S[c] + 1 + \text{rank}_c(S, \text{rank}_1(B.i))) - 1$, when $L[i] \neq c$ (e.g., when backward search), and otherwise (e.g., when reverse, sometimes backward search too) to $\text{select}_1(B', C_S[c] + \text{rank}_c(S, \text{rank}_1(B.i))) + i - \text{select}_1(B, \text{rank}_1(B.i))$.

You can apply these formulas to do reversing & backward search.

CHANGES TO FORMULAS

- Recall that we need to compute $C_T[c] + \text{rank}_c(L.i)$ in the backward search.
- Theorem:** $C[c] + \text{rank}_c(L.i)$ is equivalent to $\text{select}_1(B', C_S[c] + 1 + \text{rank}_c(S, \text{rank}_1(B.i))) - 1$, when $L[i] \neq c$ (e.g., when backward search), and otherwise (e.g., when reverse, sometimes backward search too) to $\text{select}_1(B', C_S[c] + \text{rank}_c(S, \text{rank}_1(B.i))) + i - \text{select}_1(B, \text{rank}_1(B.i))$.

But I promised that I would explain why/how these formulas actually work

RLFM Index (con't from the prev lecture)


	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	a	c	a	0	1	1	c	a	a	0	1
2	a	c	c	3	2	0	a	a	c	2	0
3	a	c	g	6	3	0	g	c	g	3	1
4	c	a	t	8	4	1	a	g	t	4	1
5	c	a			5	0	t	t			0
6	c	g			6	1					0
7	g	g			7	0					1
8	g	a			8	1					0
9	t	t			9	1					1
10	t	t			10	0					0

Suppose reverse from L[8]

RLFM Index (con't from the prev lecture)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$\text{rank}_{\underline{a}}(\text{S}, \text{rank}_1(\text{B}, 8))$
 $= 2$



RLFM Index (con't from the prev lecture)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$\text{rank}_{\underline{a}}(\text{S}, \text{rank}_1(\text{B}, 8))$
 $= 2$

RLFM Index (con't from the prev lecture)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$\text{select}_1(B', C_s[a] + \text{rank}_{\underline{a}}(S, \text{rank}_1(B, 8)))$

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	a	c	a	0	1	1	c	a	a	0	1
2	a	c	c	3	2	0	a	a	c	2	0
3	a	c	g	6	<u>3</u>	0	g	c	g	3	1
4	c	a	t	8	4	1	a	g	t	4	1
5	c	a			5	0	t	t			0
6	c	g			6	1					0
7	g	g			7	0					1
8	g	a			8	1					0
9	t	t			9	1					1
10	t	t			10	0					0

$$\text{select}_1(B', C_s[a] + \text{rank}_{\underline{a}}(S, \text{rank}_1(B, 8))) = 3$$

Good, but not good enough

RLFM Index (con't from the prev lecture)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$$\begin{aligned} & \text{select}_1(B', C_s[c] + \text{rank}_c(S, \text{rank}_1(B, 3))) \\ &= \text{select}_1(B', 2 + 1) = 4 \end{aligned}$$

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	a	c	a	0		1	c	a	a	0	1
2	a	c	c	3		0	a	a	c	2	0
3	a	c	g	6		0	g	c	g	3	1
4	c	a	t	8		1	a	g	t	4	1
5	c	a				0	t	t			0
<u>6</u>	c	g				1					0
7	g	g				0					1
8	g	a				1					0
9	t	t				1					1
10	t	t				0					0

$$\text{select}_1(B', C_s[c] + \text{rank}_c(S, \text{rank}_1(B, 3)))$$

$$= \text{select}_1(B', 2 + 1) = \underline{4} \quad ?$$

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	a	c	a	0	1	1	c	a	a	0	1
2	a	c	c	3	2	0	a	a	c	2	0
3	a	c	g	6	3	0	g	c	g	3	1
4	c	a	t	8	4	1	a	g	t	4	1
5	c	a			5	0	t	t			0
6	c	g			6	1					0
7	g	g			7	0					1
8	g	a			8	1					0
9	t	t			9	1					1
10	t	t			10	0					0

$$\begin{aligned}
 & \text{select}_1(B', C_s[c] + \text{rank}_c(S, \text{rank}_1(B, 3))) \\
 &= \text{select}_1(B', 2 + 1) = 4 + 2
 \end{aligned}$$

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	a	c	a	0	1	1	c	a	a	0	1
2	a	c	c	3	2	0	a	a	c	2	0
3	a	c	g	6	3	0	g	c	g	3	1
4	c	a	t	8	4	1	a	g	t	4	1
5	c	a			5	0	t	t			0
6	c	g			6	1					0
7	g	g			7	0					1
8	g	a			8	1					0
9	t	t			9	1					1
10	t	t			10	0					0

$$\begin{aligned}
 & \text{select}_1(B', C_s[c] + \text{rank}_c(S, \text{rank}_1(B, 3))) \\
 &= \text{select}_1(B', 2 + 1) = 4 + (i - \text{rank}_1(B, i))
 \end{aligned}$$

Another example, $LF[5] = ?$

RLFM Index (con't from the prev lecture)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$$\begin{aligned}
 & \text{select}_1(B', C_s[a] + \text{rank}_a(S, \text{rank}_1(B, 5))) \\
 &= \text{select}_1(B', 0 + 1) = 1 + (i - \text{rank}_1(B, i))
 \end{aligned}$$

RLFM Index (con't from the prev lecture)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$$\begin{aligned} & \text{select}_1(B', C_s[a] + \text{rank}_a(S, \text{rank}_1(B, 5))) \\ &= \text{select}_1(B', 0 + 1) = 1 + (i - \text{rank}_1(B, i)) \end{aligned}$$

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$$\begin{aligned} & \text{select}_1(B', C_s[a] + \text{rank}_a(S, \text{rank}_1(B, 5))) \\ &= \text{select}_1(B', 0 + 1) = 1 + (i - \text{rank}_1(B, i)) \end{aligned}$$

RLFM Index (con't from the prev lecture)

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$$\begin{aligned}
 & \text{select}_1(B', C_s[a] + \text{rank}_a(S, \text{rank}_1(B, 5))) \\
 &= \text{select}_1(B', 0 + 1) = 1 + \text{rank}_1(B, 5) - \text{select}_1(B, \text{rank}_1(B, 5))
 \end{aligned}$$

	<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	1	c	a	a	0 →	1
2	0	a	a	c	2	0
3	0	g	c	g	3	1
4	1	a	g	t	4	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					1
10	0					0

$$\begin{aligned} & \text{select}_1(B', C_s[a] + \text{rank}_a(S, \text{rank}_1(B, 5))) \\ &= \text{select}_1(B', 0 + 1) = 1 + (i - \text{select}_1(B, \text{rank}_1(B, i))) \end{aligned}$$

RLFM Index (con't from the prev lecture)

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>Fs</u>	<u>Cs</u>		<u>B'</u>
1	a	c	a	0	1	1	c	a	a	0	1
2	a	c	c	3	2	0	a	a	c	2	0
3	a	c	g	6	3	0	g	c	g	3	1
4	c	a	t	8	4	1	a	g	t	4	1
5	c	a			5	0	t	t			0
6	c	g			6	1					0
7	g	g			7	0					1
8	g	a			8	1					0
9	t	t			9	1					1
10	t	t			10	0					0

$$\begin{aligned}
 & \text{select}_1(B', C_s[a] + \text{rank}_a(S, \text{rank}_1(B, 5))) \\
 &= \text{select}_1(B', 0 + 1) = 1 + (i - \text{select}_1(B, \text{rank}_1(B, i))) \\
 & \quad \quad \quad 1 + 5 - 4
 \end{aligned}$$

RLFM Index (con't from the prev lecture)

CHANGES TO FORMULAS

- Recall that we need to compute $C_T[c] + \text{rank}_c(L.i)$ in the backward search.
- Theorem:** $C[c] + \text{rank}_c(L.i)$ is equivalent to $\text{select}_1(B', C_s[c] + 1 + \text{rank}_c(S, \text{rank}_1(B.i))) - 1$, when $L[i] \neq c$ (e.g., when backward search), and otherwise (e.g., when reverse, sometimes backward search too) to $\text{select}_1(B', C_s[c] + \text{rank}_c(S, \text{rank}_1(B.i))) + i - \text{select}_1(B, \text{rank}_1(B.i))$.

$$\begin{aligned} & \text{select}_1(B', C_s[a] + \text{rank}_a(S, \text{rank}_1(B, 5))) \\ &= \text{select}_1(B', 0 + 1) = 1 + (i - \text{select}_1(B, \text{rank}_1(B, i))) \end{aligned}$$

Backward Search

Backward search for “si”

	<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>
1	1	i	#	#	0	1
2	1	p	i	i	1	1
3	1	s	i	m	4	1
4	0	m	i	p	5	1
5	1	#	m	s	7	0
6	1	p	p			1
7	1	i	p			1
8	1	s	s			1
9	1	i	s			1
10	0					0
11	1					1
12	0					0

Backward search for “si”

	<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>
1	1	i	#	#	0	1
2	1	p	i	i	1	1
3	1	s	i	m	4	1
4	0	m	i	p	5	1
5	1	#	m	s	7	0
6	1	p	p			1
7	1	i	p			1
8	1	s	s			1
9	1	i	s			1
10	0					0
11	1					1
12	0					0

Backward search for “si”

	<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	1	i	#	#	0	1	
2	1	p	i	i	1	1	
3	1	s	i	m	4	1	
4	0	m	i	p	5	1	
5	1	#	m	s	7	0	
6	1	p	p			1	c = i
7	1	i	p			1	Fst = 2
8	1	s	s			1	Lst = 5
9	1	i	s			1	
10	0					0	
11	1					1	
12	0					0	

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>		
	1	#	i	#	0	1	1	i	#	#	0	1	
➡	2	i	p	i	1	2	1	p	i	i	1	1	
	3	i	s	m	5	3	1	s	i	m	4	1	
	4	i	s	p	6	4	0	m	i	p	5	1	
➡	5	i	m	s	8	5	1	#	m	s	7	0	
	6	m	#			6	1	p	p			1	c = i
	7	p	p			7	1	i	p			1	Fst = 2
	8	p	i			8	1	s	s			1	Lst = 5
	9	s	s			9	1	i	s			1	
	10	s	s			10	0					0	
	11	s	i			11	1					1	
	12	s	i			12	0					0	

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>		<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>
1	#	i	#	0	1	i	#	#	0	1
→ 2	i	p	i	1	2	p	i	i	1	1
3	i	s	m	5	3	s	i	m	4	1
4	i	s	p	6	4	m	i	p	5	1
→ 5	i	m	s	8	5	#	m	s	7	0
6	m	#			6	p	p			1
7	p	p			7	i	p			1
8	p	i			8	s	s			1
9	s	s			9	i	s			1
10	s	s			10					0
11	s	i			11					1
12	s	i			12					0

$c = s$
 $Fst =$
 $C[c] + Occ(c,$
 $Fst - 1) + 1$
 $= ?$

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0		1	i	#	#	0	1	
→ 2	i	p	i	1	$F_r \rightarrow$	2	p	i	i	1	1	
3	i	s	m	5		3	s	i	m	4	1	
4	i	s	p	6		4	m	i	p	5	1	
→ 5	i	m	s	8	$L_r \rightarrow$	5	#	m	s	7	0	
6	m	#				6	p	p			1	
7	p	p				7	i	p			1	
8	p	i				8	s	s			1	
9	s	s				9	i	s			1	
10	s	s				10					0	
11	s	i				11					1	
12	s	i				12					0	

$c = i$
 $Fst = 2$
 $Lst = 5$

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>		<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	i	#	#	0	1	
→ 2	i	p	i	1	2	p	i	i	1	1	
3	i	s	m	5	3	s	i	m	4	1	
4	i	s	p	6	4	m	i	p	5	1	
→ 5	i	m	s	8	5	#	m	s	7	0	
6	m	#			6	p	p			1	c = s
7	p	p			7	i	p			1	Fst = ??
8	p	i			8	s	s			1	
9	s	s			9	i	s			1	
10	s	s			10					0	
11	s	i			11					1	
12	s	i			12					0	

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	1	i	#	#	0	1	
→ 2	i	p	i	1	2	1	p	i	i	1	1	
3	i	s	m	5	3	1	s	i	m	4	1	
4	i	s	p	6	4	0	m	i	p	5	1	
→ 5	i	m	s	8	5	1	#	m	s	7	0	
6	m	#			6	1	p	p			1	c = s
7	p	p			7	1	i	p			1	<u>Fst</u>
8	p	i			8	1	s	s			1	Occ of s:
9	s	s			9	1	i	s			1	rank _s (S,
10	s	s			10	0					0	rank ₁ (B,2-1))
11	s	i			11	1					1	= 0
12	s	i			12	0					0	

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	1	i	#	#	0	1	
→ 2	i	p	i	1	2	1	p	i	i	1	1	
3	i	s	m	5	3	1	s	i	m	4	1	
4	i	s	p	6	4	0	m	i	p	5	1	
→ 5	i	m	s	8	5	1	#	m	s	7	0	
6	m	#			6	1	p	p			1	c = s
7	p	p			7	1	i	p			1	<u>Fst</u>
8	p	i			8	1	s	s			1	Occ of s:
9	s	s			9	1	i	s			1	$\text{rank}_s(S,$
10	s	s			10	0					0	$\text{rank}_1(B, 2-1))$
11	s	i			11	1					1	= 0
12	s	i			12	0					0	$\text{select}_1(B', 7+$
												$1+0)$
												So Fst = 9

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	1	i	#	#	0	1	
2	i	p	i	1	2	1	p	i	i	1	1	
3	i	<u>s</u>	m	5	3	1	s	i	m	4	1	
4	i	s	p	6	4	0	m	i	p	5	1	
5	i	m	s	8	5	1	#	m	s	7	0	
6	m	#			6	1	p	p			1	c = s
7	p	p			7	1	i	p			1	<u>Fst</u>
8	p	i			8	1	s	s			1	Occ of s:
9	<u>s</u>	s			9	1	i	s			1	$\text{rank}_s(S,$
10	s	s			10	0					0	$\text{rank}_1(B, 2-1))$
11	s	i			11	1					1	= 0
12	s	i			12	0					0	$\text{select}_1(B', 7+$
												$1+0)$
												So Fst = 9

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	1	i	#	#	0	1	
2	i	p	i	1	2	1	p	i	i	1	1	c = s
3	i	s	m	5	3	1	s	i	m	4	1	<u>Lst</u>
4	i	s	p	6	4	0	m	i	p	5	1	Occ of s:
5	i	m	s	8	5	1	#	m	s	7	0	$\text{rank}_s(S,$
6	m	#			6	1	p	p			1	$\text{rank}_1(B,5))$
7	p	p			7	1	i	p			1	= 1
8	p	i			8	1	s	s			1	
9	s	s			9	1	i	s			1	
10	s	s			10	0					0	
11	s	i			11	1					1	
12	s	i			12	0					0	

Backward search for “si”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	1	i	#	#	0	1	
2	i	p	i	1	2	1	p	i	i	1	1	c = s
3	i	s	m	5	3	1	s	i	m	4	1	<u>Lst</u>
4	i	s	p	6	4	0	m	i	p	5	1	Occ of s:
5	i	m	s	8	5	1	#	m	s	7	0	$\text{rank}_s(S,$
6	m	#			6	1	p	p			1	$\text{rank}_1(B,5))$
7	p	p			7	1	i	p			1	= 1
8	p	i			8	1	s	s			1	$\text{select}_1(B', 7+$
9	s	s			9	1	i	s			1	$1+1) = 11$
10	s	s			10	0					0	$11 - 1 = 10$
11	s	i			11	1					1	So Lst = 10
12	s	i			12	0					0	

-1: since
inclusively,
e.g., $\text{Lst} - \text{Fst} + 1$
= #matches

Backward search for “ssi”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>
1	#	i	#	0	1	1	i	#	#	0	1
2	i	p	i	1	2	1	p	i	i	1	1
3	i	s	m	5	3	1	s	i	m	4	1
4	i	s	p	6	4	0	m	i	p	5	1
5	i	m	s	8	5	1	#	m	s	7	0
6	m	#			6	1	p	p			1
7	p	p			7	1	i	p			1
8	p	i			8	1	s	s			1
9	s	s			9	1	i	s			1
10	s	s			10	0					0
11	s	i			11	1					1
12	s	i			12	0					0

c = s

Fst

Occ of s:

$\text{rank}_s(S,$
 $\text{rank}_1(B, 9-1))$

= 1

$\text{select}_1(B', 7+$
 $1+1)$

So Fst = 11

Backward search for “ssi”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	1	i	#	#	0	1	$c = s$
2	i	p	i	1	2	1	p	i	i	1	1	<u>Lst</u>
3	i	s	m	5	3	1	s	i	m	4	1	Occ of s:
4	i	s	p	6	4	0	m	i	p	5	1	$\text{rank}_s(S,$
5	i	m	s	8	5	1	#	m	s	7	0	$\text{rank}_1(B, 10))$
6	m	#			6	1	p	p			1	$= 2$
7	p	p			7	1	i	p			1	Since $L[i]=c$,
8	p	i			8	1	s	s			1	$\text{select}_1(B', C_s[c] +$
9	s	s			9	1	i	s			1	$\text{rank}_c(S, \text{rank}_1(B,$
10	s	s			10	0					0	$i))) + i -$
11	s	i			11	1					1	$\text{select}_1(B, \text{rank}_1(B, i)).$
12	s	i			12	0					0	$\text{select}_1(B', 7+2)$
												$= 11$
												$11 + 1 = 12$ So
												$\text{Lst} = 12$

Backward search for “issi”

	<u>F</u>	<u>L</u>	<u>C</u>			<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	1	i	#	#	0	1	
2	i	p	i	1	2	1	p	i	i	1	1	
3	i	s	m	5	3	1	s	i	m	4	1	
4	i	s	p	6	4	0	m	i	p	5	1	
5	i	m	s	8	5	1	#	m	s	7	0	c = i
6	m	#			6	1	p	p			1	<u>Fst</u>
7	p	p			7	1	i	p			1	Occ of i:
8	p	i			8	1	s	s			1	rank _s (S,
9	s	s			9	1	i	s			1	rank ₁ (B, 11-
10	s	s			10	0					0	1))
11	s	i			11	1					1	= 2
12	s	i			12	0					0	select ₁ (B', 1+
												1+2)
												So Fst = 4

Backward search for “issi”

	<u>F</u>	<u>L</u>	<u>C</u>		<u>B</u>	<u>S</u>	<u>F_s</u>	<u>C</u>		<u>B'</u>	
1	#	i	#	0	1	i	#	#	0	1	$c = i$
2	i	p	i	1	2	p	i	i	1	1	Lst
3	i	s	m	5	3	s	i	m	4	1	Occ of i:
→ 4	i	s	p	6	4	m	i	p	5	1	$\text{rank}_s(S,$
→ 5	i	m	s	8	5	#	m	s	7	0	$\text{rank}_1(B, 12))$
6	m	#			6	p	p			1	$= 3$
7	p	p			7	i	p			1	Since $L[i]=c$,
8	p	i			8	s	s			1	$\text{select}_1(B', C_s[c] +$
9	s	s			9	i	s			1	$\text{rank}_c(S, \text{rank}_1(B,$
10	s	s			10					0	$i))) + i -$
11	s	i			11					1	$\text{select}_1(B, \text{rank}_1(B, i)).$
12	s	i			12					0	$\text{select}_1(B', 1+3)$
										1	$= 4$
										0	$4 + 1 = 5$
										0	So Lst = 5

Therefore ...

CHANGES TO FORMULAS

- Recall that we need to compute $C_T[c] + \text{rank}_c(L, i)$ in the backward search.
- Theorem:** $C[c] + \text{rank}_c(L, i)$ is equivalent to $\text{select}_1(B', C_S[c] + 1 + \text{rank}_c(S, \text{rank}_1(B, i))) - 1$, when $L[i] \neq c$ (e.g., when backward search), and otherwise (e.g., when reverse, sometimes backward search too) to $\text{select}_1(B', C_S[c] + \text{rank}_c(S, \text{rank}_1(B, i))) + i\text{-select}_1(B, \text{rank}_1(B, i))$.