COMP9319 Web Data Compression and Search

An Occ Implementation,

RLFM (Compressed FM Index) Revisit

An example Occ implementation

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

	F	L	c
0	#	i	0
1	i	p	0
2	i	s	0
3	i	s	1
4	i	m	0
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

C	
#	0
i	1
m	5
p	6
s	8

FM Index (when reversing from L[5])

	F	L	c		C	
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])

	F	L	c
0	#	i	0
1	i	p	0
2	i	s	0
3	i	s	1
4	i	m	0
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

C	
#	0
i	1
m	5
p	6
s	8

LF[5] = 0+0 = 0, i
LF[0] = 1+0 = 1, p
LF[1] = 6+0 = 6, p
LF[6] = 6+1 = 7, i
LF[7] = 1+1 = 2, s
LF[2] = 8+0 = 8, s
LF[8] = 8+2 = 10, i
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 ($L(x) \neq c$)

	F	L	i	m	p	s
0	#	i	1	0	0	0
1	i	p	1	0	1	0
2	i	s	1	0	1	1
3	i	s	1	0	1	2
4	i	m	1	1	1	2
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

C	
#	0
i	1
m	5
p	6
s	8

FM Index (con't)

		F	L	i	m	p	s	C	
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		
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)

pssi

	F	L	i	m	p	s
0	#	i	1	0	0	0
1	i	p	1	0	1	0
2	i	s	1	0	1	1
3	i	s	1	0	1	2
4	i	m	1	1	1	2
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

C	
#	0
i	1
m	5
p	6
s	8

Fst=8+0

Lst=(8+2)-1

FM Index (con't)

		F	L	i	m	p	s	C	
pss<u>si</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+0

Lst=(8+2)-1

FM Index (con't)

pssi

	F	L	i	m	p	s
0	#	i	1	0	0	0
1	i	p	1	0	1	0
2	i	s	1	0	1	1
3	i	s	1	0	1	2
4	i	m	1	1	1	2
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

C	
#	0
i	1
m	5
p	6
s	8

Fst=8+2

Lst=(8+4)-1

FM Index (con’t)

p<u>ssi</u>

	F	L	i	m	p	s
0	#	i	1	0	0	0
1	i	p	1	0	1	0
2	i	s	1	0	1	1
3	i	s	1	0	1	2
4	i	m	1	1	1	2
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

C	
#	0
i	1
m	5
p	6
s	8

Fst=8+2

Lst=(8+4)-1

FM Index (con't)

pssi

	F	L	i	m	p	s
0	#	i	1	0	0	0
1	i	p	1	0	1	0
2	i	s	1	0	1	1
3	i	s	1	0	1	2
4	i	m	1	1	1	2
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

C

#	0
i	1
m	5
p	6
s	8

Fst=6+2

Lst=(6+2)-1

FM Index (con't)

pssi

	F	L	i	m	p	s
0	#	i	1	0	0	0
1	i	p	1	0	1	0
2	i	s	1	0	1	1
3	i	s	1	0	1	2
4	i	m	1	1	1	2
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

C	
#	0
i	1
m	5
p	6
s	8

Fst=6+2

Lst=(6+2)-1

Fst > Lst => No match

FM Index (con't)

pssi

	F	L	i m p s	C
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)

pssi

	F	L	i	m	p	s	C	
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

	F	L	i m p s	C
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)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs	
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 (con’t from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
1	1	c	a	a	0	
2	0	a	a	c	2	
3	0	g	c	g	3	
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 (con’t from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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	1
5	0	t	t			0
6	1					0
7	0					
8	1					
9	1					
10	0					

RLFM Index (con’t from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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	1
5	0	t	t			0
6	1					0
7	0					1
8	1					0
9	1					
10	0					

RLFM Index (con’t from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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 L & F, nor LF)

	B	S	B′
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)

	B	S	Fs	B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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)

	B	S	Fs	Cs		B′
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 $C_T[c]+rank_c(L,i)$ in the backward search.
- **Theorem:** $C[c]+rank_c(L,i)$ is equivalent to $select_1(B',C_S[c]+1+rank_c(S,rank_1(B,i)))-1$, when $L[i]\neq c$ (e.g., when backward search) , and otherwise (e.g., when reverse, sometimes backward search too) to $select_1(B',C_S[c]+rank_c(S,rank_1(B,i)))+i-select_1(B,rank_1(B,i))$.

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

CHANGES TO FORMULAS

- Recall that we need to compute $C_T[c]+rank_c(L,i)$ in the backward search.
- **Theorem:** $C[c]+rank_c(L,i)$ is equivalent to $select_1(B',C_S[c]+1+rank_c(S,rank_1(B,i)))-1$, when $L[i]\neq c$ (e.g., when backward search) , and otherwise (e.g., when reverse, sometimes backward search too) to $select_1(B',C_S[c]+rank_c(S,rank_1(B,i)))+ i-select_1(B,rank_1(B,i))$.

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

RLFM Index (con’t from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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

Suppose reverse from L[8]

RLFM Index (con't from the prev lecture)

	B	**S**	Fs	Cs		**B'**
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

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

RLFM Index (con’t from the prev lecture)

	B	S	Fs	Cs		B′
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

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

RLFM Index (con’t from the prev lecture)

	B	S	Fs	Cs		B′
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

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

RLFM Index (con’t from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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

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

Good, but not good enough

RLFM Index (con’t from the prev lecture)

	B	S	Fs	Cs		B’
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

$\mathrm{select}_1(B', C_s[c]+\mathrm{rank}_{\underline{c}}(S, \mathrm{rank}_1(B, 3)))$
$=\mathrm{select}_1(B', 2 + 1)=4$

RLFM Index (con't from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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

$\text{select}_1(B', C_s[c]+\text{rank}_{\underline{c}}(S, \text{rank}_1(B, 3)))$
$=\text{select}_1(B', 2 + 1) = \underline{4}$?

RLFM Index (con't from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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

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

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

RLFM Index (con’t from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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

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

$=\text{select}_1(B', 2 + 1) = 4 + (i - \text{rank}_1(B, i))$

Another example, LF[5] = ?

RLFM Index (con't from the prev lecture)

	B	S	Fs	Cs		B'
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

$\text{select}_1(B', C_s[a]+\text{rank}_{\underline{a}}(S, \text{rank}_1(B, 5)))$

$=\text{select}_1(B', 0 + 1\ )=1 + (i - \text{rank}_1(B, i))$

RLFM Index (con’t from the prev lecture)

	B	S	Fs	Cs		B′
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

$select_1(B', C_s[a]+rank_{\underline{a}}(S, rank_1(B, 5)))$

$=select_1(B', 0 + 1) = 1 + (i - rank_1(B, i))$

=3

RLFM Index (con't from the prev lecture)

	B	**S**	**Fs**	**Cs**		**B'**
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

$\text{select}_1(B', C_s[a]+\text{rank}_{\underline{a}}(S, \text{rank}_1(B, 5)))$

$=\text{select}_1(B', 0 + 1) = 1 + (i - \text{rank}_1(B, i))$ $= 3$

RLFM Index (con’t from the prev lecture)

	B	S	Fs	Cs		B′
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

$\text{select}_1(B', C_s[a]+\text{rank}_{\underline{a}}(S, \text{rank}_1(B, 5)))$

$=\text{select}_1(B', 0 + 1) = 1 + (i - \text{select}_1(B, \text{rank}_1(B, i)))$

RLFM Index (con't from the prev lecture)

	B	S	Fs	Cs		B'
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

$\text{select}_1(B', C_s[a]+\text{rank}_{\underline{a}}(S, \text{rank}_1(B, 5)))$

$=\text{select}_1(B', 0 + 1\ )=1 + (i - \text{select}_1(B, \text{rank}_1(B, i)))$

1 + 5 - 4

RLFM Index (con't from the prev lecture)

	F	L	C	
1	a	c	a	0
2	a	c	c	3
3	a	c	g	6
4	c	a	t	8
5	c	a		
6	c	g		
7	g	g		
8	g	a		
9	t	t		
10	t	t		

	B	S	Fs	Cs		B′
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

$\text{select}_1(B', C_s[a]+\text{rank}_{\underline{a}}(S, \text{rank}_1(B, 5)))$

$=\text{select}_1(B', 0 + 1) = 1 + (i - \text{select}_1(B, \text{rank}_1(B, i)))$

1 + 5 - 4

RLFM Index (con’t from the prev lecture)

CHANGES TO FORMULAS

- Recall that we need to compute $C_T[c]+rank_c(L,i)$ in the backward search.
- **Theorem:** $C[c]+rank_c(L,i)$ is equivalent to $select_1(B',C_S[c]+1+rank_c(S,rank_1(B,i)))-1$, when $L[i]\neq c$ (e.g., when backward search) , and otherwise (e.g., when reverse, sometimes backward search too) to $select_1(B',C_S[c]+rank_c(S,rank_1(B,i)))+i-select_1(B,rank_1(B,i))$.

$select_1(B', C_s[a]+rank_{\underline{a}}(S, rank_1(B, 5)))$

$=select_1(B', 0 + 1)=1 + (i -select_1(B, rank_1(B, i)))$

Backward Search

Backward search for "si"

	B	S	F_s	C		B′
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”

	B	S	F_s	C		B′
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”

	B	S	F_s	C		B′
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

c = i

Fst = 2

Lst = 5

Backward search for “si”

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

	B	S	F_S	C		B′
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

c = i

Fst = 2

Lst = 5

Backward search for "si"

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

	B	S	F_s	C		B′
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

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

Backward search for “si”

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

Fr

Lr

	B	S	F_s	C		B′
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

c = i
Fst = 2
Lst = 5

Backward search for “si”

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

	B	S	F_s	C		B′
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

c = s
Fst = ??

Backward search for “si”

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

	B	S	F_s	C		B′
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

c = s
Fst
Occ of s:
$\mathrm{rank}_s(S,$
$\mathrm{rank}_1(B,2\text{-}1))$
$= 0$

Backward search for “si”

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

	B	S	F_s	C		B′
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

c = s

<u>Fst</u>

Occ of s: $\text{rank}_s(S, \text{rank}_1(B,2-1))$ = 0

$\text{select}_1(B',7+1+0)$

So Fst = 9

Backward search for “si”

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

	B	S	F_s	C		B′
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

c = s

<u>Fst</u>

Occ of s: $\text{rank}_s(S, \text{rank}_1(B,2-1)) = 0$

$\text{select}_1(B',7+1+0)$

So Fst = 9

Backward search for “si”

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

	B	S	F_s	C		B′
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

c = s

Lst

Occ of s:

$\mathrm{rank}_s(S, \mathrm{rank}_1(B,5))$

= 1

Backward search for "si"

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

	B	S	F_S	C		B'
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

c = s
Lst
Occ of s:
$rank_s(S, rank_1(B,5))$
= 1
$select_1(B',7+1+1) = 11$
$11 - 1 = 10$
So Lst = 10

-1: since inclusively, e.g., Lst-Fst+1 = #matches

Backward search for “ssi”

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

	B	S	F_s	C		B′
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

c = s

Fst

Occ of s: $rank_s(S, rank_1(B, 9-1)) = 1$

$select_1(B', 7+1+1)$

So Fst = 11

Backward search for “ssi”

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

	B	S	F_S	C		B’
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

c = s

Lst

Occ of s: $rank_s(S, rank_1(B,10))$ = 1

Since L[i]=c, $select_1(B', C_S[c] + rank_c(S, rank_1(B, i))) + i - select_1(B, rank_1(B, i))$.

$select_1(B', 7+2) = 11$

$11 + 1 = 12$

So Lst = 12

Backward search for “issi”

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

	B	S	F_s	C		B’
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

c = i
Fst
Occ of i: $rank_s(S, rank_1(B, 11-1))$
= 2
$select_1(B’, 1+1+2)$
So Fst = 4

Backward search for “issi”

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

	B	S	F_S	C		B’
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

c = i

Lst

Occ of i: $rank_s(S, rank_1(B,12))$ = 3

Since L[i]=c, $select_1(B', C_S[c]+rank_c(S,rank_1(B, i)))+ i-select_1(B,rank_1(B,i))$.

$select_1(B’,1+3)$ = 4

4 + 1 = 5

So Lst = 5

Therefore …

CHANGES TO FORMULAS

- Recall that we need to compute $C_T[c]+rank_c(L,i)$ in the backward search.
- **Theorem:** $C[c]+rank_c(L,i)$ is equivalent to $select_1(B',C_S[c]+1+rank_c(S,rank_1(B,i)))-1$, when $L[i] \neq c$ (e.g., when backward search) , and otherwise (e.g., when reverse, sometimes backward search too) to $select_1(B',C_S[c]+rank_c(S,rank_1(B,i)))+ i-select_1(B,rank_1(B,i))$.