BIOINFORMATICS

How do we locate disease causing mutation?

Marco Beccuti

Università degli Studi di Torino

Dipartimento di Informatica

Exercise 1

Finding the Matched Patterns

Exercise

- For reference **bioinformatics** create:
  - BWT;
  - FirstOccurence;
  - CheckPointArray$_b$ with b=5;
  - SuffixArray$_a$ with a=5;
  - Search Pattern: **ioi**

Esercize

- For reference **bioinformatics** create:
  - BWT.

```
Text:
bioinformatics$

Rotations:
bioinformatics$
ioinformatics$b
oinformatics$bi
informatics$bio
nformatics$bioi
formatics$bioin
ormatics$bioinf
rmatics$bioinfo
matics$bioinfor
atics$bioinform
tics$bioinforma
ics$bioinformat
cs$bioinformati
s$bioinformatic
$bioinformatics
```

Esercize

- For reference **bioinformatics** create:
  - BWT.

```
Text:
bioinformatics$

Rotations:
bioinformatics$
ioinformatics$b
oinformatics$bi
informatics$bio
nformatics$bioi
formatics$bioin
ormatics$bioinf
rmatics$bioinfo
matics$bioinfor
atics$bioinform
tics$bioinforma
ics$bioinformat
cs$bioinformati
s$bioinformatic
$bioinformatics

Sorted Rotations:
$bioinformatics
atics$bioinform
bioinformatics$
cs$bioinformati
formatics$bioin
ics$bioinformat
informatics$bio
ioinformatics$b
matics$bioinfor
nformatics$bioi
oinformatics$bi
ormatics$bioinf
rmatics$bioinfo
s$bioinformatic
tics$bioinforma
```

Esercize

- For reference **bioinformatics** create:
  - BWT.

```
Text:
bioinformatics$

BWT matrix
$       bioinformatic   s
a       tics$bioinfor   m
b       ioinformatics   $
c       s$bioinformat   i
f       ormatics$bioi   n
i       cs$bioinforma   t
i       nformatics$bi   o
i       oinformatics$   b
m       atics$bioinfo   r
n       formatics$bio   i
o       informatics$b   i
o       rmatics$bioin   f
r       matics$bioinf   o
s       $bioinformati   c
t       ics$bioinform   a
```

Esercize

- For reference **bioinformatics** create:
  - FirstOccurence.

Text:
bioinformatics$

BWT matrix			FirstOccurence
$	bioinformatic	s	0
a	tics$bioinfor	m	1
b	ioinformatics	$	2
c	s$bioinformat	i	3
f	ormatics$bioi	n	4
i	cs$bioinforma	t	5
i	nformatics$bi	o	
i	oinformatics$	b	
m	atics$bioinfo	r	8
n	formatics$bio	i	9
o	informatics$b	i	10
o	rmatics$bioin	f	
r	matics$bioinf	o	12
s	$bioinformati	c	13
t	ics$bioinform	a	14

Esercize

- For reference **bioinformatics** create:
  - CheckPointArray$_b$ with b=5.

```
Text:
bioinformatics$
a=b=5

BWT matrix                          FirstOccurence                $ a b c f i m n o r s t
$       bioinformatic   s<--                 0                  0 0 0 0 0 0 0 0 0 0 0 0
a       tics$bioinfor   m                   1
b       ioinformatics   $                   2
c       s$bioinformat   i                   3
f       ormatics$bioi   n                   4
i       cs$bioinforma   t<--                5                  1 0 0 0 0 1 1 1 0 0 1 0
i       nformatics$bi   o
i       oinformatics$   b
m       atics$bioinfo   r                   8
n       formatics$bio   i                   9
o       informatics$b   i<--                10                 1 0 1 0 0 2 1 1 1 1 1 1
o       rmatics$bioin   f
r       matics$bioinf   o                   12
s       $bioinformati   c                   13
t       ics$bioinform   a                   14
```

Esercize

- For reference **bioinformatics** create:
  - Suffix Array$_a$ with a=5.

```
Text:
bioinformatics$
a=b=5

BWT matrix                    FirstOccurence          $ a b c f i m n o r s t        Suffix array
$       bioinformatic   s<--          0               0 0 0 0 0 0 0 0 0 0 0 0
a       tics$bioinfor   m            1
b       ioinformatics   $            2                                                    0
c       s$bioinformat   i            3
f       ormatics$bioi   n            4                                                    5
i       cs$bioinforma   t<--          5               1 0 0 0 0 1 1 1 0 0 1 0
i       nformatics$bi   o
i       oinformatics$   b
m       atics$bioinfo   r            8
n       formatics$bio   i            9
o       informatics$b   i<--         10              1 0 1 0 0 2 1 1 1 1 1 1
o       rmatics$bioin   f
r       matics$bioinf   o            12
s       $bioinformati   c            13
t       ics$bioinform   a            14                                                  10
```

Esercize

- For reference **bioinformatics** create:
  - Search Pattern: **ioi**

```
Text:
bioinformatics$
a=b=5
Search =ioi
Top=0 Bottom=14
        BWT matrix                      FirstOccurence          $ a b c f i m n o r s t         Suffix array
-->     $       bioinformatic   s               0               0 0 0 0 0 0 0 0 0 0 0 0
        a       tics$bioinfor   m               1
        b       ioinformatics   $               2                                                       0
        c       s$bioinformat   i               3
        f       ormatics$bioi   n               4                                                       5
        i       cs$bioinforma   t               5               1 0 0 0 0 1 1 1 0 0 1 0
        i       nformatics$bi   o
        i       oinformatics$   b
        m       atics$bioinfo   r               8
        n       formatics$bio   i               9
        o       informatics$b   i               10              1 0 1 0 0 2 1 1 1 1 1 1
        o       rmatics$bioin   f
        r       matics$bioinf   o               12
        s       $bioinformati   c               13
-->     t       ics$bioinform   a               14                                                      10

Top= FirstOccurance(i)+Count_i(0,sm$intobriifoca)=5
Bottom= FirstOccurance(i)+Count_i(15,sm$intobriifoca)-1=7
```

Esercize

- For reference **bioinformatics** create:
  - Search Pattern: **ioi**

Text:
bioinformatics$
a=b=5
Search =io->i
Top=5 Bottom=7

	BWT matrix			FirstOccurence	$	a	b	c	f	i	m	n	o	r	s	t	Suffix array
	$	bioinformatic	s	0	0	0	0	0	0	0	0	0	0	0	0	0	
	a	tics$bioinfor	m	1													
	b	ioinformatics	$	2													0
	c	s$bioinformat	i	3													
	f	ormatics$bioi	n	4													5
-->	i	cs$bioinforma	t	5	1	0	0	0	0	1	1	1	0	0	1	0	
	i	nformatics$bi	o														
-->	i	oinformatics$	b														
	m	atics$bioinfo	r	8													
	n	formatics$bio	i	9													
	o	informatics$b	i	10	1	0	1	0	0	2	1	1	1	1	1	1	
	o	rmatics$bioin	f														
	r	matics$bioinf	o	12													
	s	$bioinformati	c	13													
	t	ics$bioinform	a	14													10

Top= FirstOccurance(o)+Count_o(5,sm$intobriifoca)=10
Bottom= FirstOccurance(o)+Count_o(8,sm$intobriifoca)-1=10

Esercize

- For reference **bioinformatics** create:
  - Search Pattern: **ioi**

```
Text:
bioinformatics$
a=b=5
Search =i->oi
Top=10 Bottom=10
```

	BWT matrix		FirstOccurence	$ a b c f i m n o r s t	Suffix array
	$ bioinformatic	s	0	0 0 0 0 0 0 0 0 0 0 0	
	a tics$bioinfor	m	1		
	b ioinformatics	$	2		0
	c s$bioinformat	i	3		
	f ormatics$bioi	n	4		5
	i cs$bioinforma	t	5	1 0 0 0 0 1 1 1 0 0 1 0	
	i nformatics$bi	o			
	i oinformatics$	b			
	m atics$bioinfo	r	8		
	n formatics$bio	i	9		
-->	o informatics$b	i	10	1 0 1 0 0 2 1 1 1 1 1 1	
	o rmatics$bioin	f			
	r matics$bioinf	o	12		
	s $bioinformati	c	13		
	t ics$bioinform	a	14		10

```
Top= FirstOccurance(i)+Count_i(10,sm$intobriifoca)=7
Bottom= FirstOccurance(i)+Count_i(11,sm$intobriifoca)-1=7
```

Esercize

- For reference **bioinformatics** create:
  - Search Pattern: **ioi**

```
Text:
bioinformatics$
a=b=5
Search =->ioi
Top=7 Bottom=7
```

	BWT matrix		FirstOccurence	$ a b c f i m n o r s t	Suffix array
	$ bioinformatic	s	0	0 0 0 0 0 0 0 0 0 0 0	
	a tics$bioinfor	m	1		
	b ioinformatics	$	2		0
	c s$bioinformat	i	3		
	f ormatics$bioi	n	4		5
	i cs$bioinforma	t	5	1 0 0 0 0 1 1 1 0 0 1 0	
	i nformatics$bi	o			
-->	i oinformatics$	b			
	m atics$bioinfo	r	8		
	n formatics$bio	i	9		
	o informatics$b	i	10	1 0 1 0 0 2 1 1 1 1 1 1	
	o rmatics$bioin	f			
	r matics$bioinf	o	12		
	s $bioinformati	c	13		
	t ics$bioinform	a	14		