

IV121 Vybrané aplikace informatiky v biologii

Stringologie a podobnost řetězců

Katedra informačních technologií
Masarykova Univerzita Brno

Jaro 2012

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu hledaného motivu

Algoritmus využívající analýzu prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

DP - Needleman-Wunsch

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

- ▶ historie
- ▶ pojmy a objekty zájmu
- ▶ operace s řetězci
- ▶ vzdálenost a podobnost
- ▶ datové struktury
- ▶ vyhledávání řetězců
- ▶ praktické aplikace

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

- , 1960 Trie (prefixový strom)
- , 1968 Radixový strom
- , 1973 Suffixový strom
- , 1988 DAWG
- , 1990 Skiplist
- , 1991 Sufixové pole
- , 1994 BW transformace
- , 1995 On/line konstrukce uffixového stromu
- , 2000 FM-index

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

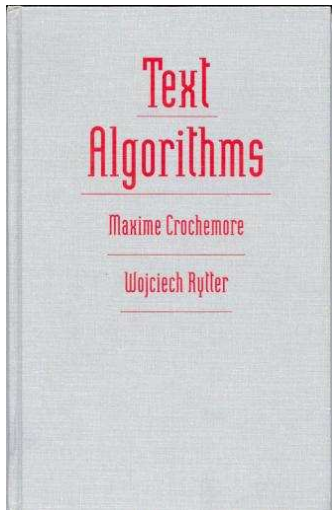
Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform



Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

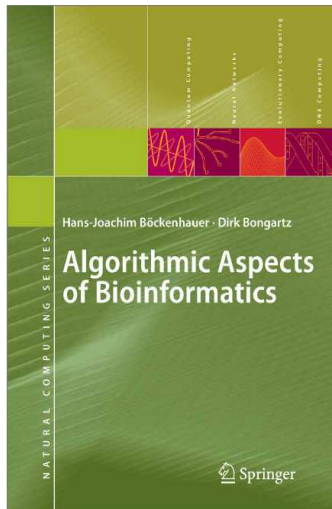
Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform



Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

abeceda	$\{\epsilon, a, c, g, t\}$
řetězec/sekvence	aaggtacgcgt
podsekvence	aag gtacg cgt
podřetězec	aag gtacg cgt
prefix	gtacg cgtggt
suffix	cgtat gtacg

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

konkatenace	$x=cggat$	$y=att$	$x.y=cggatatt$
průnik	$x=cggat$	$y=att$	$Over(x,y)=at$
sjednocení	$x=cggat$	$y=att$	$\langle x,y \rangle=cggatt$
projekce	$x=cnggatx$	$\Sigma=(\epsilon,a,c,g,t)$	$\Pi_{\Sigma}(x)=cggat$
oddělení (zprava/zleva)	$x=cggat$	$x/t=cgga$	
odstranění (zprava/zleva)	$x=cggat$	$x \div a=cggt$	

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

- ▶ seznam (prostý, skip-list)
- ▶ strom (trie, Radixový strom, suffixový strom, metrický strom)
- ▶ acyklický graf (DAWG, GADDAG)
- ▶ transformovaná pole (invertovaný soubor, suffixové pole, transf. Burrows-Wheeler, FM-index)
- ▶ automaty

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

- ▶ Editační a podobné vzdálenosti
 - ▶ Hammingova
 - ▶ Levenshtein
 - ▶ Damerau-Levenshtein
 - ▶ Jaro-Winkler
- ▶ Podobnost na bázi n-gramů
 - ▶ Jaccardův index, Diceův koeficient
 - ▶ kosínová vzdálenost
 - ▶ vzájemná informace
- ▶ Lee-ův koeficient
- ▶ Inverzní číslo
- ▶ Soundex
- ▶ Počet spol. n-gramů nebo slov
- ▶ LCS (nejdelší spol. podřetězec)

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximální k chyby

Burrows-Wheeler transform

- ▶ vyhledávání vzorů v jednom řetězci
- ▶ vyhledávání vzorů ve více řetězcích (LCP, LCS, SCS)
- ▶ vyhledávání přibližných vzorů
- ▶ komprese (slovníková, statistická)

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 3 chyby

Burrows-Wheeler transform

Výskyt sekvenčních motivů v genomech a databázích

Cílem je zjistit všechny pozice delšího řetězce, na kterých se vyskytuje kratší řetězec

- ▶ přesný výskyt
- ▶ přibližný výskyt
 - ▶ vzdálenost (k-NN, podobnostní práh)
 - ▶ vzorec (RE)

řetězec t dlouhý (n), např genomová sekvence
motiv p krátký (m), např `cgcgggctgggtggctcg`

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

```
a c t g t g t a t g a a a t c g c
1..n → t g t c a
      1..m →
```

Složitost: $O(mn)$

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 1 chybu

Burrows-Wheeler transform

a c t g t g t a t g a a a t c g c
1..n → t g t c a

$$h(1..m) = h(i..i+m-1)?$$

Porovnávání znaků je nahrazeno porovnáváním hodnot vhodné "hešovací" funkce. Např: hodnota ASCII znaků v prvočíselné bázi (101)
Složitost: $O(mn)$

```
a c t g t g t a t g a a a t c g c
  → g t g c c t
    ↑ ↑ ↑ ×
```

máme v motivu předchozí výskyt g, tg nebo gtg?

```
a c t g t g t a t g a a a t c g c
      +3 → g t g c c t
```

Složitost

konstrukce: $O(\|abeceda\|.m)$

hledání: $O(mn)$ (v praxi ale blíže k $O(n)$)

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 1: chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

a c t g t g t a t g a a a t c g c
→ g a t c a t
 x ↑ ↑ ←

máme v motivu další t?

a c t g t g t a t g a a a t c g c
+1 → g a **t** c a t

kde máme v motivu další výskyt suffixu at?

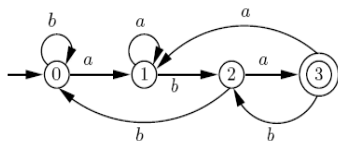
a c t g t g t a t g a a a t c g c
+3 → g **a t** c a t

Realizujeme krok, který je větší

Složitost

konstrukce: $O(\|abeceda\|.m)$

hledání: $O(mn)$ (v praxi ale blíže k $O(n)$)



Automat vytvořen z motivu p postupně čte symboly z řetězce m . Koncový stav automatu dosáhneme po načtení celého hledaného motivu.

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu hledaného motivu

Algoritmus využívající analýzu prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

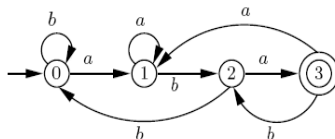
Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

t=bababaa p=aba

ε	0
b	0
ba	1
bab	2
baba	3
babab	2
bab aba	3
bababaa	1



Složitost

konstrukce: naivní $O(m^3)$; optimální $O(\|abeceda\|.m)$

hledání: $O(n)$

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

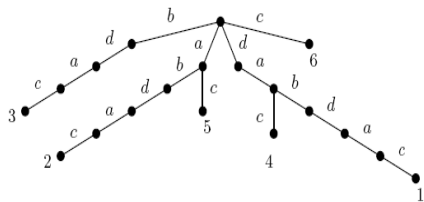
Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Suffixový strom pro řetězec dabdac



Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

**Algoritmus využívající analýzu
prohledávaného řetězce**

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Kompaktní suffixový strom pro řetězec

aaabbbc

Stringologie

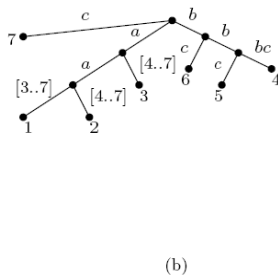
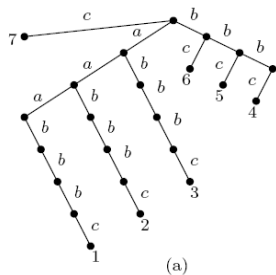
- Úvod
- Základní pojmy
- Základní algoritmy
- Algoritmus využívající analýzu hledaného motivu
- Algoritmus využívající analýzu prohledávaného řetězce**

Hledání opakování

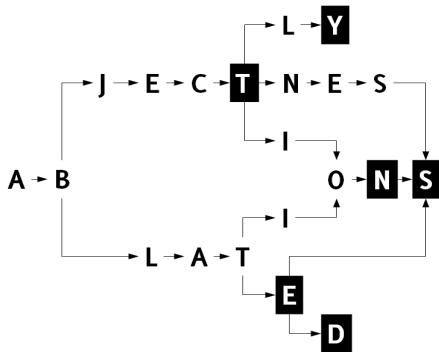
- Tandemové opakování
- Palindromy

Srovnávání dvou sekvencí

- Vylepšení pro maximálně k chyb
- Burrows-Wheeler transform



Konstrukce: $O(n \cdot \log n)$
Hledání: $O(m \cdot \|abceda\| + k)$



abject
 abjection
 abjections
 abjectly
 abjectness
 ablate
 ablated
 ablation
 ablations

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivuAlgoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 3 chyby

Burrows-Wheeler transform

Sufixové pole - ukazovatele na polohy suffixů seřazené lexikograficky

Dlouho bylo považováno za méně kvalitní datovou strukturu, protože neobsahuje přímo informace o společných prefixech. Ty lze však spočítat do lcp pole (least common prefix) tak, že konstrukce pole i stromu má stejnou složitost.

$t = \text{dabdac}$

$\text{sa}(t) = 6, 1, 4, 2, 5, 0, 3$

$\text{rank}(t) = 5, 1, 3, 6, 2, 4, 0$

$\text{lcp}(t) = 0, 0, 1, 0, 0, 0, 2$

6 0

1 0 abdac

4 1 ac

2 0 bdac

5 0 c

0 0 dabdac

3 2 dac

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu hledaného motivu

Algoritmus využívající analýzu prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

DP - Needleman-Wunsch

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Tandemová a palindromická opakování nesou biologický i praktický význam

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 1: chyb

Burrows-Wheeler transform

palindrom možná sekundární struktura DNA nebo RNA
tandem regulace genů, telomery, identifikace jedinců
z DNA

Nejdelší společný prefix dvou pozic

t g c a g a a g c a g a t c c t g a c g
↑ ↑

Složitost naivního algoritmu $O(n^3)$

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

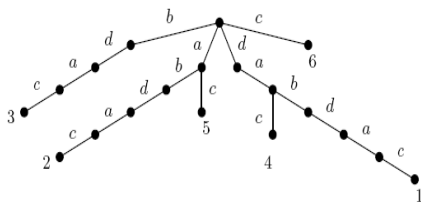
Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 3 chyby

Burrows-Wheeler transform

Posuzování tandemových opakování pomocí suffixových stromů, příp. polí (t=dabdac)



$\text{lcp}(1,4)=?$

Nalezneme větve označené 1 a 4

$\text{lcp}(1,4)=da$

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

- ▶ konstrukce stromu: $O(n \cdot \log n)$
- ▶ hledání lcp pro dvě konkrétní pozice $O(n \cdot \log n)$
- ▶ Prohledávání sekvence

Složitost: $O(n \cdot (\log n)^2 + p)$

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Palindromy - nejdelší společný prefix mezi originální a komplementární sekvencí umožňuje urychlení hledání podobně jako pro tandemové opakování

Stringologie

- Úvod
- Základní pojmy
- Základní algoritmy
- Algoritmus využívající analýzu hledaného motivu
- Algoritmus využívající analýzu prohledávaného řetězce

Hledání opakování

- Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

- Vylepšení pro maximálně k chyb
- Burrows-Wheeler transform

```

                    ↓ 8
t g c a g a a g c t t c t g t c t g a c g
a c g t c t t c g a a g a c a g a c t g c
                    ↑ 9*

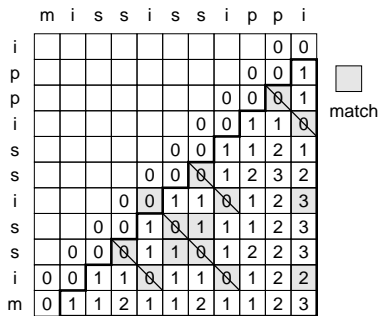
```

Složitost naivního algoritmu $O(n^3)$

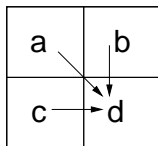
Složitost naivního algoritmu $O(nlp)$ (pro omezenou vzdálenost a délku)

Složitost s použitím suffixových struktur $O(n)$

Využití DP pro identifikaci palindromů



a)



$$d = \min \begin{cases} a & \text{match} \\ a+1 & \text{mismatch} \\ b+1 & \text{insertion} \\ c+1 & \text{deletion} \end{cases}$$

b)

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

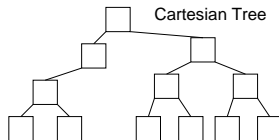
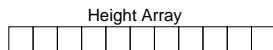
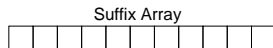
Využití SA a LCP k rychlému postupu po diagonále

	m	i	s	s	i	s	s	i	p	p	i
i										0	0
p									0	0	1
p								0	0	0	1
i						0	0	1	1	2	1
s					0	0	1	1	1	2	1
s				0	0	0	0	0	1	2	3
i			0	0	1	0	0	0	1	2	3
s		0	0	1	0	0	0	1	1	2	3
s	0	0	0	0	1	0	0	0	1	2	3
i	0	0	1	0	0	1	0	0	1	2	2
m	0	1	1	2	1	1	2	1	1	2	3

→ Longest Common Prefix

- - → Neighboring cells with worse scores

a)



b)

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu hledaného motivu

Algoritmus využívající analýzu prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

DP - Needleman-Wunsch

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Výpočet omezeného počtu buněk v tabulce DP

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Stačí počítat $2k+1$ diagonál bez ohledu na délku sekvencí

Složitost: $O(kn)$ (naproti $O(mn)$)

Tabulka pro algoritmus dynamického programování

Stringologie

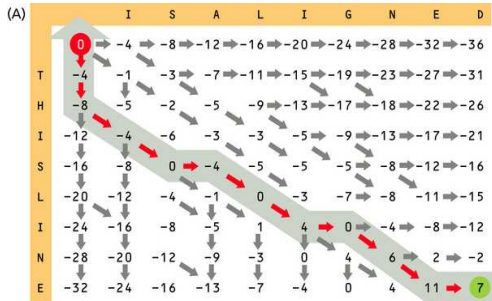
- Úvod
- Základní pojmy
- Základní algoritmy
- Algoritmus využívající analýzu hledaného motivu
- Algoritmus využívající analýzu prohledávaného řetězce

Hledání opakování

- Tandemové opakování
- Palindromy

Srovnávání dvou sekvencí

- Vylepšení pro maximálně k chyb
- Burrows-Wheeler transform



(B) THIS-LI-NE-
--ISALIGNED

Využití SA a LCP k rychlému postupu po diagonále

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Složitost: $O(k^2)$ (viz pou/vzit/'i pro palindromy)

BWT - Burrows-Wheeler transform

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 1 chybu

Burrows-Wheeler transform

Transformation				
Input	All Rotations	Sorting All Rows in Alphabetical Order by their first letters	Taking Last Column	Output Last Column
<code>^BANANA@</code>	<code>^BANANA@</code> <code>@^BANANA</code> <code>A@^BANAN</code> <code>NA@^BANA</code> <code>ANA@^BAN</code> <code>NANA@^BA</code> <code>ANANA@^B</code> <code>BANANA@^</code>	<code>ANANA@^B</code> <code>ANA@^BAN</code> <code>A@^BANAN</code> <code>BANANA@^</code> <code>NANA@^BA</code> <code>NA@^BANA</code> <code>^BANANA@</code> <code>@^BANANA</code>	<code>ANANA@^B</code> <code>ANA@^BAN</code> <code>A@^BANAN</code> <code>BANANA@^</code> <code>NANA@^BA</code> <code>NA@^BANA</code> <code>^BANANA@</code> <code>@^BANANA</code>	<code>BNN^AA@A</code>

Inverse Transformation			
Input			
BNN^AA@A			
Add 1	Sort 1	Add 2	Sort 2
B N N ^ A A @ A	A A A B N N ^ @	BA NA NA ^B AN AN @^ A@	AN AN A@ BA NA NA ^B @^

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivuAlgoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Add 7	Sort 7	Add 8	Sort 8
BANANA@ NANA@^B NA@^BAN ^BANANA ANANA@^ ANA@^BA @^BANAN A@^BANA	ANANA@^ ANA@^BA A@^BANA BANANA@ NANA@^B NA@^BAN ^BANANA @^BANAN	BANANA@^ NANA@^BA NA@^BANA ^BANANA@ ANANA@^B ANA@^BAN @^BANANA A@^BANAN	ANANA@^B ANA@^BAN A@^BANAN BANANA@^ NANA@^BA NA@^BANA ^BANANA@ @^BANANA
Output			
^BANANA@			

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivuAlgoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně k chyb

Burrows-Wheeler transform

Input	SIX.MIXED.PIXIES.SIFT.SIXTY.PIXIE.DUST.BOXES
Output	TEXYDST.E.IXIXIXSSMPPS.B..E.S.EUSFXDII OIIIT

Stringologie

Úvod

Základní pojmy

Základní algoritmy

Algoritmus využívající analýzu
hledaného motivu

Algoritmus využívající analýzu
prohledávaného řetězce

Hledání opakování

Tandemové opakování

Palindromy

Srovnávání dvou sekvencí

Vylepšení pro maximálně 1 chybu

Burrows-Wheeler transform