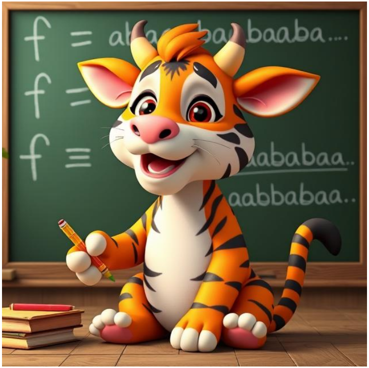
*Všechno, co jste kdy chtěli vědět o BWTu**

*(*ale báli jste se zeptat)*

Francesco Dolce



Janov nad Nisou, 31 května 2025

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

Slunce, seno, konjugáty

nose – seno

Slunce, seno, konjugáty

nose – seno

duha – hadu

Slunce, seno, konjugáty

nose – seno

duha – hadu

Karle – lekar

Slunce, seno, konjugáty

nose – seno

duha – hadu

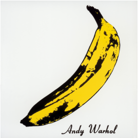
Karle – lekar

Dvě slova uv a vu jsou *sdružená*.

$[w] = \{w' = uv \mid w = vu\}$

Lásky jedné transformace

Burrowsova-Wheelerova transformace

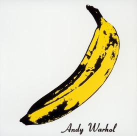


Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

`b a n a n a`

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

b	a	n	a	n	a					
	a	n	a	n	a	b				
		n	a	n	a	b	a			
			a	n	a	b	a	n		
				n	a	b	a	n	a	
					a	b	a	n	a	n

Lásky jedné transformace

Burrowsova-Wheelerova transformace

Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

a	b	a	n	a	n
a	n	a	b	a	n
a	n	a	n	a	b
b	a	n	a	n	a
n	a	b	a	n	a
n	a	n	a	b	a

Lásky jedné transformace

Burrowsova-Wheelerova transformace

Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

$$\begin{array}{cccccc}
\mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{n} \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{b} \\
\mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\
\mathtt{n} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\
\mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{b} & \mathtt{a}
\end{array}$$

$$\mathrm{BWT}_{\mathcal{A}}(\mathtt{banana}) = \mathtt{nnbaaa}$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace

Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

$$\begin{array}{cccccc}
\mathtt{a}_1 & \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \color{red}{\mathtt{n}_1} \\
\mathtt{a}_2 & \mathtt{n} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \color{red}{\mathtt{n}_2} \\
\mathtt{a}_3 & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \color{red}{\mathtt{b}_1} \\
\mathtt{b}_1 & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \color{red}{\mathtt{a}_1} \\
\mathtt{n}_1 & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{n} & \color{red}{\mathtt{a}_2} \\
\mathtt{n}_2 & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{b} & \color{red}{\mathtt{a}_3}
\end{array}$$

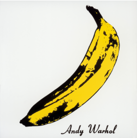
$$\mathrm{BWT}_{\mathcal{A}}(\mathtt{banana}) = \mathtt{nnbaaa}$$

Tvrzení

Výskyty stejného písmene se objevují ve stejném pořadí v prvním a posledním sloupci.

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

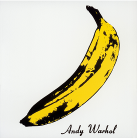
n
n
b
a
a
a

$$\mathrm{BWT}_{\mathcal{A}}(\texttt{banana}) = \texttt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{nnbaaa}) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

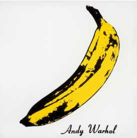
a	n
a	n
a	b
b	a
n	a
n	a

$$\mathrm{BWT}_{\mathcal{A}}\left(\texttt{banana}\right) = \texttt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}\left(\texttt{nnbaaa}\right) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

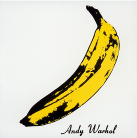
a		n
a		n
a		b
b		a
n	a	a
n		a

$$\mathrm{BWT}_{\mathcal{A}}\,(\texttt{banana}) = \texttt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}\,(\texttt{nnbaaa}) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

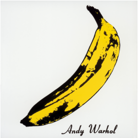
a		n
a		n
a		b
b		a
n	a	a
n	a	a

$$\mathrm{BWT}_{\mathcal{A}}(\texttt{banana}) = \texttt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{nnbaaa}) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

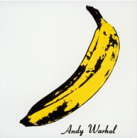
a		n
a		n
a		b
b	a	a
n	a	a
n	a	a

$$\mathrm{BWT}_{\mathcal{A}}(\texttt{banana}) = \texttt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{nnbaaa}) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

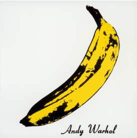
a	b	n
a		n
a		b
b	a	a
n	a	a
n	a	a

$$\mathrm{BWT}_{\mathcal{A}}(\mathtt{banana}) = \mathtt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\mathtt{nnbaaa}) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

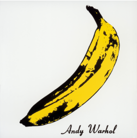
a	b	n
a	n	n
a		b
b	a	a
n	a	a
n	a	a

$$\mathrm{BWT}_{\mathcal{A}}\left(\texttt{banana}\right) = \texttt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}\left(\texttt{nnbaaa}\right) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

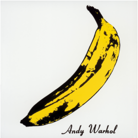
a	b	n
a	n	n
a	n	b
b	a	a
n	a	a
n	a	a

$$\mathrm{BWT}_{\mathcal{A}}(\mathtt{banana}) = \mathtt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\mathtt{nnbaaa}) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

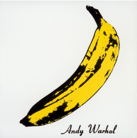
a	b	a	n	a	n
a	n	a	b	a	n
a	n	a	n	a	b
b	a	n	a	n	a
n	a	b	a	n	a
n	a	n	a	b	a

$$\mathrm{BWT}_{\mathcal{A}}(\mathtt{banana}) = \mathtt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\mathtt{nnbaaa}) = ?$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

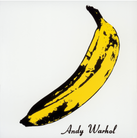
a	b	a	n	a	n
a	n	a	b	a	n
a	n	a	n	a	b
b	a	n	a	n	a
n	a	b	a	n	a
n	a	n	a	b	a

$$\mathrm{BWT}_{\mathcal{A}}(\texttt{banana}) = \texttt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{nnbaaa}) = [\texttt{banana}]$$

Lásky jedné transformace

Burrowsova-Wheelerova transformace



Nechť $\mathcal{A} = \{a_1 < a_2 < \cdots < a_d\}$ a $w \in \mathcal{A}^*$.

$$\begin{array}{cccccc}
\mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{n} \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{b} \\
\mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\
\mathtt{n} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\
\mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{b} & \mathtt{a}
\end{array}$$

$$\mathrm{BWT}_{\mathcal{A}}\left(\mathtt{banana}\right) = \mathtt{nnbaaa}$$

$$\mathrm{BWT}_{\mathcal{A}}^{-1}\left(\mathtt{nnbaaa}\right) = \left[\mathtt{banana}\right] = \left[\mathtt{abanan}\right]$$

Pro hrst dolarů

Nechť \$ je takový, že $a <$ \$ pro každé $a \in \mathcal{A}$.

a	n	a	n	a	\$	b
a	n	a	\$	b	a	n
a	\$	b	a	n	a	n
b	a	n	a	n	a	\$
n	a	n	a	\$	b	a
n	a	\$	b	a	n	a
\$	b	a	n	a	n	a

$\mathrm{BWT}_{\mathcal{A}}(\texttt{banana\$}) = \texttt{bnn\$aaa}$

Ostře sledované BWT

Tvrzení

$[u] = [v]$ tehdy a jen tehdy, když $\mathrm{BWT}_{\mathcal{A}}(u) = \mathrm{BWT}_{\mathcal{A}}(v)$

Ostře sledované BWT

Tvrzení

$[u] = [v]$ tehdy a jen tehdy, když $\mathrm{BWT}_{\mathcal{A}}(u) = \mathrm{BWT}_{\mathcal{A}}(v)$

Tvrzení

$[w] = [u^p]$ tehdy a jen tehdy, když $\mathrm{BWT}_{\mathcal{A}}(u) = b_1 \cdots b_n$ a $\mathrm{BWT}_{\mathcal{A}}(w) = b_1^p \cdots b_n^p$.

$\mathrm{BWT}_{\mathcal{A}}(\mathtt{bon}) = \mathtt{nob}$ $\qquad$ $\mathrm{BWT}_{\mathcal{A}}(\mathtt{bonbon}) = \mathtt{nnoobb}$

Ostře sledované BWT

Tvrzení

$[u] = [v]$ tehdy a jen tehdy, když $\mathrm{BWT}_{\mathcal{A}}(u) = \mathrm{BWT}_{\mathcal{A}}(v)$

Tvrzení

$[w] = [u^p]$ tehdy a jen tehdy, když $\mathrm{BWT}_{\mathcal{A}}(u) = b_1 \cdots b_n$ a $\mathrm{BWT}_{\mathcal{A}}(w) = b_1^p \cdots b_n^p$.

$$\mathrm{BWT}_{\mathcal{A}}(\mathtt{bon}) = \mathtt{nob} \qquad \mathrm{BWT}_{\mathcal{A}}(\mathtt{bonbon}) = \mathtt{nnoobb}$$

Tvrzení

$\mathrm{BWT}_{\mathcal{A}}(\cdot)$ je prostým zobrazení z množiny *Lyndonových slov* do $\mathcal{A}^*$.

Lyndonádový Joe

Slovo w je *Lyndonovo*, pokud je *primitivní* a nejmenší v $[w]$.

Definice [Lyndon (1954), Ширшов (1953)]

$w \in \mathcal{A}^+$ je *Lyndonovo slovo* (or правильное слово), pokud pro každé $p, s \in \mathcal{A}^+$ tak, že $w = ps$, je splněna jedna z následujících tří ekvivalentních podmínek:

1. $w < sp$,
2. $w < s$,
3. $p < s$.

Příklad

✓ a, b, ab, aab, aabab, ananas ✘ abab, ba, banana, kolaloka

S tebou mě baví věta

Věta [Chen-Fox-Lyndon (1958)]

Každé slovo w lze jednoznačně zapsat jako

$$w = \ell_1 \ell_2 \cdots \ell_m$$

tak, že každé ℓ_i je Lyndonovo a $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_m$.

$$\texttt{banana} = \texttt{b} \cdot \texttt{an} \cdot \texttt{an} \cdot \texttt{a}$$

$$\texttt{ananas} = \texttt{ananas}$$

Návrat do BWT

$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

b
a
b
b
a
a
a

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a	b
a	a
a	b
a	b
b	a
b	a
b	a

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a_1	b_1
a_2	a_1
a_3	b_2
a_4	b_3
b_1	a_2
b_2	a_3
b_3	a_4

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

$\mathtt{a}_1$	$\mathtt{a}_2$	$\mathtt{b}_1$
$\mathtt{a}_2$		$\mathtt{a}_1$
$\mathtt{a}_3$		$\mathtt{b}_2$
$\mathtt{a}_4$		$\mathtt{b}_3$
$\mathtt{b}_1$		$\mathtt{a}_2$
$\mathtt{b}_2$		$\mathtt{a}_3$
$\mathtt{b}_3$		$\mathtt{a}_4$

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a_1	a_2	b_1
a_2	b_1	a_1
a_3		b_2
a_4		b_3
b_1		a_2
b_2		a_3
b_3		a_4

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a_1	a_2	b_1
a_2	b_1	a_1
a_3	b_2	b_2
a_4		b_3
b_1		a_2
b_2		a_3
b_3		a_4

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a_1	a_2	b_1
a_2	b_1	a_1
a_3	b_2	b_2
a_4	b_3	b_3
b_1	a_1	a_2
b_2	a_3	a_3
b_3	a_4	a_4

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a_1	a_2	b_1	b_1
a_2	b_1		a_1
a_3	b_2		b_2
a_4	b_3		b_3
b_1	a_1		a_2
b_2	a_3		a_3
b_3	a_4		a_4

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a_1	a_2	b_1	b_1
a_2	b_1		a_1
a_3	b_2		b_2
a_4	b_3		b_3
b_1	a_1		a_2
b_2	a_3		a_3
b_3	a_4		a_4

Návrat do BWT

$$\mathrm{BWT}_{\mathcal{A}}^{-1}(\texttt{babbaaa}) = ?$$

a_1	a_2	b_1	b_1
a_2	b_1		a_1
a_3	b_2		b_2
a_4	b_3		b_3
b_1	a_1		a_2
b_2	a_3		a_3
b_3	a_4		a_4

Ne každé slovo má inverzní BWT!

Tři transformace pro Popelku

BWT, eBWT, bBWT

Nechť W je multimnožina $\{\mathtt{aab}, \mathtt{ab}, \mathtt{ab}\}$ Lyndonových slov nad $\{\mathtt{a} <_\omega \mathtt{b}\}$.

a	a	b	a	a	b	⋯
a	b	a	a	b	a	⋯
a	b	a	b	a	b	⋯
a	b	a	b	a	b	⋯
b	a	a	b	a	a	⋯
b	a	b	a	b	a	⋯
b	a	b	a	b	a	⋯

Tři transformace pro Popelku
BWT, eBWT, bBWT

Nechť W je multimnožina $\{\texttt{aab}, \texttt{ab}, \texttt{ab}\}$ Lyndonových slov nad $\{\texttt{a} <_\omega \texttt{b}\}$.

a	a	b	a	a	b	⋯
a	b	a	a	b	a	⋯
a	b	a	b	a	b	⋯
a	b	a	b	a	b	⋯
b	a	a	b	a	a	⋯
b	a	b	a	b	a	⋯
b	a	b	a	b	a	⋯

$\mathrm{eBWT}_{\mathcal{A}}(W) = \texttt{babbaaa}$

Tři transformace pro Popelku

BWT, eBWT, bBWT

Nechť W je multimnožina $\{\mathtt{aab}, \mathtt{ab}, \mathtt{ab}\}$ Lyndonových slov nad $\{\mathtt{a} <_\omega \mathtt{b}\}$.

$$\begin{array}{ccccccc}
\mathtt{a} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{a} & \mathtt{b} & \cdots \\
\mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \cdots \\
\mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{b} & \cdots \\
\mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{b} & \cdots \\
\mathtt{b} & \mathtt{a} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{a} & \cdots \\
\mathtt{b} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \cdots \\
\mathtt{b} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \mathtt{b} & \mathtt{a} & \cdots
\end{array}$$

$$\mathrm{eBWT}_{\mathcal{A}}(W) = \mathtt{babbaaa}$$

Věta [Mantaci, Restivo, Rosone, Sciortino (2007)]

$\mathrm{eBWT}_{\mathcal{A}}(\cdot)$ je bijekce mezi množinou multimnožin Lyndonových slov a $\mathcal{A}^*$.

Tvrzení [Mantaci, Restivo, Rosone, Sciortino (2007)]

Nechť W je multimnožina primitivních slov.
Pro každé $w \in W$ je $\mathrm{BWT}_{\mathcal{A}}(w)$ podposloupností $\mathrm{eBWT}_{\mathcal{A}}(W)$.

$\mathrm{BWT}_{\mathcal{A}}(\mathtt{aab}) = \mathtt{baa}$ $\qquad$ $\mathrm{eBWT}_{\mathcal{A}}(\{\mathtt{aab}, \mathtt{ab}, \mathtt{ab}\}) = \mathtt{babbaaa}$

Tři transformace pro Popelku

BWT, eBWT, bBWT

Definice [Gil, Scott (2012)]

Bijektivní BWT slova w je

$$\mathrm{bBWT}_{\mathcal{A}}(w) = \mathrm{eBWT}_{\mathcal{A}}(\mathcal{L}(w)),$$

kde $\mathcal{L}(w)$ je multimnožina Lyndonových slov Lyndonovy faktorizace w.

Tři transformace pro Popelku

BWT, eBWT, bBWT

Definice [Gil, Scott (2012)]

Bijektivní BWT slova w je

$$\mathrm{bBWT}_{\mathcal{A}}(w) = \mathrm{eBWT}_{\mathcal{A}}(\mathcal{L}(w)),$$

kde $\mathcal{L}(w)$ je multimnožina Lyndonových slov Lyndonovy faktorizace w.

$\mathcal{L}(\mathtt{banana}) = \{\mathtt{b}, \mathtt{an}, \mathtt{an}, \mathtt{a}\}$

Tři transformace pro Popelku

BWT, eBWT, bBWT

Definice [Gil, Scott (2012)]

Bijektivní BWT slova w je

$$\mathrm{bBWT}_{\mathcal{A}}(w) = \mathrm{eBWT}_{\mathcal{A}}(\mathcal{L}(w)),$$

kde $\mathcal{L}(w)$ je multimnožina Lyndonových slov Lyndonovy faktorizace w.

$\mathcal{L}(\mathtt{banana}) = \{\mathtt{b}, \mathtt{an}, \mathtt{an}, \mathtt{a}\}$

$$\begin{array}{ccccc}
\mathtt{a} & \mathtt{a} & \mathtt{a} & \mathtt{a} & \cdots \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \cdots \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \cdots \\
\mathtt{b} & \mathtt{b} & \mathtt{b} & \mathtt{b} & \cdots \\
\mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \cdots \\
\mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \cdots
\end{array}$$

Tři transformace pro Popelku

BWT, eBWT, bBWT

Definice [Gil, Scott (2012)]

Bijektivní BWT slova w je

$$\mathrm{bBWT}_{\mathcal{A}}(w) = \mathrm{eBWT}_{\mathcal{A}}(\mathcal{L}(w)),$$

kde $\mathcal{L}(w)$ je multimnožina Lyndonových slov Lyndonovy faktorizace w.

$\mathcal{L}(\mathtt{banana}) = \{\mathtt{b}, \mathtt{an}, \mathtt{an}, \mathtt{a}\}$

$$\begin{array}{ccccc}
\mathtt{a} & \mathtt{a} & \mathtt{a} & \mathtt{a} & \cdots \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \cdots \\
\mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \cdots \\
\mathtt{b} & \mathtt{b} & \mathtt{b} & \mathtt{b} & \cdots \\
\mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \cdots \\
\mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \cdots
\end{array}$$

$\mathrm{bBWT}_{\mathcal{A}}(\mathtt{banana}) = \mathrm{eBWT}_{\mathcal{A}}(\{\mathtt{b}, \mathtt{an}, \mathtt{an}, \mathtt{a}\}) = \mathtt{annbaa}$

Tři transformace pro Popelku
BWT, eBWT, bBWT

Definice [Gil, Scott (2012)]

Bijektivní BWT slova w je

$$\mathrm{bBWT}_{\mathcal{A}}(w) = \mathrm{eBWT}_{\mathcal{A}}(\mathcal{L}(w)),$$

kde $\mathcal{L}(w)$ je multimnožina Lyndonových slov Lyndonovy faktorizace w.

$\mathcal{L}(\mathrm{banana}) = \{\mathrm{b}, \mathrm{an}, \mathrm{an}, \mathrm{a}\}$

$$\begin{array}{ccccc}
\color{red}{a} & a & a & a & \cdots \\
a & \color{red}{n} & a & n & \cdots \\
a & \color{red}{n} & a & n & \cdots \\
\color{red}{b} & b & b & b & \cdots \\
n & \color{red}{a} & n & a & \cdots \\
n & \color{red}{a} & n & a & \cdots
\end{array}$$

$$\mathrm{bBWT}_{\mathcal{A}}(\mathrm{banana}) = \mathrm{eBWT}_{\mathcal{A}}(\{\mathrm{b}, \mathrm{an}, \mathrm{an}, \mathrm{a}\}) = \color{red}{\mathrm{annbaa}} \neq \mathrm{nnbaaa} = \mathrm{BWT}_{\mathcal{A}}(\mathrm{banana})$$

Tři transformace pro Popelku

BWT, eBWT, bBWT

Tvrzení [Gil, Scott (2012)]

$\mathrm{bBWT}_{\mathcal{A}}(\cdot)$ je bijekce na $\mathcal{A}^*$.

Tři transformace pro Popelku

BWT, eBWT, bBWT

Tvrzení [Gil, Scott (2012)]

$\mathrm{bBWT}_{\mathcal{A}}(\cdot)$ je bijekce na $\mathcal{A}^*$.

Tvrzení [Biagi, Cenzato, Lipták, Romana (2025)]

$\mathrm{bBWT}_{\mathcal{A}}(w) = \mathrm{BWT}_{\mathcal{A}}(w)$ tehdy a jen tehdy, když w je mocnina Lyndonova slova.

Shlukují se, má panenko

Slovo $w \in \mathcal{A}^*$ je π-*shlukové* pro $\mathcal{A}$, pokud $\mathrm{BWT}_{\mathcal{A}}(w) = a_{\pi(1)}^{k_1} \cdots a_{\pi(d)}^{k_d}$, kde $k_i = |w|_{a_{\pi(i)}}$.

Shlukují se, má panenko

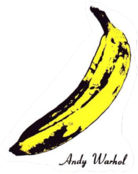
Slovo $w \in \mathcal{A}^*$ je π-*shlukové* pro $\mathcal{A}$, pokud $\mathrm{BWT}_{\mathcal{A}}(w) = a_{\pi(1)}^{k_1} \cdots a_{\pi(d)}^{k_d}$, kde $k_i = |w|_{a_{\pi(i)}}$.
w je *dokonale shlukové*, pokud π je symetrická permutace: $\pi(i) = d - i + 1$.

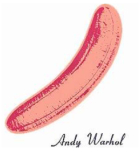
Shlukují se, má panenko

Slovo $w \in \mathcal{A}^*$ je π-*shlukové* pro $\mathcal{A}$, pokud $\mathrm{BWT}_{\mathcal{A}}(w) = a_{\pi(1)}^{k_1} \cdots a_{\pi(d)}^{k_d}$, kde $k_i = |w|_{a_{\pi(i)}}$.
w je *dokonale shlukové*, pokud π je symetrická permutace: $\pi(i) = d - i + 1$.

$$\mathcal{A} = \{\mathtt{a} < \mathtt{b} < \mathtt{n}\}, \qquad \mathcal{A}' = \{\mathtt{a} < \mathtt{n} < \mathtt{b}\}, \qquad \mathcal{A}'' = \{\mathtt{n} < \mathtt{a} < \mathtt{b}\}$$

$\mathrm{BWT}_{\mathcal{A}}(\mathtt{banana}) = \mathtt{nnbaaa}$,	$\mathrm{BWT}_{\mathcal{A}'}(\mathtt{banana}) = \mathtt{bnnaaa}$,	$\mathrm{BWT}_{\mathcal{A}''}(\mathtt{banana}) = \mathtt{aabnna}$
$\pi = (\mathtt{n\ b\ a})$	$\pi' = (\mathtt{b\ n\ a})$	ne shlukové





Vesničko má shlukovací

Tvrzení

$w = u^p \in \mathcal{A}^*$ je π-shlukové pro $\mathcal{A}$ tehdy a jen tehdy, když u je π-shlukové pro $\mathcal{A}$.

Vesničko má shlukovací

Tvrzení

$w = u^p \in \mathcal{A}^*$ je π-shlukové pro $\mathcal{A}$ tehdy a jen tehdy, když u je π-shlukové pro $\mathcal{A}$.

Věta [Mantaci, Restivo, Sciortino (2003)]

Nad binární abecedou jsou (dokonale) shluková slova přesně mocniny *Christoffelových slov* (tj. *konečných standardních sturmovských slov*) a slov s nimi sdružených.

Vesničko má shlukovací

Tvrzení

$w = u^p \in \mathcal{A}^*$ je π-shlukové pro $\mathcal{A}$ tehdy a jen tehdy, když u je π-shlukové pro $\mathcal{A}$.

Věta [Mantaci, Restivo, Sciortino (2003)]

Nad binární abecedou jsou (dokonale) shluková slova přesně mocniny *Christoffelových slov* (tj. *konečných standardních sturmovských slov*) a slov s nimi sdružených.

A nad většími abecedami?

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

Vrchní, palindromy!

Věta [Lapointe, Reutenauer (2024)]

Nechť $w \in \mathcal{A}^*$ je primitivní slovo. Následující jsou ekvivalentní:

- w je Lyndonovo dokonale shlukové
- $\begin{cases} w = a_1 p_1 a_2 p_2 \cdots p_{k-1} a_k, & p_i = \tilde{p}_i \\ w = q_1 q_2, & q_j = \tilde{q}_j \end{cases}$ (*speciální faktorizace*)
- $\begin{cases} w = a_1 p_1 a_2 p_2 \cdots p_{k-1} a_k \\ [w] = [a_k p_{k-1} \cdots a_2 p_1 a_1] \end{cases}$

$\mathrm{BWT}_{\mathcal{A}}(\mathtt{ananas}) = \mathtt{n}^2\mathtt{s}^1\mathtt{a}^3$

$\mathtt{a} \cdot \varepsilon \cdot \mathtt{n} \cdot \mathtt{ana} \cdot \mathtt{s}$
$\mathtt{anana} \cdot \mathtt{s}$

$[\mathtt{a\,n\,ana\,s}] = [\mathtt{s\,ana\,n\,a}]$

Až přijde IET

a	b	a	n	a	n
a	n	a	b	a	n
a	n	a	n	a	b
b	a	n	a	n	a
n	a	b	a	n	a
n	a	n	a	b	a

$\boxed{\texttt{a a a}}\ \boxed{\texttt{b}}\ \boxed{\texttt{n n}}$

$\boxed{\texttt{n n}}\ \boxed{\texttt{b}}\ \boxed{\texttt{a a a}}$

Až přijde IET

a	b	a	n	a	n
a	n	a	b	a	n
a	n	a	n	a	b
b	a	n	a	n	a
n	a	b	a	n	a
n	a	n	a	b	a

a a a | b | n n

n n | b | a a a

Věta [Ferenczi, Zamboni (2013)]

Nechť $w \in \mathcal{A}^*$ je primitivní slovo a $\text{Kard}\,(\mathcal{A}) = d$. Následující jsou ekvivalentní:

1. w je π-shlukové;
2. ww se vyskytuje v trajektorii minimálního d-DIET s permutací π;
3. ww se vyskytuje v trajektorii d-DIET s permutací π;
4. ww se vyskytuje v trajektorii d-IET s permutací π.

Stay tuned for the next talk!

Děkuji vám za pozornost