Observations On Compressed Pattern-Matching with Ranked Variables in Zimin Words

Radosław Głowiński, Wojciech Rytter

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University,
Department of Math/Inf. Warsaw University

Prague, 29 August 2011

We consider words over a finite alphabet $A = \{a, b, c, ..., 0, 1, 2...\}$,

We consider words over a finite alphabet $A = \{a, b, c, ..., 0, 1, 2...\}$, and patterns over a finite alphabet $\mathcal{V} = \{\alpha, \beta, \gamma, ...\}$.

We consider words over a finite alphabet $A = \{a, b, c, ..., 0, 1, 2...\}$,
and patterns over a finite alphabet $\mathcal{V} = \{\alpha, \beta, \gamma, ...\}$.

We are interested in occurences of a pattern in a text, ie.

$$a\,\overbrace{bba}^{\alpha}\,\overbrace{ca}^{\beta}\,\overbrace{bba}^{\alpha}\,cab$$

We consider words over a finite alphabet $A = \{a, b, c, ..., 0, 1, 2...\}$, and patterns over a finite alphabet $\mathcal{V} = \{\alpha, \beta, \gamma, ...\}$.

We are interested in occurences of a pattern in a text, ie.

$$a\,\overbrace{bba}^{\alpha}\,\overbrace{ca}^{\beta}\,\overbrace{bba}^{\alpha}\,cab$$

Definition

Pattern p is avoidable if there exist an infinite word that doesn't encounter p.

We consider words over a finite alphabet $A = \{a, b, c, ..., 0, 1, 2...\}$, and patterns over a finite alphabet $\mathcal{V} = \{\alpha, \beta, \gamma, ...\}$.

We are interested in occurences of a pattern in a text, ie.

$$a\,\overbrace{bba}^{\alpha}\,\overbrace{ca}^{\beta}\,\overbrace{bba}^{\alpha}\,cab$$

Definition

Pattern p is avoidable if there exist an infinite word that doesn't encounter p.

Example: pattern $\alpha\alpha$ is avoidable, because there exist square-free infinite word.

There exists an exponential algorithm solving pattern avoidability problem (Bean, Ehrenfeucht, McNulty 1979; Zimin 1982).

There exists an exponential algorithm solving pattern avoidability problem (Bean, Ehrenfeucht, McNulty 1979; Zimin 1982).

Definition (Zimin Words)

$\mu: \; 1 \to 121, \; i \to i+1 \; \forall \, i > 1$

$Z_1 = 1$

$Z_2 = 1 \; 2 \; 1$

$Z_3 = 1 \; 2 \; 1 \; 3 \; 1 \; 2 \; 1$

$Z_4 = 121312141213121$

There exists an exponential algorithm solving pattern avoidability problem (Bean, Ehrenfeucht, McNulty 1979; Zimin 1982).

Definition (Zimin Words)

$\mu: \ 1 \to 121, \ i \to i+1 \ \forall \ i > 1$

$Z_1 = 1$

$Z_2 = 1 \ 2 \ 1$

$Z_3 = 1 \ 2 \ 1 \ 3 \ 1 \ 2 \ 1$

$Z_4 = 121312141213121$

Theorem

Pattern p is unavoidable $\Leftrightarrow$ *p occurs in* $Z_{\#alph(p)}$.

Definition

Rank of a variable is the highest number in Zimin subword to which this variable morphs.

$$\overbrace{1}^{\alpha}\overbrace{21}^{\beta}\overbrace{31}^{\gamma}\overbrace{21}^{\beta} \qquad = Z_3$$

Problem

Input: given a pattern π with k variables and the ranking sequence.

Problem

Input: given a pattern π with k variables and the ranking sequence.
Output: an instance of an occurrence of π in Z_k with the given ranking function, or information that there is no such valuation.

4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$

4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$

$$\overbrace{\lambda_{\downarrow}}^{1}$$

4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$

$$\overbrace{\lambda_{\downarrow}}^{}$$

$$1$$

$$\overbrace{\lambda}^{}$$

$$121$$

4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$

$$\overbrace{1}^{\lambda_\downarrow}$$

$$\overbrace{121}^{\delta_\downarrow \lambda \delta_\downarrow}$$

$$\overbrace{1}^{\lambda_{\downarrow}}$$

$$\overbrace{121}^{\delta_{\downarrow}\lambda\delta_{\downarrow}}$$

$$\overbrace{121}^{\delta}\overbrace{3}^{\lambda}\overbrace{121}^{\delta}$$

4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$

$$4\ 1\ 3\ 2\ 5\ 3\ 1\ 4\ 1\ 3\ 2\ 1$$

$$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$$

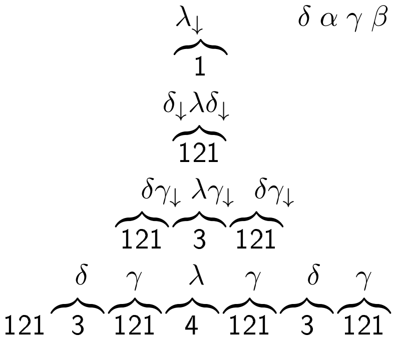
$$\overbrace{1}^{\lambda_{\downarrow}}$$

$$\overbrace{121}^{\delta_{\downarrow}\lambda\delta_{\downarrow}}$$

$$\overbrace{121}^{\delta\gamma_{\downarrow}}\overbrace{3}^{\lambda\gamma_{\downarrow}}\overbrace{121}^{\delta\gamma_{\downarrow}}$$

4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$



4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$

$$\overbrace{1}^{\lambda_\downarrow}$$

$$\overbrace{121}^{\delta_\downarrow\lambda\delta_\downarrow}$$

$$\overbrace{121}^{\delta\gamma_\downarrow}\ \overbrace{3}^{\lambda\gamma_\downarrow}\ \overbrace{121}^{\delta\gamma_\downarrow}$$

$$121\ \overbrace{3}^{\delta}\ \overbrace{121}^{\gamma\ \beta_\downarrow}\ \overbrace{4}^{\lambda}\ \overbrace{121}^{\gamma}\ \overbrace{3}^{\delta}\ \overbrace{121}^{\gamma\beta_\downarrow}$$

4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$

$$\overbrace{1}^{\lambda_\downarrow}$$

$$\overbrace{121}^{\delta_\downarrow\lambda\delta_\downarrow}$$

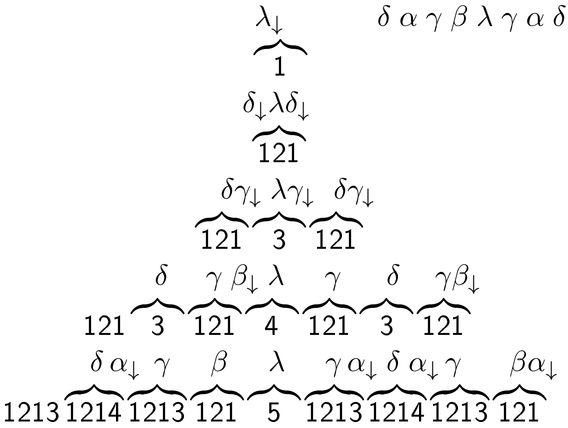
$$\overbrace{121}^{\delta\gamma_\downarrow}\ \overbrace{3}^{\lambda\gamma_\downarrow}\ \overbrace{121}^{\delta\gamma_\downarrow}$$

$$121\ \overbrace{3}^{\delta}\ \overbrace{121}^{\gamma\ \beta_\downarrow}\ \overbrace{4}^{\lambda}\ \overbrace{121}^{\gamma}\ \overbrace{3}^{\delta}\ \overbrace{121}^{\gamma\beta_\downarrow}$$

$$1213\ \overbrace{1214}^{\delta}\ \overbrace{1213}^{\gamma}\ \overbrace{121}^{\beta}\ \overbrace{5}^{\lambda}\ \overbrace{1213}^{\gamma}\ \overbrace{1214}^{\delta}\ \overbrace{1213}^{\gamma}\ \overbrace{121}^{\beta}$$

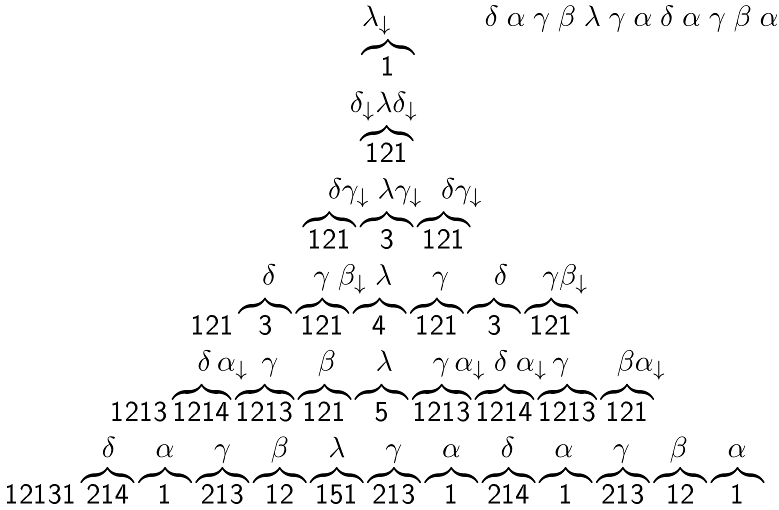
4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$



4 1 3 2 5 3 1 4 1 3 2 1

$\delta\ \alpha\ \gamma\ \beta\ \lambda\ \gamma\ \alpha\ \delta\ \alpha\ \gamma\ \beta\ \alpha$



$$Z_3 = 1213121$$

$$Z_3 = 1213121$$

Application of 2SAT - Example

- $\beta\ \alpha\ \gamma\ \beta$
  $(\beta^{last} \vee \alpha^{first}) \wedge (\neg\beta^{last} \vee \neg\alpha^{first}) \wedge \ldots \wedge \alpha^{first} \wedge \alpha^{last}$

$$Z_3 = 1213121$$

Application of 2SAT - Example

- $\beta\ \alpha\ \gamma\ \beta$
  $(\beta^{last} \vee \alpha^{first}) \wedge (\neg\beta^{last} \vee \neg\alpha^{first}) \wedge \ldots \wedge \alpha^{first} \wedge \alpha^{last}$
- $\beta\ \alpha\ \gamma\ \delta\ \beta\ \gamma$

Definition

For a given Zimin subword we partition it into $w_1 m w_2$, where m is the highest number. Then we remove every element i of w_1 (resp. w_2) such that there exists larger element to the left (resp. to the right). We call obtained sequence compact representation.

$$121312\overbrace{141213121512131 21}^{\ldots\ \alpha\ \ldots}41213121 \qquad 1213121\overbrace{41213121}^{\ldots\ \beta\ \ldots}$$

Definition

For a given Zimin subword we partition it into $w_1 m w_2$, where m is the highest number. Then we remove every element i of w_1 (resp. w_2) such that there exists larger element to the left (resp. to the right). We call obtained sequence compact representation.

$$121312\overbrace{14121312151213121}^{\dots\ \alpha\ \dots}41213121 \qquad 1213121\overbrace{41213121}^{\dots\ \beta\ \dots}$$

$\alpha \rightarrow 145321$

$\beta \rightarrow 4321$

Definition

For a given Zimin subword we partition it into $w_1 m w_2$, where m is the highest number. Then we remove every element i of w_1 (resp. w_2) such that there exists larger element to the left (resp. to the right). We call obtained sequence compact representation.

$$121312\overbrace{1412131215121312 1}^{\dots\ \alpha\ \dots}41213121 \qquad 1213121\overbrace{41213121}^{\dots\ \beta\ \dots}$$

$\alpha \rightarrow 145321$

$\beta \rightarrow 4321$

Compact representation of any subword of Z_k has at most $2 * k - 1$ letters.

Theorem

The compressed ranked pattern matching in Zimin words can be solved in time $O(n * k)$ and (simultaneously) space $O(n + k^2)$, where n is the size of the pattern and k is the highest rank of a variable. A compressed instance of the pattern can be constructed within the same complexities, if there is any solution.

Thank you for your attention!