

# Nová lepší horní mez pro časovou složitost problému zaklínadla

Albert Bakoč      Matěj Smetana

Říjen 2026

## Abstrakt

Na cvičení P2X [2] vedené Mgr. Martinem Marešem PhD. byla zadán problém „Zaklínadlo“ (viz sekci problém). Prezentované řešení mělo asymptotickou časovou složitost v  $\mathcal{O}(N \cdot K \cdot |\Sigma|)$ . Přicházíme s novým řešením s rychlejší asymptotickou časovou složitostí v  $\mathcal{O}(N \cdot \log K)$ .

## 1 Problém:

Zadání problému má následující znění:

„Mějme abecedu  $\Sigma$ , přirozené číslo  $N$  a řetězec  $Z$  (zaklínadlo) délky  $K$ . Spočítejte, kolik existuje řetězců délky  $N$ , které neobsahují zaklínadlo jako podřetězec. Výsledek může být hodně velké číslo, tak ho počítejte modulo zadané prvočíslo  $p$ .“

Pozn.: V původním zadání je  $\Sigma = \{A, B, \dots, Z\}$ , v naší formulaci problému jsme dovolili abecedu obecnou a její velikost je dalším parametrem časové složitosti.

## 2 Řetězce:

Písmena značíme malým písmem, řetězce značíme velkým, množiny řetězců kaligraficky. Indexujeme od 1.

**Definice 1** (Relace býti podřetězcem). Necht'  $R, S \in \Sigma^*$ ,  $|R| < |S|$ . Pak definujeme relaci

$$R \text{ je podřetězcem } S \iff \exists R_1, R_2 \in \Sigma^* : S = R_1.R.R_2$$

Budeme ji značit jako  $R \sqsubseteq S$ .

**Definice 2** (Podřetězce). Necht'  $S \in \Sigma^*$  a  $i, j \in \mathbb{N}_0, i < j$ . Definujeme následující značení:

$$\begin{aligned} S[i] &:= \text{i-tý znak v } S \\ S[i : j] &:= \text{podřetězec } S \text{ začínající na } i \text{ a končící na } j \text{ včetně} \\ \text{pref}_i S &:= \text{prefix } S \text{ délky } i, \text{ tj. } S[1 : i] \\ \text{suff}_i S &:= \text{sufix } S \text{ délky } i, \text{ tj. } S[|S| - i + 1 : |S|] \end{aligned}$$

**Definice 3** (Okraj). Necht'  $O, S \in \Sigma^*, |O| \leq |S|$ . Potom:

$$O \text{ je okraj } S \iff O = \text{pref}_{|O|} S = \text{suff}_{|O|} S,$$

**Definice 4** (Vlastní Okraj). Necht'  $O, S \in \Sigma^*, |O| < |S|$ . Potom:

$$O \text{ je vlastní okraj } S \iff O = \text{pref}_{|O|} S = \text{suff}_{|O|} S$$

**Lemma 5.** Necht'  $S, O \in \Sigma^*, O \text{ okraj } S$ . Potom

$$\forall P \in \Sigma^*, |P| \leq |O| : P \text{ okraj } O \iff P \text{ okraj } S.$$

*Důkaz.* Z definice je  $P$  prefix  $O$  a  $O$  prefix  $S$ . Prefix prefixu je také prefix. U zpětné implikace jsou  $O, P$  prefixy  $S$  a  $|P| \leq |O|$ , tedy  $P$  je prefix  $O$ . Pro sufixy to platí analogicky.  $\square$

### 3 Vzorec:

**Definice 6** (Řetězce obsahující zaklínadlo). Necht'  $n \in \mathbb{N}_0$ . Potom  $\mathcal{F}_n$  značí množinu všech řetězců délky  $n$  obsahujících zaklínadlo jako podřetězec.

$$\mathcal{F}_n = \{S \in \Sigma^n \mid Z \subseteq S\},$$

její velikost označíme  $f(n) = |\mathcal{F}_n|$ .

**Definice 7** (Řetězce obsahující zaklínadlo jako sufix). Necht'  $n \in \mathbb{N}_0$ . Potom  $\mathcal{G}_n$  značí množinu všech řetězců délky  $n$  obsahujících zaklínadlo jako sufix.

$$\mathcal{G}_n = \{S \in \Sigma^n \mid \text{suff}_K S = Z\}.$$

Lze triviálně vidět, že její velikost je  $|\mathcal{G}_n| = |\Sigma|^{n-k}$ , pokud je  $n \geq k$ , jinak 0.

**Definice 8** (Řetězce obsahující zaklínadlo právě jako sufix). Necht'  $n \in \mathbb{N}_0$ . Potom  $\mathcal{H}_n$  značí množinu všech řetězců délky  $n$  obsahujících jediný výskyt zaklínadla a to jako sufix.

$$\mathcal{H}_n = \{S \in \Sigma^n \mid \text{suff}_K S = Z \wedge Z \not\subseteq \text{pref}_{n-1} S\},$$

její velikost označíme  $h(n) = |\mathcal{H}_n|$ .

**Lemma 9.** *Funkci  $f(n)$  lze počítat následovně:*

$$f(n) = |\Sigma| \cdot f(n-1) + h(n)$$

*Důkaz.*  $f(n-1)$  nám udává počet řetězců délky  $n-1$  obsahujících  $Z$  jako podřetězec, každý takový můžeme prodloužit o libovolné písmeno  $a \in \Sigma$  a vznikne nový řetězec délky  $n$  obsahující  $Z$ . Je jich  $|\Sigma| \cdot f(n-1)$  a každý takový leží v  $\mathcal{F}_n$ . To ale nemusí nezapočítat všechny řetězce délky  $n$  jejichž suffix je  $Z$ . Abychom nezapočítali žádný dvakrát, tak se omezíme na řetězce, ve kterých se  $Z$  nevyskytuje jinde než jako prefix. To je přesně definice  $\mathcal{H}_n$  a takových je  $h(n)$ .  $\square$

**Lemma 10.** *Pro  $n \geq K$ , platí, že množiny*

- $\{S \in \mathcal{G}_n \mid Z \not\sqsubseteq \text{pref}_{n-1} S\},$
- $\{S \in \mathcal{G}_n \mid Z \sqsubseteq \text{pref}_{n-K} S\}$  a
- $\{S \in \mathcal{G}_n \mid Z \sqsubseteq \text{pref}_{n-1} S \wedge Z \not\sqsubseteq \text{pref}_{n-K} S\}$

*tvoří rozklad  $\mathcal{G}_n$ .*

*Důkaz.* Označme si

$$\begin{aligned} \mathcal{A} &= \{S \in \mathcal{G}_n \mid Z \not\sqsubseteq \text{pref}_{n-1} S\}, \\ \mathcal{B} &= \{S \in \mathcal{G}_n \mid Z \sqsubseteq \text{pref}_{n-K} S\}, \\ \mathcal{C} &= \{S \in \mathcal{G}_n \mid Z \sqsubseteq \text{pref}_{n-1} S \wedge Z \not\sqsubseteq \text{pref}_{n-K} S\}. \end{aligned}$$

1. Chceme dokázat  $\mathcal{G}_n = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ .

I.  $\mathcal{G}_n \supseteq \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ , z definic  $\mathcal{A}, \mathcal{B}, \mathcal{C}$ .

II.  $\mathcal{G}_n \subseteq \mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \iff$  každé  $S \in \mathcal{G}_n$  patří alespoň do jedné z  $\mathcal{A}, \mathcal{B}, \mathcal{C}$ .  
Označíme  $S'$  jako prefix  $S$  délky  $n-1$ . Pokud  $Z \not\sqsubseteq S'$ , potom patří do  $\mathcal{A}$ . V opačném případě  $S'$  obsahuje  $Z$  jako podřetězec.

Nechť  $i$  označuje poslední písmeno posledního výskytu zaklínadla v  $S'$ . Pokud je  $i \leq n-K$ , tak  $Z \sqsubseteq \text{pref}_{n-K} S$  a proto  $S \in \mathcal{B}$ .

Jinak musí platit, že  $i > n-K$  a tedy  $Z \not\sqsubseteq \text{pref}_{n-K} S$ . Zároveň však víme, že  $Z \sqsubseteq \text{pref}_{n-1} S$  a tak  $S \in \mathcal{C}$ .

Z I. a II.  $\implies \mathcal{G}_n = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ .

2. Chceme dokázat, že množiny  $\mathcal{A}, \mathcal{B}, \mathcal{C}$  jsou po dvou disjunktní.

I.  $\mathcal{A} \cap \mathcal{B} = \emptyset$ . Sporem. Nechť  $S \in \mathcal{A} \cap \mathcal{B}$ , potom

$$Z \not\sqsubseteq \text{pref}_{n-1} S = S_1 \wedge Z \sqsubseteq \text{pref}_{n-K} S = S_2.$$

Protože  $S_1$  je prefix  $S_2$ , tak každý podřetězec  $S_1$  je také podřetězcem  $S_2$ , včetně  $Z$  a to je spor.

II.  $\mathcal{A} \cap \mathcal{C} = \emptyset$ . Také sporem. Necht'  $S \in \mathcal{A} \cap \mathcal{C}$ , potom  $Z \not\sqsubseteq \text{pref}_{n-1} S$  a  $Z \sqsubseteq \text{pref}_{n-1} S$ .

III.  $\mathcal{B} \cap \mathcal{C}$  analogicky.

Z I., II. a III.  $\implies \mathcal{A}, \mathcal{B}, \mathcal{C}$  disjunktní.

Proto

$$\boxed{\mathcal{G}_n = \mathcal{A} \dot{\cup} \mathcal{B} \dot{\cup} \mathcal{C}.}$$

□

**Důsledek 11.** *Z minulého důkazu plyne, že*

$$|\mathcal{G}_n| = |\mathcal{A}| + |\mathcal{B}| + |\mathcal{C}|$$

Dle definice je  $\mathcal{A} = \mathcal{H}_n$  a tedy  $|\mathcal{A}| = |\mathcal{H}_n| = h(n)$ . Protože existuje bijekce mezi  $\mathcal{B}$  a  $\mathcal{F}_{n-K}$  (přidání/odebrání kouzla na konci), tak  $|\mathcal{B}| = |\mathcal{F}_{n-K}| = f(n-K)$ .

Tedy

$$h(n) = |\Sigma|^{n-K} - f(n-K) - |\mathcal{C}|.$$

Označme si  $\mathcal{C}$  pro dané  $n$  jako  $\mathcal{C}_n$ .

**Lemma 12.** *Necht'  $n \in \mathbb{N}$  a  $\mathcal{S}$  je množina vlastních okrajů  $Z$  bez  $\varepsilon$ , potom*

$$|\mathcal{C}_n| = \sum_{S \in \mathcal{S}} h(n-K+|S|).$$

*Důkaz.* Připomeňme, že

$$\mathcal{C}_n = \{S \in \Sigma^n \mid \text{suff}_K S = Z \wedge Z \sqsubseteq \text{pref}_{n-1} S \wedge Z \not\sqsubseteq \text{pref}_{n-K} S\}.$$

Dále víme, že pro každé  $S_1, S_2 \in \mathcal{S}$  platí, že  $S_1 \neq S_2$ , kdyby to neplatilo, tak by existovali dva různé prefixy  $Z$  stejné délky, ale to je spor.

Z minulého argumentu plyne, že každé  $h(n-K+|S|)$  značí velikost jiné množiny  $\mathcal{H}_{n-K+|S|}$ , a proto

$$\sum_{S \in \mathcal{S}} h(n-K+|S|) = \sum_{S \in \mathcal{S}} |\mathcal{H}_{n-K+|S|}| = \left| \bigcup_{S \in \mathcal{S}} \mathcal{H}_{n-K+|S|} \right|,$$

označme si poslední sjednocení jako  $\mathcal{H}$ .

Stačí ukázat, že  $f: \mathcal{H} \rightarrow \mathcal{C}_n, f(X) = X.\text{suff}_{n-|X|} Z$  je bijekce.

I.  $f(\mathcal{H}) \subseteq \mathcal{C}_n$ : Necht'  $X \in \mathcal{H}$ . Potom  $f(X) = X.\text{suff}_{n-|X|} Z$  a  $|X| = n-K+|S|$  pro právě jedno  $S \in \mathcal{S}$ , takže délka sufixu  $Z$ , který přidáme bude  $K-|S|$ .

Navíc platí, že  $\text{suff}_{|S|} X = S$ , protože  $\text{suff}_K X = Z$  a  $S$  je okraj a tedy i sufix  $Z$ . A  $\text{suff}_K f(X) = S.\text{suff}_{K-|S|} Z = Z$ , protože  $S$  je okraj a tedy i prefix  $Z$ .

Z definice víme, že  $\text{suff}_K X = Z$  a také víme, že  $|S| < K$ , protože  $S$  je vlastní okraj, takže máme jeden výskyt  $Z$  v  $\text{pref}_{n-1} f(X)$ ,  $\text{suff}_K f(X) = Z$  a protože  $S \neq \varepsilon$ , tak  $|S| \neq 0$  a proto  $Z \not\sqsubseteq \text{pref}_{n-K} S$ . Tudíž  $f(X) \in \mathcal{C}_n$ .

II.  $\forall X, Y \in \mathcal{H}: f(X) = f(Y) \iff x = y$ , ( $f$  je prostá):

(a)  $X = Y \implies f(X) = f(Y)$ . Necht'  $X = Y$  Potom i  $|X| = |Y|$  a tedy i

$$f(X) = X.\text{suff}_{n-|X|} Z = Y.\text{suff}_{n-|Y|} Z = f(Y).$$

(b)  $X \neq Y \implies f(X) \neq f(Y)$ . Necht'  $X \neq Y$ , potom buď  $|X| = |Y|$  a pak sice platí, že přidané sufixy budou stejné, ale

$$f(X) = X.\text{suff}_{n-|X|} Z \neq Y.\text{suff}_{n-|Y|} Z = f(Y).$$

Nebo je  $|X| \neq |Y|$ . Necht' BÚNO  $X$  je delší. Potom necht' sporem platí, že  $f(X) = f(Y)$ . To by znamenalo, že  $Y = \text{pref}_{|Y|} X$ . Protože  $\text{suff}_K Y = Z$ , tak by platilo, že  $Z \subseteq \text{pref}_{|X|-1} X$ , ale to je spor s definicí, takže  $f(X) \neq f(Y)$ .

III.  $\forall Y \in \mathcal{C}_n: \exists X \in \mathcal{H}: f(X) = Y$ , ( $f$  je „na“): Z definice víme, že  $Y$  obsahuje alespoň dva výskyty zaklínadla, jeden jako  $\text{suff}_K Y$  a druhý někde předtím. Označme si  $i$  jako index posledního znaku prvního výskytu  $Z$ . Z definice je  $i > n - K$ .  $\text{pref}_i Y \in \mathcal{H}_i$ , protože dle definice  $Z \not\subseteq \text{pref}_{n-K} Y$ , tak ani  $Z \not\subseteq \text{pref}_{i-1} Y$ . Proto také  $\text{pref}_i Y \in \mathcal{H}$ , a tedy hledané  $X = \text{pref}_i Y$ .  $\square$

**Důsledek 13.**  $Z$  toho plyne, že

$$h(n) = |\Sigma|^{n-K} - f(n-K) - \sum_{S \in \mathcal{S}} h(n-K+|S|).$$

## 4 Kvadratický algoritmus:

Algoritmus přímo plyne z lemmatu 9 a důsledku 13. Tyto dvě funkce jdou počítat dynamicky, protože v  $i$ -tém kroku závisí pouze na hodnotách z kroků  $< i$ . To platí, protože sčítáme přes všechny vlastní prefixy a každý má délku  $< K$ . Stejně tak se dá dynamicky počítat  $|\Sigma|^n$ .

To bude mít asymptotickou časovou složitost v  $\mathcal{O}(N \cdot K)$ , protože v každém z  $N$  kroků počítáme sumu přes  $\mathcal{O}(K)$  vlastních okrajů zaklínadla. Paměťová složitost bude  $\mathcal{O}(K)$ , protože nám stačí si pamatovat pouze posledních  $K$  hodnot u každé DP tabulky.

Množinu vlastních okrajů zaklínadla sestrojíme následovně: pomocí algoritmu Knutha, Morrise a Pratta [1] najdeme pro každý prefix zaklínadla jeho nejdelší vlastní okraj. Následně iterativně projdeme vzniklý automat od konce k začátku pouze po zpětných hranách. Stavy které jsme navštívili, až na stav odpovídající celému  $Z$ , odpovídají všem vlastním okrajům zaklínadla.

*Důkaz.* Necht' máme automat vzniklý z předpočtu KMP. Potom z každý stav odpovídá jednomu prefixu  $Z$  a z každého vede zpětná hrana do jeho nejdelšího vlastního okraje.

I. Každý vrchol který navštívíme je vlastní okraj  $Z$ :

Matematickou indukcí podle počtu kroků:

V prvním kroku jsme ve stavu odpovídajícímu celému  $Z$ , tudíž se po zpětné hraně přesuneme do jeho nejdelšího vlastního okraje, a to vlastní okraj.

Indukční krok: Jsme ve stavu odpovídajícímu vlastnímu okraji  $\implies$  zpětná hrana vede do vlastního okraje.

Z lemmatu 5 je každý okraj vlastního okraje  $Z$  také vlastní okraj  $Z$ . Proto hrana vede také do vlastního okraje  $Z$ .

II. Neexistuje vlastní okraj  $Z$ , který jsme nenavštívili:

Nechť sporem  $S$  je vlastní okraj  $Z$ , který jsme nenavštívili, potom se podívejme na nejkratší delší okraj  $S'$ , který jsme navštívili. Opět z lemmatu 5 vidíme, že  $S$  je vlastní okraj  $S'$ . Z  $S'$  vede hrana do nejdelšího vlastního okraje  $S'$ , ten bude nutně kratší, než  $S$ , což je ale spor.

□

Tento předpočet běží v  $\mathcal{O}(K)$ , protože KMP běží v lineárním čase a prefixů které můžeme projít je také lineárně mnoho. Předpočet se vejde do  $\mathcal{O}(N \cdot K)$  a má paměťovou složitost  $\mathcal{O}(K)$ .

## 5 Vlastnosti okrajů řetězce:

**Definice 14** (Vzálenost okrajů řetězce). Necht'  $S \in \Sigma$ ,  $R_1, R_2$  jsou okraje  $S$  délek  $i, j$ . Potom

$$\text{vzdálenost } R_1 \text{ a } R_2 := |j - i|.$$

**Lemma 15.** *Necht' řetězec  $T$  má okraje  $R, S$  délek  $i, j$ ,  $i < j$  se vzdáleností  $d = |i - j|$ , pak má  $i$  okraje délek  $j - d \cdot k$  pro všechna  $k \in \mathbb{N}, k \leq \frac{j-1}{d}$ .*

*Důkaz.* Označme si  $W_l = Z[j - d \cdot (l + 1) : j - d \cdot l]$  pro každé  $l \in \mathbb{N}_0$  takové, že  $l \leq \frac{j-1}{d} - 1$ ,  $l_{max} = \lfloor \frac{j-1}{d} \rfloor - 1$  a  $V = Z[1 : j - d \cdot (l_{max} + 1)]$ , potom

$$S = V.W_{l_{max}} \dots W_1.W_0,$$

$$R = V.W_{l_{max}} \dots W_1.$$

Podle lemmatu 5 je  $R$  okrajem  $S$  a tedy  $R$  je sufix  $S$ .

Protože  $\forall l_1, l_2: |W_{l_1}| = |W_{l_2}|$ , tak vzniknou rovnosti

$$\forall l: W_k = W_{l+1},$$

$$V = \text{suff}_{|V|} W_{l_{max}}.$$

Z indukce podle  $l$  přes první rovnosti plyne, že  $\forall l_1, l_2: W_{l_1} = W_{l_2}$ . Označme si  $W = W_l, \forall l$ .

Pro každé  $k \in \mathbb{N}, k < j/d$  pak platí, že  $V.W^{\lfloor \frac{j-1}{d} \rfloor - k}$  je okraj  $S$  a díky lematu 5 i okraj  $T$ .

$$\begin{aligned} |V.W^{\lfloor \frac{j-1}{d} \rfloor - k}| &= |V| + |W| \cdot (\lfloor \frac{j-1}{d} \rfloor - k) \\ &= j - d \cdot \lfloor \frac{j-1}{d} \rfloor + d \cdot \lfloor \frac{j-1}{d} \rfloor - d \cdot k \\ &= j - d \cdot k \end{aligned}$$

A tedy  $T$  má okraje délek  $j - d \cdot k$ .  $\square$

**Definice 16** (Posloupnost mezer řetězce). Necht'  $S \in \Sigma^*$  a  $(o_i)_{i=1}^n$  je posloupnost všech délek okrajů řetězců seřazená vzestupně, potom definujeme posloupnost mezer  $(\mu_i)_{i=1}^{n-1}$  řetězce  $S$  následovně:

$$\mu_i := o_{i+1} - o_i.$$

**Lemma 17.** *Posloupnost mezer je neklesající.*

*Důkaz.* Necht'  $S$  je řetězec a  $(\mu_i)_{i=1}^n$  je jeho posloupnost mezer.

Pro spor uvažujme, že  $\exists i \in [n]: \mu_i > \mu_{i+1}$ . Potom

$$\begin{aligned} &\exists R_1, R_2, R_3 \text{ okraje } S \text{ takové, že:} \\ &\quad |R_1| < |R_2| < |R_3|, \\ &\quad |R_2| - |R_1| = \mu_i, \\ &\quad |R_3| - |R_2| = \mu_{i+1}, \\ &\forall R' \text{ okraj } S: |R_1| < |R'| < |R_3| \implies R' = R_2. \end{aligned}$$

Dle lematu 15 existuje  $R'$  okraj  $S$ , délky

$$|R'| = |R_2| - \mu_{i+1} > |R_2| - \mu_i = |R_1|.$$

Protože z definice  $\forall j: \mu_j > 0$ , tak  $|R_1| < |R'| < |R_3|$ , ale  $R' \neq R_2$ , což je spor.  $\square$

**Definice 18** (Posloupnost period řetězce). Necht'  $S \in \Sigma^*$  a  $(\mu_i)_{i=1}^n$  je jeho posloupnost mezer. Potom definujeme posloupnost period  $(\tau_i)_{i=1}^m, m \leq n$  řetězce  $S$  jako podposloupnost  $(\mu_i)$ , která vznikne odebráním duplicitních hodnot.

**Lemma 19.** *Pro každou periodu platí, že je aspoň tak velká jako součet předchozích.*

*Důkaz.* Necht'  $S$  je řetězec a  $(\tau_i)_{i=1}^n$  jeho posloupnost period. Všimněme si,  $(\tau_i)$  je ostře rostoucí z lematu 17 a protože nemá duplicity.

Mějme libovolnou z period  $\tau_i$ . Pokud  $i = 0$ , lemma zřejmě platí. Jinak nalezneme v řetězci první mezeru takové délky, a ohraničující okraje označíme  $R_1, R_2$ , platí tedy  $|R_1| < |R_2|, |R_2| - |R_1| = \tau_i$ .

Nyní dokažme, že  $|R_1| \leq \tau_i$ . Pro spor předpokládejme, že  $|R_1| > \tau_i$ , pak  $R_0 = \text{pref}_{|R_1| - \tau_i} S$  je podle lematu 15 je také okrajem  $S$ .  $R_1, R_2$  ale ohraničují první mezeru délky  $\tau_i$ , takže existuje okraj  $R: |R_0| < |R| < |R_1|$ , označme si  $a = |R| - |R_0|$ . Z důkazu lematu 15 víme, že okraje můžeme rozepsat na  $R_0 = V.W^n, R_1 = R_0.W = V.W^{n+1}, R_2 = R_1.W = R_0.W^2 = V.W^{n+2}$ , kde

$|W| = \tau_i$  a  $V = \text{suff}_{|V|} W$ . Pak  $R = V.W^n.Y$ , kde  $Y = \text{pref}_a W$ . Podle lemmatu 5 je  $R$  okrajem  $R_1$ , a z toho vyplývá  $W.Y = \text{suff}_a W.W$  a tedy  $\forall i \in [\tau_i] : W[i] = W[(i+a) \bmod \tau_i]$ .

Jelikož rovnost znaků je tranzitivní, tak obecně v mod  $\tau_i$  platí:  $\forall i, j \in [\tau_i] : (\exists k : i \equiv j + k \cdot a) \implies (W[i] = W[j])$ . Označme si  $g = \gcd(a, \tau_i)$ , pak podle Bézoutovy identity existují  $x, y \in \mathbb{Z}$ :

$$\begin{aligned} x \cdot a + y \cdot \tau_i &= g \\ x \cdot a &= g - y \cdot \tau_i \\ x \cdot a &\equiv g \pmod{\tau_i} \end{aligned}$$

Z toho už snadno nahlédneme, že  $\forall m \in \mathbb{Z} \exists k \in \mathbb{N} : k \cdot a \equiv m \cdot g \pmod{\tau_i} \implies \forall i \in [\tau_i], m \in \mathbb{Z} : W[i] = W[i + m \cdot g]$  (pro všechny  $i + m \cdot g$  ležící v  $[1, \tau_i]$ ), tedy, znaky v  $W$  se opakují s periodou  $g$ , a jelikož  $g \mid \tau_i$ , tak  $W = w^x$ , kde  $x = \frac{\tau_i}{g}$ . Pak  $R_2$  lze zapsat jako  $V.w^{(n+2) \cdot x}$ . Zjevně  $R' = V.w^{(n+2) \cdot x - 1}$  je jak prefixem tak suffixem, a tedy i okrajem,  $R_2$ , a podle lemmatu 5 i celého řetězce. Zároveň ale  $|R_1| < |R'| < R_2$ , což je spor s tím, že  $R_1, R_2$  tvoří první mezeru délky  $\tau_i$ . Tím jsme dokázali, že  $|R_1| \leq \tau_i$ .

Nyní už si stačí uvědomit, že součet velikostí všech mezer před  $R_1$  musí být menší nebo rovno  $|R_1|$ , a dostaneme

$$\sum_{\tau_j < \tau_i} \tau_j \leq \sum_{\mu_j < \tau_i} \mu_j \leq |R_1| \leq \tau_i$$

□

**Důsledek 20.** *Z minulého lemmatu plyne, že period bude maximálně logaritmicky mnoho a tím pádem se seřazená posloupnost velikostí okrajů řetězce délky  $n$  dá rozdělit do několika aritmetických posloupností s různými diferencemi a takových bude  $\mathcal{O}(\log n)$ .*

## 6 Optimalizovaný algoritmus:

Pomocí důsledku 20 můžeme optimalizovat počítání sumy z kvadratické verze algoritmu. Můžeme si totiž představit, že přes pole, ve kterém dynamicky počítáme  $h(n)$ , budeme posouvat děrným štítkem, který má díry na místech kde končí vlastní okraje  $Z$ . V každé iteraci posuneme štítek o jedna doprava a chceme vědět součet v dírách. Když si rozdělíme díry do aritmetických posloupností s různými periodami, tak můžeme každou posloupnost počítat zvlášť a pak je sečíst.

Pro každou periodu a každou fázi té periody si budeme pamatovat aktuální součet, navíc si pro každou periodu budeme pamatovat aktuální fázi. Při každém kroku si pro každou periodu zvýším fázi modulo perioda. K odpovídajícímu součtu přičteme hodnotu která se ve štítku nově objevila, tento součet přičteme k celku. Na konci odečteme hodnotu v první díře a tím uvolníme místo pro další iteraci.

Takto budeme v konstantním čase na krok přepočítávat součet pro každou periodu. Period je  $\mathcal{O}(\log K)$ , takže jsme tím snížili časovou složitost.

Celková paměť kterou na součty využijeme je rovna součtu všech period. Díky lemmatu 19 víme, že ten bude nejvýše tak velký jako dvojnásobek největší periody. Největší perioda je v  $\mathcal{O}(K)$ , takže celé se to vejde do  $\mathcal{O}(K)$  paměti.

Předpočet v  $\mathcal{O}(K)$  se stále vejde do  $\mathcal{O}(N \log K)$ , protože kdyby  $N < K$ , tak vrátíme  $|\Sigma|^N$ , takže uvažujeme, že  $K \geq N$ .

## Odkazy

- [1] Donald E. Knuth, James H. Morris Jr. a Vaughan R. Pratt. „Fast Pattern Matching in Strings“. In: *SIAM Journal on Computing* 6.2 (1977), s. 323–350. DOI: 10.1137/0206024.
- [2] Mgr. Martin Mareš PhD. *P2X*. <https://mj.ucw.cz/vyuka/2425/p2x/>.