

# MI-FME Cvičení 5

Tomáš Chvosta

Březen 2020

## Vyhledávání řetězce z 1. sady úkolů

### Zadání:

Navrhněte input/output specifikaci pro vyhledávání řetězce.

### Vstup:

Pole znaků *text* délky  $n$

Pole znaků *hledany* délky  $k$

### Výstup:

Pravda (resp. 1) - v případě, že se hledaný řetězec vyskytuje v zadaném textu.

Nepravda (resp. 0) - v případě, že se hledaný řetězec nevyskytuje v zadaném textu.

$$(\exists i \in \{0, \dots, n - k\})(\forall j \in \{0, \dots, k - 1\})(text[i + j] = hledany[j])$$

### Poznámky:

V této úloze vlastně hledám index v textu, kde začíná hledaný řetězec. Mělo by se mi to do toho textu vejít, proto je tam  $i \leq n - k$ . Když ten index mám, tak už jen stačí zkontrolovat jednotlivé znaky, jestli jsou stejné.

## Cvičení 5a

### Zadání:

Rozšiřte specifikaci uvedenou výše o certifikát pro případ, že je návratová hodnota 1. Popište certifikát formálně, využijte predikátovou logiku.

## Řešení:

Certifikát:  $i$  kdy platí:

$$(i \in \{0, \dots, n - k\}) \wedge ((\forall j \in \{0, \dots, k - 1\})(text[i + j] = hledany[j]))$$

## Poznámka:

Certifikát představuje index v textu, kde hledaný řetězec začíná. Kdokoliv může zkontrolovat, že se na daném indexu skutečně vyskytuje hledaný řetězec.

## Cvičení 5b

### Zadání:

Rozšiřte specifikaci uvedenou výše o certifikát pro případ, že je návratová hodnota 0. Popište certifikát neformálně, využijte přirozený jazyk.

### Úvaha:

V případě, že návratová hodnota byla 1, to bylo snadné. Stačilo spolu s odpovědí ano vrátit certifikát s indexem, kde se hledaný řetězec nachází. V případě, že je návratová hodnota 0, už to není tak snadné. Znamená to, že takový index neexistuje. Certifikát by měl tuto skutečnost dokazovat, tedy že pro každý index v původním textu platí, že to nemůže být začátek hledaného řetězce. Toho můžeme dosáhnout tak, že pro každý index v původním textu vrátíme hodnotu indexu v hledaném řetězci, kde se písmena liší. Tím certifikát jasně dokáže, že žádný index v původním textu nemůže být začátkem hledaného řetězce. Zároveň pro některé indexy je zbytečné tuto hodnotu hledat, u některých indexů víme, že nikdy nemohou být začátkem hledaného řetězce. Například mějme text  $text = \text{"abcdaccddbc"}$  a hledejme  $hledany = \text{"acd"}$ . Můžeme si všimnout, že  $text[9]$  a  $text[10]$  nemohou být počáteční písmena pro  $hledany$ , protože by se tam pak hledaný výraz nemohl vejít vzhledem k jeho délce.

### Příklad:

Pojďme si nyní názorně ukázat, jak takový certifikát bude vypadat. Mějme text  $text = \text{"abcdaccddbc"}$  a hledejme  $hledany = \text{"acd"}$ . Text můžeme reprezentovat jako pole znaků takto:

a	b	c	d	a	c	c	d	d	b	c
---	---	---	---	---	---	---	---	---	---	---

Nyní budeme pro každý znak v textu kontrolovat, zda může být začátkem pro hledaný řetězec. Poslední dva znaky můžeme rovnou vyloučit, na těch nemůže hledaný řetězec začínat, neboť by se tam nevešel. Začneme tím, že zkontrolujeme, zda se znak shoduje s prvním znakem v hledaném řetězci. Pokud ne

uložíme si 0 jakožto hodnotu indexu v hledaném řetězci, kde se neshoduje s původním textem:

	0	0	0		0	0	0	0	X	X
--	---	---	---	--	---	---	---	---	---	---

Pokračujeme tím, že zkontrolujeme, zda se znak shoduje s druhým znakem v hledaném řetězci. Pokud ne uložíme si 1:

1	0	0	0		0	0	0	0	X	X
---	---	---	---	--	---	---	---	---	---	---

A to samé s třetím znakem:

1	0	0	0	2	0	0	0	0	X	X
---	---	---	---	---	---	---	---	---	---	---

Nakonec tedy vrátíme tento certifikát:

1	0	0	0	2	0	0	0	0
---	---	---	---	---	---	---	---	---

## Řešení

Certifikát je tedy pole délky  $n - k + 1$ , které má na každé  $i$ . té pozici uložený index z hledaného řetězce, na kterém se hledaný řetězec neshoduje s původním textem za předpokladu, že  $i$  představuje v původním textu potenciální začátek hledaného řetězce.

## Cvičení 5c

### Zadání:

Rozšiřte specifikaci uvedenou výše o certifikát pro případ, že je návratová hodnota 0. Popište certifikát formálně, využijte predikátovou logiku.

### Řešení:

Pole  $P$  délky  $n - k + 1$  takové, že platí:

$$(\forall i \in \{0, \dots, n - k\})(hledany[P[i]] \neq text[i + P[i]])$$