

1 Vyhledávání v řetězcích

Řekněme, že vaším úkolem je zjistit, na kterých indexech uvnitř textu se vyskytuje nějaké slovo. Například pokud máme text $T = abaababcaba$ o délce $t = 11$ a $P = ab$ o délce $p = 2$, tak výsledkem má být $\{0, 3, 5, 8\}$, protože se slovo ab vyskytuje právě na těchto indexech uvnitř T .

1.1 Naivní algoritmus

Jeden způsob jak vyřešit tento úkol je podívat se na každou pozici, na které se může vzorek vyskytnout, a zkusit každou pozici porovnat s textem. Tento způsob má složitost $O(tp)$, protože pro každou pozici v textu t provedeme maximálně p operací porovnání.

1.2 KMP algoritmus

Pro rychlejší vyhledávání v řetězcích se nejčastěji setkáme s algoritmem Morris-Pratt, případně Knuth-Morris-Pratt (mírné vylepšení). Algoritmus funguje na základě poznatku, že vzorek obrahuje opakující se části tzv. bordery.

Prvně spočítáme border array – pole, ve kterém jsou uloženy nejdelší bordery prefixu vzorku, které nám pomáhá v běhu vyhledávacího algoritmu. Potom můžeme za pomoci border array rychle vyhledávat v libovolném textu.