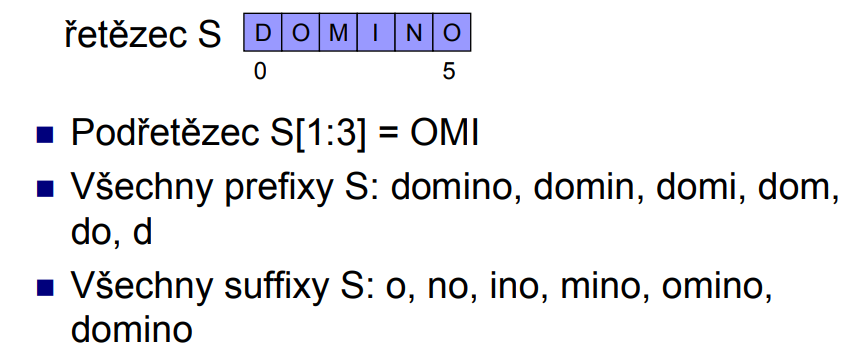
26. Prohledávání řetězců - terminologie, princip, přirozené prohledávání, KMP, chybová funkce

# Co je prohledávání řetězců

* Je dán řetězec T (text) a vyhledávaný řetězec P (vzor, pattern)
* Hledáme vzor P uvnitř textu T
* Aplikace: textové editory, částečně vyhledávání na webu, zpracování obrazu, strukturní rozpoznávání aj.

# Základní pojmy

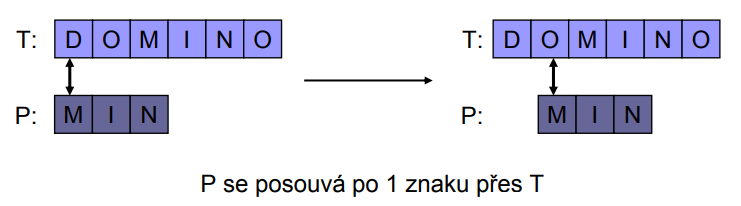
* Označme S řetězec o velikosti m
* Podřetězec S[i:j] je část řetězce S mezi index i a j
* Prefix (predpona) S je podřetězec S[0:i]
* Suffix (přípona) S je podřetězec S[i:m-1], kde i je index mezi 0 a m-1



* Každé slovo je prefixem i sufixem sebe sama – takový prefix nazýváme nevlastní
* Abeceda A je konečná množina znaků, ze kterých tvoříme T a P
* Velikost abecedy je během algoritmu konstantní

# Přirozené vyhledávání

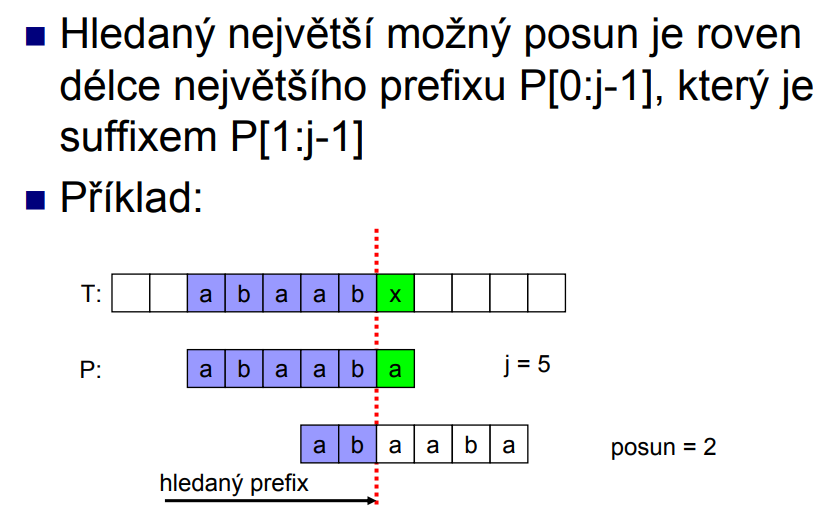
* Postupně procházíme celý řetězec T a pro každou pozici testujeme, zda na ní nezačíná hledaný řetězec P

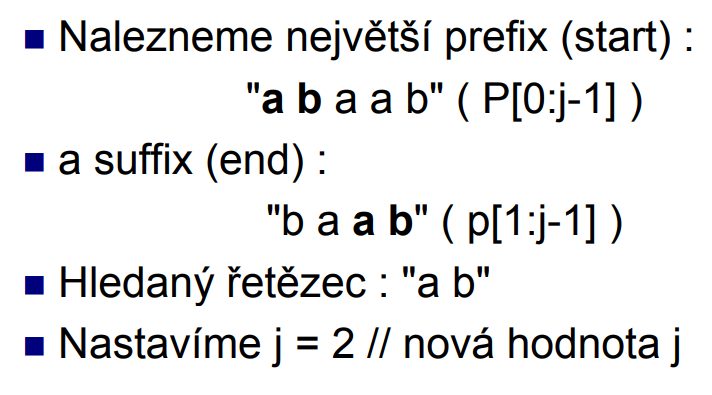


* Algoritmus je rychlý, pokud máme velkou abecedu a naopak pomalý, pokud máme malou abecedu (například jen 0 a 1)
* Nejhorší případ:
  + T: 1111111111111111110
  + P: 110

# Algoritmus KMP

* https://www.youtube.com/watch?v=pu2aO\_3R118&ab\_channel=GBhat
* Knuth-Morris-Pratt (KMP)
* Princip je stejný jako u přirozeného prohledávání
* Řízení procesu:
  + Nastupuje pokud se vyskytne neshoda mezi textem a vzorem v pozici P[j]
  + Jaký je největším možný posun vzoru, abychom se vyhnuli opakovanému porovnávání





# KMP preprocesing – chybová funkce

* Protože prefix i sufix hledáme ve vzoru, který známe předem, můžeme celou analýzu prefixů a sufixů provést předem
* Označme k jako pozici před neshodou
* Chybová funkce F(k) je definována jako nejdelší prefix P[0:k], který je také suffixem P[1:k]

