

Алгоритми конструкције суфиксног низа

Тражење једне ниске (шаблона) у другој (бази) основни је проблем који решавају многобројни алгоритми текста. Традиционални приступи, попут Бојер-Муровог и Кнут-Морис-Пратовог алгоритма, засновани су на припреми шаблона. Шаплони су, међутим, променљиви, док база углавном није, па су нови приступи засновани управо на њеној предобради. Припрема базе такође омогућава ефикасно решавање многих сродних задатака. Позната структура података која добро описује структуру базе јесте суфиксно стабло. Ипак, стабла су просторно захтевна и могу бити сложена. Уместо њих, ефикасније је чувати суфиксни низ, што је сортирани низ суфикса, односно низ индекса сортиран према лексикографском поретку суфикса који почињу тим индексима. Уз помоћни низ најдужих заједничких префикса, суфиксни низ ефикасно решава све проблеме претраге.

У раду је размотрен појам суфиксног низа, са посебном пажњом посвећеном његовој конструкцији, те су описана, имплементирана и упоређена четири алгоритма изградње. На првом месту је наивни алгоритам, који једноставно сортира индексе према лексикографском поретку одговарајућих суфикса, па је временске сложености $O(n^2 \log n)$ за ниску дужине n . Следеће је дуплирање префикса (*prefix doubling*), које искључује многа поновљена поређења, чиме постиже сложеност реда $O(n \log n)$. Напоследку, реализовани су напреднији алгоритми *DC3 (difference cover modulo 3)* и *SA-IS (suffix array-induced sorting)*, који су линеарне временске сложености, дакле $O(n)$. Упореджени су како теоријски, тако и практично, табеларним и графичким приказом рада имплементација.