
Advanced Text Indexing Techniques

Besprechung in den Übungen am 30.6.2009

Exercise 1

Let $T = \text{CTTCTTCTCTACTTC\$}$.

- Show all structures for representing ψ_0 of the CSA. Assume $h = 4$, $\epsilon = 1/2$, and $s = 5$. Table G should only be *sketched*.
- Show in detail how to retrieve $\psi_0[8]$.

Exercise 2

- Derive the Elias ϵ -code from the δ -code in the same way as we derived the δ - from the γ -code.
- What is the length of an ϵ -encoded codeword $(x)_\epsilon$?
- When is the ϵ -code better than the δ -code?

Exercise 3

Consider a bit-sequence $S = s_1s_2 \dots s_n$ of fixed size n , and let $\rho_S(i, j)$ be the operation that reverses the bits from i to j in S : applying $\rho_S(i, j)$ yields $s_1 \dots s_{i-1}s_j s_{j-1} \dots s_{i+1}s_i s_{j+1} \dots s_n$. A naive implementation of ρ runs in $O(j - i)$ time, which can be as bad as $O(n)$. Show how to lower the worst-case running time to $O(n/\log n)$.

Exercise 4

Still open from the previous assignment: Show that the structures of the CSA encapture enough information also for computing an arbitrary entry of the *inverse* suffix array A^{-1} in $O(\log^\epsilon n)$ time.