

Übungsblatt 5

Vorlesung Algorithmen II im WS 12/13

Ausgabe 12. Dezember 2012

Besprechung 20. Dezember 2012

Problem 1: Einstieg

- Berechnen Sie für das Wort `mississippi` sowohl den Suffixbaum, als auch das Suffix-Array.
- Wenden Sie den Rabin-Karp-Algorithmus auf $T = 3141592653589793$ und $P = 26$ an. Nehmen Sie hierzu $q = 11$ an.
- Konstruieren Sie einen String-Matching-Automat für das Muster $P = \text{aababb}$.

Problem 2: Kantenbeschriftung

In der Vorlesung wurde für Suffixbäume gefordert, dass die Kantenbeschriftung B einer Kante e nicht explizit gespeichert wird, sondern nur ein Paar (i, j) mit $1 \leq i \leq j \leq n$, sodass $B = T[i, j]$.

Zeigen Sie, dass, wenn man die Kantenbeschriftungen explizit speichert, der Speicherverbrauch größer als $O(n)$ sein kann.

Problem 3: Kürzester eindeutiger Teilstring

Ein Teilstring eines Strings T heißt *eindeutig*, falls er genau einmal in T vorkommt. Wie kann man alle kürzesten eindeutigen Teilstrings von T mithilfe eines Suffixbaumes bestimmen? Geben Sie auch die Laufzeit Ihres Verfahrens an.

Beispiel: Die kürzesten eindeutigen Teilstrings von $T = \text{badadb}$ sind `ba`, `da`, `db`.

Problem 4: Längster gemeinsamer Teilstring

Im Jahr 1970 stellte Donald Knuth die Vermutung auf, dass es nicht möglich ist in linearer Zeit den längsten gemeinsamen Teilstring zweier Strings T_1 und T_2 zu bestimmen. Zeigen Sie, dass diese Vermutung nicht stimmt.

Verwenden Sie hierzu einen *generalisierten Suffixbaum* \mathcal{S} , der nicht nur einen Text T sondern eine Menge von Texten T_1, \dots, T_ℓ repräsentiert: Ein solcher Suffixbaum entsteht aus dem String $T_1\$_1T_2\$_2 \dots \$_{\ell-1}T_\ell$, wobei $\$_1, \dots, \$_{\ell-1}$ paarweise unterschiedliche Zeichen sind, die nicht im Alphabet vorkommen. Die Blätter von \mathcal{S} können dann den unterschiedlichen Texten und ihren Suffixen zugeordnet werden (siehe Abbildung 1).

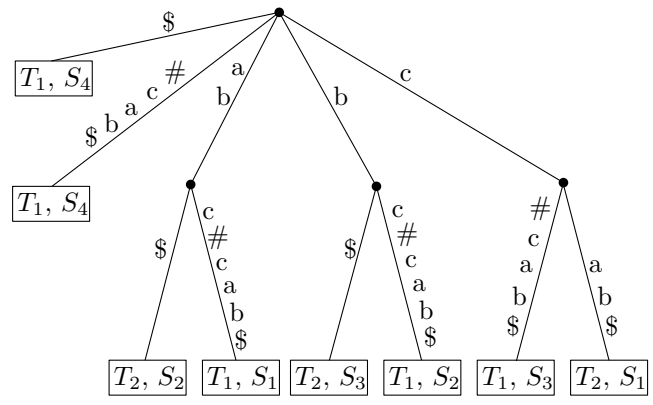


Abbildung 1: Generalisierter Suffixbaum für $T_1 = abc$ und $T_2 = cab$: $T = abc\#cab\$_$, wobei $\#$ und $\$_$ Trennzeichen sind.