

Foundations of Sequence Analysis
Winter 2005/2006

Exercises

Übung 9, Besprechung am 30.01.2006 (14 – 16h in E0-160)
bzw. 19.01.2006 (14 – 16h in C01-148(!!!)).

1. Ukkonens Algorithmus.

Illustrieren sie die einzelnen Schritte von Ukkonens Algorithmus zur Online-Konstruktion des Suffixbaumes für den String $S_1 = ABAABAABAB$.

2. Algorithmus – Längster gemeinsamer Teilstring.

Sei $\mathcal{S} = \{S_i \mid i \in \{1 \dots k\}\}$ eine Menge von k Strings, $S_i = S_{i,1}S_{i,2} \dots S_{i,n_i}$. Der verallgemeinerte Suffixbaum repräsentiert die Menge aller Suffixe der Sequenzen aus \mathcal{S} und kann definiert werden als der Suffixbaum des verallgemeinerten Strings S , $S = S_{1,1}S_{1,2} \dots S_{1,n_1} \$1 S_{2,1}S_{2,2} \dots S_{2,n_2} \$2 \dots S_{k,1}S_{k,2} \dots S_{k,n_k} \k .

Seien nun S_1 und S_2 zwei Strings. Entwickeln sie einen Algorithmus zur Bestimmung des längsten gemeinsamen Teilstrings von S_1 und S_2 in $O(|S_1| + |S_2|)$ Zeit. Benutzen sie dazu den verallgemeinerten Suffixbaum von S_1 und S_2 .

3. Maximale eindeutige Übereinstimmungen – Maximal Unique Matches.

Lesen und verstehen Sie Kapitel 4.9.3 im Skript. Gegeben seien die Strings $S_1 = TTCCTAC$ und $S_2 = TCACTAC$.

- Konstruieren Sie den verallgemeinerten Suffixbaum für S_1 und S_2 .
- Berechnen sie unter Benutzung des verallgemeinerten Suffixbaumes alle maximalen eindeutigen Übereinstimmungen.