

Algorithmen zur Sequenzanalyse

Wintersemester 2018/2019
Besprechung am 30.11.2018

Übungsblatt 3

Prof. Dr. E. Ohlebusch,
Institut für Theoretische Informatik

Für dieses Übungsblatt gelte $S^1 = \text{annaassannasananas}\$$. In Tabelle 1 ist das Suffix Array und das LCP-Array von S^1 gegeben.

Aufgabe 3.1.

Definition 1.3.1 des Vorlesungsskripts führt den Begriff des *lcp*-Intervalls ein. Bestimmen Sie alle *lcp*-Intervalle von S^1 , die keine Singleton-Intervalle (Definition 1.3.4) sind. Bestimmen Sie den *lcp*-Intervallbaum von S^1 .

Aufgabe 3.2.

1. Bestimmen Sie, mit Hilfe von Lemma 1.3.5, alle Kindintervalle des *lcp*-Intervalls $1 - [10..15]$ von S^1 .
2. Bestimmen Sie, mit Hilfe von Lemma 1.3.9, die Elternintervalle aller *lcp*-Intervalle von S^1 , die den *lcp*-Wert 3 haben.

Aufgabe 3.3.

Modifizieren Sie Algorithmus 5 des Vorlesungsskripts so, dass auch Singleton-Intervalle behandelt werden.

Aufgabe 3.4.

Gesucht seien alle Teilstrings eines Strings S , die höchstens q mal in S vorkommen. Geben Sie einen Algorithmus an, der dieses Problem in linearer Zeit löst.

Aufgabe 3.5.

Wir betrachten die Menge M aller Teilstrings eines Strings S , die genau einmal in S vorkommen. Gesucht ist ein kürzester String aus M . Geben Sie einen Algorithmus an, der dieses so genannte *shortest unique substring* Problem in linearer Zeit löst.

i	SA	LCP	$S_{SA[i]}$
1	19	-1	\$
2	4	0	aassannasananas\$
3	13	1	anas\$
4	15	3	anas\$
5	1	2	annaassannasananas\$
6	8	4	annasananas\$
7	17	1	as\$
8	11	2	asananas\$
9	5	2	assannasananas\$
10	3	0	naassannasananas\$
11	14	2	nanas\$
12	16	2	nas\$
13	10	3	nasananas\$
14	2	1	nnaassannasananas\$
15	9	3	nasananas\$
16	18	0	s\$
17	12	1	sananas\$
18	7	3	sannasananas\$
19	6	1	ssannasananas\$
20		-1	

Tabelle 1: Das Suffix Array und das LCP-Array des Strings $S^1 = \text{annaassannasananas}\$$.