

Text Indexing and Information Retrieval

Übungsblatt 2

Besprechung: 7. 11. 2013

Aufgabe 1 (Praxis)

Implementieren Sie den Konstruktionsalgorithmus für Suffix Arrays aus der Vorlesung. Legen Sie hierbei besonderen Wert auf Platzeffizienz, indem Sie allokierten Speicherplatz (wieder)verwenden, wenn er nicht (mehr) benutzt wird.

Im Optimalfall benutzt Ihre Implementierung nur den Platz für den Text und den für das finale Suffix Array, also insgesamt $5n$ Bytes für ASCII-Alphabete. Ideen für solche Platz-sparenden Tricks finden sich u.a. im folgenden Artikel:

- Ge Nong: Practical Linear-Time $O(1)$ -Workspace Suffix Sorting for Constant Alphabets. *ACM Transactions on Information Systems* **31**(3), Article No. 15, 2013.

Testen Sie Ihren Algorithmus für Texte der Größenordnung 100MB und mehr. Solche Texte können Sie etwa auf der folgenden Seite finden:

<http://pizzachili.dcc.uchile.cl/texts.html>