Aufgabenblatt 6 Sommersemester 2006 Markus Lohrey

Übungen zur Vorlesung Algorithmische Fragestellungen für komprimierte Daten

---

1. BISECTION ist ein Grammatik-basierter Kompressionsalgorithmus, der wie folgt arbeitet:

Sei s der zu komprimierende String. Wir definieren eine Menge $B(s)$ von Faktoren von s induktiv wie folgt: Falls $|s| = 1$, so ist $B(s) = \{s\}$. Sei nun $|s| > 1$ und sei j die größte Zahl mit $2^j < |s|$. Zerlege s als $s = s_1 s_2$ mit $|s_1| = 2^j$ und $|s_2| = |s| - 2^j$. Dann ist $B(s) = \{s\} \cup B(s_1) \cup B(s_2)$. Sei außerdem $\ell(s) = s_1$ und $r(s) = s_2$.

Führe nun für jedes $u \in B(s)$ mit $|u| > 1$ ein Nichtterminal A_u ein und definiere

$$\gamma(u) = \begin{cases} A_u & \text{falls } u \in B(s) \text{ und } |u| > 1 \\ u & \text{falls } u \in B(s) \text{ und } |u| = 1. \end{cases}$$

Schließlich führen wir für jedes Nichtterminal A_u die Produktion

$$A_u \to \gamma(\ell(u))\, \gamma(r(u))$$

ein.

Komprimieren Sie den String $s = 1110111010011$ mittels BISECTION.

2. Zeigen Sie, dass für die Approximationsrate $\alpha(n)$ von BISECTION gilt: $\alpha(n) \in \Omega(\sqrt{n}/\log(n))$.

Hinweis: Untersuchen Sie, wie BISECTION sich auf dem String $s_k = a(b^{2^k}a)^{2^k-1}$ für $k > 0$ verhält. Beachten Sie, dass s_k durch ein SLP der Größe $O(k)$ erzeugt werden kann.

3. Zeigen Sie, dass für die Approximationsrate $\alpha(n)$ von BISECTION gilt: $\alpha(n) \in O(\sqrt{n/\log(n)})$.

Hinweis: Berechnen Sie eine obere Schranke für $|B(s)|$ in Abhängigkeit von $|s| = n$. Sei j die größte Zahl mit $2^j < n$. Dann gibt es in $B(s)$ höchstens j viele Strings, deren Länge keine 2er-Potenz ist (warum?). Die Anzahl der Strings in $B(s)$ mit einer Länge von 2^i für $1 \leq i \leq \frac{1}{2}(j - \log(j))$ können Sie mittels Lemma 7 aus der Vorlesung abschätzen. Die Anzahl der Strings in $B(s)$ mit einer Länge von 2^i für $\frac{1}{2}(j - \log(j)) \leq i \leq j$ können Sie durch $O(2^{j-i})$ abschätzen (warum?).