International Workshop On
Combinatorial Algorithms
IWOCA

http://www.iwoca.org—Problems section.

Grammar-Based Compression

Katrin Casel Henning Fernau
Markus L. Schmid
Trier University, Fachbereich IV – Abteilung Informatikwissenschaften,
D-54286 Trier, Germany,
{Casel, Fernau, MSchmid@}uni-trier.de

presented at IWOCA 2016
18 Aug. 2016

A context-free grammar $G = (N, \Sigma, R, S)$ is a *singleton grammar* if R is a total function $N \to (N \cup \Sigma)^+$ and the relation $\{(A, B) \mid (A, w) \in R, |\alpha|_B \geq 1\}$ is acyclic. The language corresponding to such a grammar G contains only a single word, denoted by $\mathfrak{D}(G)$. The size of $G = (N, \Sigma, R, S)$ is given by $|G| = \sum_{(A,w)\in R} |w|$.

Shortest Grammar Problem (SGP)
Instance: A word $w \in \Sigma^+$ and a $k \in \mathbb{N}$.
Question: Does there exist a grammar G with $\mathfrak{D}(G) = w$ and $|G| \leq k$?

Known results, open for improvement:

- NP-hardness for alphabet-size $|\Sigma| = 24$, possibly improvable to $|\Sigma| = 18$ (see [1]) but especially binary alphabet open.
- $\log(|w|)$-approximation for any, especially unbounded, terminal alphabet (see [2, 3]). Constant-factor approximation (for bounded alphabet)?
- W[1]-hardness for parameter $|N|$ (number of non-terminals) for unbounded terminal alphabet Σ. Fixed-parameter tractability with parameter $|N|$ for bounded Σ only known to be in XP (see [1]) but otherwise open.

References

[1] K. Casel, H. Fernau, S. Gaspers, B. Gras and M. L. Schmid: On the Complexity of Grammar-Based Compression over Fixed Alphabets. *to appear in proceedings of ICALP 2016*

[2] P. Berman, M. Karpinski, L. L. Larmore, W. Plandowski, and W. Rytter: On the complexity of pattern matching for highly compressed two-dimensional texts. *Journal of Computer and System Sciences* 65(2):332350, 2002.

[3] W. Rytter: Application of Lempel-Ziv factorization to the approximation of grammar-based compression. *Theoretical Computer Science*, 302:211222, 2003.