Weekplan: Compression

Patrick Hagge Cording

References and Reading

[1] Introduction to Data Compression, Guy E. Blelloch.

[2] Notes on Compression Schemes, Patrick Hagge Cording.

We recommend reading 4.1, 5, and 6 of [1] in detail, and 1.1 of [2].

Exercises

1 [w] **LZ77 and LZ78.** Encode the string `bcabccabccab` using LZ77 and LZ78.

2 [w] **Compression rates.** Suppose you are given a text of size N. What is the best possible compressed file sizes achievable by LZ77, LZ78, and grammar-compression?

3 **Reversal of the Burrows-Wheeler Transform.** Explain how to reverse the Burrows-Wheeler Transform to the original string. *Hint:* The BWT is the last column of the lexicographically sorted cyclic rotations of S. Which other column can easily be derived from the BWT?

4 **From LZ78 to grammar.** Show that given a LZ78 encoding of size n compressing a string S, it can be converted to a grammar of size $O(n)$ compressing S.

5 **Reverse string without decompression.** Suppose you are given a grammar $\mathcal{G}$ producing the string S. You may assume that the grammar has degree 2.

5.1 Give an algorithm that produces a grammar producing the string S^R (the reverse of S).

5.2 What is the time and space complexity of your algorithm?

5.3 Prove that your algorithm is correct.

6 **LZ77 and grammars.** Let z be the number of factors in the LZ77 parse of a string S and n the size of a grammar producing S.

6.1 [$*$] Prove that $z \leq n$.

6.2 A professor is claiming that he has found a polynomial-time algorithm that given a string S can produce a grammar of size z compressing S. Argue that either the professor is wrong or $P = NP$.