

Abstract

Efficient Construction of Hierarchical Overlap Graphs

Sung Gwan Park¹, Bastien Cazaux², Kunsoo Park¹, Eric Rivals²¹*Department of Computer Science, Seoul National University, Korea*²*LIRMM, Montpellier University, Montpellier, France****Corresponding author:** rivals@lirmm.fr

Abstract

The hierarchical overlap graph (HOG for short) is an overlap encoding graph that efficiently represents overlaps from a given set P of n strings. An existing algorithm constructs the HOG in $O(\|P\| + n^2)$ time and $O(\|P\| + n \times \min(n, \max\{|s| : s \in P\}))$ space, where $\|P\|$ is the sum of lengths of the n strings in P . We present a new algorithm of $O(\|P\| \log n)$ time and $O(\|P\|)$ space to compute the HOG, which exploits the segment tree data structure. We also propose an alternative algorithm using $O(\|P\| \frac{\log n}{\log \log n})$ time and $O(\|P\|)$ space in the standard word RAM model of computation.

Work published in SPIRE 2020 conference: [1].

References

- [1] Sung Gwan Park, Bastien Cazaux, Kunsoo Park, and Eric Rivals. Efficient construction of hierarchical overlap graphs. In Christina Boucher and Sharma V. Thankachan, editors, *String Processing and Information Retrieval*, pages 277–290, Orlando, FL, Oct. 2020. Springer International Publishing.