Flexible Grammar-based Indexes⋆

Travis Gagie*

Dalhousie University, Halifax, Nova Scotia, Canada

Abstract

For highly repetitive texts such as pangenomic databases, indexes based on grammars (or Lempel-Ziv parses, string attractors, etc.) are usually significantly smaller than indexes based on the Burrows-Wheeler Transform. Nevertheless, they are not widely used in practice, probably because fast implementations are complicated and support only exact pattern matching. In this talk we will review theoretical and practical advances in the past ten years towards computing matching statistics, MEMs, k-MEMs and MUMs, for example, quickly with grammar-based indexes. We will close with a survey of some future advantages of challenges of using grammar-based indexes for pangenomics.