Abstract

The fields of succinct data structures and compressed text indexing have seen quite a bit of progress over the last two decades. An important achievement, primarily using techniques based on the Burrows-Wheeler Transform (BWT), was obtaining the full functionality of the suffix tree in the optimal number of bits. A crucial property that allows the use of BWT for designing compressed indexes is *order-preserving suffix links*. Specifically, the relative order between two suffixes in the subtree of an internal node is same as that of the suffixes obtained by truncating the first character of the two suffixes. Unfortunately, in many variants of the text-indexing problem, for e.g., parameterized pattern matching, 2D pattern matching, and order-isomorphic pattern matching, this property does not hold. Consequently, the compressed indexes based on BWT do not directly apply. Furthermore, a compressed index for any of these variants has been elusive throughout the advancement of the field of succinct data structures. We achieve a positive breakthrough on one such problem, namely the *Parameterized Pattern Matching* problem. Let T be a text that contains n characters from an alphabet Σ, which is the union of two disjoint sets: Σ_s containing static characters (s-characters) and Σ_p containing parameterized characters (p-characters). A pattern P (also over Σ) matches an equal-length substring S of T iff the s-characters match exactly, and there exists a one-to-one function that renames the p-characters in S to that in P. The task is to find the starting positions (occurrences) of all such substrings S. Previous index [Baker, STOC 1993], known as *Parameterized Suffix Tree*, requires $\Theta(n \log n)$ bits of space, and can find all *occ* occurrences in time $O(|P| \log \sigma + occ)$, where $\sigma = |\Sigma|$. We introduce an $n \log \sigma + O(n)$-bit index with $O(|P| \log \sigma + occ \cdot \log n \log \sigma)$ query time. At the core, lies a new BWT-like transform, which we call the *Parameterized Burrows-Wheeler Transform* (pBWT). The techniques are extended to obtain a succinct index for the *Parameterized Dictionary Matching* problem of Idury and Schäffer [CPM, 1994].