

## Abstract

We present a distributed full-text index for big data applications in a distributed environment. Our index can answer different types of pattern matching queries (existential, counting and enumeration). We perform experiments on inputs up to 100 GiB using up to 512 processors, and compare our index with the distributed suffix array by Arroyuelo et al. [Parall. Comput. 40(9): 471–495, 2014]. The result is that our index answers counting queries up to 5.5 times faster than the distributed suffix array, while using about the same space. We also provide a succinct variant of our index that uses only one third of the memory compared with our non-succinct variant, at the expense of only 20% slower query times.