

Wheeler Graphs: Variations on a Theme by Burrows and Wheeler

Giovanni Manzini

Computer Science Institute, DiSIT, University of Eastern Piedmont, Alessandria, Italy; and
IIT-CNR, Pisa, Italy
giovanni.manzini@uniupo.it

Abstract

The famous Burrows-Wheeler Transform was originally defined for single strings but variations have been developed for sets of strings, labelled trees, de Bruijn graphs, alignments, etc. In this talk we propose a unifying view that includes many of these variations and that we hope will simplify the search for more.

Somewhat surprisingly we get our unifying view by considering the Nondeterministic Finite Automata related to different pattern-matching problems. We show that the state graphs associated with these automata have common properties that we summarize with the concept of a Wheeler graph.¹ Using the notion of a Wheeler graph, we show that it is possible to process strings efficiently even if the automaton is nondeterministic. In addition, we show that Wheeler graphs can be compactly represented and traversed using up to three arrays with additional data structures supporting efficient rank and select operations. It turns out that these arrays coincide with, or are substantially equivalent to, the output of many Burrows-Wheeler Transform variants described in the literature.

This is joint work with Travis Gagie and Jouni Sirén.

1998 ACM Subject Classification E.1 Data Structures, F.2.2 Nonnumerical Algorithms and Problems, H.3 Information Storage and Retrieval

Keywords and phrases compressed data structures, pattern matching

Digital Object Identifier 10.4230/LIPIcs.CPM.2017.1

Category Invited Talk

¹ On many occasions Mike Burrows stated that the original idea of the transformation is due to David Wheeler. We therefore decided to name this graph class after this pioneer of computer science.

