

# Grammar-Compressed Indexes with Logarithmic Search Time

Francisco Claude<sup>1</sup>, Gonzalo Navarro<sup>2,3</sup>, and Alejandro Pacheco<sup>3</sup>

<sup>1</sup> LinkedIn, USA

<sup>2</sup> Millennium Institute for Foundational Research on Data (IMFD), Chile

<sup>3</sup> Department of Computer Science, University of Chile

**Abstract.** Let a text  $T[1..n]$  be the only string generated by a context-free grammar with  $g$  (terminal and nonterminal) symbols, and of size  $G$  (measured as the sum of the lengths of the right-hand sides of the rules). Such a grammar, called a *grammar-compressed representation* of  $T$ , can be encoded using  $G \lg G$  bits. The first grammar-based self-index was proposed by Claude and Navarro, and it was shown to be implementable and practical. In this work, we present and evaluate a practical implementation of a second grammar-self-index proposed by Claude and Navarro, which is a theoretical improvement over the first index but had not been implemented. The results show that the self-index is practical and offers a competitive trade-off between time and space to various self-indexes alternatives in the literature on highly repetitive text collections.