On the reproducibility of experiments of indexing repetitive document collections

Fariña, Antonio

Martínez-Prieto, Miguel A.

Claude, Francisco

Navarro, Gonzalo

Lastra-Díaz, Juan J.

Prezza, Nicola

Seco, Diego

This work introduces a companion reproducible paper with the aim of allowing the exact replication of the methods, experiments, and results discussed in a previous work Claude et al., (2016). In that parent paper, we proposed many and varied techniques for compressing indexes which exploit that highly repetitive collections are formed mostly of documents that are near-copies of others. More concretely, we describe a replication framework, called uiHRDC (universal indexes for Highly Repetitive Document Collections), that allows our original experimental setup to be easily replicated using various document collections. The corresponding experimentation is carefully explained, providing precise details about the parameters that can be tuned for each indexing solution. Finally, note that we also provide uiHRDC as reproducibility package.