

August 26, 2025

Editor
The R Journal

Dear Professor Hyndman,

Please consider our article titled “blocking: An R Package for Blocking of Records for Record Linkage and Deduplication” for publication in the R Journal.

The manuscript introduces a new package called **blocking**, which aims to provide probabilistic blocking methods for record linkage and deduplication using approximate nearest neighbour (ANN) algorithms and graph techniques. It supports multiple ANN implementations via `rnndescent`, `RcppHNSW`, `RcppAnnoy`, and `mlpack` packages, and provides integration with the `reclin2` package.

The package generates shingles from character strings and similarity vectors for record comparison, and includes evaluation metrics for assessing blocking performance, including false positive rate and false negative rate estimates.

The package is already available on CRAN, has been included in the ‘Awesome Official Statistics Software’ list, and was presented at the Use of R in Official Statistics (uRos) 2024 conference in Greece.

We believe readers of the R Journal will find this article helpful for their work, particularly those who integrate data from various sources.

Regards,

Maciej Beręsewicz
Department of Statistics
Poznań University of Economics and Business
Poznań, Poland
maciej.beresewicz@ue.poznan.pl