

## Exploring Research Pathways in Record Deduplication and Record Linkage

Authors: Vaishali Wangikar, Sachin Deshmukh, Sunil Bhirud

Publisher: Springer Singapore

Published in: Recent Trends in Image Processing and Pattern Recognition

### Abstract

This paper provides a detailed introduction, significance and research progression of record de-duplication (RDD) as well as record linkage (RL) process. The basic study starts with the experimental analysis of various Blocking and Indexing techniques for Record de-duplication process, where Sorted Neighborhood Method (SNM) is found to be the best choice among all the methods. SNM is further improved using Adaptive variants of SNM. The advancements in record de-duplication are further explored and various methods for it are reviewed and implemented. The major two contributions in the unsupervised record de-duplication, FDJ and OATF are implemented and compared where it is observed that OATF which is a completely automated and unsupervised approach performs equally well as compared to unsupervised FDJ approach, where limited automation is achieved.