Abstract

Record linkage is a momentous process in data soundness which is used in combining, matching and duplicate removal from more than two databases that refer to the same entities. Deduplication is the process of taking off duplicate records in a united database. Now a day, data cleaning and standardization becomes a pompous process. Due to yielding capacity of today's database, discovering matching records in united database is a crucial one. Indexing technique specifically suffix array is used to efficiently implement record linkage and deduplication.

References

- Gog, Simon, Alistair Moffat, J. Culpepper, Andrew Turpin, and Anthony Wirth. “Large-scale pattern search using reduced-space on-disk suffix arrays.” IEEE


Keywords
Record linkage  suffix array  blocking