PPRL:
PRIMAT Toolbox

Di Mario Cristiano

Università di Modena e Reggio Emilia

Sistemi Informativi e Web Semantico

The problem of PPRL

In many cases, data owners are allowed to provide their data for data integration only if there is sufficient protection of sensitive information to ensure the privacy of individuals (such as patients of a hospital or clients of a facility).

For example, in medical research, data from different sources (e.g., data from different hospitals) must be matched to study possible correlations between diseases without revealing the identity of individual patients.

Privacy Preserving Record Linkage (PPRL) addresses this problem by providing techniques for matching different records while preserving their privacy and allowing data from different sources to be combined to improve data analysis and research.

PRIMAT

PRIMAT is an open source toolbox for the definition and execution of PPRL workflows. It offers several components for data owners and the central linkage unit that provide state-of-the-art PPRL methods, including Bloom-filter-based encoding and hardening techniques, LSH-based blocking, metric space filtering, post-processing and more.

PRIMAT is developed by the Database Group of the University of Leipzig, Germany.

Tests

Input Dataset A: 5000 record Dataset B: 5000 record

Esempio
rec_id, given_name, surname, street_number, address_1, address_2, suburb, postcode, state, date_of_birth, soc_sec_id

rec-1070-org, michaela, neumann, 8, stanley street, miami, winston hills, 4223, nsw, 19151111,5304218
rec-1016-org, courtney, painter, 12, pinkerton circuit, bega flats, richlands, 4560, vic, 19161214,4066625

Data Cleaning Accent Remover Special Character Remover Lower Case Normalizer Umlaut Normalizer Trim Normalizer

Test
Funzione di Similarità: Jaccard Similarity
Threshold: 0.8
True Positive (TP): 4950
False Positive (FP): 50
Precision: 0.9990
Recall: 1
F-measure: 0.9994
Match completati correttamente: 4950 / 5000

References

- Scientific paper: Privacy Preserving Record Linkage (Rainer Schnell)

- Scientific paper: PRIMAT: A Toolbox for Fast Privacy-preserving Matching (Martin Franke, Ziad Sehili, Erhard Rahm)

- PRIMAT: https://git.informatik.uni-leipzig.de/dbs/pprl/primat

- PRIMAT Application: https://github.com/gen-too/primat

- https://www.boozallen.com/insights/ai/privacy-preserving-record-linkage.html

- https://www.sciencedirect.com/science/article/abs/pii/S0306437921001526

- https://github.com/data61/anonlink-entity-service

- My Code + Presentation: https://github.com/mariocris/SIWS.git