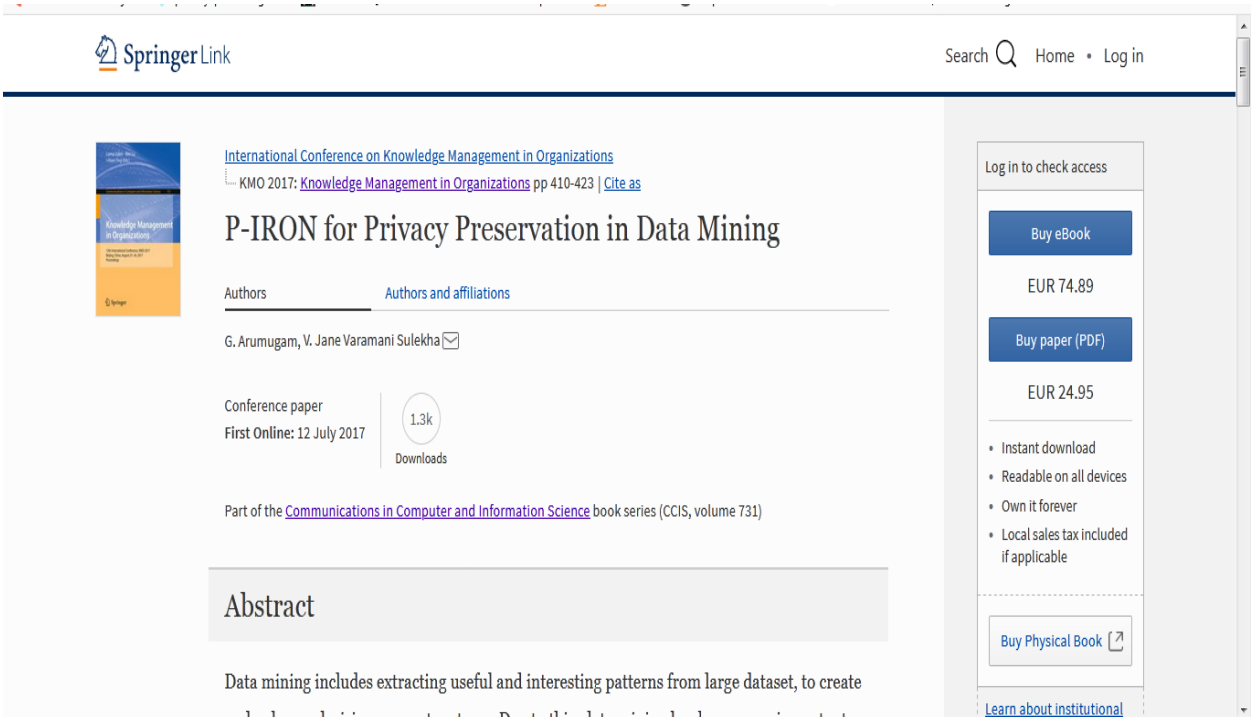
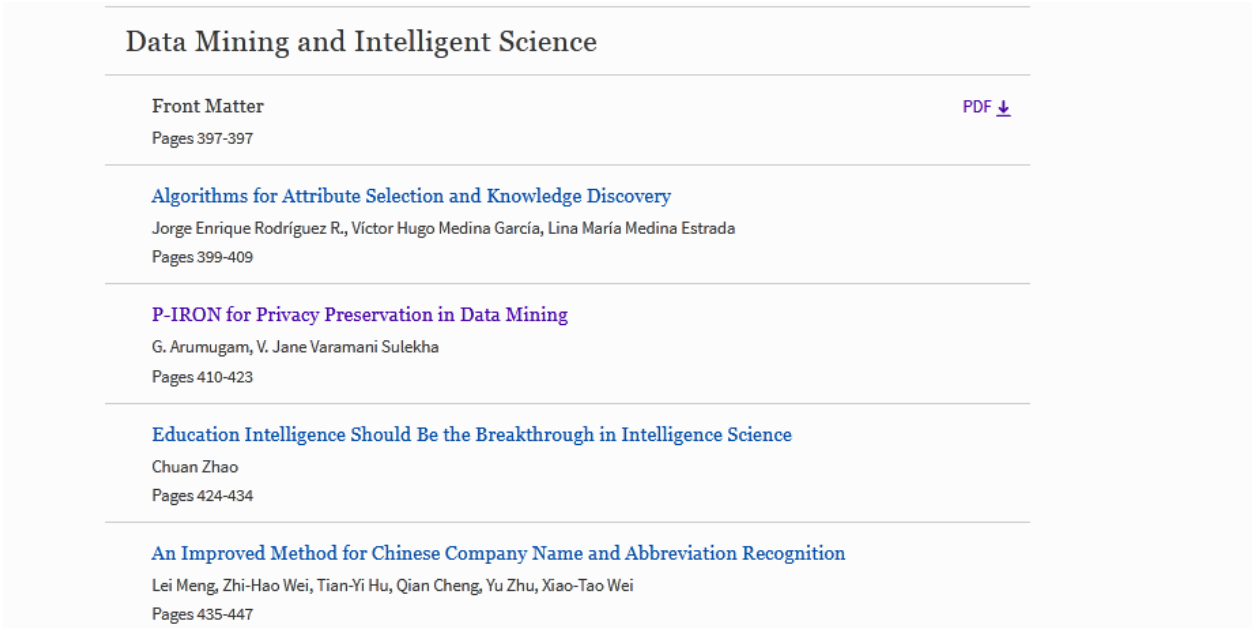
https://link.springer.com/chapter/10.1007/978-3-319-62698-7_34





P-IRON for Privacy Preservation in Data Mining

G. Arumugam
Senior Professor and Head, Department of Computer Science
Madurai Kamaraj University
Madurai, Tamilnadu, India.
gurusamyarumugam@gmail.com

V. Jane Varamani Sulekha
Assistant Professor, Fatima College
Madurai, Tamilnadu, India.
sulekhaisaac@gmail.com

Abstract. Data mining includes extracting useful and interesting patterns from large dataset, to create and enhance decision support systems. Due to this, data mining has become an important component in various fields of day-to-day life including medicine, business, education, science and so on. Numerous data mining techniques have been developed. These techniques make the privacy preservation an important issue. When applying privacy preservation techniques, importance is given to the utility and information loss. In this paper we propose Preference Imposed Individual Ranking based microaggregation with Optimal Noise addition technique (P-IRON) for anonymizing the individual records. Through the experimental results, our proposed technique is validated to prevent the disclosure of sensitive data without degradation of data utilization. Our work highlights some discussions about future work and promising directions in the perspective of privacy preservation in data mining.

Keywords: PPDM, Microaggregation, Privacy, Perturbation, Differential Privacy, Individual Ranking, Optimal Noise

1 Introduction

Privacy preservation is an essential need for all the data mining applications where there exists a large dataset which needs to be analyzed without the analyst or third party data miner obtaining the data directly. To eliminate this problem, researchers have developed many techniques to hide or anonymize the data before analysis. Privacy Preserving Data Mining (PPDM) made of two parts. The simplest technique is anonymization that is deidentification of the data, whereby sensitive raw data (identifiers, quasi identifiers, sensitive attributes) such as name, age, address, phone number, income, disease, SSN (social security number), SIN (social insurance number) is transformed, modified, or eliminated from the data records. In this method, non-sensitive attributes are transmitted without any change. In some cases, anonymized data can be reconstructed. Next, sensitive information mined from a database by using