

Linked Data Sanitization: utility and privacy

The PhD is funded by the ANR (french national research agency) in the context of the national SENDUP project. The PhD will be co-supervised with the Laboratoire d'Informatique de Grenoble (LIG) and will start on September 1, 2019 (may be postponed upon candidates request until December). The deadline for applications is May 30, 2019.

Scientific Context: The amount of data produced by individuals and corporations has dramatically increased during the last decades. This generalized gathering of data brings opportunities (e.g., building new knowledge using this "Big Data") but also new privacy challenges. The general public express a growing distrust over personal data exploitation , which has been met with successive strengthened regulations (e.g. EU general data protection regulation).

This has led to a growing interest for data sanitization, the art of disclosing personal data without jeopardizing privacy, and data-set anonymisation. An anonymized dataset is a dataset which is difficult, costly, or impossible to relate to real individuals. Both domains aim to maintain a certain data quality in order to produce information as useful as possible.

The LIFO (Orléans/Bourges) and LIG (Grenoble) laboratories are working on an innovative ANR project for efficient sanitization and anonymization for semantic graph data-bases. The Semantic Networks of Data: Utility and Privacy (SEND UP) project aims at introducing new utility and privacy concepts for data stored as graphs with an underlying semantic (e.g., RDF). A PhD grant is available for an applicant in the context of SEND UP.

Objectives: The PhD applicant will integrate and collaborate with SEND UP's team. Its main objectives will be:

- introduce new knowledge- and usage-based utility metrics for graph data-bases.
- introduce new privacy guaranties and metrics (e.g. k-anonymity, differential privacy) for graph data-bases.
- contribute to the suite of software modules implementing the proposed algorithms.

Keywords: Open linked data, sanitization, utility, privacy, graph & graph rewriting

Requirements

- Master degree in computer science or equivalent
- Fluent written and spoken english
- Good coding skills (java or C++)
- Autonomous and team working

Knowledge of topics related to databases and/or privacy is an advantage. Knowledge of French language is a plus, but not required. French courses are offered to PhD students if they desire to learn French during their thesis.

Duration and start: 3 years, starting around September 2019 (flexible).

Laboratory & Team: LIFO, Systems and Data Security team

Host institution and workplace: INSA Centre Val de Loire, 88 boulevard Lahitolle 18022 Bourges

To apply: Please submit your application by email to Cdric Eichler (cedric.eichler@insa-cvl.fr), Benjamin Nguyen (benjamin.nguyen@insa-cvl.fr) and Frdric Prost (frederic.prost@univ-grenoble-alpes.fr) before May 30, 2019, including:

- cover letter
- CV
- copies of the relevant certificates
- list of references

- relevant publications, if exist

For inquiries please contact Cdric Eichler (cedric.eichler@insa-cvl.fr).

Relevant references

- 1) Shiva Prasad Kasiviswanathan, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. "Analyzing graphs with node differential privacy". Proceedings of the 10th Theory of Cryptography Conference, TCC 2013.
- 2) Shouling Ji, Weiqing Li, Prateek Mittal, Xin Hu, and Raheem Beyah. "SecGraph: a uniform and open-source evaluation system for graph data anonymization and de-anonymization". Proceedings of the 24th USENIX Conference on Security Symposium (SEC'15), 2015.
- 3) A. A. Mubark, E. Elabd, and H. Abdulkader. "Semantic anonymization in publishing categorical sensitive attributes". Proceedings of the 8th International Conference on Knowledge and Smart Technology, 2016.
- 4) Remy Delanaux, Angela Bonifati, Marie-Christine Rousset, and Romuald Thion. "Query-Based Linked Data Anonymization". Proceedings of the International Semantic Web Conference, 2018.
- 5) World Wide Web Consortium, RDF <https://www.w3.org/RDF/>