

## Compte-rendu de la journée « Données personnelles, vie privée et éthique (EGC&IA 2018) »

Cette journée était organisée par Jérôme Azé (Université de Montpellier/LIRMM) et Thierry Charnois (Université Paris 13/LIPN). Elle a été ouverte par Yves Demazeau (CNRS, Président de l'AFIA) qui a présenté les activités et perspectives de l'AFIA et Fabrice Guillet (Université de Nantes/Polytech'Nantes, Président de EGC) qui a présenté les activités de l'association EGC. Suite au faible nombre de soumissions, la journée initialement prévue a été réduite à une demijournée. Après la présentation des associations AFIA et EGC, trois orateurs se sont succédés, en ciblant différents champs de la discipline :

« *A Declarative Approach to Linked Data Anonymization* », Angela Bonifati, **Rémy Delanaux**, Marie-Christine Rousset and Romuald Thion.

**Résumé:** We present a declarative approach for privacy-preserving data publishing in Linked Open Data, in which privacy and utility policies along with basic data transformation operators are encoded as SPARQL queries. We believe that our approach can be useful in many settings in which data providers are responsible for the non-disclosure of information that could serve as quasi-identifiers whenever crossed with external data sources. Data producers need their own means to specify the privacy policies they want to enforce on their datasets to be published, but also to specify the utility policies that tell which parts of the data should be kept. They also need tools for computing the anonymization operations to apply to their datasets prior to their publication with the guarantee that the specified privacy and utility policies are verified by the resulting dataset. Our approach is designed to meet these requirements by leveraging the expressive power of SPARQL queries and the effectiveness of SPARQL query engines.

« *Co-clustering for differentially private synthetic data generation* », **Tarek Benkhelif**, Françoise Fessant, Fabrice Clérot and Guillaume Raschia.

**Résumé:** We propose a methodology to anonymize microdata (i.e. a table of  $n$  individuals described by  $d$  attributes). The goal is to be able to release an anonymized data table built from the original data while meeting the differential privacy requirements. The proposed solution combines co-clustering with synthetic data generation to produce anonymized data. First, a data independent partitioning on the domains is used to generate a perturbed multidimensional histogram; a multidimensional co-clustering is then performed on the noisy histogram resulting in a partitioning scheme. This differentially private co-clustering phase aims to form attribute values clusters and thus, limits the impact of the noise addition in the second phase. Finally, the obtained scheme is used to partition the original data in a differentially private fashion. Synthetic individuals can then be drawn from the partitions. We show through experiments that our solution outperforms existing approaches and we demonstrate that the produced synthetic data preserve sufficient information and can be used for several datamining tasks.

« *Système de gestion de l'obsolescence dans une base d'informations personnelles* », Salma Chaieb, **Véronique Delcroix**, Ali Benmrad and Emmanuelle Grislin.

**Résumé :** L'évaluation régulière du risque de chute des personnes âgées requiert des informations fiables et nombreuses. Comme il n'est pas possible de recueillir régulièrement toutes ces informations, les observations sont faites au fil du temps et conservées, ce qui entraîne une problématique liée au vieillissement des informations. Cet article traite de la détection des

informations obsolètes dans une base d'informations sur une personne âgée. Nous proposons un agent chargé de maintenir une base d'informations et capable de fournir à la demande des informations fiables et cohérentes sur la personne. L'agent est équipé d'un modèle de connaissances sur les personnes âgées, sous forme d'un réseau bayésien et d'un module de raisonnement chargé de la détection et de la gestion des contradictions et des doutes sur les informations.

Puis, après cette session de présentation d'articles, Karën Fort (Université Paris-Sorbonne) a donné une conférence d'une heure sur la thématique « TAL et éthique ». Karën Fort a dans un premier temps présenté les spécificités des problématiques éthiques en TAL (Traitement Automatique des Langues), puis elle a élargi la problématique aux enjeux éthiques de la communauté scientifique. Elle a ensuite présenté les actions du groupe « Ethique et TAL ». Elle a conclu en montrant la nécessité d'une réflexion éthique à différents niveaux : chercheurs, sociétales, politique, ...