



## Traitement intelligent des données

Meghyn Bienvenu (équipe MF) a obtenu une chaire IA pour son projet intitulé « Intended »

De nos jours, l'énorme quantité de données disponibles recèle un potentiel d'exploitation inédit, pour des applications variées telles que la médecine ou encore pour les outils d'aide à la décision. Cependant, il est souvent difficile d'obtenir des informations pertinentes et fiables à partir de données réelles en raison de leur hétérogénéité et des problèmes de qualité (faits manquants ou incorrects). L'ambition de la chaire est de développer des méthodes intelligentes pour traiter ces données imparfaites, en vue de permettre une prise de décision fiable.

Le projet s'articule autour de l'accès aux données à base d'ontologies (en anglais : ontology-based data acces, OBDA). C'est une approche déclarative prometteuse qui exploite les connaissances sémantiques et le raisonnement automatique pour combler l'écart entre les besoins d'informations des utilisateurs et la manière dont les données sont réellement stockées. Alors que ces systèmes OBDA gagnent en maturité, ils ne parviennent toujours pas à résoudre le problème de la qualité des données, se limitant souvent à émettre des alertes quand les incohérences sont découvertes. Pour pallier cette limitation, le programme de recherche (i) développera des méthodes pragmatiques de gestion d'incohérences qui permettent de traiter des cadres plus expressifs (ontologies plus riches, présence de mappings, dimension temporelle), actuellement non résolues au niveau de l'état de l'art (ii) exploitera des informations de fiabilité qualitatives et quantitatives pour les faits et les contraintes afin d'affiner les résultats des requêtes et de les annoter avec des scores de confiance, (iii) de résoudre un plus large éventail de problèmes de données et obtenir ainsi de meilleurs résultats globaux en développant une approche qui intègre étroitement les méthodes existantes de nettoyage des données (liaison d'entités, analyse statistique) et (iv) développera une approche personnalisée adaptée aux besoins des utilisateurs en les intégrant dans le processus. Cela leur permettra de déterminer la marche à suivre pour traiter certains types d'erreurs, en fonction de leurs connaissances et de la manière dont les données seront utilisées. Durant le projet, nous veillerons à ce que les approches développées adoptent une sémantique claire et qu'il soit possible de retracer les résultats de la requête. Le but est de comprendre ainsi comment les différentes données ont contribué aux résultats.

Pour valider notre approche, nous allons implémenter et tester les algorithmes les plus prometteurs et les mettre à la disposition de la communauté scientifique. De plus, nous démontrerons l'intérêt pratique à travers un cas d'usage en milieu hospitalier. L'objectif est d'ajouter des fonctionnalités de recherche sémantique ainsi que des indicateurs de fiabilité à une interface en cours de développement pour afficher des informations pertinentes dans le cas d'urgences médicales.

# ACTUS LaBRI 13-janvier/février 2020

Je m'attends à ce que les solutions pragmatiques développées dans le cadre du projet soient intégrées dans les systèmes OBDA existants ou futurs, ce qui constituerait un énorme pas en avant pour rendre ces systèmes suffisamment robustes pour s'attaquer à l'intégration des données du monde réel qui sont en désordre, élargissant ainsi les applications potentielles de l'approche OBDA.

Grâce à nos travaux et aux événements organisés, la chaire encouragera les collaborations entre les chercheurs en IA travaillant sur l'OBDA et le traitement des incohérences et les chercheurs en BD travaillant sur la qualité des données et la cohérence des réponses aux requêtes. De façon plus générale, la chaire aidera à relever le défi majeur de l'IA qui consiste à combiner les méthodes symboliques et numériques.

Contact : [Meghyn Bienvenu](#)