

Construction d'arrimages ontologiques de données historicisées

Record number : OPR-1296

Overview

RESEARCH DIRECTION

Christina Khnaisser, Professeure -
Department of Medicine

INFORMATION

christina.khnaisser@usherbrooke.ca

ADMINISTRATIVE UNIT(S)

Faculté des sciences
Département d'informatique

LEVEL(S)

2e cycle
3e cycle

LOCATION(S)

Campus de la santé
Université de Sherbrooke, campus
principal
Pavillon de santé de précision et de
recherche translationnelle (PSPRT)
Département d'informatique

Project Description

Une grande variété des données est générée à différents endroits à différent moment. En plus, des systèmes indépendants utilisent diverses méthodes d'analyse de données pour les activités de tous les jours afin d'améliorer leurs prises de décisions menant à des «silos d'applications». Cependant, pour échanger des données ou des résultats pour prendre des décisions d'une façon éclairée au bon moment, ces systèmes doivent se parler et se comprendre.

Dans le cadre de ce projet de recherche, la personne étudiante contribuera à l'élaboration d'un atelier logiciel afin de permettre : (O1) l'arrimage des données qui proviennent de plusieurs sources à un modèle de connaissances garantissant une interprétation uniforme; (O2) le suivi temporel de l'évolution des données offrant une traçabilité complète des changements et (O3) la formulation de requête sur le modèle arrimé avec un langage intuitif rendant les données facilement accessibles.

L'interprétation des données et des résultats dépend du contexte d'origine ainsi que des hypothèses connues uniquement par l'analyste ou le concepteur du système. Le contexte des données et les hypothèses de travail sont rarement adéquatement documentés. Par exemple, dans un contexte d'une étude sur la pollution dans les grandes villes, les indices de la qualité de l'air à l'heure de pointe ou durant un congé ne peuvent être interprétés de la même façon. C'est encore plus complexe lorsque des données provenant de différents capteurs à différente fréquence sont utilisées ensemble pour une analyse. L'absence d'un modèle de données formel, concis, et évolutif cause plusieurs problèmes d'interopérabilité rendant les données et les résultats difficilement explicables et partageables. L'analyse produit un résultat, mais, en cas d'erreur d'interprétation, cela peut mener à une conclusion inexacte sans que l'équipe d'analyse ait la possibilité de le détecter. La disponibilité d'informations sémantiques, et en particulier de liens avec des concepts du monde réel, permet à des personnes ayant une connaissance limitée des sources de données de comprendre et d'interpréter les données fournies.

Pour assister les analystes et les gestionnaires à prendre des décisions justes au bon moment, il est indispensable de mettre en place une solution intégrée qui permet de documenter les différentes sources, de les mettre en correspondance et de les interroger. L'approche de

solution intégrée est innovante, puisque la résolution des problèmes d'hétérogénéité sémantique et temporelle sera fondée sur un modèle unique et des algorithmes qui s'occupent de la génération de codes appropriés. Ainsi, l'approche permettra de conduire à moindre coût des projets interopérables et de meilleure qualité contribuant tant à l'avancement de la science qu'à la mise en place de systèmes d'aide à décision.

Expertise théorique et pratique. Dans plusieurs domaines, la gestion et l'intégration des données de plusieurs sources sont devenues incontournables pour être compétitives et améliorer les prises de décisions. Grâce à ce projet, la personne étudiante va acquérir des compétences théoriques et pratiques qui lui permettront d'élaborer des méthodes de modélisation pour la manipulation et l'interrogation des données provenant de plusieurs sources. Au niveau théorique, la personne étudiante acquerra des compétences en modélisation (théorie relationnelle, théorie des graphes, logique descriptive et logique des intervalles) nécessaires au développement d'algorithmes innovants pour l'arrimage sémantique des données et l'historicisation. De plus, plusieurs technologies et bonnes pratiques seront explorées, notamment les bases de données RDF avec SPARQL ainsi que les bases de données de graphes (Neo4J avec Cypher), en plus des bases de données relationnelles avec un SQL avancé de plusieurs SGBD, comme PostgreSQL, Microsoft Server et DB2. La connaissance de ces technologies leur permettra d'enrichir leurs compétences pratiques qui les rendront attrayants pour leurs futurs employeurs.

Milieu de formation. La personne recrutée effectuera ses travaux au sein du groupe de recherche interdisciplinaire en informatique de la santé (GRIIS). Le GRIIS réalise de la recherche dans le but d'améliorer les soins et le système de santé en alliant la science, la technologie, la médecine, la philosophie, le droit et l'éthique. La personne pourra valider l'aspect théorique et pratique de ses travaux en lien avec les enjeux de modélisation de données et du génie logiciel et acquérir des compétences transversales. De plus, grâce à mes collaborations avec l'Université Paris Cité, l'Institut de recherche en informatique de Toulouse et la Free University of Bozen-Bolzano (Italie), elles pourront effectuer une thèse en cotutelle.

Veuillez envoyer votre candidature à :
christina.khnaisser@usherbrooke.ca
Joindre : votre CV et vos derniers relevés de notes et une lettre de référence.

Discipline(s) by sector	Funding offered
Sciences naturelles et génie	Yes

The last update was on 17 November 2025. The University reserves the right to modify its projects without notice.