Université de Rennes

Appel à projets - Année 2025

Contrats doctoraux «Ordinaires»

Direction de la Recherche et de l'Innovation

ECOLE DOCTORALE : Matisse **POLE DE RECHERCHE : Mathématiques-Numérique**

Unité de recherche d'accueil

Unité de Recherche : Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA)

Equipe : Shaman

Encadrement

Directeur de thèse : GOASDOUÉ FRANCOIS **Unité de recherche** : IRISA

Employeur/Établissement d'origine : Université de Rennes

HDR : X oui □ non **Nb de thèses en cours** : 0

Mail : fg@irisa.fr **Tél**. : 02 96 46 90 45

Co-directeur de thèse : JAUDOIN HÉLÈNE **Unité de recherche** : IRISA

Employeur/Établissement d'origine : Université de Rennes

HDR : □ oui X non mais HDR à soutenir en 2025 **Nb de thèses en cours** : 0

Mail : helene.jaudoin@irisa.fr **Tél.** : 02 96 46 90 58

Sujet

Titre : CODEX – COmparaison de Descriptions vis-à-vis d'un domaine d'eXpertise

Mots clés : bases de données, raisonnement automatique, représentation des connaissances, web sémantique

Cotutelle internationale (oui/non) : non **Si oui, préciser l'établissement et le pays de rattachement** :

Descriptif (15-30 lignes) :

Dans de nombreux domaines métiers, la compréhension ou l'analyse d'objets d'intérêt par des experts s'effectue via un raisonnement par comparaison, qui consiste à exhiber leurs points communs et leurs différences. La notion d'objet est volontairement abstraite car le raisonnement par comparaison se veut générique et donc indépendant du domaine métier où il est utilisé. Néanmoins, en pratique, ces objets peuvent prendre des formes diverses comme des requêtes sur des bases de données, des faits journalistiques ou politiques, des modèles de structures chimiques ou biologiques, des articles législatifs, des résultats d'examens de dossiers médicaux de patients, etc.

L'objectif de la thèse CODEX est l'automatisation du raisonnement par comparaison entre objets d'intérêt. Il s'agit d'élaborer une solution théorique générique qui soit à la fois indépendante d'un domaine métier particulier, mais qui, lorsqu'elle est appliquée à la comparaison d'objets d'un domaine, est capable d'exploiter des connaissances expertes de celui-ci. En effet, la comparaison d'objets n'a de sens que si elle est capable d'interpréter de façon précise ce que les objets en jeu représentent (c'est-à-dire leur sémantique) au sens des experts du domaine métier.

De façon plus spécifique, le sujet de thèse CODEX vise à étudier des opérateurs de calcul de points communs et de calcul de différences entre objets pour les standards du Web Sémantique du W3C. Ceux-ci permettent de représenter des données hétérogènes avec RDF & OWL ou des requêtes hétérogènes avec SPARQL, soient les objets d'intérêt à comparer, dans un cadre formel homogène : des structures de graphes associées à des connaissances d'un domaine métier exprimées par une ontologie, c'est-à-dire une modélisation d'un domaine dans un langage pragmatique de la logique du premier ordre.

De tels opérateurs ont été proposés pour RDF et SPARQL dans la littérature sur le calcul de points communs [1,2,3,4,5] et dans une moindre mesure sur le calcul de différences [3], sous certaines conditions.

La contribution attendue du sujet de thèse CODEX est de proposer des opérateurs de calcul de points communs et de calcul de différence les plus généraux possibles pour les graphes RDF et les requêtes SPARQL, puis de porter ses résultats sur le standard plus expressif OWL2-QL pour la gestion de grands volumes de données à l'aide d'ontologies. Le gain en expressivité d'OWL2-QL permet une description plus riche et donc plus précise des connaissances expertes d'un domaine d'application.

Les résultats obtenus seront appliqués à au moins l'un des deux problèmes suivants : la découverte et l'explication de manipulation (transformation) d'opinions dans le domaine du journalisme ou l'optimisation d'un workflow de requêtes par sélection de vues à matérialiser dans les systèmes de gestion de données, notamment le système OptiRef de l'équipe Shaman de l'IRISA [6,7].

[1] Simona Colucci, Francesco M. Donini, Eugenio Di Sciascio. Defining and computing Least common subsumers in RDF. Journal of Web Semantic 39, 2016.

[2] Sara El Hassad, François Goasdoué, and Hélènene Jaudoin. Learning commonalities in SPARQL. International Semantic Web Conference (ISWC), 2017.

[3] Alina Petrova, Evgeny Sherkhonov, Bernardo Cuenca Grau, Ian Horrocks. Entity Comparison in RDF Graphs. International Semantic Web Conference (ISWC), 2017.

[4] Sara El Hassad, François Goasdoué, and Hélènene Jaudoin. Learning commonalities in RDF. Extended Semantic Web Conference (ESWC), 2017.

[5] Simona Colucci, Francesco M. Donini, Eugenio Di Sciascio. Computing the Commonalities of Clusters in RDF: Computational Aspects. Journal on Data in Science (DATA), 2024.

[6] Wafaa El Husseini, Cheikh-Brahim El Vaigh, François Goasdoué, and Hélène Jaudoin. Query Optimization for Ontology-Mediated Query Answering. The ACM Web Conference (WWW), 2024.

[7] Wafaa El Husseini, Cheikh-Brahim El Vaigh, François Goasdoué, and Hélène Jaudoin. OptiRef: Query Optimization for Knowledge Bases. The ACM Web Conference (WWW), 2023.