

Avis de Soutenance

Monsieur Jey PUGET GIL

Informatique

Soutiendra publiquement ses travaux de thèse intitulés

Interrogation de graphes de connaissances RDF multiversions à l'aide de représentations condensées

Soutenance prévue le **lundi 22 juin 2026** à 14h30

Lieu : Université Lyon 1 - Bâtiment Gabriel Lippmann - amphithéâtre Gouy au 16 rue Enrico Fermi à Villeurbanne

Composition du jury proposé

M. John SAMUEL	Maître de conférences	CPE Lyon	Co-encadrant de thèse
M. Mohand-Said HACID	Professeur des universités	Université Lyon 1	Examineur
Mme Zoubida KEDAD	Maître de conférences	Université de Versailles Saint-Quentin-en-Yvelines	Examinatrice
Mme Claudia RONCANCIO	Professeure des universités	Grenoble INP	Examinatrice
Mme Salima BENBERNOU	Professeure des universités	Université Paris Descartes	Rapporteure
Mme Ouassila LABBANI NARSIS	Maître de conférences	Université Bourgogne Europe	Rapporteure
M. Gilles GESQUIERE	Professeur des universités	Université Lyon 2	Directeur de thèse
M. Emmanuel COQUERY	Université Lyon 1	Invité	

Mots-clés : graphe, connaissances, évolution, données, urbaines

Résumé :

L'évolution constante des informations du monde réel nécessite des graphes de connaissances dynamiques capables de représenter des ensembles de données multiversionnés. Les systèmes d'archivage traditionnels basés sur le cadre de description des ressources ont du mal à stocker et à interroger efficacement des historiques de versions complexes et non linéaires, tels que ceux comportant des ramifications et des fusions simultanées. Afin de remédier aux limites théoriques et pratiques de la gestion des graphes de connaissances multiversionnés, cette thèse formalise un modèle de « représentation condensée » pour le versionnage RDF. Plutôt que de conserver des copies de données indépendantes ou des chaînes delta séquentielles, le modèle condensé proposé mappe chaque quadruplet RDF unique à un ensemble mathématiquement défini d'instantanés valides, réduisant ainsi la redondance structurelle. Afin de permettre des requêtes temporelles expressives sur ce modèle, cette recherche introduit une nouvelle algèbre de requêtes SPARQL

spécialement conçue pour les données condensées. Cette algèbre est prise en charge par un système de types spécialisé qui suit les représentations des variables (en identifiant si les variables sont liées à des termes, des identifiants ou des ensembles d'instantanés condensés) tout au long du cycle de vie de l'exécution. Une contribution théorique centrale de cette thèse est la preuve formelle de l'équivalence entre les modèles d'évaluation plats et condensés. Cette démonstration rigoureuse établit que les évaluations de requêtes sur la représentation condensée préservent une exactitude sémantique stricte, donnant des résultats identiques aux évaluations plates traditionnelles sans nécessiter de coûteux aplatissement des données. Enfin, ce cadre formel est mis en œuvre de manière pratique dans "Concurrent Versioning of knowledge Graphs", un système de bout en bout qui compile l'algèbre SPARQL condensée en SQL pour l'exécuter sur un backend de base de données relationnelle. Des tests de performance expérimentaux ont été réalisés afin de valider l'approche proposée.

Summary:

The continuous evolution of real-world information necessitates dynamic knowledge graphs capable of representing multi-versioned datasets. Traditional Resource Description Framework archiving systems struggle to efficiently store and query complex, non-linear version histories, such as those featuring concurrent branching and merging. To address both the theoretical and practical limitations of multi-version knowledge graph management, this thesis formalizes a "Condensed Representation" model for RDF versioning. Rather than maintaining independent data copies or sequential delta chains, the proposed condensed model maps each unique RDF quad to a mathematically defined set of valid snapshots, reducing structural redundancy. To enable expressive temporal querying over this model, this research introduces a new SPARQL query algebra specifically designed for condensed data. This algebra is supported by a specialized type system that tracks variable representations—identifying whether variables bind to terms, identifiers, or condensed snapshot sets—throughout the execution lifecycle. A central theoretical contribution of this thesis is the formal proof of equivalence between the flat and condensed evaluation models. This demonstration establishes that query evaluations on the condensed representation preserve strict semantic correctness, yielding identical results to traditional flat evaluations without requiring costly data flattening. Finally, this formal framework is practically implemented in Concurrent Versioning of knowledge Graphs, an end-to-end system that compiles the condensed SPARQL algebra into SQL for execution on a relational database backend. Experimental benchmarking has been conducted to validate the proposed approach.