



Master Informatique

Proposition de sujet de Stage Recherche 2018-2019

Titre : Méthodes avancées de recherche des causes d'échec d'une requête RDF

Laboratoire : LIAS

Encadrant(s) : Dr. Brice CHARDIN (brice.chardin@ensma.fr)

Dr. Stéphane JEAN (stephane.jean@ensma.fr)

Prof. Allel HADJALI (allel.hadjali@ensma.fr)

Référent technique: Mickael BARON (Ingénieur de Recherche au LIAS / ENSMA et Responsable des rubriques Java du site Developpez.com), baron@ensma.fr

Mots clés : données, bases de connaissance, RDF, Yago, DBPedia

Sujet :

Une base de connaissances est une représentation formelle de connaissances sous la forme d'entités et de faits sur ces entités. Ces dernières années, de nombreuses bases de connaissances ont été développées dans des contextes académiques et industriels. Nous pouvons citer Yago, DBPedia, Nell, DeepDive, Google's Knowledge Vault, Wikidata ou Linked Geo Data. Ces bases de connaissances sont généralement représentées sous la forme de graphes à l'aide du langage RDF et interrogées avec le langage de requêtes SPARQL. Les bases de connaissances sont utilisées dans de nombreux domaines comme, par exemple, la traduction automatique des textes, la recherche d'information ou l'intégration de données.

Une caractérisation essentielle de ces bases de connaissances est qu'elles contiennent des données incertaines. Cette incertitude peut être due à l'intégration de sources de données de différents niveaux de fiabilité, à un manque d'informations, ou encore à la limitation des équipements de mesures.

Problématique

La complexité des bases de connaissances en termes de structure et de contenu peut amener les utilisateurs à formuler des requêtes qui ne retournent aucun résultat, ce qui est considéré comme insatisfaisant. Dans le cadre de la thèse d'Ibrahim Dellal, menée au laboratoire LIAS, un algorithme a été développé pour extraire les causes d'échec d'une requête de ce type afin de les présenter à l'utilisateur. Cet algorithme a été comparé avec des méthodes de bases sur plusieurs bancs d'essai qui sont constitués de données générées.

Objectif

L'objectif de ce stage est dans un premier temps de valider la solution proposée par le laboratoire sur des bases de connaissances réelles telles que Yago ou DBPedia.

Dans une seconde partie, nous nous intéresserons à la définition de nouveaux algorithmes (dire ce qui les caractérisent ou leurs idées clés) et à leur évaluation expérimentale. Le travail mené devra être à la fois théorique pour caractériser la performance des algorithmes



Master Informatique

développés (preuve d'optimalité) et empirique pour montrer l'intérêt de ces propositions dans des cas d'application réels.

Organisation

Le travail sera réalisé dans les locaux du LIAS à l'ENSMA. Le stagiaire travaillera en étroite collaboration avec Brice CHARDIN, Stéphane JEAN et Allel HADJALI pour la définition des algorithmes proposés et avec Mickaël BARON (Ingénieur de Recherche au LIAS/ENSMA) pour les développements logiciels associés.

Lieu du stage : LIAS, ENSMA

Parcours conseillé : Données

UEs optionnelles conseillées :

Etudiant pressenti (éventuellement) :