APPROCHES VERS DES MODÈLES UNIFIÉS POUR L'INTÉGRATION DE BASES DE CONNAISSANCES PAR MARIA KOUTRAKI

Présentée par : Maria Koutraki Discipline : informatique Laboratoire : DAVID

Résumé :
Une des idées de base du web sémantique est que les données peuvent être partagées au-delà des limites des applications et des sites web. Une application particulière du Web sémantique est celle des données liées, constituant des ressources dans le web des données et généralement publiées au format RDF. Elle permet le partage et l'accès aux données d'une manière décentralisée. Ceci a conduit à l'émergence de nombreuses initiatives de données liées, dont le projet phare Linked Open Data (LOD). On compte aujourd'hui des milliers de jeux de données formant des milliards de triplets RDF. Cependant, le LOD est encore à ses débuts, car s'il a permis de rendre les données accessibles sur le Web, il est loin d'avoir atteint son objectif initial qui est d'accéder de manière uniforme à travers le Web à des données de sources différentes. Dans cette thèse, nous identifions plusieurs défis qui entravent la réalisation de l'objectif initial du web sémantique, visé par l'initiative Linked Data. La question est : Comment pouvons-nous accéder uniformément à toutes les ressources publiées dans le cadre de LOD

indépendamment de leur source, du type de données, du mode d'accès ou de la structure de données ? L'objectif principal de ce travail est de proposer des approches qui abordent ces défis, en mettant en avant un modèle permettant d'interroger de manière uniforme et d'intégrer des données provenant de sources ayant des structures hétérogènes, à savoir des sources de données RDF ou des services Web. Dans ce contexte, la première contribution de cette thèse est la proposition du système DORIS pour un accès uniforme aux services Web et qui permet d'enrichir une base de connaissances cible. La seconde contribution est une approche d'alignement basée sur les instances des relations des sources RDF. Partant d'une relation donnée et de la base de connaissances, cette approche permet de découvrir les liens de subsomption avec des relations d'une base de connaissances de cible. Pour ce faire, elle utilise des techniques d'apprentissage.

Abstract :
One of the core visions of the Semantic Web is that data can be shared across the boundaries of applications and websites. A particular application of Semantic Web is Linked Data, which publishes data as Web Data in RDF format. It enables sharing and accessing data in a decentralised manner. This is has led to the rise of many Linked Data initiatives, with its most successful project Linked Open Data (LOD), which at this time numbers thousands of datasets, and with a magnitude of billions of triples. The LOD is still in its beginnings and apart from making data accessible across the Web, however, it has a long way to go to accomplish its original goal of uniformly accessing data across the Web and datasets. In this work we identify several challenges that hinder the fulfilment of the original goal of the Semantic Web, respectively the Linked Data initiative. The question is: How do we uniformly accessing all the resources published as part of the LOD independent of their source, data type, access mode or data structure? The main focus of this work is at understanding and proposing approaches that tackle these challenges, towards a model for uniformly accessing and integrating data coming from different datasets with heterogeneous structures, i.e. RDF datasets or Web services. In this context DORIS system is proposed in order to enable a uniform access to Web services with the purpose of enriching a target knowledge base. Secondly, we propose an on-line instance-based relation alignment approach between RDF datasets. For a given relation and knowledge base, we uncover super-relations on a target knowledge base.

INFORMATIONS COMPLÉMENTAIRES

Mme Karine ZEITOUNI, Professeur des Universités, Université de Versailles Saint-Quentin-en-Yvelines - Laboratoire DAVID - Directeur de these
Mme Marie-Christine ROUSSET, Professeur des Universités, Université de Grenoble - Rapporteur
M. Vassilis CHRISTOPHIDES, Professeur des Universités, Université de Crete (Grèce) - Rapporteur
M. Bernd AMANN, Professeur des Universités, Université de Paris 6 - Examinateur
M. François GOASDOUE, Professeur des Universités, Université de Rennes 1 - Examinateur
Mme Fatiha SAIS, Maître de conférences, Université Paris-Sud 11 - Examinateur
M. Dan VODISLAV, Professeur des Universités, Université de Cergy Pontoise - CoDirecteur de these
Mme Nicoleta PREDA, Maître de conférences, Université de Versailles Saint-Quentin-en-Yvelines - Laboratoire DAVID - Co-encadrant de these

Contact : dredval service FED : theses@uvsq.fr