

LKSA : un algorithme de sélection de clés de liage dans des données RDF guidée par des paires de classes

Nacira Abbas*, Jérôme David**
Amedeo Napoli*

* Université de Lorraine, CNRS, Inria, Loria
Nancy, F-54000, France
nacira.abbas@inria.fr,
amedeo.napoli@loria.fr

** Université de Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG
Grenoble, F-38000, France
jerome.david@inria.fr

Résumé. Dans cet article, nous nous intéressons à la sélection de clés de liage dans des jeux de données RDF. Une clé de liage permet d'établir des liens d'identité entre deux entités dans le web des données. Le couple d'entités liées détermine à son tour un couple de classes qui peut guider la recherche et la sélection de clés de liage de meilleure qualité. Pour ce faire, nous proposons un algorithme de sélection de clés de liage s'appuyant sur des mesures de qualité adaptées. Une série d'expérimentations sur des jeux de données RDF diversifiés montre le potentiel et l'efficacité de l'approche.

1 Introduction

Nous nous intéressons à la découverte de liens d'identité entre les ressources de deux jeux de données RDF (Resource Description Framework). Cette tâche est cruciale pour le web des données car les liens d'identité autorisent une meilleure interopérabilité entre différentes applications ainsi que l'amélioration de la qualité des données (recherche de "doublons"). Plusieurs approches existent pour trouver les liens d'identité parmi lesquelles l'approche "logique" qui s'appuie sur des axiomes exprimant des conditions suffisantes pour que deux ressources soient identiques (Saïs et al., 2007; Al-Bakri et al., 2015, 2016). Suivant cette ligne, cet article se concentre sur les *clés de liage* qui généralisent la notion de clé sur deux jeux de données (Atencia et al., 2014a). Un exemple de clé de liage est :

$$k = (\{(designation, titre)\}, \{(designation, titre), (author, auteur)\}, (Book, Livre))$$

qui signifie que si une instance a_1 de la classe `Book` et une instance b_1 de la classe `Livre`, possèdent les mêmes valeurs pour la propriété `designation` et pour la propriété `titre`, et que a_1 et b_1 partagent au moins une valeur pour les propriétés `author` et `auteur`, alors a_1 et b_1 désignent la même entité. En particulier la clé de liage k instancie la paire de classes $\langle Book, Livre \rangle$ et génère un lien d'identité ("same-as") entre a_1 et b_1 .