



Université Claude Bernard



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **19 Septembre 2019**

Nom de famille et prénom de l'auteur : **M. COMIGNANI Ugo.**

Titre de la thèse : «**Spécification et réparation interactive de mappings en présence de polices de sécurité.** ».



La migration de données entre des sources aux schémas hétérogènes est un domaine en pleine croissance avec l'augmentation de la quantité de données en accès libre, et le regroupement des données à des fins de fouilles et d'apprentissage automatisé.

Cependant, la description du processus de transformation des données d'une instance source vers une instance définie sur un schéma différent est un processus complexe même pour un utilisateur expert dans ce domaine.

Cette thèse aborde le problème de la définition de mapping par un utilisateur non expert dans le domaine de la migration de données, ainsi que la vérification du respect par ce mapping des contraintes d'accès ayant été définies sur les données sources.

Pour cela, dans un premier temps nous proposons un système dans lequel l'utilisateur fournit un ensemble de petits exemples de ses données, et est amené à répondre à des questions booléennes simples afin de générer un mapping correspondant à ses besoins.

Dans un second temps, nous proposons un système permettant de réécrire le mapping produit de manière à assurer qu'il respecte un ensemble de vues de contrôle d'accès définis sur le schéma source du mapping.

Plus précisément, le premier grand axe de cette thèse est la formalisation du problème de la définition interactive de mappings, ainsi que la description d'un cadre formel pour la résolution de celui-ci. Cette approche formelle pour la résolution du problème de définition interactive de mappings est accompagnée de preuves de bonnes propriétés.

À la suite de cela, basés sur le cadre formel défini précédemment, nous proposons des algorithmes permettant de résoudre efficacement ce problème en pratique.

Ces algorithmes visent à réduire le nombre de questions auxquelles l'utilisateur doit répondre afin d'obtenir un mapping correspondant à ces besoins.

Pour cela, les mappings possibles sont ordonnés dans des structures de treillis imbriqués, afin de permettre un élagage efficace de l'espace des mappings à explorer.

Nous proposons également une extension de cette approche à l'utilisation de contraintes d'intégrité afin d'améliorer l'efficacité de l'élagage.

Le second axe majeur vise à proposer un processus de réécriture de mapping qui, étant donné un ensemble de vue de contrôle d'accès de référence, permet d'assurer que le mapping réécrit ne laisse l'accès à aucune information n'étant pas accessible via les vues de contrôle d'accès.

Pour cela, nous définissons un protocole de contrôle d'accès permettant de visualiser les informations accessibles ou non à travers un ensemble de vues de contrôle d'accès.

Ensuite, nous décrivons un ensemble d'algorithmes permettant la réécriture d'un mapping en un mapping sûr vis-à-vis d'un ensemble de vues de contrôle d'accès.

Comme précédemment, cette approche est complétée de preuves de bonnes propriétés.

Afin de réduire le nombre d'interactions nécessaires avec l'utilisateur lors de la réécriture d'un mapping, une approche permettant l'apprentissage des préférences de l'utilisateur est proposée, cela afin de permettre le choix entre un processus interactif ou automatique.

L'ensemble des algorithmes décrits dans cette thèse ont fait l'objet d'un prototypage et les expérimentations réalisées sur ceux-ci sont présentées.