

N° d'ordre : D -

THESE

présentée

devant l'Institut National des Sciences Appliquées de Rennes

en vue de l'obtention du

DOCTORAT

spécialité : Informatique

par M Mouhamadou Ba

Intitulé : Composition guidée de services - Application aux workflows d'analyse de données en bio-informatique

Directeur de Thèse : Mireille Ducassé

Date, heure et lieu de soutenance : 04/12/2015 à 14h à l'IRISA Rennes

Membres du jury (nom, prénom, titre et établissement de rattachement, fonction)

1. Cohen-Boulakia, Sarah, F, Maître de conférence à l'Université de Paris Sud, France (Rapporteuse)
2. Hacid, Mohand-Said, H, Professeur à l'Université Claude Bernard Lyon 1, France (Rapporteur)
3. Lavenier, Dominique, H, Directeur de recherche CNRS, France (Examineur)
4. Lo, Moussa, H, Professeur à l'Université Gaston Berger de Saint-Louis, Sénégal (Examineur)
5. Ducassé, Mireille, F, Professeur à l'INSA de Rennes, France (Directrice de thèse)
6. Ferré, Sébastien, H, Maître de conférence à l'Université Rennes 1, France (Co-Directeur de thèse)

RESUME DE LA THESE

Dans les domaines scientifiques, particulièrement en bio-informatique, des services élémentaires sont composés sous forme de workflows pour effectuer des expériences d'analyse de données complexes. À cause de l'hétérogénéité des ressources, la composition de services est une tâche difficile. Les utilisateurs, en composant des workflows, manquent d'assistance pour retrouver et interconnecter les services compatibles. Les solutions existantes utilisent des services spéciaux définis de manière manuelle pour gérer les conversions de formats de données entre les entrées et sorties des services dans les workflows. Cela est pénible pour un utilisateur final. Gérer les incompatibilités des services avec des convertisseurs manuels prend du temps et c'est lourd. Il existe des solutions automatisées pour faciliter la composition de workflows mais elles sont généralement limitées dans le guidage et l'adaptation des données entre services.

La première contribution de cette thèse propose de détecter systématiquement la convertibilité des sorties vers les entrées des services. La détection de convertibilité repose sur un système de règles basé sur une abstraction des types d'entrée et sortie des services. L'abstraction de types permet de considérer la nature et la composition des données d'entrée et sortie. Les règles permettent la décomposition et la composition ainsi que la spécialisation et la généralisation de types. Elles permettent également de générer des convertisseurs de données à utiliser entre services dans les workflows.

La deuxième contribution propose une approche interactive qui permet de guider des utilisateurs à composer des workflows en fournissant des suggestions de services et de liaisons compatibles basées sur la convertibilité de types d'entrée et sortie des services. L'approche est basée sur le modèle des Systèmes d'Information Logiques (LIS) qui permettent des requêtes et une navigation guidées et sûres sur des données représentées avec une logique uniforme. Avec notre approche, la composition de workflows est sûre et complète vis-à-vis de propriétés désirées.

Les résultats et les expériences, effectués sur des services et des types de données en bio-informatique, montrent la pertinence de nos approches. Nos approches offrent des mécanismes adaptés pour gérer les incompatibilités de services dans les workflows, en prenant en compte la structure composite des données d'entrée et sortie. Elles permettent également de guider, étape par étape, des utilisateurs à définir des workflows bien formés à travers des suggestions pertinentes.