

# Explication de décisions de réconciliation : approche fondée sur les réseaux de Petri colorés

Souhir Gahbiche\*, Nathalie Pernelle\*, Fatiha Saïs\*

\* LRI (CNRS & Université Paris-Sud) et INRIA Saclay

2-4 rue J. Monod, Parc-Club Orsay Université, F-91893 Orsay, France

{Souhir.Gahbiche@limsi.fr, Nathalie.Pernelle@lri.fr, Fatiha.Sais@lri.fr}

**Résumé.** L'objectif des systèmes d'intégration de données est de faciliter l'exploitation et l'interprétation d'informations hétérogènes provenant de différentes sources. Lorsque l'on doit intégrer de grands volumes de données, le recours à un expert n'est pas envisageable mais l'exploitation de processus d'intégration automatiques peut introduire des approximations ou des erreurs. Nous nous focalisons sur les résultats fournis par les méthodes de réconciliation de données. Ces dernières comparent les données entre elles et détectent celles qui réfèrent à la même entité du monde réel. Pour renforcer la confiance des utilisateurs dans les résultats retournés par ces méthodes, nous proposons dans cet article une approche d'explication graphique fondée sur les réseaux de Petri colorés qui est particulièrement adaptée aux approches de réconciliation globales, numériques et guidées par une ontologie.

## 1 Introduction

De nos jours, nous disposons de plus en plus d'informations provenant du web ou de différentes sources distantes. Ces informations sont créées à différents moments, par différentes personnes, pour répondre à divers besoins applicatifs, ce qui les rend inéluctablement hétérogènes. L'objectif des systèmes d'intégration de données est de faciliter l'exploitation et l'interprétation de ces informations hétérogènes en les intégrant dans un cadre uniforme de façon à donner l'illusion à l'utilisateur qu'il interroge une seule source. Pour réaliser de tels systèmes, il est possible de s'appuyer sur les ontologies afin d'intégrer les données sémantiquement. Les ontologies fournissent alors un vocabulaire structuré servant de support à la représentation des données et des requêtes.

Lorsque l'on s'intéresse à l'intégration sémantique de données, il faut faire face à deux problèmes de réconciliation. Le premier correspond au problème de réconciliation de schémas qui consiste à trouver des appariements entre les éléments (e.g concepts et relations) de deux schémas ou deux ontologies. Le deuxième problème est la réconciliation de données (ou réconciliation de références) qui consiste à mettre en correspondance des données provenant de différentes sources, et à détecter celles qui représentent la même entité du monde réel (e.g. la même personne, le même lieu, le même article). Ce problème est souvent critique. Ainsi, chaque année, de nombreuses entreprises doivent intégrer des milliers de références d'articles