
Editorial

La popularité du web et l'essor de XML ont contribué à l'émergence des bases de données semi-structurées et des bases de documents. Les modèles classiques de bases de données relationnelles ou objets supportent difficilement les données semi-structurées. En effet, ces données sont complexes, hétérogènes, distribuées, parfois incomplètes... Des modèles comme le modèle OEM (Object Exchange Model) ont été introduits dans la communauté des bases de données pour supporter des données complexes de structures de graphes irrégulières comme les documents XML. La force des modèles semi-structurés est de ne plus imposer de structure a priori dans le schéma mais de la définir a posteriori dans les données elles-mêmes.

Pour réaliser des systèmes de données semi-structurés et multimédias, il est nécessaire que ces systèmes fournissent les mêmes fonctionnalités que les SGBD classiques : le stockage efficace des données dans des « entrepôts » (repository) et leur interrogation. En ce qui concerne le stockage, on distingue deux approches principales : la première s'appuie sur un système de gestion de bases de données existant et est construit en surcouche de ce dernier, la seconde propose un stockage de documents semi-structurés dans une structure native (souvent inspirée par le langage XML). De plus, afin de pouvoir interroger des données semi-structurées, de nombreux langages de requêtes ont été proposés : OEM-QL, Lorel, XML-QL, etc., et la norme XQUERY du W3C. Pour l'exécution des requêtes, le problème de disponibilité des données se pose. Les documents XML sont par nature répartis et de formats variés. D'où l'introduction du concept de médiation pour l'interrogation de sources de données hétérogènes et distribuées. La médiation consiste à décomposer une requête en sous-requêtes réparties sur plusieurs sources de données. Plus récemment, les chercheurs s'intéressent à l'extraction de connaissances à partir de données semi-structurées du web (web mining).

L'objectif de ce numéro spécial est de faire le point sur les avancées actuelles dans le domaine des bases de données semi-structurées d'un point de vue théorique et de recenser les applications originales qui s'appuient sur ce formalisme. Nous avons réuni dans ce numéro différents points de vue sur des modèles, systèmes, applications qui ne prétendent pas répondre à toutes les questions évoquées plus haut.

L'article de T.T. Dang-Ngoc et G. Gardarin décrit le système libre XMedia pour la médiation de sources de données hétérogènes. L'architecture du système ainsi que l'algorithme de traitement des requêtes y sont présentés en détail. Le médiateur s'appuie sur une nouvelle algèbre de traitement de XML appelée XAlgebra. Les auteurs apportent également des idées novatrices sur l'optimisation de requêtes.

M.C. Fauvet et S. Baina proposent également un modèle pour l'évaluation coopérative de requêtes sur des sources de données semi-structurées distribuées. Le modèle est illustré sur un scénario issu des services web.

L'article de P.A Laur, M. Teisseire et P. Poncelet s'intéresse à l'extraction des connaissances dans des bases de données semi-structurées du web et présente une approche originale pour la recherche de structures régulières. De plus, leur proposition est étendue pour prendre en compte l'évolution des données sources.

O. Boussaid, F. Bentayeb, J. Darmont et S. Rabaseda mènent une réflexion sur les aspects multidimensionnels dans le contexte des données complexes. Leur article survole la réutilisation des techniques classiques dans le contexte des données complexes (données hétérogènes et multimédias comme texte, vidéo ou bases de données stockées dans différents formats) en mettant en évidence certains des problèmes posés. En particulier, ils proposent une démarche pour le stockage de données complexes.

Dans leur article, F. Laforest et Y. Badr présentent un système de couplage entre documents semi-structurés et bases de données : le système DRUID. Ce système permet d'analyser des documents semi-structurés pour en extraire les informations pertinentes qui seront ensuite mémorisées dans une base de données relationnelle (les documents sont ainsi centrés données). Les règles d'extraction sont basées sur XSL et des transducteurs à états finis.

K. Khrouf, F. Ravat et C. Soulé-Dupuy se placent dans le cadre des entrepôts de documents qui permettent le stockage de documents hétérogènes et leur classification selon des structures logiques génériques. L'objectif de leur article est de présenter une méthode de comparaison de structures logiques. Cette méthode est basée sur la définition d'un calcul de similarité d'arborescences hétérogènes d'éléments ordonnés et étiquetés.

I. Amous, A. Jedidi et F. Sèdes proposent une méthode d'extraction automatique de métadonnées dans des documents multimédias semi-structurés. Leurs travaux se situent dans le contexte de la recherche d'information : reconstruction dynamique de documents répondant aux besoins des utilisateurs.

Les travaux présentés par I. Mbaye, J. Martinez et R. Oulad Haj Thami s'intègrent dans la conception d'un outil de navigation dans une base de vidéos et d'images. Les auteurs présentent l'élaboration d'un schéma de données générique de la partie audiovisuelle du SGBD. Leur proposition met en avant deux aspects : la structuration d'une vidéo et son indexation.

Nous tenons à remercier tous les auteurs qui ont soumis un article pour ce numéro spécial ainsi que tous les membres du comité de lecture qui ont effectué un excellent travail de relecture.

Sylvie Calabretto
Jean-Marie Pinon
INSA de Lyon

COMITE DE LECTURE

Laurent Amsaleg
Université de Rennes

Mohand Boughanem
Université de Toulouse

Sylvie Calabretto
INSA de Lyon

Anne Doucet
Université de Paris

Mohand-Saïd Hacid
Université de Lyon

Frédérique Laforest
INSA de Lyon

Jacques Le Maitre
Université de Toulon et du Var

Philippe Lopisteguy
Université de Pau et des Pays de l'Adour

Serge Miranda
Université de Nice

Noureddine Mouaddib
Université de Nantes

Jean-Marie Pinon
INSA de Lyon

Pascal Poncelet
Ecole des Mines d'Alès

Claudia Roncancio
INP Grenoble

Farouk Toumani
Université de Clermont-Ferrand

Gilles Zurfluh
Université de Toulouse