

Modélisation multidimensionnelle d'entrepôts de documents XML répartis

Ines Ben Messaoud*, Jamel Feki*
Gilles Zufluh**

*Laboratoire Mir@cl, Faculté des Sciences Economiques et de Gestion, Université de Sfax
Route de l'Aéroport km 4, 3018 Sfax, BP. 1088 - Tunisie
{ines.benmessaoud ; jamel.feki}@fsegs.rnu.tn

**IRIT, Institut de Recherche en Informatique de Toulouse, Université Toulouse 1,
2 Rue du Doyen Gabriel Marty, 31042 Toulouse Cedex 9 – France
zurfluh@univ-tlse1.fr

Résumé. De nos jours, et avec l'ouverture des organisations sur Internet, les documents constituent une source intéressante pour les analyses décisionnelles; ils aident les décideurs à mieux comprendre l'évolution des processus métier de leur organisation. Généralement, ces documents existent sous format XML, sont géographiquement répartis et décrits par des structures différentes. Cet article propose une méthode de construction d'entrepôts de documents distribués comportant deux étapes : *i) Unification des structures des documents XML*, et *ii) Modélisation multidimensionnelle des arbres unifiés*. Plus précisément, il focalise sur l'étape de modélisation multidimensionnelle.

1 Introduction

Face à la progression rapide des technologies de l'information, les organisations se situent dans un environnement caractérisé par un volume croissant de données qui transitent par leur système d'information. Pour faire face aux différentes pressions, d'ores et déjà internes et externes, elles doivent mettre à la disposition de leurs dirigeants des supports d'assistance à la bonne prise de décisions en respectant des contraintes de délais raisonnables. Les entrepôts de données et les technologies OLAP (« *On-Line Analytical Processing* ») soutiennent de tels besoins. En effet, ils permettent d'analyser de grandes volumétries de données structurées que les organisations stockent dans des bases de données Pérez et al. (2008). Ces bases sont construites à partir de données transactionnelles extraites des systèmes d'information des entreprises. Cependant, seuls 20% des données d'un système d'information sont des données transactionnelles et peuvent être traitées par un système OLAP et les 80% restants sont constitués de documents (rapports, articles, etc.) Tseng et Chou (2006).

Par ailleurs, les documents constituent une capitalisation importante des connaissances du système de production et sont utiles pour le système de pilotage. Généralement, le contenu de ces documents est peu structuré d'où la difficulté de les intégrer dans les systèmes d'information décisionnels Ronan (2007). En conséquence, il résulte lors du processus décisionnel, que les analystes-décideurs n'arrivent pas à explorer facilement, rapidement et effi-