
Editorial

Depuis l'essor des technologies de l'information et de la communication, le volume d'information stockée électroniquement ne cesse de s'accroître (bibliothèques numériques, web, intranets, extranets...). Il est entendu que plus de la moitié des informations utilisées au quotidien à des fins de veille économique, stratégique, scientifique ou technique sont à ce jour contenues dans des documents. De fait, ces documents et données semi-structurées doivent être intégrés et stockés afin d'être manipulés et analysés en utilisant conjointement des modèles et techniques issus de plusieurs domaines de l'informatique et notamment de la recherche d'information et des bases de données. Le concept d'entreposage de documents permet de construire cette mémoire documentaire « métier » flexible et partageable. Ces répertoires ou entrepôts de documents doivent permettre d'une part l'intégration et le stockage de documents de données semi-structurées issus de sources différentes et de structures hétérogènes, et, d'autre part, une recherche personnalisée comme collaborative, ainsi qu'une analyse multidimensionnelle des informations extraites de ces documents.

L'objectif de ce numéro spécial est de rassembler une sélection de travaux de recherche récents et représentatifs des courants actuels dans le domaine de l'entreposage de documents et de données semi-structurées. Même s'ils ne couvrent que partiellement le domaine et les recherches en cours, ces articles présentent un échantillon représentatif des verrous à lever, comme des différents aspects à prendre en compte en matière d'entreposage de documents.

Les deux premiers articles se focalisent sur les aspects structure sémantique et multistructuralité des documents, les deux suivants sur la modélisation et l'analyse multidimensionnelles des documents et des données semi-structurées ; enfin les deux derniers traitent de la préservation des documents et des usages qui en découlent, puis de la dimension temporelle des informations géographiques.

Le premier article, « Accès au contenu des thèses numériques par leur structure sémantique » (R. Abascal-Mena et B. Rumpler), étudie la structuration logique et sémantique d'un corpus de thèses scientifiques en informatique afin de définir des tags sémantiques qui permettront eux-mêmes par la suite d'affiner la description sémantique de nouvelles thèses.

Le deuxième article, « Une approche multivue pour la gestion des documents multistructurés » (K. Djemal, M. Mbarki et N. Vallès-Parlangeau) aborde le problème de la modélisation et de l'entreposage de documents multistructurés. Un métamodèle ainsi que des exemples d'exploitation de cette multistructuralité sont décrits.

Le troisième article, « Entreposage et exploitation de documents multidimensionnels évolutifs : le cas des tableaux statistiques » (V. Detienne, F. Vesentini et J-L. Hainaut), s'intéresse tout particulièrement à l'extraction, et à la représentation dans un même entrepôt, d'informations issues de tableaux statistiques (sous forme d'images numérisées) montrant des évolutions de tendances dans le temps, et décrivant des faits mesurés selon différentes dimensions. Des méthodes et outils sont proposés et discutés.

Le quatrième article, « Analyse multidimensionnelle de documents via des dimensions OLAP » (F. Ravat, O. Teste et R. Tournier), présente une solution basée sur XML pour intégrer des documents textuels dans un environnement OLAP. Plus précisément, les auteurs proposent une nouvelle définition du modèle en étoile pour exploiter des documents textuels structurés (utilisation de mesures adaptées au texte).

Le cinquième article, « Le document numérique entre préservation et usage » (Y. Keraron) aborde la problématique de la préservation dans le temps de l'accessibilité et de l'interprétabilité des informations archivées. Cet article apporte une solution par le biais d'un modèle de référentiel de documents numériques et discute du cas particulier des données et des documents techniques.

Le sixième article, « La composante temps dans l'information géographique textuelle » (A. Le Parc-Lacayrelle, M. Gaio et C. Sallaberry), propose un modèle de description d'expressions temporelles, une méthodologie d'annotation de ces expressions en vue de leur indexation automatique et une méthode de recherche d'information géographique.

Nous remercions les auteurs ayant proposé leur contribution, ainsi que les membres du comité de rédaction et du comité de lecture dont les recommandations ont permis la sélection et la relecture des articles de ce numéro.

SYLVIE CALABRETTO
LIRIS – INSA Lyon

CHANTAL SOULÉ-DUPUY
IRIT – Université Toulouse 1