

# Interrogation de systèmes d'intégration de données

## Etat de l'art et perspectives

Cheikh Niang<sup>\*,\*\*</sup> — Béatrice Bouchou<sup>\*</sup> — Moussa Lo<sup>\*\*</sup>

<sup>\*</sup> Laboratoire d'Informatique (LI)  
Université François Rabelais de Tours - Antenne Universitaire de Blois  
3 place Jean Jaures, 41000 Blois(FRANCE)  
{cheikh.niang - beatrice.bouchou}@univ-tours.fr

<sup>\*\*</sup> Laboratoire d'Analyse Numérique et d'Informatique (LANI)  
Université Gaston Berger de Saint-Louis  
BP 234 Saint-Louis(SENEGAL)  
lom@ugb.sn

**RÉSUMÉ.** Le domaine de l'intégration de données est étudié depuis longtemps notamment par la communauté base de données, mais il y reste néanmoins des problèmes ouverts. Les dernières approches proposées intègrent les standards du web sémantique et proposent des solutions pour faciliter l'automatisation du processus d'intégration. Le but de cet article est d'abord de passer en revue les approches et méthodes existante pour réaliser l'intégration des données, ensuite de présenter un modèle particulier de système d'intégration, réalisé dans le cadre du projet Sic-Sénégal, et enfin de donner les directions des travaux que nous comptons réaliser afin d'apporter nos contributions à l'amélioration de la qualité des résultats fournis par les systèmes d'intégration de données.

**ABSTRACT.** Data Integration is a domain which has been extensively studied, in particular by the Database community, but some open problems remain. Recently proposed approaches make use of Semantic Web standards and offer solutions to facilitate integration process automation. The aim of the paper is firstly to review existant approaches and methods for data integration, secondly to present a concrete model of data integration, achieved in the context of SIC-Senegal project, and finally to sketch future work directions that we intend to study in order to improve result's quality of data integration systems.

**MOTS-CLÉS :** Intégration de données, Web sémantique, Interrogation

**KEYWORDS :** Data integration, Semantic web, querying

---

## 1. Introduction

Aujourd'hui, les organisations utilisant les systèmes d'information devront de plus en plus faire face à la distribution et à l'évolution croissante de leurs sources de données. En général, ces sources de données, qui peuvent être des documents et des données structurées ou semi-structurées, sont autonomes et gérés par des systèmes différents. Par ailleurs, beaucoup de ces sources sont reliées tout en étant sémantiquement hétérogènes : elles modélisent la même réalité externe tout en utilisant des concepts et des structures distincts. Or, pour un traitement efficace et rapide des informations, l'interopérabilité entre ces différentes sources distribuées et hétérogènes est nécessaire. Cette interopérabilité passe par ce que l'on nomme un processus d'intégration de données. C'est un processus par lequel plusieurs sources d'information, distribuées, hétérogènes et autonomes sont intégrées sous forme d'une source unique, par l'intermédiaire d'un système d'intégration de données.

L'objectif d'un système d'intégration de données est donc de proposer une interface homogène pour interroger plusieurs sources, qui peuvent être distribuées, hétérogènes et autonomes. Pour mettre en place un tel système plusieurs défis sont à relever, notamment ceux de l'hétérogénéité structurelle et sémantique des sources de données et du traitement efficace des requêtes.

Durant ces dernières années beaucoup de projets de recherche [6, 7, 9, 10, 14, 16, 23, 34, 35, 39, 42] se sont intéressés à cette problématique. Ces travaux présentent des solutions distinctes pour l'intégration des données, et chaque solution proposée est spécifique à chacun des problèmes évoqués. Néanmoins, des contributions restent à être apportées pour l'amélioration des performances de ces systèmes, relatives notamment à leur interrogation par un utilisateur quelconque et aux traitements des requêtes soumises par ce dernier. C'est ce qui motive le travail que nous présentons dans cet article.

Le reste de l'article est organisé comme suit. La première partie présente un état de l'art sur les différentes approches proposées pour la mise en œuvre des systèmes d'intégration de données. La deuxième partie décrit le processus d'intégration adopté dans le cadre du projet Sic-Sénégal et son positionnement par rapport aux autres approches déjà existantes. Enfin, dans la dernière partie nous donnons les grandes lignes de nos travaux futurs.

---

## 2. Système d'intégration de données : un état de l'art

### 2.1. Bref historique

L'intégration de données a été étudiée depuis longtemps par la communauté base de données [4, 11]. Les premières approches d'intégration ont été réalisées dans le cadre de système de bases de données (BDs) relationnelles, objets/relationnelles ou objets [3, 36, 43]. Ainsi, la plupart d'entre elles ont adapté des outils standards autour du modèle relationnel, comme les langages d'interrogations et les vues. Cependant, l'émergence de XML comme nouveau format standard pour la publication et l'échange de données a un peu changé la donne. Beaucoup de systèmes et d'applications exportent leurs données en XML et publient leur contenu sous forme de DTD et de schémas XML. Ainsi, indépen-

damment du format de stockage des données qui peut être XML, relationnel, orienté-objet ou autre, la vue présentée à l'utilisateur est XML [7]. Toutefois, les nombreux avantages présentés par XML et la richesse de son modèle de schéma ont aussi permis à de nouveaux projets de recherche d'émerger pour faire face à ces nouveaux défis d'intégration de données. Les systèmes d'intégration proposés [9, 24, 26, 30] permettent d'interroger des sources de données hétérogènes d'une manière uniforme en se basant sur le modèle d'échange XML. Certains de ces systèmes comme [9, 30] sont même destinés à l'intégration des nombreuses et diverses ressources accessibles via le web.

Dans ce contexte d'intégration à grande échelle apparaissent de nouveaux challenges relatifs notamment au nombre et à la diversité des sources d'information accessibles via le web, qui peuvent être de simples sources de données et/ou des systèmes intégrant d'autres services. Actuellement, les nouvelles initiatives prises pour lever ces verrous scientifiques s'orientent vers les solutions proposées dans le cadre du web sémantique. En effet, l'expression Web sémantique, attribuée à Tim Berners-Lee au sein du W3C, fait référence à la vision du Web de demain comme un vaste espace d'échange de ressources entre êtres humains et machines permettant une exploitation, qualitativement supérieure, de grands volumes d'informations et de services variés. Il repose sur des langages et une infrastructure dont l'objectif est de se donner la possibilité d'enrichir le Web à l'aide d'informations dites " sémantiques ", qui peuvent être traitées par des machines non seulement pour être affichées et lues par des humains, mais également pour être réutilisées par des applications qui faciliteront l'intégration, la recherche et l'usage des ressources Web .

## 2.2. Entrepôts vs Médiateurs

Un des points communs des travaux évoqués précédemment est l'utilisation, soit d'une approche virtuelle, soit d'une approche matérialisée. Il s'agit de deux approches utilisées pour interroger de façon simultanée plusieurs sources d'information. Elles préconisent la définition d'un schéma global et l'objectif est de permettre aux utilisateurs d'interroger les données comme une seule source avec ce schéma. L'approche matérialisée [9, 27, 43] applique le principe des vues matérialisées et intègre les données en accord avec le schéma global. Toutes les données sources conformes au schéma global sont alors copiées vers un emplacement unique appelé entrepôt, ce qui est un peu problématique lorsqu'on dispose d'une énorme masse de données. En outre, cette migration massive des données oblige une maintenance permanente de l'entrepôt afin de garder la cohérence entre les données sources et celles transformées. Par contre, elle facilite l'optimisation et le traitement des requêtes, en tirant profit du fait que les données sont stockées localement en fonction d'un seul schéma global. Cette gestion facile des requêtes n'est pas aussi évidente dans l'approche médiateur. En effet, dans cette approche [7, 10, 39] toutes les données restent au niveau de leur source locale et elles sont interrogées par l'intermédiaire du schéma global. Ainsi, une requête exprimée sur le schéma global doit être décomposée et réécrite en requêtes locales adressées aux sources. Les résultats obtenus à partir des différentes sources sont ensuite combinés conformément au schéma global avant d'être présentés à l'utilisateur. L'avantage avec cette approche est que l'on n'a pas une matérialisation des données mais seulement une migration des requêtes, ce qui permet aux sources de garder leur autonomie.

### 2.3. Différentes architectures de collaboration

L'approche médiateur peut être mise en œuvre dans deux configurations différentes : dans une architecture centralisée [7, 10, 16, 23, 42] et dans une autre qui est celle des systèmes peer-to-peer [6, 13, 14]. Dans une architecture centrale, les requêtes sont formulées par l'utilisateur dans les termes du schéma global, elles sont ensuite réécrites en requêtes qui sont évaluées par les sources. Les réponses retournées par chacune des sources sont combinées et transformées afin d'être compatibles avec le schéma global et conformes à la requête posée par l'utilisateur. Toutes ces tâches sous-jacentes d'intégration c'est-à-dire le schéma global, la traduction des requêtes et l'intégration des résultats, sont accomplies par un médiateur. Il offre ainsi une interface uniforme à l'utilisateur en lui permettant d'accéder de manière transparente aux informations stockées dans les sources locales.

La configuration peer-to-peer permet, au contraire, à chaque pair, contenant une source de données, de recevoir une requête et de la traiter. Le schéma du pair interrogé est en effet l'interface virtuelle à travers laquelle les requêtes sont formulées. Ces requêtes sont par la suite reformulées sur le schéma des autres pairs à travers des correspondances définies entre les différents pairs.

### 2.4. Différentes définitions des correspondances (LAV vs GAV)

Afin de répondre aux requêtes, des correspondances (mappings) doivent être définies, d'une part, entre le schéma global et les schémas sources, pour l'architecture centrale, et, d'autre part, entre les pairs, pour l'architecture peer-to-peer. Les deux grandes approches utilisées pour construire de telles correspondances sont l'approche GAV (Global As View) et l'approche LAV (Local As View). Dans l'approche GAV [35, 23], le schéma global est défini comme une vue sur les schémas sources. L'avantage majeur de cette approche est la simplicité du traitement des requêtes. La définition des mappings est assez explicite pour permettre une traduction facile d'une requête utilisateur, en remplaçant les prédicats du schéma global de la requête par leur définition. Par contre, la définition de la vue globale prend en compte l'ensemble des schémas sources et tout changement de schéma au niveau des sources nécessite une vérification générale du schéma global pour la détection des modifications nécessaires. Au contraire, avec l'approche LAV [7, 10] les changements effectués sur les schémas des sources n'affecteront pas le schéma global, car les schémas sources sont considérés comme des vues sur le schéma global. Seule la définition de la vue sera donc adaptée. Cependant, l'évaluation des requêtes peut s'avérer un peu complexe.

### 2.5. Les ontologies dans l'intégration de données

Beaucoup des travaux déjà cités sont des réponses au problème de l'hétérogénéité structurelle des données. Cependant, un autre aspect essentiel à prendre en compte dans un système d'intégration est le traitement de l'hétérogénéité sémantique des données. En général, les données à intégrer sont mises en place par des acteurs différents, chacun avec son propre vocabulaire. Ainsi, des différences sémantiques liées à la représentation du sens associé à la description des données peuvent subsister. Pour réduire cette hétérogénéité, de plus en plus d'approches procèdent à la représentation conceptuelle des données et de leurs relations. Le schéma global ne représente plus des contraintes de structure (relations ou schéma XML) mais des contraintes sémantiques. Cette conceptualisation est formalisée à l'aide d'une ontologie. L'ontologie qui, elle-même, est définie comme étant

une spécification explicite et formelle d'une conceptualisation d'un domaine de connaissance [37, 38]. Elle définit des concepts (principes, idées, catégories d'objets, notions potentiellement abstraites) et des relations. Elle inclut généralement une organisation hiérarchique des concepts pertinents et des relations qui existent entre ces concepts, ainsi que des rôles et axiomes qui les contraignent [17].

En raison de leurs potentiels de description de la sémantique des connaissances, les ontologies peuvent jouer plusieurs rôles dans le processus d'intégration. Elles peuvent être utilisées pour établir les liens sémantiques entre des éléments dans des sources différentes [16]. Elles peuvent aussi servir de modèle d'interrogation pour le système intégré lorsqu'elles sont utilisées pour décrire le schéma global [10, 42]. Toutefois la méthode de conception des ontologies et la manière dont elles sont utilisées peuvent être différentes.

Classiquement, il existe trois architectures en fonction de la façon dont les ontologies sont utilisées au sein d'une infrastructure d'intégration [21] :

**L'approche mono-ontologie** : Toutes les sources de données sont reliées directement à une ontologie globale, qui fournit un vocabulaire partagé pour la spécification de la sémantique et qui offre aussi une interface uniforme aux utilisateurs. Cette approche est surtout utilisée lorsque toutes les sources d'information à intégrer présentent une même vision d'un certain domaine. [7] et [42] sont des exemples typiques d'implémentation de cette approche.

**L'approche multi-ontologies** : Chaque source de données est décrite par sa propre ontologie. Au lieu de l'utilisation d'une ontologie globale, des correspondances entre les différentes ontologies locales sont établies afin de disposer d'une interprétation commune des données. Ces correspondances peuvent se révéler parfois très complexes, notamment à cause des niveaux de granularité différents entre les ontologies. Le système OBSERVER [16] s'inscrit dans cette approche.

**L'approche hybride** : C'est une combinaison des deux approches précédentes. Comme dans l'approche multi-ontologies, une ontologie locale est définie pour chaque source. En plus de cela, comme dans l'approche mono-ontologie, des correspondances sont définies entre les ontologies locales et une ontologie globale. L'avantage de cette approche est que de nouvelles sources peuvent être ajoutées facilement sans modifier les mappings déjà existant. [22] est un exemple de cette approche.

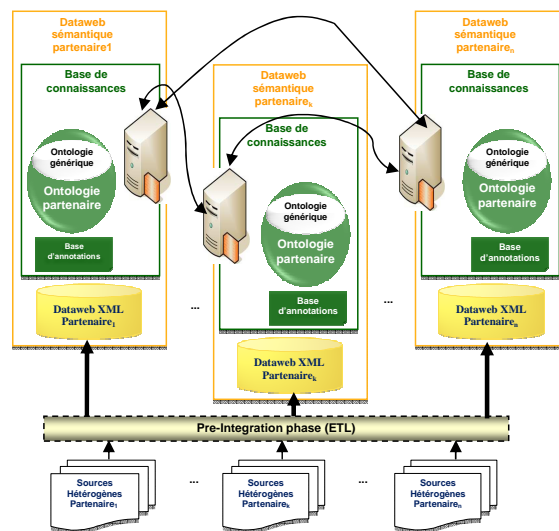
## 2.6. Traitement des requêtes dans un système d'intégration de données

L'objectif principal d'un système d'intégration est de répondre efficacement aux requêtes des utilisateurs. Le traitement des requêtes dans de tels systèmes n'est pas toujours une tâche facile. Les requêtes sont généralement formulées dans les termes du schéma médiateur, elles sont ensuite décomposées en une série de requêtes traduites et évaluées sur les sources de données locales. Ce processus de traduction, plus connu sous les termes de reformulation de requête, dépend de l'approche GAV ou LAV utilisée pour établir les mappings. Dans le cas de l'approche GAV, elle s'effectue très simplement, par simple dépilement de la requête. Il s'agit là de remplacer les vues par leur définition. La reformulation est beaucoup plus complexe dans l'approche LAV, puisqu'il faut réécrire la requête en termes de vues. D'un aspect modulaire, le traitement d'une requête dans un système d'intégration de données peut passer par les étapes suivantes [19] :

- Analyse syntaxique et sémantique de la requête, y compris sa reformulation.
- Décomposition de la requête.
- Traitement de la requête dans les sources locales, c'est-à-dire la transformation de la requête dans les termes du langage d'interrogation de chaque source locale puis la transformation du résultat dans le modèle de données commun du système d'intégration.
- Recomposition des résultats partiels en considérant la cohérence des données.

### 3. Le projet SIC-Sénégal : Un cadre applicatif

Le travail que nous effectuons s'inscrit dans le contexte du projet SIC-Web Sénégal, initié à l'Université de Saint-Louis en juin 2004. Ce projet a pour objectif de spécifier et réaliser une plate-forme logicielle à la disposition des producteurs de données (experts, organismes) et des consommateurs de données (décideurs, bailleurs) pour faciliter l'intégration, la gestion, l'organisation, la diffusion, et l'exploitation des ressources (données et applicatifs) utilisées dans le processus de mise en valeur de la région du fleuve Sénégal. D'autres travaux [8, 15, 32, 33] ont été déjà réalisés dans le cadre de ce projet et une architecture d'intégration des données (Figure 1) a été proposée. Le processus d'intégration se déroule en deux phases : la première phase consiste en la création d'un entrepôt de documents XML, construit à partir de données issues des sources hétérogènes (structurées et/ou semi-structurées) et appelé dataweb [2, 28, 29, 30]. Dans la deuxième phase, il s'agit d'intégrer ces différents dataweb.



**Figure 1.** Architecture du système d'intégration.

L'architecture repose sur un système à base de hubs [18]. C'est une architecture N-M c'est-à-dire une architecture dans laquelle N systèmes d'intégration permettent d'interroger M sources, contrairement aux approches classiques de types base de données où l'objectif est de construire un système intégré à partir de N sources. Ces genres de système [12, 31] sont apparus avec l'émergence du système décentralisé qu'est l'Internet. Le

but est de définir des approches qui permettent à chaque acteur de se fabriquer sa propre vue intégrée des données. Ici, comme dans [31], il existe une ontologie partagée (ontologie de domaine) qui est référencée par chaque source, disposant elle aussi de sa propre ontologie (ontologie locale). Des relations sémantiques existantes sont alors définies entre les concepts des ontologies locales et ceux de l'ontologie de domaine (articulation). Ces articulations sont exprimées par un ensemble de relations de subsomption. Dans notre contexte :

- l'ontologie locale de chaque source est construite automatiquement à partir des documents XML contenus dans son dataweb.
- Les correspondances entre l'ontologie locale et les documents XML sont définies dans une base d'annotations. Chaque concept de l'ontologie locale est instancié dans la base d'annotations avec une propriété permettant de spécifier le chemin XPATH indiquant la source XML d'où a été extrait ce concept.
- Chaque source garde l'articulation entre son ontologie et l'ontologie de domaine dans une ontologie dite générique. Contrairement à [12] et [13], l'ontologie globale, qui fournit aux utilisateurs le vocabulaire de base approprié pour formuler leurs requêtes, n'est pas entièrement l'ontologie de domaine, mais une sous partie de cette ontologie. Elle est constituée par la fusion de l'ensemble des ontologies génériques des sources participant au processus d'intégration. L'ontologie globale est donc une ontologie de domaine qui ne couvre que les données concernées par l'intégration.

---

## 4. Conclusion et perspectives

L'une des difficultés rencontrées pour la validation applicative et notamment l'évaluation des performances de ce système d'intégration est la mise en place d'un langage et d'une interface de requête. Cela fera l'objet de notre travail de thèse. En effet, dans le cadre de notre travail, nous nous intéressons aux traitements des requêtes dans un système d'intégration de données.

Nous pensons que les performances d'un système d'intégration de données peuvent être évaluées suivant deux facteurs : le temps que met le système pour répondre aux requêtes et la pertinence des résultats restitués aux utilisateurs. Si l'optimisation du temps de traitement des requêtes est largement abordée dans beaucoup d'approches [1, 20, 40], qui proposent des techniques d'optimisation, de transformation et de réécriture de requête, d'indexation de données, de copie (ou cache) des résultats de requêtes pour améliorer les performances du système, il n'en est pas de même concernant la satisfiabilité de la demande formulée par l'utilisateur. En effet, dans un système d'intégration l'utilisateur n'a pas une connaissance assez explicite de la structure des données, ni de leurs sémantiques et encore moins de leur répartition. En conséquence, les différents éléments de réponse nécessaires à la satisfaction de sa requête peuvent se trouver dans plusieurs sources différentes, formulés différemment.

Pour apporter une réponse pertinente, le système doit être capable au minimum : (1) d'identifier les sources susceptibles d'avoir des éléments de réponse ; (2) d'interroger les sources distantes sur les éléments qu'elles connaissent ; (3) de décomposer la requête et router les sous-requêtes vers les éventuelles sources concernées ; (4) de recomposer les résultats à partir des réponses partielles, tout en gardant la cohérence entre les données.

Ainsi, dans un premier temps nous allons donc développer une méthode d'interrogation de données XML distribuées et hétérogènes. Ces données sont interrogeables au niveau de chaque source par le langage de requête XQUERY [41], le standard du W3C pour l'interrogation de documents XML. Cependant, la démarche de médiation adoptée a nécessité la mise en place d'une base de connaissances associées à chaque source de données à intégrer. Cette base de connaissances se situe au-dessus des sources de données et contient notamment une ontologie globale, qui sert de support aux utilisateurs pour l'expression de leurs requêtes et qui également permet d'établir une connexion entre les différentes sources. Les formalismes de représentation des connaissances utilisés, pour l'implémentation de la base de connaissances, sont les standards du web sémantique recommandés par le W3C. Il s'agit du langage OWL, pour l'implémentation des ontologies, du langage RDF, pour la définition des annotations et du langage de requête SPARQL, pour l'interrogation de la base de connaissances.

Nous allons ainsi proposer une approche d'interrogation permettant d'abord d'exploiter la base de connaissances pour ensuite interroger efficacement les données XML distribuées. La requête initiale sera formulée en SPARQL, selon les termes de l'ontologie globale. Il s'agira ensuite, pour le médiateur sollicité, (1) de trouver les sources impliquées par cette requête ; (2) d'envoyer à chacune d'elle la partie de la requête qui le concerne ; (3) de recevoir les réponses partielles, de les fusionner et de présenter une réponse globale à l'utilisateur. Pour donner sa contribution, chaque médiateur participant à la résolution de la requête devra donc interroger ses sources XML avec la sous-requête SPARQL reçue. Nous procéderons alors à la traduction des sous-requêtes SPARQL en requêtes XQUERY adressées aux sources. La base d'annotations permettra à chaque médiateur, par simple interrogation, de trouver les expressions XPATH correspondant aux concepts de la requête SPARQL. Il devient dès lors facile de réécrire la requête SPARQL en requête XQUERY, avec un algorithme optimal que nous proposerons à cet effet.

Nous avons l'intuition que le système doit aussi pouvoir profiter des mécanismes de raisonnement automatique offerts par les technologies du web sémantique pour inférer sur les données afin d'améliorer la précision des résultats et de déduire d'autres résultats, qui peuvent être aussi pertinents pour l'utilisateur que ceux qui sont directement issus de la requête qu'il a formulée. Dans cet ordre d'idée, nous étudierons les possibilités d'adapter des résultats de référence dans le domaine de la recherche d'information, nous pensons en particulier aux systèmes d'expansion de requêtes [5, 25].

"Une question bien posée est à moitié résolue". Pour bien exprimer sa requête, l'utilisateur doit disposer d'outils qui lui sont adaptés. En général, les langages de requêtes utilisés dans les systèmes d'intégration ne sont pas tellement connus des utilisateurs finaux, leur utilisation peut se révéler contraignante pour eux. Il est donc nécessaire de les rendre un peu naturels. Le système doit aussi disposer d'interfaces intelligentes assistant l'utilisateur dans la formulation de ses requêtes. La dernière grande tâche de notre travail consistera à adapter le système aux utilisateurs finaux, non expert du domaine. Pour cela nous nous appuierons sur des travaux réalisées dans le domaine du traitement automatique de la langue naturelle.

C'est à ces différentes problématiques que nous allons nous attaquer afin d'apporter nos contributions sur l'interrogation de systèmes d'intégration de données, en particulier celui proposé dans le cadre du projet SIC-Web Sénégal. Il faut ajouter qu'au-delà de leur cadre applicatif, nos recherches visent à proposer de nouvelles solutions d'orchestration



des ressources et des connaissances distribuées sur le web, en vue de leur restitution personnalisée aux utilisateurs finaux.

---

## 5. Bibliographie

- [1] A. HALEVY, A. RAJARAMAN, J. ORDILLE, « Data integration : the teenage years », *LVLDB '06 : Proceedings of the 32nd international conference on Very large data bases*, pages 9-16. VLDB Endowment.
- [2] A. HOCINE, M. LO, « Modeling and information retrieval on XML-dataweb », *Lecture Notes in Computer Science*, vol. 1909, Springer, pp. 398-408, ADVIS'00, 25-27 Octobre, 2000.
- [3] A. LEVY, A. RAJARAMAN, J. ORDILLE, « Querying Heterogeneous Information Sources Using Source Descriptions », *Proc. VLDB*, Mumbai (Bombay), India, September 1996, p. 251-262.
- [4] A. P. SHETH, J. LARSON, « Federated database systems for managing heterogeneous, distributed and autonomous Databases », *9ACM Computing Surveys*, vol. 22, n° 3, 1990.
- [5] A. VENTRESQUE, S. CAZALENS, P. LAMARRE, P. VALDURIEZ, « Query Expansion and Interpretation to Go Beyond Semantic P2P Interoperability », *OTM Confederated International Conferences CoopIS, DOA, ODBASE, GADA, and IS 2007*, Springer, pp. 870-877.
- [6] A. Y. HALEVY, Z. G. IVES, Z. G. IVES, I. TATARINOV, « Piazza : Data Management Infrastructure for Semantic Web Applications », *International World Wide Web Conference (WWW)*, pages 556-567, 2003.
- [7] B. AMANN, C. BEERI, I. FUNDULAKI, M. SCHOLL, « Ontology-Based Integration of XML Web Resources », *International Semantic Web Conference (ISWC)*, pages 117-131, 2002.
- [8] C. A. T. NIANG, « Atelier d'Intégration de Données Environnementales Issues de Sources Hétérogènes », Mémoire de Master 2 Recherche Informatique, Université Gaston Berger, Saint-Louis, Juillet 2008.
- [9] C. DELOBEL, C. REYNAUD, G. GIRALDO, C. M. ROUSSET, J. P. SIROT, D. VODISLAV, « (2003), Semantic Integration in Xyleme : a Uniform Tree-Based Approach », *Data and Knowledge Engineering Review*, 44(2), pp. 267-298.
- [10] C. M. ROUSSET, A. BIDAULT, C. FROIDEVAUX, H. GAGLIARDI, F. GOASDOUE, C. REYNAUD, B. SAFAR, « Construction de médiateurs pour intégrer des sources d'information multiples et hétérogènes : le projet PICSEL », *Revue I3*, vol. 2, n° 1, p.5-59.
- [11] C. PARENT, S. SPACCAPIETRA, « Intégration de bases de données : Panorama des problèmes et des approches », *Ingénierie des systèmes d'information*, vol. 4, n° 3, Lavoisier (1996)
- [12] C. REYNAUD, G. GIRALDO, « An application of the mediator approach to services over the web », *Special track, Data Integration in Engineering, Concurrent Engineering (CE'2003) - the vision for the Future Generation in Research and Applications*, pp. 209-216, July 2003.
- [13] C. SARTIANI, P. MANGHI, G. GHELLI, G. CONFORTI, « XPeer : A Self-organizing XML P2P Database System », *Proceedings of the First EDBT Workshop on P2P and Databases (P2PDB 2004)*, Crete, Greece July, 2004.
- [14] D. BENEVENTANO, S. BERGAMASCHI, S. CASTANO, A. CORNI, R. GUIDETTI, G. MALZEVEZZI, M. MELCHIORI, M. VINCINI, « Information integration : The MOMIS project demonstration », *VLDB 2000 proceedings of 26th International Conference*.
- [15] D. C. FAYE, G. NACHOUKI, P. VALDURIEZ, « SenPeer : Un système Pair-à-Pair de médiation de données », *ARIMA 2006*, vol. 4, pp. 24-52.
- [16] E. MENA, V. KASHYAP, A. P. SHETH, A. ILLARRAMENDI, « OBSERVER : An Approach for Query Processing in Global Information Systems Based on Interoperation Across Pre-

- existing Ontologies », *IFCIS International Conference on Cooperative Information Systems (CoopIS)*, pages 14-25, 1996.
- [17] F. GANDON , « Graphes RDF et leur manipulation pour la gestion de connaissances », Habilitation à Diriger les Recherches, Edelweiss, Inria, novembre 2008.
- [18] F. GANDON, M. LO, C. NIANG, « Un modèle d'index pour la résolution distribuée de requêtes sur un nombre restreint de bases d'annotations RDF », *9èmes Journées Francophones d'Ingénierie des Connaissances*, Nancy, Juin 2008.
- [19] G. GARDARIN, F. DRAGAN, L. YEH, « P2p semantic mediation of web sources », *ICEIS (I)*, pages 7-15, 2006.
- [20] G. Z. IVES, A. Y. HALEVY, D. S. WELD, « Adapting to source properties in processing data integration queries », *In SIGMOD '04 : Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 395-406, New York, NY, USA. ACM.
- [21] H. WACHE, T. VÖGELE, H. STUCKENSCHMIDT, G. SCHUSTER, H. NEUMANN, S. HÜBNER, « Ontology-Based Integration of Information-A Survey of Existing Approaches », *IJCAI Workshop on Ontologies and Information Sharing*, 2001.
- [22] I. F. CRUZ, H. XIAO, « Ontology Driven Data Integration in Heterogeneous Networks », *Complex Systems in Knowledge-Based Environments*, eds. Andreas Tolk and Lakhmi Jain, pages 75-97, Springer, 2009.
- [23] I. F. CRUZ, H. XIAO, F. HSU, « An Ontology-based Framework for Semantic Interoperability between XML Sources », *International Database Engineering and Applications Symposium (IDEAS)*, pages 217-226, 2004.
- [24] I. MANOLESCU, D. FLORESCU, D. KOSSMANN, « Answering XML Queries over Heterogeneous Data Sources », *27th Very Large Data Bases*, pp. 241-250, Roma, Italy, Sept. 2001.
- [25] J. BHOGAL, A. MACFARLANE, P. SMITH, « A review of ontology based query expansion », *Information processing management*, vol. 43, n° 4, pp. 866-886, 2007.
- [26] J. SHANMUGASUNDARAM, J. KIERNAN, E. SHEKITA, C. FANandnameJ. FUNDERBURK, « Querying XML Views of Relational Data », *Proc. Of the 27th International Conference on Very Large Data Bases*, pp. 261-270, Roma, Italy, Sept. 2001.
- [27] L. GERAT, « Xedix, un système de gestion de masses de données », *La Recherche*, n° 408, pp. 15-17, 2007.
- [28] M. LO, « Dataweb basés sur XML : modélisation et recherche d'informations pertinentes », thèse de Doctorat de l'Université de Pau et des Pays de l'Adour, décembre 2002.
- [29] M. LO, A. HOCINE, « Modélisation de dataweb : une approche basée sur l'intégration de la sémantique des données et XML », *actes du congrès Conférence Africaine de Recherche en Informatique (CARI)*, Antanarivo, Madagascar, pp. 31-38, 16-19 octobre, 2000.
- [30] M. LO, A. HOCINE, « ISYWEB : an XML-based Architecture for web information systems », *IEEE SITIS Conference*, pp. 116-121, Yaounde, 27 Novembre - 1er décembre, 2005
- [31] N. BELAID, L. BELLATRECHE, Y. A. AMEUR, G. PIERRA, « Intégration de sources à base ontologique : architecture en réseau VS architure en étoile », *Plate-Forme AFIA : Atelier Thématique GDR I3 sur Ontologies et Gestion de l'hétérogénéité sémantique (OGHS)*, Juillet, 2007, pp. 9-20
- [32] O. SALL, M. LO, « Intégration de données environnementales : une approche basée sur les entrepôts de documents XML et les ontologies », *Pactes des 3èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA)*, Revue des Nouvelles Technologies de l'Information (RNTI B-3), Editions Cepadues, Juillet, 2007, pp. 9-20
- [33] O. SALL, M. LO, F. GANDON, C. NIANG, I. DIOP, « Using XML data integration and ontology reuse to share agricultural data », *International Journal of Metadata, Semantics and Ontologies 2009*, vol. 4, n° 1/2, pp. 93 - 105, Inderscience Publishers, Mai 2009.

- [34] S. ABITRBOUL, O. BENJELLOURN, I. MANOLESCU, R. WEBER, « Active XML : peer-to-peer data and web services integration », *Proceedings of the 28th international conference on Very Large Data Bases*, p.1087-1090, August 20-23, 2002, Hong Kong, China.
- [35] S. CHAWATHE, H. GARCIA-MOLINA, J. HAMMER, K. IRELAND, Y. PAKONSTANTINO, J. ULLMAN, J. WIDOM, « The TSIMMIS project : Integration of heterogeneous information sources », *In proceedings of IPSI conference*, Tokyo Japan.
- [36] T. MILLAN, F. MULATÉRO, M. LAMOLLE, « Design, Share and Re-use of Data and Applications into a Federate Database System », *Onzièmes Journées Internationales le Génie Logiciel et ses Applications*, Paris, France(1998).
- [37] T. R. GRUBER, « A Translation Approach to Portable Ontology Specifications », *Knowledge Acquisition*, 5(2) :199-220, 1993.
- [38] T. R. GRUBER, G. R. OLSEN, « An Ontology for Engineering Mathematics », *International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 258-269, 1994.
- [39] T. T. DANG-NGOC, C. JAMARD, N. TRAVERS, « XLive :An XML Light Integration Virtual Engine », *In Proc. of Bases de Données Avancées (BDA)'2005*, Saint-Malo, France.
- [40] T. T. DANG-NGOC, G. GARDARIN, N. TRAVERS, « Tree graph views : On efficient evaluation of xquery in an xml mediator », *BDA*, 2004.
- [41] W3C., « An XML Query Language (XQuery 1.0) », Recommendation, <http://www.w3.org/TR/2006/CR-xquery-20060608/>, June 2006.
- [42] Y. ARENS, C. A. KNOBLOCK, C. HSU, « Query Processing in the SIMS Information Mediator », *Tate, editor, Advanced Planning Technology. The AAAI Press*, Menlo Park, CA, 1996.
- [43] Y. PAKONSTANTINO, H. GARCIA-MOLINA, J. WIDOM, « Object Exchange Across Heterogeneous Information Sources », *proc. ICDE*, Taipei, Taiwan, March 1995, p 251-260.
- [44] Y. PAKONSTANTINO, V. BORKAR, M. ORGIYAN, K. STATHATOS, L. SUTA, V. VASSALOS, P. VELIKHOV, « XML queries and algebra in the Enosys integration platform », *Data Knowledge Engineering*, vol. 44, n° 3, pp. 299-322.