



République Algérienne Démocratique et Populaire  
Université Abou Bakr Belkaid– Tlemcen  
Faculté des Sciences  
Département d'Informatique

**Mémoire de fin d'études**  
**pour l'obtention du diplôme de Master en Informatique**

*Option: Système d'Information et de Connaissances (S.I.C)*

## *Thème*

**Construction d'une identité sociale  
fédérée représenté en XML par une  
approche d'intégration de données**

**Réalisé par :**

- M<sup>r</sup> DIB Sidi Mohammed Fazil
- M<sup>r</sup> LARABI Nor El Islam

*Présenté le 14 Octobre 2014 devant le jury composé de :*

- Dr Abdelkrim BENAMAR (Président)
- Dr Abdeljalil KHELASSI (Encadreur)
- Mr Yaghmorassen BENZIAN (Examineur)
- Mm Zeineb BOUCLI HCEN (Examineur)

Année universitaire : 2013-2014

## Sommaire

Introduction général.....	5
<b>Chapitre 1 : Intégration de données</b> .....	7
I.I. Intégration de données.....	8
I.1. Introduction.....	8
I.1.1. Problématique.....	8
I.1.2. Contexte.....	11
I.2. Etat de l’art .....	16
I.2.1. L’intégration de schémas.....	16
I.2.2. Les approches virtuelles et matérialisées.....	17
I.2.2.1. Les approches virtuelles.....	17
I.2.2.1.1.La fédération de bases de données .....	17
I.2.2.1.2.La médiation.....	18
I.2.2.2. L’approche matérialisée.....	19
I.3. Comparaison entre Architecture Matérialisé et Architectures Virtuelle.....	20
I.4. Processus d’intégration.....	21
I.4.1. Pré-intégration.....	21
I.4.2. Comparaison.....	21
I.4.3. Fusion.....	21
I.4.4. Restriction.....	22
I.II-Base de donnée fédérée.....	22
I.II 1. introduction.....	22
I.II 2 .Architecture fédérée.....	23
I.II.2.1. Les fédérations faiblement couplées.....	23
I.II.2.2. Les fédérations fortement couplées.....	24
I.II 2.2.1.Processeurs de transformation .....	25
I.II 2.2.2.Processeurs filtres.....	25
I.II 2.2.3. Processeurs de construction.....	26
I.II 2.2.4. Schémas locaux.....	26
I.II 2.2.5. Schéma fédéré.....	26
I.II 2.2.6.Schémas externes.....	26
I. Conclusion.....	26
<b>Chapitre 2 : Réseaux sociaux</b> .....	27
II.1.Introduction.....	28
II.2.Présentation.....	28
II.3.Etat du web social en 2014 .....	28
II.4.Les différents réseaux sociaux.....	30
II.4.1.Facebook, c’est quoi?.....	31
II.4.1.1.Face book, ça sert à quoi? .....	31
II.4.2.Google+.....	31
II.4.2.1.s’inscrire à Google+ et commencer à resauter.....	32
II.4.2.2.Fonctionnement général des cercles Google+.....	32
II.4.2.3.Le principe de cercles étendus et du filtre « mode public ».....	33
II.4.2.4.Créer une communauté Google+.....	33
II.4.3.Twitter.....	33
II.4.3.1.Identité Visuelle (logo).....	34
II.4.3.2.Fonctionnalités .....	34
II.5.Web social et profils utilisateur.....	34
II.6.L’influence des RS sur le commerce électronique .....	35

II.7.Les standards pour la description des profils tels que oAuth.....	36
II.8.Conclusion.....	37
<b>Chapitre 3 : Base de données XML.....</b>	<b>38</b>
III.1.Définition d'une base de données XML native .....	39
III.2.XML et les bases de données.....	39
III.2.1.XML, une base de données ?.....	39
III.2.2.Données versus document.....	40
III.2.2.1.XML orienté données.....	40
III.2.2.2.XML orienté documents.....	40
III.2.3.Stockage de données XML.....	41
III.2.3.1.Mapping des documents.....	41
III.2.3.2.Stockage dans une base relationnelle.....	42
III.2.3.3.Stockage à l'aide d'une base de données native XML.....	43
III.3.Avantages-inconvénients des BD natives et relationnelles.....	44
III.4.Exemples de choix de bases de données.....	45
III.5.Conclusion.....	46
<b>Chapitre4 : Construction d'une identité sociale fédérée.....</b>	<b>47</b>
IV.1.Introduction .....	48
IV.2.Parti1 Conception.....	48
IV. 2.1.Extraction des schémas locaux à partir des réseaux sociaux.....	49
IV.2.1.1.Facebook.....	49
IV.2.1.2.Twiter .....	50
IV 2.1.3.Google+ .....	50
IV 2.2.Résultat.....	51
IV .2.2.1.Schéma local de Facebook.....	51
IV.2.2.2.Schéma local de Twiter.....	51
IV.2.2.3.Schéma local de Google+.....	51
IV.3.Partie II : Intégration de données.....	51
IV.3.1. Méthode utilisée .....	51
IV.3.2. Schéma fédérée.....	54
IV.4. Partie III Implémentation.....	54
IV .4.1.Les moyens Utilisés pour faire la connexion avec les réseaux sociaux.....	54
IV.4.2. La Base de données.....	55
IV.4.3. Forme principale de l'application.....	55
IV.4.4. Extraction de l'identité Facebook.....	56
IV.4.5. Extraction de l'identité Twitter.....	57
IV.4.6. Extraction de l'identité Google+.....	58
IV.4.7. Intégration .....	59
IV.4.8. Représentation de l'identité fédérer en xml.....	60
IV.4.9.Conclusion.....	62
<b>Conclusion général.....</b>	<b>63</b>

## Liste des Figures

- Figure I.1: Problématique de l'intégration de données
- Figure I.2: Exemple de catalogues électroniques hétérogènes
- Figure I.3: Exemple de problème d'intégration
- Figure I.4: Infrastructures de médiation
- Figure I.5 : Taxonomie de SGBD hétérogènes et répartis
- Figure I.6 : Localisation de données
- Figure I.7 : Niveau intégration
- Figure I.8 : Architecture d'un système médiateur
- Figure I.9 : Architecture d'entrepôt de données
- Figure I.10 : Modèle d'architecture fédérée faiblement couplée
- Figure I.11 : fédération fortement couplées
- Figure II.1: web social et profile d'utilisateur
- Figure IV.1 : fédération fortement couplées
- Figure IV.2 : l'interface de La Base de données
- Figure IV.3 : l'interface de l'Application
- Figure IV.4 : l'interface d'extraction de l'identité Facebook
- Figure IV.5 : exemple de table de profil Facebook
- Figure IV.6 : L'interface d'extraction de l'identité Twitter
- Figure IV.7 : Exemple de table de profil Twitter
- Figure IV.8 : L'interface d'extraction de l'identité Google+
- Figure IV.9 : Exemple de table de profil Google +
- Figure IV.10 : L'interface d'extraction de l'identité fédérer
- Figure IV.11 : Exemple d'identité fédérer
- Figure IV.12 : Exemple d'identité fédérer

## **Liste des Tables**

Tab1 - Sémantique de données de trois catalogues dans la Figure I. 2

Tab.2 - Comparaisent entre Architectures Matérialisé et Architectures virtuelles

## Introduction

Les réseaux sociaux sont des sites Internet qui offrent la possibilité aux internautes de communiquer entre eux, de partager des informations, des photos, des vidéos, parmi ces réseaux sociaux Facebook, Twitter et Google+ les plus populaires.

Facebook est un réseau social très connu, qui a été fondé en 2004 par Mark Zuckerberg. La création d'un compte Facebook est gratuite et vous permet de créer un profil ainsi que de poster des statuts : messages, vidéos, liens, ou photos.

Twitter est un réseau social américain lancé en 2006 en concurrence directe avec Facebook. Bien que moins utilisé que ce dernier, il reste l'un des réseaux enregistrant le plus trafic au monde, en se classant régulièrement dans le top 10 des sites les plus visités de la planète.

Google+, un réseau social très proche de Facebook Comme son nom l'indique, Google+ est le réseau social lancé par Google au cours de l'année 2011, après les échecs d'Orkut, de Google Wave, et de Google Buzz. À l'inverse de Twitter, Google+ est un réseau social très proche de Facebook, dont le fonctionnement est presque identique.

Il ya une grande importance des informations collectée des réseaux sociaux pour les réutiliser dans les domaines de commerce, médical ou politique.

L'intégration de données regroupe les processus par lesquels les données provenant de différentes parties du système d'information sont déplacées, combinées et consolidées. Ces processus consistent habituellement à extraire des données de différentes sources (bases de données, fichiers, applications, Services Web, emails, etc.), et leur appliquent des transformations (jointures, Wordnet, etc.), et à envoyer les données résultantes ainsi le schéma fédérée.

Dans notre cas on va prendre un compte Facebook, Twitter et Google+ d'un même utilisateur La vue globale d'un internaute permet une reconnaissance forte d'utilisateurs de l'application web et par conséquent une interaction intelligente peut être envisagé.

En a utilisé aussi le dictionnaire Wordnet pour optimiser le schéma fédéré en éliminant la doublure d'un même attribut qui existe dans les comptes d'un seul utilisateur. Le schéma fédéré a été représenté en XML pour faciliter la réutilisation des informations dans d'autres applications web exemple le e-commerce.

La propagation des identités dans le système hétérogènes est devenue un véritable défi pour le grand public, le principal avantage de la fédération de l'identité est de disposer d'une identité unifiée, utilisée sur divers sites web indépendants.

Le présent mémoire se compose de quatre chapitres :

**Le premier chapitre** : Intégration des données

**Le deuxième chapitre** : Réseaux sociaux

**Le troisième chapitre** : Base de données XML

**Le quatrième chapitre** : Application

Le mémoire se termine par une conclusion et perspectives.

# Chapitre 1

**Intégration des données**

## I-Intégration de données :

### I.1 – Introduction :

#### I.1.1 Problématique de l'intégration de données :



Figure I.1 : Problématique de l'intégration de données (URL1)

Avant de faire un état de l'art des systèmes d'intégration de données, il est important de définir la nature du problème à résoudre.

#### Schéma Global



#### Schémas des sources intégrées

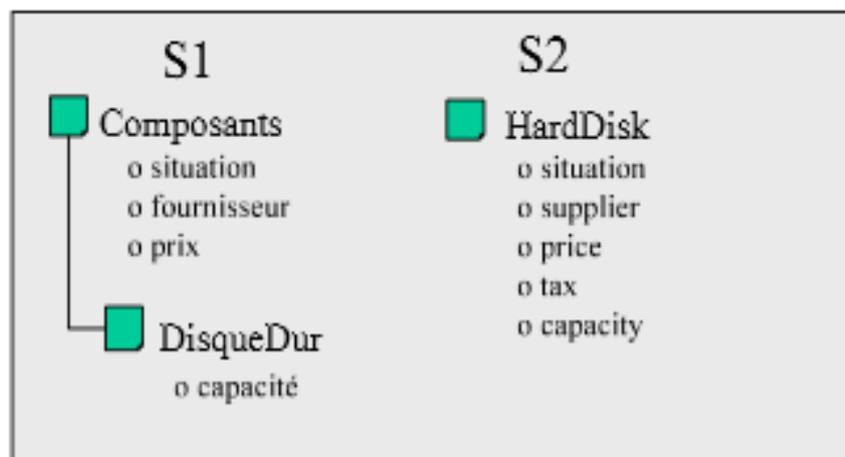


Figure I.2 - Exemple de catalogues électroniques hétérogènes (URL2)

Les systèmes d'intégration doivent permettre à l'utilisateur d'accéder, via une interface d'accès unique, à des données stockées dans plusieurs sources de données. Ces sources de

données ont été conçues indépendamment par des concepteurs différents. Cela entraîne l'hétérogénéité de données, c'est-à-dire que les données relatives à un même sujet sont représentées différemment sur des systèmes d'information distincts. Cette hétérogénéité provient des choix différents qu'ils sont faits pour représenter des faits du monde réel dans un format informatique. En effet, les données de sources sont structurellement indépendantes mais sont toujours supposées relevées de domaines similaires.

Exemple1 Pour illustrer ce problème d'hétérogénéité de sources, nous étudions l'exemple dans la Figure I.2. Supposons qu'une entreprise de commerce électronique vend des disques durs sur l'Internet. Les disques durs mis en vente sont fournis par des fournisseurs différents, à chacun organise ses produits dans son catalogue selon ses propres critères décidés localement. Cette entreprise doit donc traduire les catalogues de différents fournisseurs à son format, appelé schéma global. Pour des raisons de simplicité, nous considérons deux catalogues S1 et S2 de deux fournisseurs et le catalogue de l'entreprise SG (Figure I.2). Remarquons que les trois catalogues décrivent un disque dur différemment. Cette différence concerne le nombre de concepts utilisés pour définir chaque source, ainsi que l'aspect sémantique de chaque concept. Le tableau I.1 décrit les différentes propriétés de SG, S1 et S2.

	SG	S1	S2
Classes		Composants : composant informatique	HardDisk : disque dur
	DisqueDur : disque dur sous garantie	DisqueDur : disque Dur	
Propriété		situation (domaine :boolean) : neuf ou occasion	situation (domaine boolean) : disponible ou non
	marque (domaine : string) : marque de fabrique	fournisseur (domaine:string):marque de fabrique	supplier(domaine string) : marque de fabrique
		prix (domaine:number) : le prix total (qui inclut la TVA) d'un matériel neuf	price(domaine number) : le prix hors taxe d'un disque dur
			Tax (domaine number) : la TVA d'un disque dur
	capacité (domaine : integer): total de méga octets	capacité (domaine : integer) : total de méga octets	capacity (domaine number) : total de giga octets

## Tab I.1 - Sémantique de données de trois catalogues dans la Figure I.2

La question fondamentale lorsque l'on veut faire inter opérer des bases de données hétérogènes est d'une part, l'identification de conflits entre les concepts dans des sources différentes qui ont des liens sémantiques, d'autre part, la résolution des différences entre les concepts sémantiquement liés.

Une taxonomie des conflits sémantiques a été proposée dans : (1) conflits de représentations, (2) conflits de noms, (3) conflits de contextes, et (4) conflits de mesure de valeur que nous allons détailler dans les sections suivantes.

### **1 Conflits de représentation**

Ces conflits se trouvent dans le cas où on utilise des propriétés différentes ou des schémas différents pour décrire le même concept. Cela signifie que le nombre de classes a propriétés et les classes&propriétés représentant un concept C dans les sources ne sont pas égaux.

Exemple 2 Reprenons la Figure I.2, où le fournisseur S1 utilise deux classes Composants et Disque Dur et 4 propriétés situation, fournisseur, prix, et capacité pour décrire Un disque dur. Tandis que le fournisseur S2 utilise une seule classe HardDisk et 5 propriétés situation, supplié, pris, tax, et capacity.

Un autre exemple de conflit de représentation entre les deux fournisseurs à mettre en évidence, c'est le cas où le fournisseur S2 utilise deux propriétés Price et tax pour calculer le prix d'un disque dur, tandis que S1 n'en utilise qu'une seule, à savoir prix.

### **2 Conflits de noms (termes)**

Ces conflits se trouvent dans le cas où on utilise soit des noms différents pour le même concept ou propriété (synonyme), soit des noms identiques pour des concepts (et des propriétés) différents (homonyme).

Exemple 3 Le même concept de disque dur est nommé par Disque Dur dans la S1, et par HardDisk dans la S2. La propriété Situation se trouve dans les deux sources, mais avec deux significations différentes (voir Tab I.1).

### **3 Conflits de contextes :**

Le contexte est une notion très importante dans les systèmes d'information répartis. En effet, un même objet du monde réel peut être représenté dans les sources de données par plusieurs représentations selon un contexte local à chaque source. Ces conflits de contexte se trouvent dans le cas où les concepts semblent avoir la même signification, mais ils sont évalués dans différents contextes.

Exemple 4 La propriété prix de disque dur ne s'applique que pour Un disque dur neuf dans S1, mais peut l'être pour tous les disques soit neufs ou d'occasions dans S2.

#### 4. Conflits de mesure de valeur

Ces conflits sont liés à la manière de coder la valeur d'un concept du monde réel dans différents systèmes se trouvant dans le cas où on utilise des unités différentes pour mesurer la valeur de propriétés.

Par exemple, l'unité de mesure de la capacité dans la S1 est le méga octet tandis que celle dans la S2 est le giga octets.

##### I.1.2. Contexte :

- Les sources d'informations nombreuses et très diversifiées (SGBDR, SGBDO, XML, fichiers texte, pages Web, etc.)[1]
- Différents modes de consultation
  - Langages et modes de requêtes différents (SQL, moteurs de recherche, programme d'applications...)
  - Différentes façons de répondre (différentes présentations du résultat) : pages Web, tableaux, relations...
- Différents interactions avec la source
  - Protocoles de communication (JDBC, ODBC, IIOP)
  - Différentes interfaces

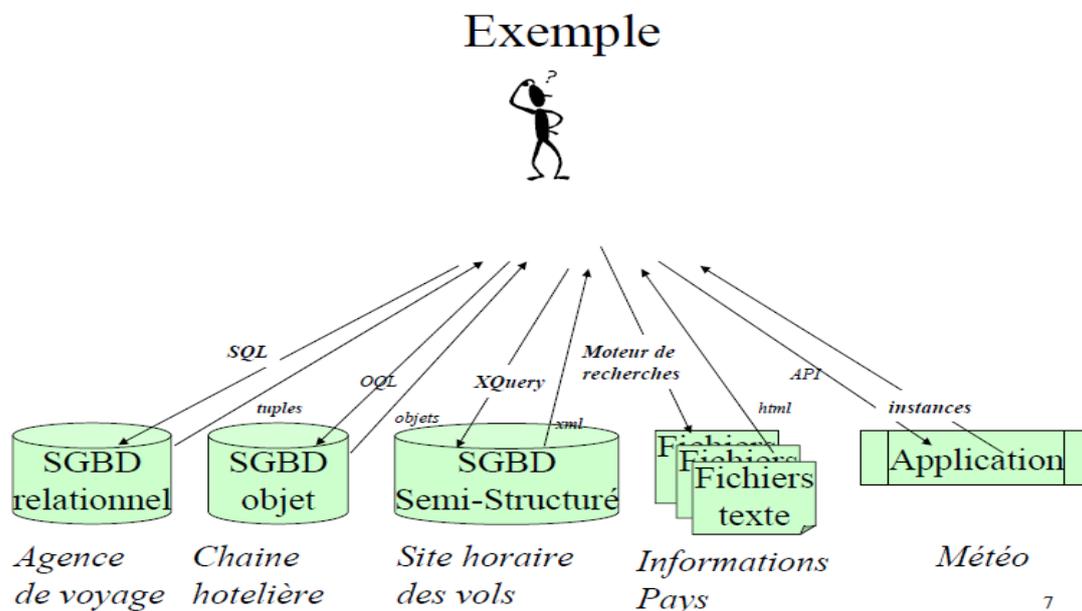


Figure I.3 – Exemple de problème d'intégration (URL3)

### Intégration de données Fourni :

- un accès (requêtes, éventuellement mises à jour)
- uniforme (les sources sont transparentes à l'utilisateur)
- à des sources (pas seulement des BD)
- multiples (même 2 est un problème)
- autonomes (sans affecter le comportement des sources)
- hétérogènes (différents modèles de données, schémas)
- structurées ( ou au moins semi-structurées)

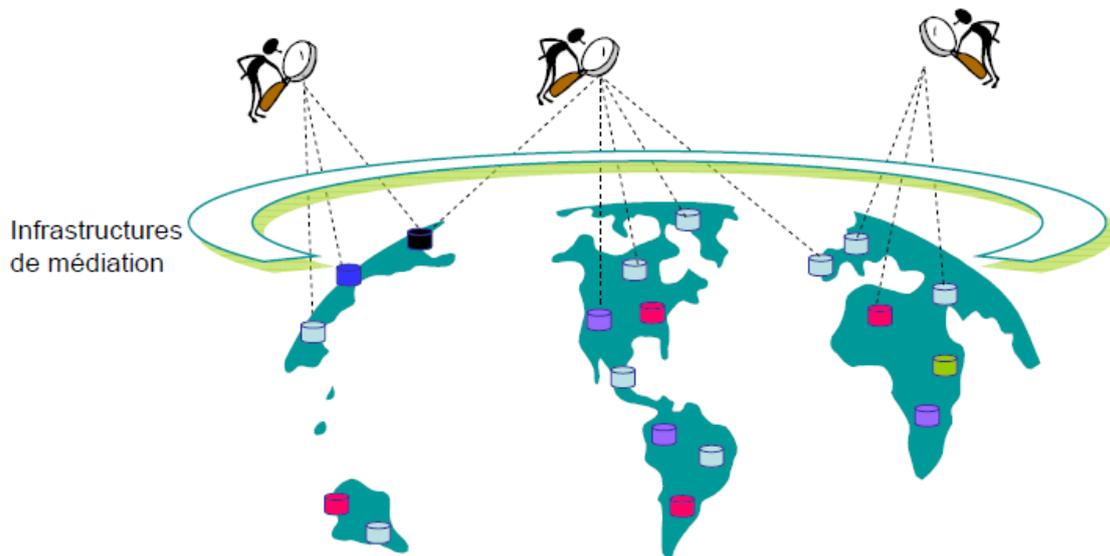
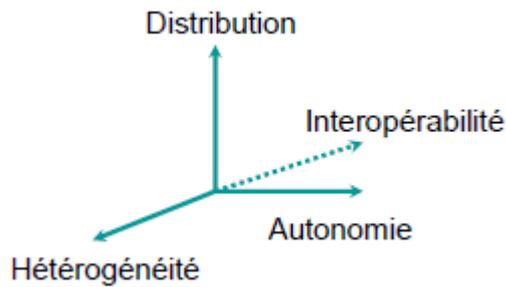


Figure I.4 – infrastructures de médiation (URL4)

### Caractéristiques des sources

- Distribution
- Hétérogénéité
- Autonomie
- Interopérabilité



➤ **Distribution :**

- Les données sont stockées sur des supports répartis géographiquement.
- Offre disponibilité et amélioration des temps d'accès.
- Pbs:
  - Localiser la (ou les) source(s) contenant les données pertinentes.
  - Tenir compte de la puissance des sources et de leur charge
  - Les sources peuvent être temporairement Indisponibles

➤ **Hétérogénéité :**

- L'hétérogénéité concerne les données, les modèles et les langages.
- Système homogène :
  - même logiciel gérant les données sur tous les sites
  - même modèle de données
  - même univers de discours
- Système hétérogène :
  - n'adhère pas à toutes les caractéristiques d'un système homogène
  - langages de programmation et d'interrogation différents, modèles différents, SGBD différents

➤ **Hétérogénéité des données :**

- Sémantique
  - Signification, interprétation ou utilisation différente de la même donnée
  - Types de relations sémantiques
  - R1 identique à R2 : même constructeur, même concept
  - R1 équivalente à R2 : constructeurs différents, même concept
  - R1 compatible avec R2 : ni identiques, ni équivalents
  - R1 incompatible avec R2 : contradictoires
- Structurelle
  - Représentation différente des mêmes concepts dans des bases différentes
  - Conflits de noms, types de données, attributs, unités

➤ **Autonomie :**

- Conception : sources locales avec des modèles de données propres, langage d'interrogation  
Interprétation sémantique des données, contraintes, fonctions ...
  - Communication : les sources de données locales décident quand et comment répondre aux questions d'autres sources
  - Exécution : pas d'information provenant des sources locales sur  
L'ordre d'exécution des transactions locales ou des opérations externes  
pas de distinction entre les opérations locales et globales
  - Association : Connexion et déconnexion des sources
- Partage de données et des fonctions

➤ **Interopérabilité :**

- Systèmes interopérables [1] :
  - Echange de messages et de requêtes
  - Fonctionnent comme une unité pour une tâche commune
  - Partagent les fonctions
  - Communiquent même avec des composants internes incompatibles
- Propriétés fondamentales à tout système interopérable :
  - Distribution
  - Hétérogénéité
  - Autonomie

**Taxonomie de SGBD hétérogènes et répartis**

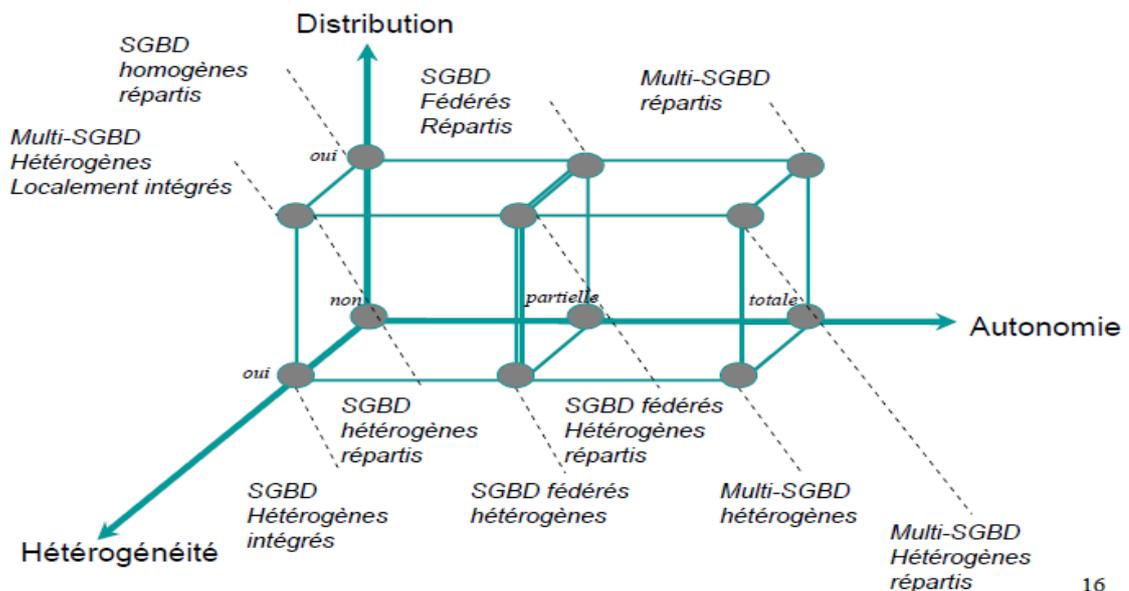


Figure I.5 - Taxonomie de SGBD hétérogènes et répartis (URL5)

## Localisation de données :

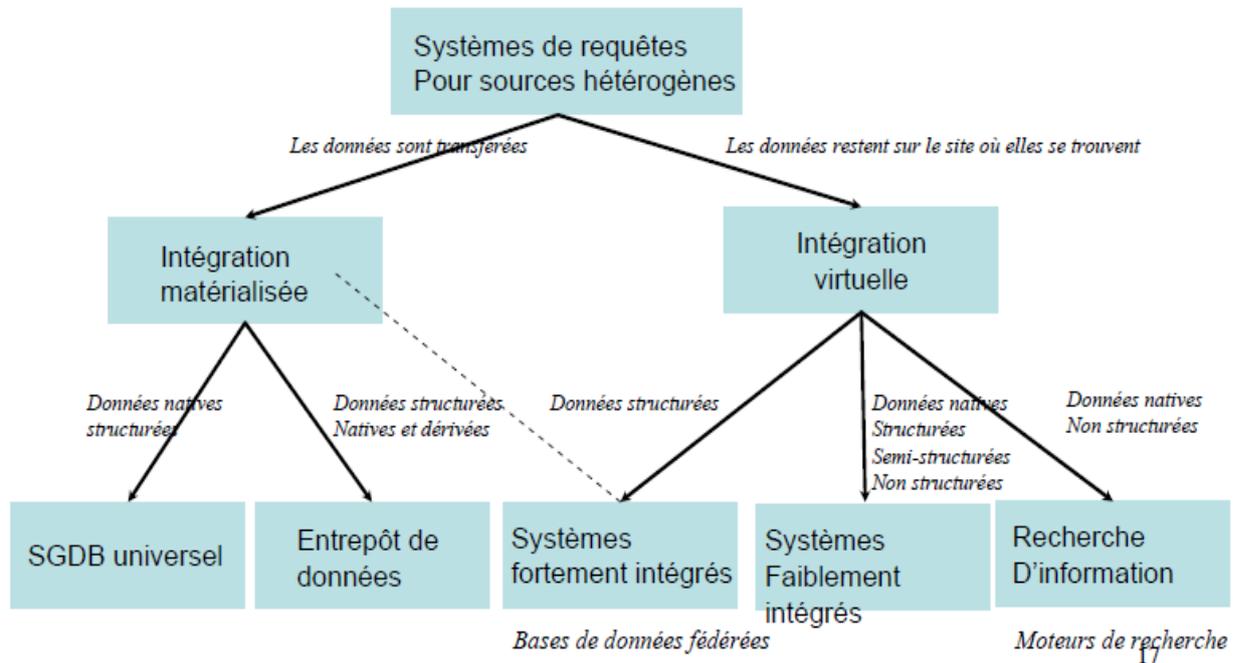


Figure I.6 - Localisation de données (URL6)

## Niveau intégration :

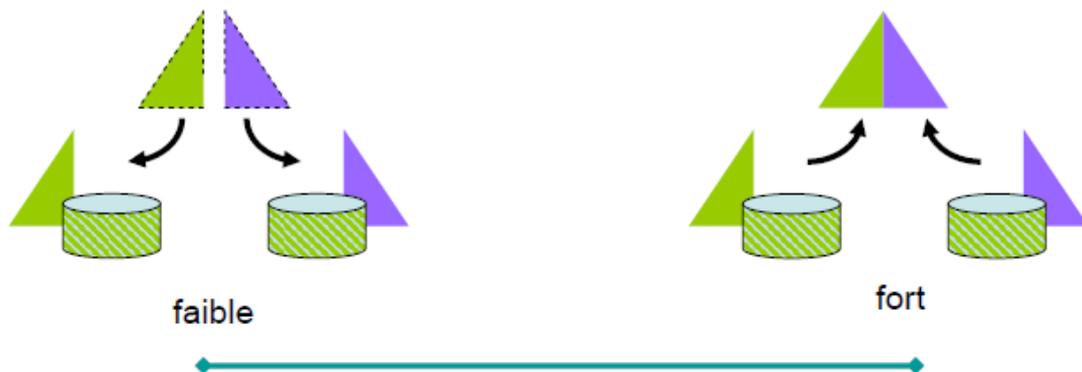


Figure I.7 - Niveau intégration (URL7)

- Bases de données réparties
- Fédérations de bases de données
- Multi bases de données avec schéma global
- Bases de données inter opérables ...

## **I.2. Etat de l'art :**

Dans notre mémoire en a menées dans le domaine de l'intégration de sources d'information structurées (bases de données). Ces recherches proposent des solutions pour construire des systèmes basés sur l'exploitation de plusieurs bases de données (BD). Le problème majeur que pose la construction de tels systèmes est l'hétérogénéité. Ce problème est triple. Il concerne (i) les plate-formes matérielle et logicielle sur lesquelles ces systèmes reposent, (ii) les modèles de données et schémas conceptuels définissant la structure logique des données stockées (nous parlons alors d'hétérogénéité sémantique), (iii) les types de données ou d'informations stockées (i.e, enregistrements, textes, vidéos). C'est sur le problème de l'hétérogénéité sémantique de données structurées que nous nous sommes focalisés. L'hétérogénéité sémantique est due à la présence de données sémantiquement proches ou équivalentes situées dans des BD dont les schémas sont différents. Les problèmes qu'elle engendre sont les suivants : (i) comment fournir une vue globale intégrée de données représentées au travers de conceptualisations différentes, (ii) comment mettre à jour les données de différentes bases étant donnée une telle vue globale intégrée, (iii) comment identifier et spécifier les relations entre des données sémantiquement liées.

Un aperçu des travaux portant plus particulièrement sur ce type d'hétérogénéité en BD, et réalisés durant ces vingt dernières années, est présenté dans [9]. Deux types d'approches s'opposent :

- La définition d'un modèle de données global permettant de représenter la structure logique et les liens sémantiques de données de différentes BD, à ce modèle correspond une base de données distribuée ou une base de données, souvent appelée entrepôt de données, regroupant des données issues de différentes BD.

- Les approches visant l'interopérabilité entre BD, sans schéma global intégré.

Dans ce cadre, nous présentons dans ce chapitre les différentes approches d'intégration de bases de données hétérogènes, après avoir présenté rapidement les principales phases du processus d'intégration.

### **I.2.1. L'intégration de schémas**

Les premiers travaux ont été réalisés dès le début des années 80. Ils ont été effectués dans le cadre de recherche portant sur l'intégration de schémas. Le but est de construire un schéma global intégré à partir des schémas, et non des données elles mêmes, de différentes BD dans le contexte du modèle Entité/Association, du modèle fonctionnel ou relationnel. Une méthode

générale d'intégration de schémas est présentée dans [7]. Elle est adaptée à l'intégration de différentes BD existantes (problème d'intégration de BD) [8], mais aussi utilisable lorsque l'on veut construire une BD à partir de vues d'utilisateurs (problème d'intégration de vues) [Bat84]. Les différentes phases du processus d'intégration général défini dans [7] sont les phases (i) de pré-intégration où l'ordre d'intégration des schémas est choisi, (ii) de comparaison où les problèmes d'hétérogénéité sémantique sont détectés semi-automatiquement (i.e., en coopération avec la personne chargée de l'intégration), (iii) de mise en conformité où la structure des schémas est modifiée en vue de leur intégration, (iv) de fusion où les schémas sont appariés, et (v) de restructuration où le schéma global est ajusté afin de le rendre minimal, complet et compréhensible. La phase de pré-intégration est parfois soumise, c'est le cas dans les méthodes proposées par [Bat84] et [8].

## **I.2.2. Les approches virtuelles et matérialisées**

### **I.2.2.1 .Les approches virtuelles**

Plus récemment, des approches, parfois qualifiées de virtuelles, ont été proposées. Elles se distinguent des approches précédentes par le fait qu'elles ne reposent pas sur la définition d'un schéma global de BD. Ainsi, pour répondre à une requête soumise dans son langage, le système agit comme un « moteur de requête » en décomposant la requête initiale de l'utilisateur, en interrogeant les différentes bases de données dans leur langage à l'aide de sous-requêtes, et en appariant les résultats trouvés.

Dans Intégration virtuelle :

- Les données restent dans les sources
- Les requêtes sont faites sur un schéma global, puis décomposées en sous-requêtes sur les sources. Les différents résultats des sources sont de la requête sont combinés pour former le résultat final.

#### **I.2.2.1.1 La fédération de bases de données**

Les premiers travaux abordant l'interopérabilité entre BD ont été effectués dans le cadre de recherches portant sur les BD fédérées. Une description complète de l'architecture et des méthodes de conception de Systèmes de Bases de Données Fédérées (FDDBS) est présentée dans [10]. Selon cette architecture, chaque schéma d'une base de données à fédérer est transformé en schéma « Composant » basé sur le modèle de données commun du FDDBS.

Cette transformation est semi-automatisée et permet d'avoir tous les schémas des bases à fédérer exprimés dans un même modèle de données. Des schémas d' « Export » représentant des sous-ensembles d'un schéma « Composant » sont construits. Ils représentent l'ensemble des données mises à la disposition de la fédération. Ceci permet d'occulter des données qui ne

doivent pas être rendues publiques. Enfin, des schémas « Fédérés » sont bâtis à partir de plusieurs schémas d' « Export » pour répondre aux besoins des différentes applications liées à la fédération. Ces schémas « Fédérés » résultent de l'intégration de schémas d'Export». Ainsi, même si l'approche BD fédérée ne repose pas sur la définition d'un unique modèle global intégré, nous constatons qu'une telle architecture résulte quand même de l'intégration de divers schémas. Des problèmes d'hétérogénéité sémantique se posent donc également dans le cadre de cette approche. De même nature que ceux rencontrés lors de la définition d'un schéma global, ils sont traités à l'aide des mêmes techniques. L'essentiel des recherches effectuées en BD fédérées a alors plutôt porté sur la définition de langages adaptés à l'interrogation de BD distantes et autonomes en l'absence de schéma global intégré.

Les recherches effectuées sur les BD fédérées ont petit à petit évolué. Une première tendance peut être illustrée à l'aide des travaux de Kashyap et Seth réalisés en 94 ([11]). Sensibles au problème de l'hétérogénéité sémantique pouvant rendre difficile le traitement de requête ils ont introduit la notion de contexte. Le contexte sert à capturer la signification et/ou le mode d'utilisation des objets modélisés dans des schémas de BD. Ainsi, ils proposent de comparer le contexte d'une requête utilisateur aux contextes des objets des différentes bases de données fédérées, afin de localiser plus facilement les sources d'information pertinentes lors du traitement d'une requête. Cette évolution s'est poursuivie. Elle a donné naissance aux approches dites «à base de médiateurs ».

### **I.2.2.1.2 L'approche médiatrice**

#### **Présentation générale**

L'approche médiatrice consiste à définir une interface entre l'agent (humain ou logiciel) qui pose une requête et l'ensemble des sources accessibles via le Web potentiellement pertinentes pour répondre.

L'objectif est de donner l'impression d'interroger un système centralisé et homogène alors que les sources interrogées sont réparties, autonomes et hétérogènes.

Un médiateur si dessous comprend un schéma global, ou ontologie, dont le rôle est central. C'est un modèle du domaine d'application du système. L'ontologie fournit un vocabulaire structuré servant de support à l'expression des requêtes. Par ailleurs, elle établit une connexion entre les différentes sources accessibles.

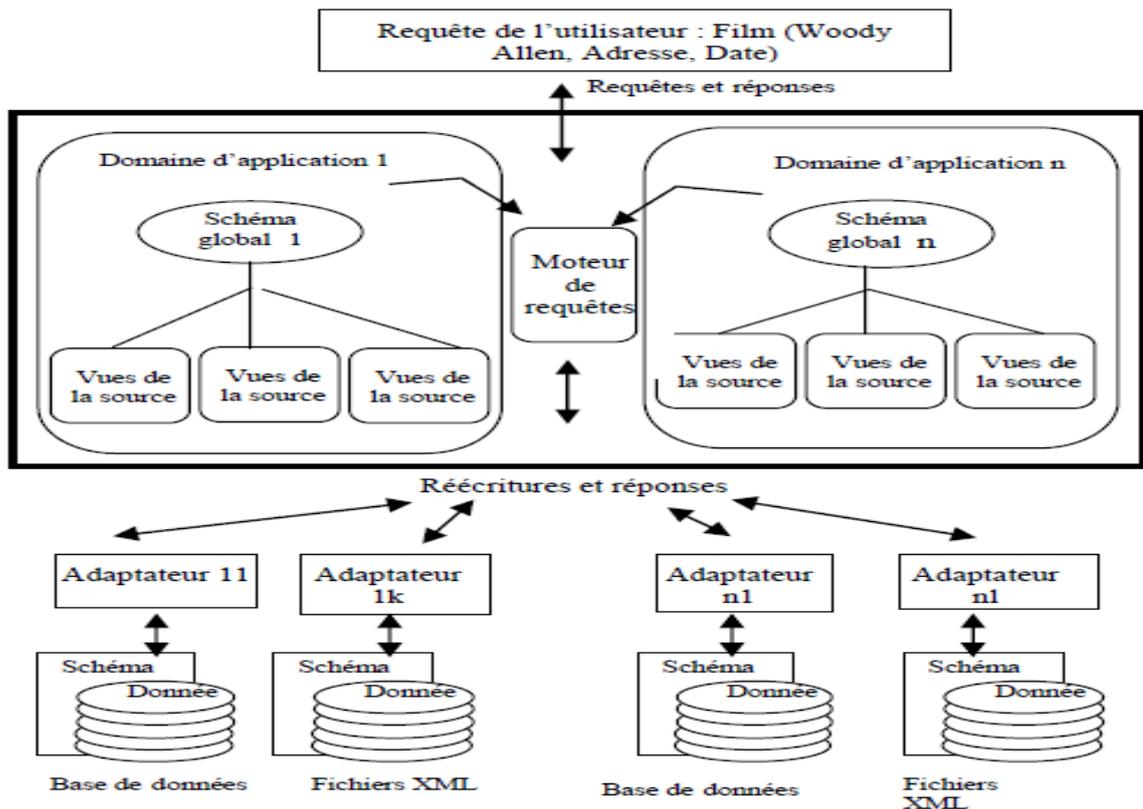


Figure I.8 : Architecture d'un système médiateur (URL8)

### I.2.2.2. L'approche matérialisée

Le troisième type d'approche que nous exposons est l'approche dite matérialisée (par opposition aux approches virtuelles présentées précédemment. Ainsi, étant donné un schéma intégré ou fédérée d'un système, l'approche matérialisée consiste à stocker localement des données issues des différentes bases [12].

L'avantage est un gain de temps considérable lors de l'accès à ces données, mais l'inconvénient réside dans la mise à jour de celles-ci. Le problème est de garder la synchronisation entre les copies des données stockées dans les entrepôts et les données originales stockées dans les bases de données [13].

Parmi ceux-ci se trouvent des problèmes d'hétérogénéité sémantique entre des données issues de bases différentes. Toutefois, au travers de la littérature, ce problème ne semble pas être au centre des préoccupations des chercheurs du domaine, plus intéressés par des problèmes nouveaux et plus spécifiques à l'architecture entrepôts de données : des problèmes d'accès aux données ou d'exploitation des données recueillies, par exemple.

Dans Intégration matérialisée :

- Les données provenant des sources à intégrer sont stockées sur un support spécifique (entrepôt de données).
- L'interrogation s'effectue comme sur une BD classique (Relationnelle).

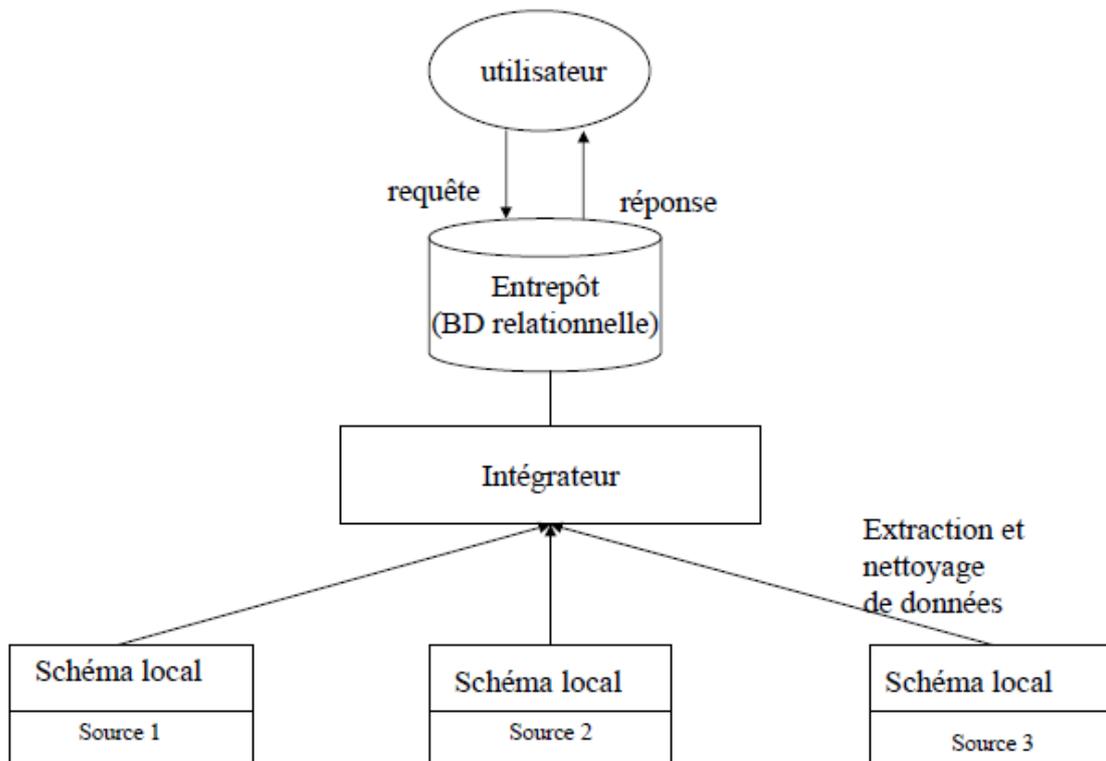


Figure I.9 : Architecture d'entrepôt de données (URL9)

### I.3. Comparaison entre les architectures Matérialisé et Architectures virtuelles :

Architectures matérialisées	Architectures virtuelles
<ul style="list-style-type: none"> <li>– Bonnes performances</li> <li>– Données pas toujours fraîches</li> <li>– Nettoyage et filtrage des données</li> </ul>	<ul style="list-style-type: none"> <li>– Les données sont toujours fraîches</li> <li>– Traitement de requêtes peut être coûteux</li> <li>– Défi principal : performances</li> </ul>

Tab I.2 – Comparaison entre les Architectures Matérialisé et Architectures virtuelles

#### **I.4. Processus d'intégration**

Les bases de données contiennent des représentations d'objets du monde réel, avec leurs liens et leurs propriétés. L'intégration de bases de données toutefois, dépasse les représentations, pour considérer en premier lieu ce qui est représenté plutôt que comment il est représenté. [2]

La plupart des approches décomposent le processus d'intégration de schémas de bases de données en plusieurs phases, dont les principales sont.

**Pré-intégration:** traduction des schémas source dans un modèle de données commun,

**Comparaison:** recherche de correspondances inter-schémas,

**Fusion :** création d'un schéma global à partir des schémas source, des correspondances inter-schémas et de règles d'intégration,

**Restructuration:** modification des schémas source.

##### **I.4.1. Pré-intégration**

Le but de cette phase est d'extraire un maximum d'informations sur les éléments (attributs, classes, méthodes, etc.) composant un schéma, ainsi que sur les relations et les règles existant entre ces éléments.

La pré-intégration est rendue nécessaire à la fois par la relative pauvreté sémantique exprimée par les schémas de bases de données et par l'hétérogénéité des modèles de données utilisés.

##### **I.4.2. Comparaison**

Lorsque les schémas initiaux ont atteint le niveau de conformité souhaité, l'étape suivante consiste à identifier les éléments communs des bases existantes.

Rappelons que deux bases de données ont un concept commun si les portions de monde réel qu'elles représentent ont des éléments communs soit une intersection non vide, ou ont des éléments en relation entre eux par un lien sémantique pour des applications futures [15].

Parmi les travaux dédiés à cette phase de comparaison, citons les approches suivantes :

Calcul d'une fonction de similitude entre paires d'éléments,

Calcul d'une pertinence sémantique entre éléments à l'aide de la logique floue,

Méthodes manuelles: les paires de noms sont proposées à l'expert de la BD (ou au DBA), qui doit déterminer lui-même les correspondances à retenir.

##### **I.4.3. Fusion**

Une fois effectué le travail de description sémantique des schémas (pré-intégration) et de recherche de correspondances entre leurs éléments (comparaison), le processus d'intégration proprement dit peut être effectué. Il consiste, de manière semi-automatique en général, à construire un schéma intégré à partir:

Des schémas source (éventuellement traduits dans un modèle canonique),  
Des correspondances inter-schémas (déterminées lors de la comparaison),  
De règles d'intégration (propres à la méthode et aux modèles de données utilisés).

#### **I.4.4. Restructuration**

Cette phase se retrouve plutôt dans les cas d'intégration de vues ou lorsque la modification ultérieure des schémas source est admise. Les approches récentes, de type fédérées, n'admettent pas de telles modifications, ceci afin de préserver l'environnement local (données, applications) des bases composant la fédération et de garantir leur autonomie [14].

Le principe d'intégration a été développé dans divers projets de recherche.

### **I.II- Bases de données fédérer**

#### **I.II.1 Introduction**

L'architecture fédérée est une forme d'architecture centrée sur les données. Le terme de système de base de données fédérée a été introduit par Heimbigner et McLeod(1). Ils définissent l'essentiel de ce qu'est une base de données fédérée faiblement couplée : « collection of components to unite loosely coupled fédération in order to share and exchange information » et une base de données fédérée fortement couplée : « an organisation model base d'on lequel, autonomes data bases, with sharing controlled by explicit interfaces. »

L'objectif de la création d'une base fédérée est de donner aux utilisateurs une vue unique des données présentes sur plusieurs systèmes a priori hétérogènes. Il s'agit d'une problématique typique lors de la concentration ou de la fusion d'entreprises : faire cohabiter les différents systèmes tout en leur permettant d'inter-opérer d'une manière harmonieuse [3]. Il ne s'agit pas ici de reconstruire une base répartie à partir des anciennes (ré-ingénierie), mais de conserver les anciennes bases et leurs autonomies locales tout en permettant l'intégration de l'ensemble.

Sheth et Larson(2) divisent les systèmes fédérés en deux catégories : les systèmes faiblement couplés réalisant l'intégration des différentes bases de données à l'aide des mécanismes des bases de données multibases et réparties (et donc la tâche de maintenir la fédération est laissée aux administrateurs locaux(3)) et les systèmes fortement couplés réalisant l'intégration à l'aide d'un SGBD fédéré (le maintien de la fédération est sous le contrôle de l'administrateur de la fédération).

## I.II.2 .Architecture fédérée

### I.II.2.1. Les fédérations faiblement couplées

Pour réaliser une telle intégration, quatre schémas sont utilisés dans les bases de données composant la fédération; le schéma d'exportation (ES) pour les données permises à l'exportation, le schéma d'importation (IS) pour les données importées, le schéma privé (PS) pour l'ensemble de données privées (ES et IS inclus) et le schéma commun (global, canonique ou pivot) qui dans le cas des bases de données faiblement couplées n'est qu'une abstraction se trouvant dans le dictionnaire de données de la fédération (FD).

Dans le diagramme, nous avons quatre bases de données distinctes, offrant la portion de leurs données qui est nécessaire à la fédération via leurs ES. Les informations (métadonnées) concernant le schéma global se retrouvent dans la FD. Nous retrouvons également dans la FD les informations indiquant comment effectuer l'importation des données pour réaliser des vues sur le schéma global (liens en pointillés).

Nous voyons dans le diagramme que le schéma global n'existe que d'une façon virtuelle, les différents administrateurs de données locaux doivent créer des vues à la main sur les différents schémas d'exportation pour réaliser une vue totale ou partielle (mais virtuelle) du schéma global. Les schémas d'exportation (ES) sont en général des vues sur une portion de la base locale de manière à permettre la correspondance entre cette portion de la base et le schéma global. Nous constaterons que les schémas originaux restent inchangés et que les utilisateurs locaux peuvent continuer à utiliser leurs anciennes bases comme si de rien n'était. La seule nouveauté pour un usager local est l'accès possible aux données de la fédération via son IS local.

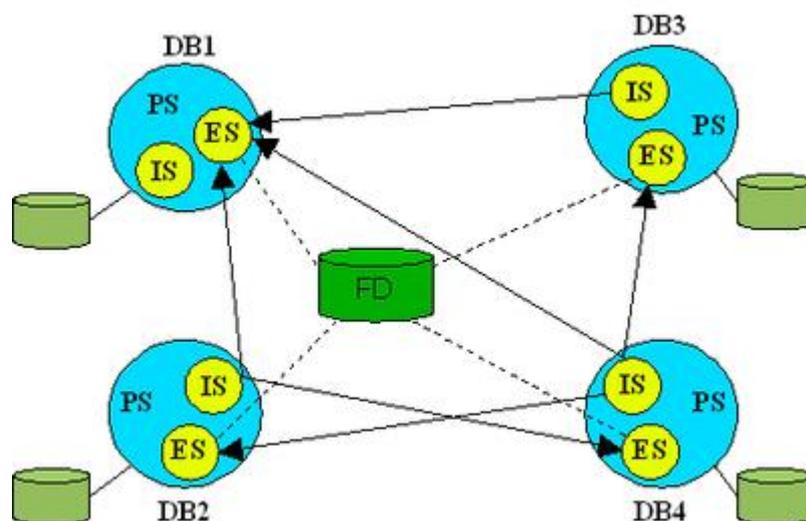


Figure I.10 : Modèle d'architecture fédérée faiblement couplée (URL 10)

### **I.II.2.2.Les fédérations fortement couplées**

Pratiquement tous les auteurs dans le domaine de la médiation de données attribuent à [Sheth & Larson, 1990] le concept de fédération de données et en général, le concept de fédération se ramène à l'approche fortement couplée. Pourtant, ces auteurs ont pour seul mérite d'avoir synthétisé et catégorisé les différentes approches fédératives existantes à l'époque et d'avoir proposé une architecture de référence pour l'approche fédérée fortement couplée (sans concevoir de système opérationnel). Voici la liste des systèmes existants à l'époque (selon Sheth et Larson) : ADDS, CALIDA, DQS, DDTS, IISS, Mermaid, MRDSM, Multibase, OMNIBASE.

Il s'agit essentiellement du même diagramme que celui présenté par Sheth et Larson. Seule la ligne pointillée a été rajoutée, celle-ci sépare la partie supérieure que nous baptiserons intégrateur et qui est responsable de l'intégration des différentes BD de la partie inférieure que nous appellerons traducteur, qui permet la traduction d'un schéma d'une BD dans un modèle canonique. Il est naturel de concevoir un déploiement de cette architecture avec les traducteurs sur les mêmes hôtes que les SGBD locaux et l'intégrateur sur un hôte particulier. Un tel système peut être baptisé SGBDF pour Système de Gestion de Bases de Données Fédérées et possède de multiples points en commun avec les SGBD traditionnels et répartis. En fait, nous pouvons considérer ces systèmes comme des SGBD répartis pour lesquels les schémas internes sont des SGBD traditionnels. Il est donc naturel de concevoir un SGBDF comme une évolution des SGBD répartis.

Tous les systèmes de type fédéré fortement couplé suivent d'une manière plus ou moins heureuse l'architecture générique établie par Sheth et Larson. Cette architecture est composée essentiellement de processeurs et de schémas. Voici les différentes catégories de ces composantes dégagées par les deux auteurs et se retrouvant dans le diagramme.

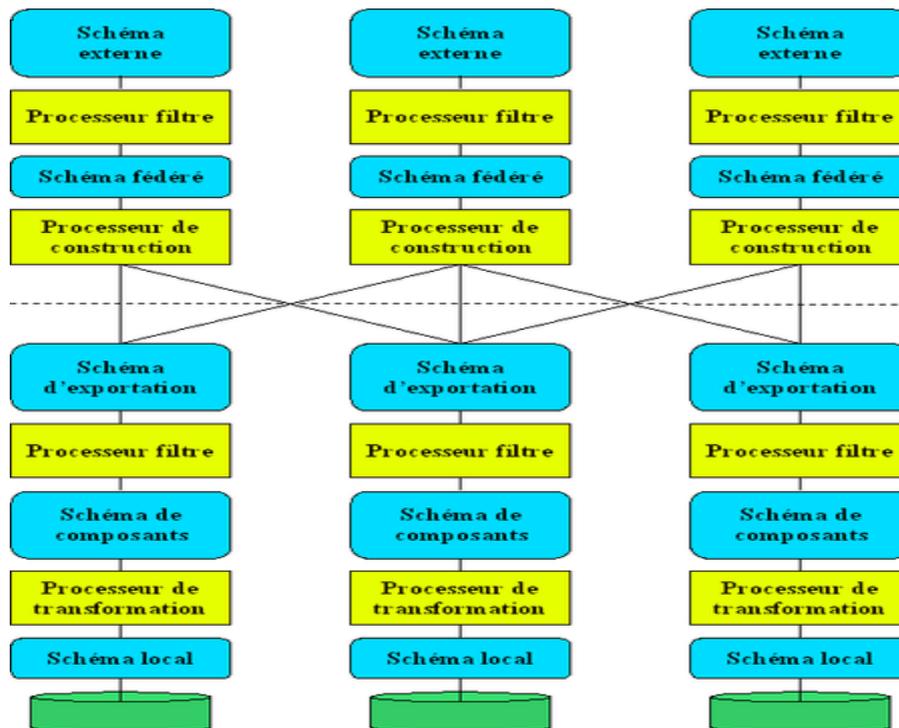


Figure I.11- fédération fortement couplées (URL11)

### I.II.2.2.1 Processeurs de transformation

Ces processeurs traduisent une commande effectuée dans un langage (ex.: SQL) en une requête dans un autre langage (ex. : CODASYL). Pour pouvoir effectuer cette tâche, ils offrent une vision transparente des modèles de données. Ils convertissent donc un schéma conçu selon un modèle en un autre (ex. : modèle réseau en modèle relationnel et vice versa).

### I.II.2.2.2 Processeurs filtres

Ces processeurs vérifient les commandes envoyées à un processeur ou les résultats retournés par un processeur pour s'assurer qu'ils ne violent pas des règles d'intégrité syntaxique et/ou sémantique. Ces processeurs peuvent s'occuper également de la vérification des droits d'accès aux données. Concrètement, ils sont responsables du maintien de la fédération en offrant des mécanismes explicites permettant de résoudre les contraintes du schéma global.

### I.II.2.2.3 Processeurs construction

Ces processeurs divisent les commandes ou les données envoyées par un processeur en commandes ou données destinées à plusieurs processeurs. Ou encore, à l'inverse, intègrent des commandes ou des données envoyées par plusieurs processeurs. Les tâches pouvant être accomplies par ces processeurs sont les suivantes : intégration de schémas, négociation de

protocoles entre processeurs, optimisation de requêtes et décomposition des requêtes pour différents processeurs, gestion des transactions (atomicité et concurrence). Processeurs d'accès : Ces processeurs acceptent les commandes effectuées par différents processeurs et manipulent directement les données en fonction des commandes. Ils correspondent à des SGBD locaux.

#### **I.II.2.2.4 Schémas locaux**

Ils correspondent aux schémas des bases de données participantes, exprimés dans leurs modèles de données propres (relationnel, réseau, hiérarchique). Schémas des composants : Ils correspondent aux schémas des bases de données participantes une fois que les schémas sont traduits dans un modèle de données commun (pivot). Schémas d'exportation : Ils correspondent à la portion des schémas des composants destinés à l'intégration au sein de la fédération.

#### **I.II.2.2.5 Schéma fédéré**

Il s'agit du schéma de la fédération proprement dit.

#### **I.II.2.2.6 Schémas externes**

Équivalents aux schémas externes de l'architecture Ansi/Sparc mais au niveau de la fédération.

### **Conclusion :**

Dans les différents systèmes décrits ci-dessus, l'intégration des données hétérogènes doit passer par une étape essentielle qui est la transformation des données interrogées. Cette transformation dépend de l'architecture du schéma fédérée. Cela suppose un prétraitement spécifique sur les données.

Dans ce contexte, notre approche a pour but de passer le prétraitement en s'intéressant à la sémantique des données quelles que soient leurs formes, autrement dit à la sémantique portée par les structures des bases de données et non pas aux données elles-mêmes.

# Chapitre 2

Réseaux Sociaux

## **Introduction :**

L'évolution d'Internet a permis la création de nouveaux outils de communication et de travail pour les entreprises ainsi que pour les particuliers.

Le Web 2.0 a amené l'évolution des réseaux sociaux. Ils sont de plus en plus utilisés par les internautes qui naviguent sur Internet, et ils touchent un public extrêmement large. Les étudiants et les adolescents ont été les premiers utilisateurs de ce genre de sites, faisant d'eux les précurseurs des réseaux sociaux actuels. Il faut également savoir qu'actuellement, pour beaucoup d'internautes, utiliser ces sites est considéré comme une activité sociale à part entière. Aujourd'hui, plus que 1.43 milliards d'utilisateurs des réseaux sociaux dans le monde en 2012 (19% de plus qu'en 2011), et on prévoit une croissance allant jusqu'à 1.85 milliards d'utilisateurs actifs au moins une fois par mois en 2014.

Les réseaux sociaux sont inspirés par la « théorie de 6 degrés » qui veut que chaque individu soit à six intermédiaires de toute autre personne sur la planète. Leur principe de fonctionnement s'inspire des sites de rencontre et les sites de retrouvailles de camarades de classe. Ainsi, ces réseaux permettent de mettre en relation des personnes entre elles et créent ce que l'on appelle des « graphes sociaux » qui représentent les liens qui s'établissent entre chaque individu d'une communauté.

## **II.2.Présentation:**

Le Web social est un ensemble de relations qui lient les gens sur le Web. Alors que les sites de réseautage social actuels les plus connus sur le Web se limitent aux relations entre les personnes disposant de comptes sur un seul site, le Web social devrait permettre aux gens de créer des réseaux de relations à travers l'ensemble du Web, tout en donnant aux gens la possibilité de contrôler leur propre vie privée et des données.

Le Web social n'est pas seulement sur les relations, mais sur les applications et les innovations qui peuvent être construits au-dessus de ces relations.

Bien que le Web social est un espace d'innovation, il n'est pas encore un citoyen "de première classe" du Web: applications sociales a largement évolué comme les silos et donc mises en œuvre et l'intégration sont incompatibles, avec peu de garanties de confidentialité et de respect des termes de service.

## **II.3.Etat du web social en 2010 :**

2010 a été une année tumultueuse pour le Web social. Cependant, le Web social n'est pas un phénomène nouveau qui n'a pas de précédent, mais le résultat d'une vulgarisation des

technologies existantes. De nombreuses fonctionnalités sociales étaient disponibles sur Internet avant le Web, depuis les fonctions de blog, la messagerie via e-mail.

Le Web depuis sa création a été conçu pour inclure des connexions entre non seulement les documents hypertextes, mais les relations entre les gens.

Depuis les premiers jours du peuple Web qui ont maintenu leurs propres pages d'accueil ont enregistré des mises à jour sur les activités de leurs sites, ce qui a été poussé dans le grand public avec le développement de convivial logiciel de blogging, des innovations dans cet espace ont permis au grand public de plus en plus aptes à les blogs et les sites d'information indépendants comme **Indymédia** (1999) pionnier de la notion de gestion de contenu généré par l'utilisateur. Cependant, ces services sont restés assez expérimental jusqu'à après l'effondrement de la "dot-com" bulle initiale. Après cette éruption de sites de réseaux sociaux comme **Friendster** (2002), **LinkedIn** (2003), et **Facebook** (2004) ont décollé, et sont finalement devenus des sites les plus populaires sur le Web. À partir de **Flickr** (2004) et **Youtube** (2005), le contenu généré par l'utilisateur a pris au cours de cette nouvelle revigoré Web social.

Le lancement de **Twitter** (2007), un site de micro-blogging, qui s'est propagé mises à jour de réseaux sociaux de l'utilisateur, via le bureau et les appareils mobiles. Il était à cette époque que le concept du Web social est devenu associé à la fois avec les entreprises mentionnées ci-dessus et avec le paradigme plus large "Web 2.0". Aujourd'hui, le Web social devient partie des portefeuilles de communication d'entreprise et sociétés du Web 2.0 commencent la commercialisation des données de et sur leurs utilisateurs.

De nombreux pays en développement leurs propres sites de réseautage social les plus populaires tels que **Hi5** au Japon et **QQ** en Chine, il ya eu une tendance générale à déplacer des utilisateurs leurs profils entre les services, tels que les utilisateurs se déplaçant leurs profils de **Friendster** à **MySpace** par exemple. **Facebook** rapidement augmenté pour devenir un leader mondial dans les réseaux sociaux.

Les sites de réseautage social encouragent les utilisateurs à mettre leurs données dans la plateforme exclusive donnée, et ont tendance à faire de la portabilité des propres données de l'utilisateur à un autre site ou même à leur ordinateur est difficile, si ce n'est pas impossible. La première technologie créée spécifiquement pour un graphe social portable était le projet **ami-de-un-ami** (FOAF) pour le Web sémantique en 2001.

Beaucoup de sites de réseautage social considérés la vie privée et la portabilité d'être contradictoire, dans la mesure où **Facebook** utilisé pour refuser aux utilisateurs la possibilité de laisser les données soient portable en dehors de son système en raison de préoccupations

quant à la confidentialité de l'utilisateur, Aujourd'hui, la vie privée est un argument secondaire pour stimuler nouvelles inscriptions.

2010 a été l'année où les questions de la vie privée sur le Web social ont augmenté au-delà d'un problème de niche et sont entrés dans la conscience populaire. En Décembre 2009, **Facebook** a modifié ses paramètres de confidentialité par défaut de certains paramètres de confidentialité qui à son tour fait partie de l'information profil public d'un utilisateur. Les utilisateurs sont encouragés à utiliser des "contrôles de confidentialité" pour fournir un contrôle d'accès à leurs données, mais de nombreux utilisateurs ont trouvé ces contrôles pour être source de confusion et les paramètres par défaut ont conduit à révéler des listes d'amis.

Le développement de **Facebook OpenGraphProtocol** et d'autres services plus distribués mené les conditions d'utilisation de **Facebook** à un inconnu encore plus ouvert avec les données des utilisateurs.

#### **II.4. Les réseaux sociaux libres comme :**

Google+, Facebook et Twitter sont les leaders du réseau social; depuis plusieurs années ils ont



le monopole planétaire sur Internet. Mais depuis quelques années ont voit émerger une nouvelle conception des réseaux sociaux; les réseaux sociaux libres.

Encore méconnus du grand public, les réseaux sociaux libres tissent leur toile sur internet, par le biais des logiciels libres. L'émergence de ces nouveaux moyens de communication libre offrent une alternative aux internautes qui souhaiteraient ne pas ou ne plus utiliser les « géants des réseaux sociaux » tout en permettant de créer son propre réseau social.

Le but de ces réseaux sociaux personnalisés et libres est de préserver sa vie privée et son anonymat si on le souhaite, tout en communiquant facilement. Ces logiciels libres peuvent aussi intéresser des projets de collectivités locales, institutions, associations, TPE-PME ou projets qui souhaitent mettre en place un réseau social avec une maîtrise de A à Z sur le processus de création, de paramétrage et de sauvegarde des données.

Il existe plusieurs sortes de réseaux sociaux libres; les réseaux sociaux décentralisés, les réseaux sociaux distribués, ainsi que les réseaux sociaux centralisés.

#### **II.4.1-Facebook, c'est quoi?**

Face book est définie sur sa page d'accueil de la façon suivante :

« Face book vous permet de rester en contact avec les personnes qui comptent dans votre vie. »

D'une manière plus générale Facebook est ce qu'on appelle un réseau social et permet à toutes personnes de rester en contact avec sa sphère familiale, amicale et professionnelle. Qu'elles soient proches ou éloignées de nous, il est parfois utile de réunir toutes nos communautés pour en retirer des avantages dont le premier est le « réseau ». Aujourd'hui Facebook compte plus de 600 millions d'utilisateurs dans le monde et plus de 20 millions en France.

Ouvrir un compte Facebook est d'une simplicité remarquable et d'une rapidité inouïe. La force de Facebook et sa popularité grandissante résident surtout dans le fait que cette plateforme est totalement gratuite.

##### **II.4.1.1-Face book, ça sert à quoi?**

Les usages de Facebook sont à la fois simples et variés. Vous pouvez avoir un « Profil » pour échanger des informations, des messages, des vidéos, des photos avec vos amis ou encore avoir une « Page » ou un « Groupe » pour faire connaître vos services, vos produits, vos projets...à toute la communauté des internautes. Votre profil est personnel et ne peut-être consulté que par vos amis ou les amis de vos amis. En revanche, lorsqu'il s'agit d'une page institutionnelle, celle-ci est visible par tout le monde et aussi par les moteurs de recherche.

D'un point de vue personnel, Facebook peut vous aider à renouer avec des amis perdus de vue, vous rapprocher de certains membres de votre famille, échanger des opinions, inviter à des évènements, partager des souvenirs de vacances et j'en passe.

D'un point de vue professionnel, Facebook est une opportunité qu'il ne faut pas délaissé. Créer une «Page» ou un «Groupe» à l'image de sa marque est un excellent moyen de publicité gratuite. Ce média vous permet de recruter de nouveaux clients, de fidéliser les anciens, d'échanger avec eux des idées et des nouvelles, de présenter une offre....

#### **II.4.2 : Google+**

Google Plus, futur carrefour d'influence pour les marques et les entreprises adeptes du marketing sur les réseaux sociaux ? Concurrent direct de Twitter et Facebook, le nouveau « phénomène social » de Google propose une nouvelle façon d'échanger des contenus sur le web. Outre son organisation en « cercles », qui donne à l'utilisateur le contrôle sur l'information qu'il partage, Il propose plusieurs outils et fonctionnalités, qui intégrés à

l'écosystème Google, offrent des perspectives intéressantes aux professionnels : pour développer sa marque personnelle, travailler en mode collaboratif, ou encore organiser une veille d'informations. Alors, comment exploiter au mieux les possibilités de cette plateforme pour le travail ? Voici un point complet sur le fonctionnement et les subtilités de Google Plus, et des pistes et conseils pour utiliser ce nouveau réseau social en contexte professionnel.

#### **II.4.2.1 : S'inscrire à Google Plus et commencer à resauter**

L'inscription à Google Plus n'est pour l'heure possible que sur invitation, dans l'attente d'une disponibilité totale du service. Les invitations peuvent être distribuées par tous les membres une fois qu'ils sont inscrits au réseau :

Soit par e-mail

Soit en partageant un lien groupé (ex : sur Facebook, Twitter, etc), qui est généré à partir de l'interface Google Plus

Si dans votre entourage, personne n'est inscrit de Google Plus, vous pouvez toujours vous y inscrire, par une astuce très simple : faites une recherche par mots-clés dans Twitter (ex : « invitation Google Plus », ou en anglais « Google Plus invites »). De très nombreux membres postent des invitations groupées permettant de s'inscrire au réseau. Une fois que vous êtes inscrit au réseau, vous pouvez donc à votre tour envoyer des invitations groupées (dans la limite de 150 unités). Vous pouvez également importer vos contacts depuis Yahoo Mail ou Hotmail pour simplifier la procédure d'envois d'invitations.

#### **II.4.2.2 : Fonctionnement général des cercles Google Plus**

Google Plus permet à ses membres d'organiser leur réseau, et les informations qu'ils partagent, à l'intérieur de « cercles », c'est-à-dire des groupes de contacts indépendants les uns des autres. L'objectif est de créer des groupes séparés et de laisser la main à l'utilisateur sur l'information qu'il partage (lien, photo, vidéo, données de géolocalisation), et avec qui il la partage (pour chaque élément posté). Plusieurs groupes peuvent donc coexister à l'intérieur du même réseau, sans risque qu'une membre non autorisé par l'utilisateur n'ait accès à une information qui ne le concerne pas.

A retenir :

Vous pouvez créer des cercles, c'est-à-dire des groupes d'utilisateurs, pour classer tous vos contacts par niveau de relation (amicale, familiale, professionnelle, etc.)

Le libellé des cercles n'est connu que de vous : une personne ajoutée dans votre réseau ne peut pas savoir à quel cercle elle appartient dans votre réseau

Vous pouvez ajouter une même personne à plusieurs cercles

Lorsque vous partagez du contenu sur votre "mur" (photo, vidéo, lien), vous pouvez choisir au cas par cas les cercles (et donc les contacts liés à ces cercles) qui seront autorisés à voir ce contenu (quand ils visualisent leur profil, par exemple)

Ce contenu apparaîtra sur leur propre mur, s'ils vous ont ajouté à l'un de leurs cercles

#### **II.4.2.3 : Le principe de cercles étendus et du filtre « mode public »**

Google Plus donne la possibilité de définir un niveau de visibilité pour chaque information publiée. Chaque flux d'informations (reçues ou partagées) correspond à un cercle, l'utilisateur pouvant naviguer d'un cercle à l'autre pour afficher des flux différents. Par défaut, une information publiée est visible seulement par les membres du cercle sélectionné. Ainsi, un message posté dans le cercle « ami », n'est visible par défaut que par les contacts du cercle « ami».

Pour plus de flexibilité, vous pouvez également choisir d'étendre la visibilité d'une publication en appliquant manuellement (pour chaque élément posté) deux autres filtres :

Cercles étendus : la publication est alors visible par une relation de « second degré », c'est-à-dire par une personne qui figure dans le cercle d'une personne figurant dans l'un de vos cercles.

En mode public : la publication est visible par tous les utilisateurs qui vous ont ajouté à un cercle, ainsi que dans votre profil public.

#### **II.4.2.4 : Créer une communauté Google plus**

Les communautés Google Plus sont des espaces de communication thématiques situés à l'intérieur du réseau social de Google : elles ont vocation à susciter la discussion et les échanges entre membres de Google Plus.

Elles s'opposent ainsi aux pages, qui sont des lieux de communication privilégiés pour les marques et leurs clients/prospects.

Une communauté thématique peut cependant être adossée à l'activité d'une entreprise, en tant que lieu de discussions autour d'un domaine particulier : marketing, référencement, nouvelles technologies, etc.

#### **II.4.3 : Twitter**

Twitter est un outil de microblogage géré par l'entreprise Twitter Inc. Il permet à un utilisateur d'envoyer gratuitement de brefs messages, appelés tweets (« gazouillis »), sur internet, par messagerie instantanée ou par SMS. Ces messages sont limités à 140 caractères.

Twitter a été créé le 21 mars 2006 par Jack Dorsey, Evan Williams, Biz Stone et Noah Glass, et lancé en juillet de la même année. Le service est rapidement devenu populaire, jusqu'à réunir plus de 500 millions d'utilisateurs en février 2012. [17].

Le siège social de Twitter Inc. se situe aux États-Unis à San Francisco, elle dispose de bureaux supplémentaires et de serveurs informatiques à New York. [17].

#### **II.4.3.1 : Identité visuelle (logo)**

La mascotte de Twitter est un oiseau stylisé, nommé Larry34 en hommage au basketteur américain Larry Bird 35. Le 6 juin 2012, Twitter présente un nouvel oiseau36, désormais utilisé comme logo unique, sans typographie37,38.

Logo de Twitter de sa création en 2006 à mai 2009. 

Logo de Twitter du 15 mai 2009 à septembre 2010. 

Logo de Twitter de septembre 2010 à juin 2012. 

Logo de Twitter depuis le 6 juin 2012. [17]. 

#### **II.4.3.2-Fonctionnalités**

Twitter est un service de microblogage ou microblogging, qui permet à ses utilisateurs de bloguer grâce à de courts messages, des « tweets ». Outre cette concision imposée, la principale différence entre Twitter et un blog traditionnel réside dans le fait que Twitter n'invite pas les lecteurs à commenter les messages postés. La promesse d'origine de Twitter, « What are you doing? », le définit comme un service permettant de raconter ce qu'on fait au moment où on le fait. Prenant acte de l'utilisation du service pour s'échanger des informations et des liens, Twitter le remplace par « What's happening ? » (« Quoi de neuf ? » ou encore « Que se passe-t-il ? » dans la version française), puis par « Compose new Tweet... » dans la dernière version de septembre 2011. De surcroît, Twitter est généralement utilisé comme plateforme de réseau social. Par contre, son interface et son format sont très différents d'un univers comme celui de Facebook39 ou Instagram, quoiqu'il existe certaines similarités. La culture très particulière et propre à sa communauté font en sorte que Twitter peut demander un peu plus d'effort de la part des nouveaux utilisateurs. Plusieurs guides existent d'ailleurs dans le but d'aider les nouveaux usagers à mieux s'immiscer dans l'univers des tweets. [17].

#### **II.5. Web social et profils utilisateur**

La figure ci-dessous montre comment un utilisateur unique (une personne) peut avoir plusieurs profils qui partagent des attributs communs. Un utilisateur peut alors associer son / ses profil au niveau du profil avec des applications particulières sociaux, leur contrôle en quelque sorte de vue agrégé que l'utilisateur peut avoir un accès aux applications de bureau

via un agrégateur. Les profils sont exposés à et / ou synchronisés avec différentes plateformes sociales. Dans la figure II.1 un profil est associé à la "lumière bleue" et des applications sociales "rouges", un profil à la demande sociale "gris", et un profil à le "bleu", "vert", et les applications sociales « orange».

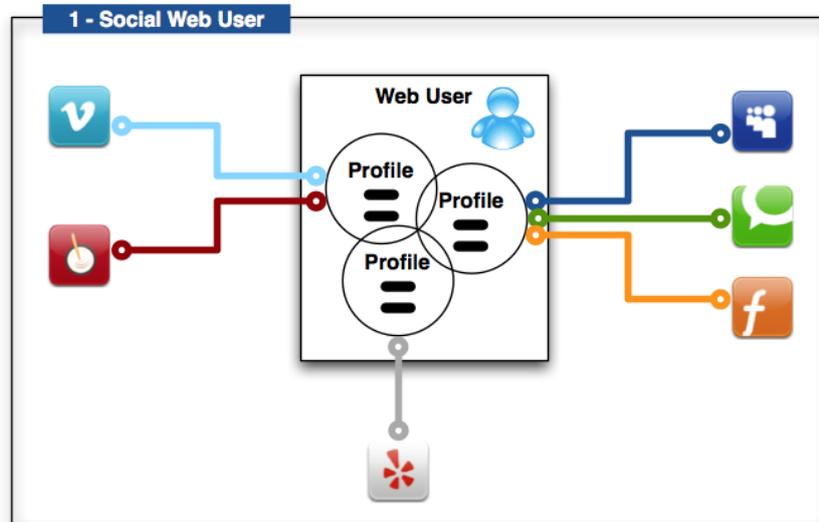


Figure II.1: web social et profile d'utilisateur (URL12)

## II.6.L'influence des RS sur le commerce électronique

### Les réseaux sociaux,

Facebook en tête, ont profondément modifié notre façon de surfer et plus encore notre façon de consommer. Ils sont effectivement venus se greffés au web à une période où le **e-commerce** était en plein essor. Rapidement adoptés par les particuliers, ils ont de ce fait mis très peu de temps à également séduire les professionnels qui ni une ni deux les ont adoptés pour en faire de véritables leviers marketing. Au delà de leur fonctionnalité première qu'est la communication, ils sont ainsi de plus en plus fréquemment utilisés pour suivre l'actualité des marques et profiter de bons plans commerciaux, ce qui se ressent indéniablement sur le chiffre d'affaires de celles qui savent les exploiter efficacement. [19].

Toutefois, malgré l'importance croissante qu'ont pris les **réseaux sociaux** au cours des derniers mois, voir même des dernières années, il reste très compliqué de mesurer leur impact direct sur les ventes réalisées par les sites marchands. Autrement dit, la fameuse question du ROI (retour sur investissement) se pose toujours, si bien que les spécialistes préfèrent désormais parler de ROE (retour sur engagement), indice qui lui peut être calculé facilement. Dans cette optique, il convient de s'intéresser avant tout à la manière dont Facebook, Twitter et autre Pinter est influencent le comportement des consommateurs, en

mettant l'accent sur certains aspects de leur potentiel social qui indirectement génèrent des achats.

C'est pourquoi a réalisé l'**infographie** ci-dessous. Celle-ci dévoile une série de statistiques qui prouvent que, sans pour autant lever le voile sur leur faculté à générer des ventes directes. On y apprend par exemple que 75% des membres de Facebook aiment au moins une marque et que les internautes passent quasiment un quart de leur temps sur les plateformes citées précédemment. Par conséquent, il n'y a rien d'étonnant à ce que les partages et les recommandations soient en constante augmentation... Des activités sociales qui ont le vent en poupe et dont on estime qu'elles seraient aujourd'hui à l'origine de près de 90% des commandes enregistrées par les boutiques en ligne. Un chiffre à prendre tout de même avec des pincettes. Quoi qu'il en soit, les **réseaux sociaux** influent bel et bien le commerce électronique, en attestent les autres données présentées dans l'**infographie** à l'image du fait que 51% des fans d'une marque sont d'avantage prédisposés à acheter [19].

## **II.7 : Les standards pour la description des profils tels que OAuth**

**OAuth** est un protocole libre, créé par Blaine Cook et Chris Messina. Il permet d'autoriser un site web à utiliser l'API sécurisée d'un autre site web pour le compte d'un utilisateur. OAuth *n'est pas* un protocole d'authentification.

OAuth permet aux utilisateurs de donner, à un site « consommateur », l'accès à des informations personnelles provenant d'un site « fournisseur » de service ou de données, ceci tout en protégeant le pseudonyme et le mot de passe des utilisateurs. Par exemple, un site de manipulation de vidéos pourra éditer les vidéos enregistrées sur Dailymotion d'un utilisateur des deux sites, à sa demande.

La partie principale de ce protocole, OAuth Core 1.0, a été finalisée le 3 octobre 2007 [20].

## **II.8. Conclusion**

Comme on a pu le constater, les réseaux sociaux ont influencé notre société sur tous les plans. Tout d'abord d'un point de vue politique, les réseaux sociaux ont su conquérir les pays touchés par «le printemps arabe». En effet c'est par le biais de ces derniers que tunisiens et égyptiens ont pu enfin s'exprimer et montrer leur désaccord .On peut ainsi affirmer que les réseaux sociaux ont joués un rôle essentiel.

Puis d'un point de vue culturel et social, nous pouvons dire que les réseaux sociaux tel que Facebook, ont bouleversé notre société, notamment en modifiant les comportements et les habitudes des individus. Par ailleurs ces derniers ont également influencé le domaine du recrutement et de l'emploi. Ainsi, il est indéniable que les réseaux sociaux ont eu un impact tant culturel que social.

Enfin, d'un point de vue commercial, les réseaux sociaux se sont imposés comme un instrument moderne et efficace de publicité. En effet, réaliser de la publicité via les réseaux sociaux est un excellent moyen d'attirer un maximum de clients et ainsi d'accroître son chiffre d'affaire. De ce fait, nous pouvons dire que les réseaux sociaux exercent une influence sur les stratégies commerciales.

Nous pouvons donc conclure que les réseaux sociaux sont devenus une composante indispensable de notre quotidien et du monde dans lequel nous vivons. Ces derniers ont exercé et continuent encore d'exercer une influence politique, sociale, culturelle et même sur les stratégies commerciales, en devenant un support efficace pour la publicité.

# Chapitre 3

**Base de données XML**

### **III.1. Définition d'une base de données XML native**

Une base de données doit répondre à plusieurs critères pour prétendre être de type XML natif [21, 22,23] : Elle doit définir un modèle logique pour tout document XML et se baser sur ce modèle pour stocker et extraire des documents. Le modèle doit au minimum inclure les éléments, les attributs, les PCDATA et l'ordre interne du document. Quelques exemples de tels modèles sont : le modèle de données de XQuery et le glossaire XML Infoset[24]. De plus, elle doit considérer le document comme unité fondamentale de stockage. Cela change radicalement des BD relationnelles, pour qui une unité de stockage correspond à une ligne d'enregistrement d'une table donnée. Enfin, le format de stockage physique des données composant les documents n'a pas d'importance. Il pourra être de type objet, relationnel, propriétaire, .... Peu importe, tant que ce dernier n'influence pas l'intégrité des données traitées.

Cette définition exprime trois idées principales [22] :

La base de données est spécialisée pour le stockage de documents XML, et elle stocke tous les composants du modèle XML d'une manière intacte.

Les documents peuvent être stockés et retrouvés.

La BDXN n'est pas destinée à remplacer les bases de données classiques.

### **III.2. XML et les bases de données**

Depuis peu, XML est devenu le standard incontournable pour l'échange de données sur le réseau. Les éditeurs de système de gestion de bases de données (SGBD) se tournent logiquement vers ce nouveau format.

#### **III.2.1 XML, une base de données ?**

Peut-on dire que XML et les technologies le concernant est une base de données ? XML fournit des fonctions proposées par les bases de données comme :

Le stockage de données

Les schémas (DTD, XML Schemas, etc.)

Les langages de requêtes (XQuery, XPath, XQL, etc.)

Les interfaces de programmation (SAX, DOM, JDOM)

Mais il lui manque aussi beaucoup de fonctions tel que :

Le stockage efficace

L'indexation

La sécurité

Les transactions

L'intégrité des données

L'accès multi-utilisateurs

Les requêtes sur plusieurs documents.

Nous pouvons donc utiliser XML comme une base de données pour peu que l'environnement d'utilisation comporte peu de données et que le besoin de performance soit faible.

Dans les autres cas, il faudra associer le document XML avec un système de gestion de bases de données ou alors tout simplement d'utiliser un SGBD classique.

### **III.2.2 Données versus document**

Un critère important dans le choix d'une base de données est le type d'informations que l'on désire stocker. On peut distinguer deux catégories :

Les « données brutes » mises au format XML

Les « documents » (tels que textes, articles, rapports, ...) formatés en XML

Nous allons voir les deux sortes d'XML : l'orienté données et l'orienté documents. Nous montrerons les différences entre ces deux types.

#### **III.2.2.1 XML orienté données**

Les contenus orientés données utilisent XML en tant que vecteur de données. Ils sont conçus pour être exploités par une machine. Le cycle de vie du fichier se limite au temps de transfert entre les applications. Le stockage de ces données au format XML présente donc peu d'intérêt.

Les exemples de ce type donné, sont les données scientifiques collectées par un système de mesure ou les informations sur un vol.

*Exemple d'informations d'un vol :*

```
<Vols>
<Compagnie>Swiss</Compagnie><Départ>Genève</Depart>
<Destination>New-York</Destination><Vol>
<HeureDepart>09: 15</HeureDepart><HeureArrivee>1 3:1 6</HeureArrivee></Vol>
<Vol>
<HeureDepart>11:15</HeureDepart><HeureArrivee>1 5:1 6</HeureArrivee></Vol>
<Vol>
<HeureDepart>13:15</HeureDepart><HeureArrivee>1 7:1 6</HeureArrivee></Vol>
</Vols>
```

#### **III.2.2.2 XML orienté documents**

Les contenus orientés documents sont habituellement conçus pour être utilisés par les humains. Les livres, les messages électroniques en sont des exemples, car ils sont caractérisés par une structure irrégulière. Les documents XML orientés documents sont donc peu structurés

de façon à être stockés dans sa forme originelle dans la base de données pour être retrouvés et réutilisés.

*Exemple :*

```
<description>
```

```
Les ordinateurs <constructeur> Dell </constructeur><modèle>Inspiron 505 </modele> sont
pourvus d'un processeur <cpu> Pentium 4 </cpu> et d'un grand écran. Vous pouvez les
commander sur Internet au prix de <prix>2000.-</prix>
```

```
</description>
```

La distinction entre les deux types d'XML n'est pas toujours claire en pratique.

### **III.2.3 Stockage de données XML**

On peut utiliser certains outils fournis avec les bases de données relationnelles tel que Oracle 9i ou IBM DB2. Ces outils peuvent traduire en XML les données structurées ou non.

Mais il existe une autre méthode de stockage : les bases de données XML natives.

#### **III.2.3.1 Mapping des documents**

Afin de transférer les données d'un document XML à une base de données, nous avons besoin de mapper le schéma XML (DTD, XML Schema, etc.) à un schéma de base de données. Ce mapping est basé sur le type d'éléments, les textes et les attributs. Seules les données du document XML sont concernées. Voici les deux techniques de mapping :

La première technique est le mapping basé sur les tables. Le document est structuré comme des tables d'une base de données relationnelle.

*Exemple de structure :*

```
<BD>
```

```
<table>
```

```
<ligne>
```

```
<colonne1>...</colonne 1><colonne2>...</colonne 2><colonne3>.. </colonne 3>
```

```
</ligne>
```

```
<ligne>
```

```
<colonne1>...</colonne 1><colonne2>...</colonne 2><colonne3>.. </colonne 3>
```

```
</ligne>
```

```
<ligne>
```

```
<colonne1>...</colonne 1><colonne2>...</colonne 2><colonne3>.. </colonne 3>
```

```
</ligne>
```

```
</table>
<table>
<ligne>
<colonne1>...</colonne 1><colonne2>...</colonne 2></ligne>
</table>
</BD>
```

La deuxième technique de mapping est basée sur un modèle objet relationnel. Elle modélise le document XML en un arbre d'objets spécifiques au document. Les types des éléments avec leurs attributs et leur contenu sont représentés par des classes. Le modèle est ensuite traduit en une base de données relationnelle. Les classes sont mappées en tables, les propriétés scalaires en colonnes etc

### III.2.3.2 Stockage dans une base relationnelle

Il existe deux façons de stocker un document XML orienté données et pouvoir le récupérer en XML.

L'utilisation d'un Middleware qui assure la conversion entre SGBD et XML.

L'utilisation d'un SGBD où XML est déjà intégré.

Voici deux exemples de tels SGBD connus :

IBM's DB2 avec DB2 XML Extender, cet outil permet d'extraire et de stocker le document XML entièrement associé avec son DTD. Mais aussi, il le partage en plusieurs tables et colonnes. Le langage de requête utiliser est le XPATH.

Oracle 9i où les documents peuvent être stockés entièrement ou de façon décomposée en utilisant le mapping orienté objets. Pour la récupération des données, on utilise XPATH.

Pour les documents XML orientés documents, le *Système de Gestion de Contenu* permet l'extraction, la modification et la mise à jour de documents XML dans un environnement multi-utilisateurs.

Voici des exemples de tels outils :

SyCOMAX

SiberSafe

L'inconvénient de ce type de stockage est principalement le recours à la technique de mapping. Ceci pourrait générer une grande quantité de tables et de colonnes dans la base de données relationnelle. De plus, si ces outils sont utilisés pour traduire des données en direct lors d'un échange XML, cela risque d'augmenter les temps de traitement des transactions XML et de ralentir également les autres applications qui accèdent à la base de données relationnelle.

### III.2.3.3 Stockage à l'aide d'une base de données native XML

Ces bases de données sont apparues récemment sur le marché et elles prennent de plus en plus de terrain bien qu'elles ne remplacent pas les bases relationnelles.

#### Définition

Une base de données doit répondre à plusieurs critères pour prétendre être de type XML natif. Elle doit déjà définir un modèle logique pour tout document XML et se baser sur ce modèle pour stocker et extraire des documents. De plus, elle doit considérer un document comme unité fondamentale de stockage. Cela change radicalement des bases de données relationnelles, pour qui une unité de stockage correspond à une ligne d'enregistrement d'une table donnée. L'unité physique de stockage n'est pas spécifique, elle pourrait être reliée à une BD relationnelle, BD orienté objet, etc. Peu importe, tant que ce dernier n'influence pas l'intégrité des données traitées.

Stockage et extraction de « documents »

Ces bases de données sont conçues spécialement pour stocker des documents XML entiers. Comme les bases de données traditionnelles, elles permettent de gérer les transactions, la sécurité, l'accès multi-utilisateurs, les langages de requêtes, ....

Le grand avantage est que ces bases de données conservent les données dans leur ordre et leur structure originale, sans toucher à leur intégrité. Ce n'est pas forcément le cas des bases de données relationnelles qui peuvent décomposer entièrement un document pour pouvoir le stocker dans des tables et des colonnes. De plus, les requêtes pouvant être effectuées sur des bases de données XML sont bien plus performantes et peuvent être nettement plus complexes que ce qu'offre le langage SQL traditionnel.

Les bases de données XML peuvent également servir à stocker des documents qui ne sont pas totalement structurés ou même des documents n'ayant pas du tout de DTD ou de Schema. Cela ne leur posera aucun problème, alors qu'avec une base relationnelle, cela serait relativement difficile à effectuer.

Langages de requêtes

Les plus populaires des langages de requête sont XPath et XQL. Toutes les bases de données ne les proposent pas, certaines offrant des langages de type propriétaires. Le W3C, quant à lui, a édicté une norme se nommant XQuery qui se base d'ailleurs sur XPath. Elle permet de construire des requêtes relativement complexes. Un chapitre dédié à XPath et un autre à XQuery en présentent les fonctionnalités.

En attendant que XQuery soit passé au statut de recommandation (cette norme est encore en développement), certains fournisseurs proposent XPath auquel ils ont ajouté des extensions propriétaires afin de le rendre plus performant.

Le fait que certaines fonctions ne sont pas forcément disponibles dans toutes les bases de données natives peut renforcer les arguments portés en faveur des bases de données relationnelles.

### **III.3. Avantages-inconvénients des BD natives et relationnelles**

Nous allons voir maintenant les avantages et les inconvénients de l'utilisation de documents XML entre les bases de données XML natives et les bases de données relationnelles.

L'inconvénient des SGBD relationnelles est premièrement l'utilisation d'un middleware (si le SGBD n'intègre pas le XML) qui complique la mise en utilisation et la maintenance de la base de données. De plus ce type de base nécessite le recours à la technique du mapping, ce qui pourrait générer une grande quantité de tables et de colonnes dans la base de données relationnelle. La base pourrait donc ne plus être efficace.

L'avantage des bases de données natives est la mise-à-jour des données efficace, le chargement efficace de gros documents. Toutefois, elle présente deux principaux inconvénients. Tout d'abord, lorsque nous désirons obtenir différentes vues sur des données, elle est moins performante que la base de données relationnelle. Ensuite, elle ne retourne les données que sous la forme XML; si une application a besoin d'un autre format de document, elle serait obligée de le passer avant d'utiliser les données.

Enfin le fait que certaines fonctions des bases de données relationnelles ne soient pas forcément disponibles dans une base de données XML peut renforcer les arguments portés en faveur des bases de données relationnelles.

De plus XML étant une technologie relativement récente, les bases de données natives n'ont pas encore été largement testées. Leur coût n'est non plus pas en leur faveur, bien que l'on voit toutefois arriver des bases à prix raisonnable. La plupart ont été développées par la communauté open-source et arrivent désormais sous forme commerciale.

*Avantages et inconvénients des BD natives XML par rapport au BD relationnelles*

Spécifiquement conçu pour XML

Chargement efficace de gros documents

Mise-à-jour efficace

Retourne uniquement des données XML

Base de données assez récente donc peu connue

Leurs fonctions ne sont pas toujours complètes- Le prix

### **III.4 Exemples de choix de bases de données**

La plupart des utilisateurs de bases de données natives XML travaille dans les secteurs de la finance ou de la production et les utilise principalement pour leur grande vitesse de transaction.

Voici deux exemples de choix du type de base de données. Supposons par exemple que la base de données soit exploitée par une application de e-commerce dans laquelle XML est utilisé comme vecteur de données. Il y a fort à parier que vos données possèdent une structure hautement régulière et soient utilisées par des applications non XML. De plus, des notions telles que les entités et les techniques d'encodage utilisées dans les documents XML ne vous seront probablement pas importantes. Après tout, ce sont les données qui vous intéressent, pas la façon dont elles sont stockées dans un document XML.

Dans ce cas de figure, je vous conseillerai l'utilisation d'une base de données relationnelle et d'une couche logicielle pour transférer les données entre les documents XML et la base.

Supposons par contre que vous disposiez d'un site Web constitué d'un grand nombre de documents orientés texte. Vous souhaitez alors non seulement gérer ce site, mais également fournir aux utilisateurs un moyen de recherche sur le contenu. Vos documents ont toutes les chances d'avoir une structure bien moins régulière, et des notions telles que les entités seront probablement importantes pour vous parce qu'il s'agit là d'un aspect fondamental de la manière dont les documents sont structurés. Dans ce cas, vous souhaitez utiliser une base de données XML native ou un système de gestion de contenu. Cela vous permettra de préserver la structure physique des documents, de supporter les transactions au niveau du document, et d'effectuer des interrogations dans un langage de requête XML.

### **III.5 Conclusion**

Le XML étant de plus en plus utilisé (pour la rédaction de documents, la structuration de données, ...), il s'est rapidement fait ressentir le besoin de trouver un moyen de stocker ces informations de la manière la plus efficace possible.

Les bases de données XML ne pourront pas remplacer les bases de données relationnelles. Ce n'est d'ailleurs pas leur but, puisqu'elles ont toutes deux des objectifs différents et sont parfaitement complémentaires. Alors que les bases relationnelles sont plus adaptées au stockage de "données", les bases de données XML sont plus adaptées au stockage de "documents".

La base de données XML facilite l'utilisation des données dans les applications web et en a utiliser la base de donnée fédérer pour facilité à l'utilisateur les opérations de consultation des donnée.

# Chapitre 4

Construction d'une identité sociale fédérée

## **IV.1 : Introduction**

L'importance de l'intégration de plusieurs schéma d'identité social en un seul c'est d'assurer une interaction intelligente avec les utilisateurs en proposons le mieux d'info utile extraite des réseaux sociaux.

La masse d'utilisation des applications web rendre la reconnaissance automatique des internautes prés nécessaire pour utilisation efficace.

Dans ce chapitre en va expliquer les étapes de la réalisation du projet.

En a 3 parties, partie 1 la conception d'application, partie 2 extraction des donnée et la partie3 implémentation.

### **IV .2.Parti1 : Conception :**

Pour réaliser une telle intégration, quatre schémas sont utilisés dans les bases de données composant la fédération; le schéma d'exportation (ES) pour les données permises à l'exportation, le schéma d'importation (IS) pour les données importées, le schéma privé (PS) pour l'ensemble de données privées (ES et IS inclus) et le schéma commun (global, canonique ou pivot) qui dans le cas des bases de données faiblement couplées n'est qu'une abstraction se trouvant dans le dictionnaire de données de la fédération (FD).

Dans le diagramme, nous avons quatre bases de données distinctes, offrant la portion de leurs données qui est nécessaire à la fédération via leurs ES. Les informations (métadonnées) concernant le schéma global se retrouvent dans la FD. Nous retrouvons également dans la FD les informations indiquant comment effectuer l'importation des données pour réaliser des vues sur le schéma global (liens en pointillés).

Nous voyons dans le diagramme que le schéma global n'existe que d'une façon virtuelle, les différents administrateurs de données locaux doivent créer des vues à la main sur les différents schémas d'exportation pour réaliser une vue totale ou partielle (mais virtuelle) du schéma global. Les schémas d'exportation (ES) sont en général des vues sur une portion de la base locale de manière à permettre la correspondance entre cette portion de la base et le schéma global. Nous constaterons que les schémas originaux restent inchangés et que les utilisateurs locaux peuvent continuer à utiliser leurs anciennes bases comme si de rien n'était. La seule nouveauté pour un usager local est l'accès possible aux données de la fédération via son IS local.

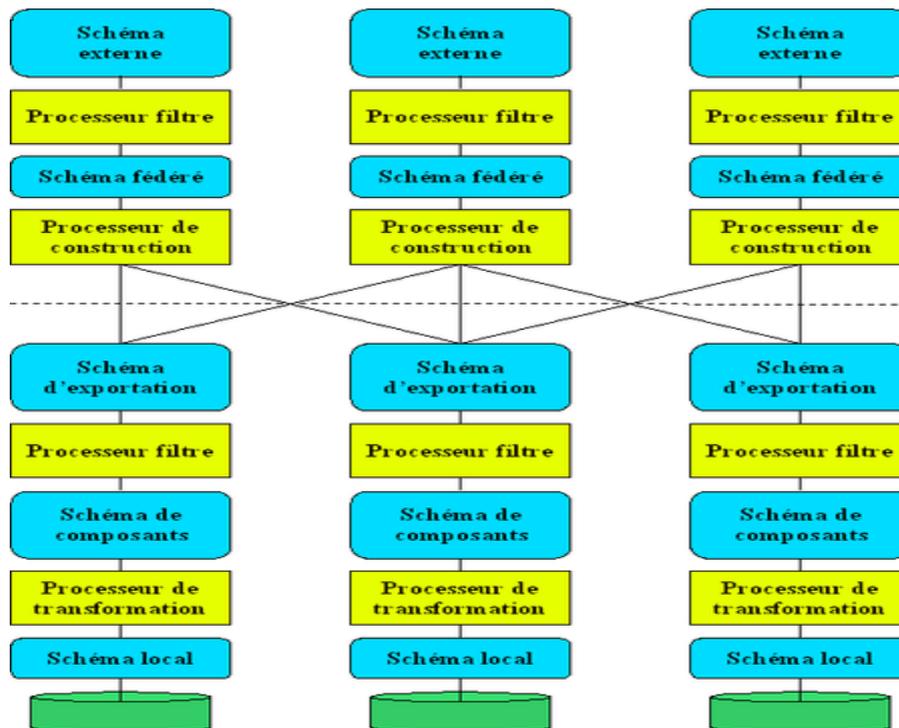


Figure IV. 1 : fédération fortement couplées (URL13)

## IV.2.1 : Extraction des schémas locaux à partir des réseaux sociaux

### IV.2.1.1 Facebook:

Ce que nous allons faire, c'est se connecter à Facebook depuis notre site, récupérer quelques informations sur notre compte.

**Étape1 :** L'application Facebook

En crée l'application sur Facebook !

Pour ceci, en un compte Facebook, puis aller sur :

<https://developers.facebook.com/apps>

En haut il y'a un bouton : <<créer une application>>, cliquez dessus (vraiment besoin de le préciser), un nom pour l'application, une langue, vous acceptez bien sur les conditions d'utilisation de Facebook et vous valider le formulaire.

**Étape2 :** Téléchargement du SDK et mise en place de l'environnement de travail

Pour cela, en va sur le site github afin de télécharger SDK PHP de Facebook (<https://github.com/facebook/php-sdk>)

Une fois votre archive obtenu, naviguez à l'intérieur jusqu'à trouver le dossier src en le désarchive ce dossier, et mettez le à la racine de notre espace de travail.

**Étape 3 :** On ouvre l'EDI

Créons deux fichiers facebook\_actions.php (qui nous servira à faire tous les appels au SDK),

avez les bases pour récupérer les informations d'une personne, et même poster sur un mur !

En a exploiter le mode offline

En remarque que la demande de la permission offline-access, grâce a cela, l'utilisateur ne doit se connecter qu'une seule et unique fois pour que je puisse utiliser l'API Facebook en son nom.

Fql : Facebook Query Langage, ou FQL, vous permet d'utiliser une interface de style SQL pour interroger les données récupérées de Facebook, Le FQL est interprété et traité par l'API Facebook.

#### **IV.2.1.2 : Twitter :**

L'api Twitter permet d' 'accéder' à la base de données Twitter et de récupérer/poster plusieurs informations. L'API se décompose en quatre classes :

SEARCH : permet d'interroger Twitter pour récupérer des données simples, essentiellement des tweets.

REST : extension avancée de REST qui permet d'accéder à des fonctionnalités avancées de Twitter : chercher des utilisateurs, des followers, voir les status, éditer des informations sur son compte, etc.

STREAMING : permet de communiquer avec Twitter en mode streaming. La particularité de cette api avancée est de permettre l'accès à de gros volume de données Twitter et d'être moins contrainte par les limites d'accès et d'interrogations de Twitter. Cependant, cet api requiert la mise en place de mécanisme plus complexe pour l'accès aux données.

WEBSITES : permet d'intégrer des fonctions de base Twitter dans des sites webs.

#### **IV.2.1.3 : Google+ :**

API g+ : L'API Google+ est l'interface de programmation sur Google+. Cette API permet d'intégrer votre application ou votre site Web à Google+. Elle permet aux utilisateurs de se connecter les uns aux autres à l'aide des fonctionnalités Google+ présentes dans votre application pour susciter un intérêt maximal de leur part.

#### Remarque :

Les problèmes rencontrés dans la transformation des schémas locaux sont :

-Problème (Interopérabilité)

-les applications change

-problème de permission de l'utilisateur en doit avoir la permission pour récupérer le profil.

## IV.2.2. Résultat

### IV.2.2.1 : schéma locale de Facebook

Voila le schéma local de Facebook :

**Profil Facebook** (id\_p, Nom, PageFacebook, Email1, Sexe, id, bio, education\_school\_id1, education\_school\_n1ame, education\_type1, year, education\_school\_id2, education\_school\_name2, education\_type2, email2, favorite\_teams\_id1, favorite\_teams\_name1, favorite\_teams\_id2, favorite\_teams\_name2, first\_name, gender, hometown\_id, hometown\_name, last\_name, link, location\_id, location\_name, locale, name, timezone, updated\_time, username, verified, work\_employer\_id, work\_employer\_name)

### IV.2.2.2 : schéma locale de Twitter

Voila le schéma local de Twitter :

**profil Twitter** (id\_p, Nom d'utilisateur, Adresseemail, Langue, Fuseau horaire, Pays, Nom, Localisation, Site Web, Biographie)

### IV.2.2.3. schéma locale de Google+

Voila le schéma local de Google+ :

**profil Google+** (id\_p, Adresse\_email\_principale, E-mail de récupération de mot de passe, nom, Téléphone, Formation, Sexe, Anniversaire, Situation amoureuse, Autres noms, Autres profils, Liens, Domicile, Professionnel, Adresses, Profession, Compétences, Emplois, langue)

## IV.3. Partie II : Intégration de données

### IV.3.1. Méthode utilisé :

**Pour générer le schéma fédérée nous avons appliqué l'algorithme suivant :**

Algorithme :

*Les entrées : deux schémas S1, S2*

*Sortie : le schéma fédérée S3*

*variables : x, y string ;*

*i, j int;*

*nouveau boolean*

*s1[n], s2[p], s3[]*

*Début*

*Nouveau = vrai ;*

*Pour i = 0 à n faire*

*Pour j = 0 à p faire*

*Si syno(s1[i], s2[j]) = faux alors*

*Nouveau = faux ;*

*Sinon*

*Y = s2[j]*

*Fin*

*Fpour*  
*Si nouveau*  
*Ajouter (S3,s1[i])*  
*Sinon*  
*X=Regrouper (s1[i],y)*  
*Ajouter (S3,x)*

*Fpour*  
Explication :

Dans cette algorithm il ya :

Ajouter (S,x) : ajouter l'attribut X au schéma S

Regrouper(x,y) :fonction qui renvoi le nom d'attribut pré d'être utilise dans le schéma fédérée a partir des -deux attributs synonyme x et y

syno(x,y): renvoi la valeur -vrais :si x est synonyme de y et faux sinon la synonymie est calculé a partie de la base de données -lexical wordnet<sup>1</sup>

Exemple :

En entrer :

***S1(nom, prénom ,date de naissance, affiliation)***

***S2 (nom, surnom, prénom, anniversaire, travail)***

*En sortie(résulta) :*

***S3(nom, surnom, prénom ,date de naissance, affiliation)***

Algorithme :

*variables : x,y string ;*

*i,j int;*

*nouveau Boolean ;*

*s1[n],s2[p],s 3[]*

1

*Début*

*Pour i= 0 à n faire*

*Pour j= 0 à p faire*

*Si syno(s1[i], s2[j])=faux alors*

*Nouveau= vrai ;*

*Sinon*

*Y=s2[j]*

*Finsi*

*Fin pour*

*Si nouveau alors*

*Ajouter (S3,s1[i])*

*Sinon*

*X=Regrouper (s1[i],y)*

*Ajouter (S3,x)*

*Fin pour*

Dans cette exemple le résultat sera *S3(nom, surnom, prénom ,date de naissance, affiliation)*

<sup>1</sup> Wordnet (WordNet est une base de données lexicale développée par des linguistes du laboratoire des sciences cognitives de l'université de Princeton depuis une vingtaine d'année<sup>1</sup>. Son but est de répertorier, classifier et mettre en relation de diverses manières le contenu sémantique et lexical de la langue anglaise. Des versions de *WordNet* pour d'autres langues existent, mais la version anglaise est cependant la plus complète à ce jour(<http://garr.dl.sourceforge.net/project/wnsq1/wnsq13/mysql/3.1/mysql-3.0.0-31-wn-1.zip>).

## Exemple de synonyme

Cette requête cherche les synonymes dans la base de donnée wordnet est donne les résultats si dessus

*Requête :* SELECT synsetid,dest.lemma,SUBSTRING(src.definition FROM 1 FOR 60)  
FROM wordsXsensesXsynsets AS src INNER JOIN wordsXsensesXsynsets AS dest  
USING(synsetid) WHERE src.lemma = 'name' AND dest.lemma <> 'name'

*Résultat :*

lemma	SUBSTRING(src.definition FROM 1 FOR 60)
Epithet	a defamatory or abusive word or phrase
Gens	family based on male descent
Figure	a well-known or notable person
public figure	a well-known or notable person
Diagnose	determine or distinguish the nature of a problem or an illne
Describe	identify as in botany or biology, for example
Discover	identify as in botany or biology, for example
Distinguish	identify as in botany or biology, for example
Identify	identify as in botany or biology, for example
Key	identify as in botany or biology, for example
key out	identify as in botany or biology, for example
List	give or make a list of; name individually; give the names of
Advert	make reference to
bring up	make reference to
Cite	make reference to
Mention	make reference to
Refer	make reference to
Identify	give the name or identifying characteristics of; refer to by
be known as	assign a specified (usually proper) proper name to
Call	assign a specified (usually proper) proper name to
lknow as	assign a specified (usually proper) proper name to
Appoint	create and charge with a task or function
Constitute	create and charge with a task or function
Nominate	create and charge with a task or function
Make	charge with a function; charge to be
Nominate	charge with a function; charge to be

### IV.3.2.Schéma fédérer

On appliquant l'algorithme de l'élimination des synonymes cités si dessus. On a eu le schéma suivant :

schémas fédérée (id, Nom\_f, PageFacebook, Email1, Sexe, id\_sexe, bio, education\_school\_id1, education\_school\_name, education\_type1, education\_school\_id2, education\_school\_name2, education\_type2, email2, favorite\_teams\_id1, favorite\_teams\_name1, favorite\_teams\_id2, favorite\_teams\_name2, first\_name, hometown\_id, hometown\_name, last\_name, link, location\_id, location\_name, name, timezone, updated\_time, username, verified, work\_employer\_id, work\_employer\_name, Nom d'utilisateur, Adresseemail, Langue, Fuseau horaire, Pays, Nom\_t, Localisation, Site Web, Biographie, Adresse\_email\_principale, E-mail de récupération de mot de passe, nom\_g, Téléphone, Formation, Anniversaire, Situation amoureuse, Autres noms, Autres profils, Liens, Professionnel, Adresses, Profession, Compétences, Emplois)

### IV.4.Partie III : Implémentation

#### IV.4.1.Les moyen Utiliser pour faire la connexion avec les réseaux sociaux

- API Facebook
- SDK (facebook-php-sdk-3.2.3)
- API Google+
- API Twitter
- MySQL

#### IV.4.2. La Base de données

Notre interface de base de donnée comme suit :

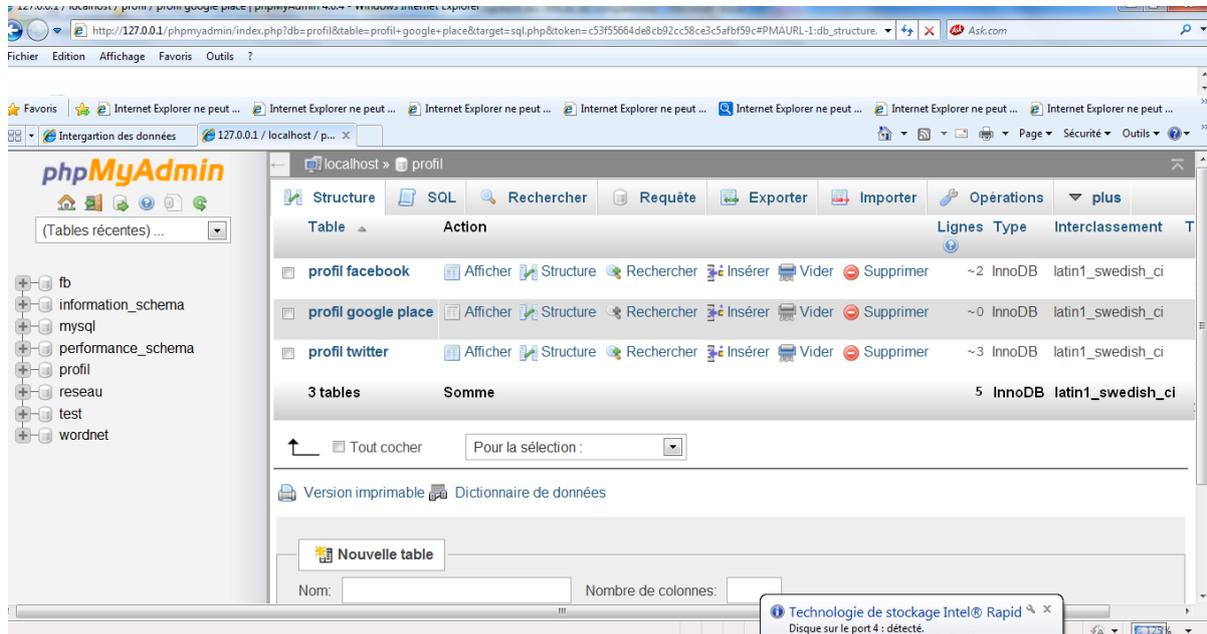


Figure IV.2 : La Base de données

#### IV.4.3. forme principale de l'application

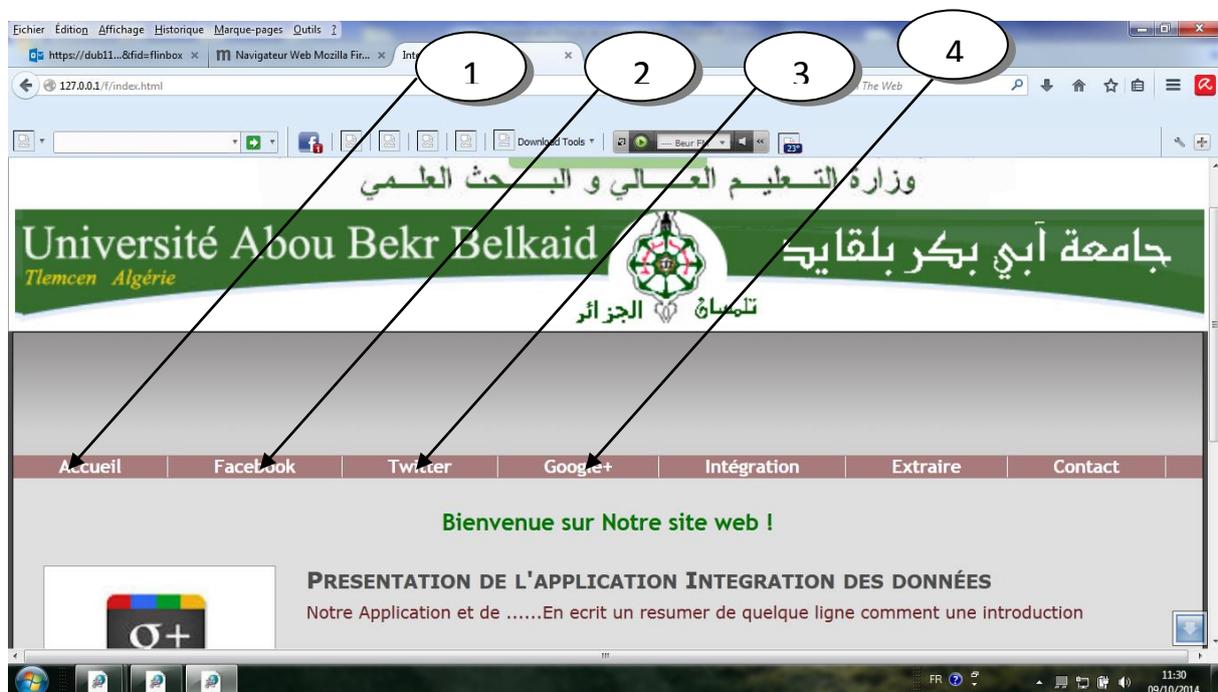


Figure IV.3 : l'interface de l'Application

Explication :

- 1 : bouton permet de retourner à la page d'accueil
- 2 : bouton permet à La page de Facebook

- 3 : bouton permet à La page de Twitter
- 4 : bouton permet à La page de Google+

#### IV.4.4. Extraction de l'identité Facebook

Notre interface de Facebook se présente comme suit :

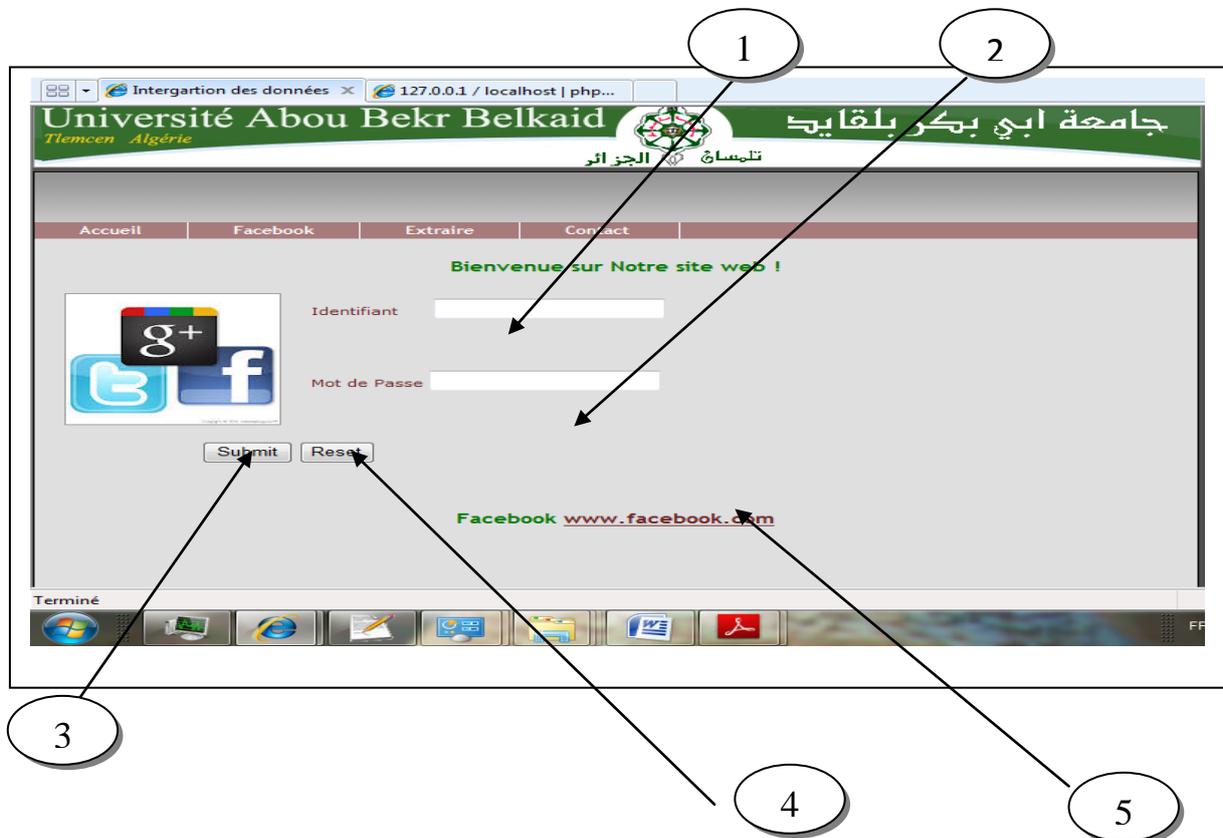


Figure IV.4 : l'interface d'extraction de l'identité Facebook

Explication :

- 1 : permet de remplir l'identifiant
- 2 : permet de remplir le mot de passe
- 3 : permet d'enter au résultat
- 5 : permet d'entrer a la page de Facebook

#### Exemple :

Voila un exemple de table de profil Facebook

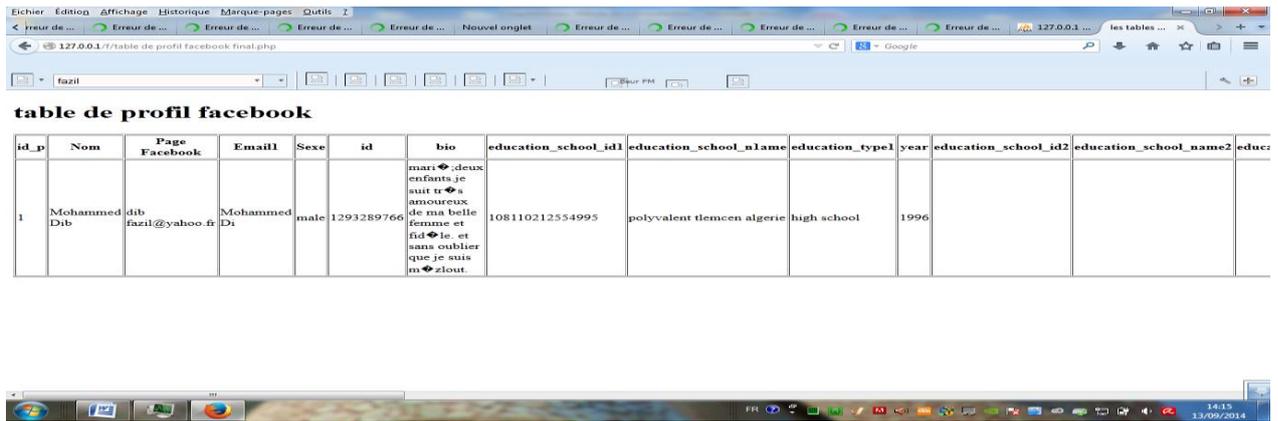


Figure IV.5 : exemple de table de profil Facebook

#### IV.4.5. Extraction de l'identité Twitter

Notre interface de Twitter présente comme suit :

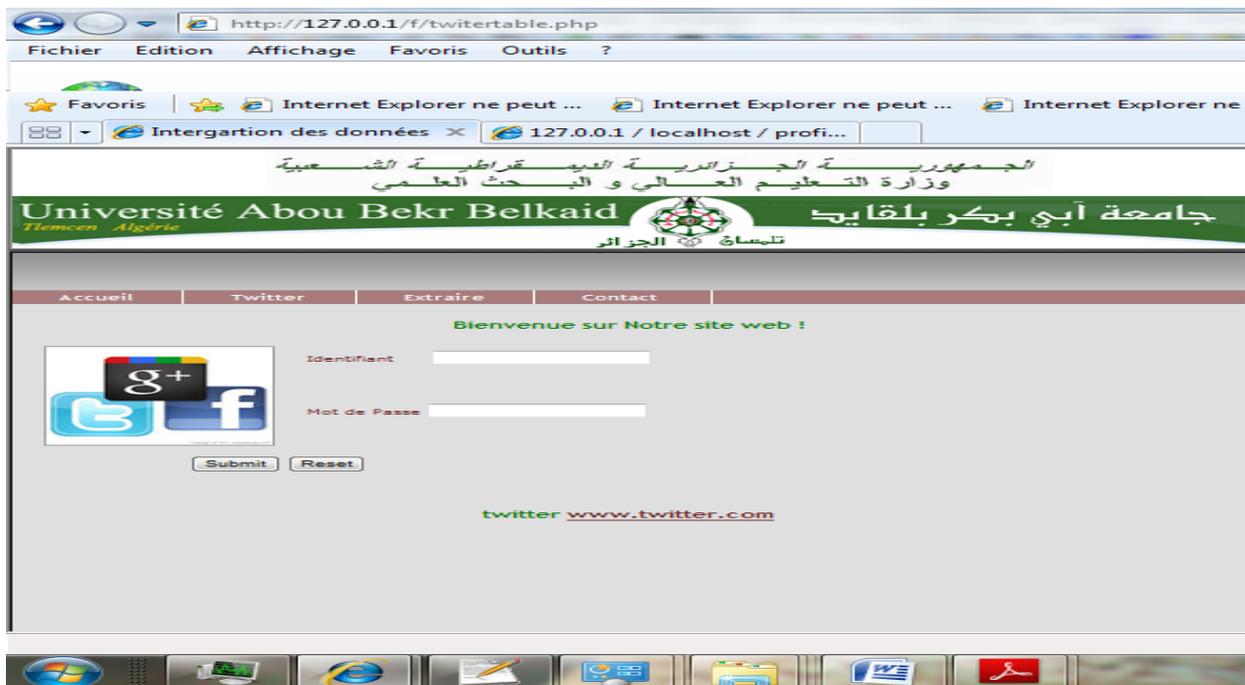
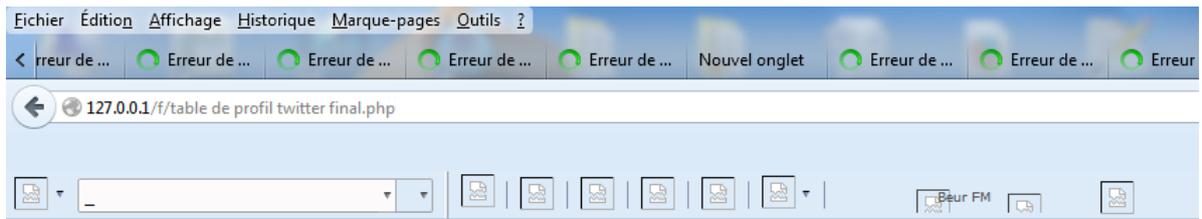


Figure IV.6 : L'interface d'extraction de l'identité Twitter

#### Exemple :

Voila un exemple de table de profil Twitter



### table de profil twiter

id_p	Nom d utilisateur	Adresse email	Langue	Fuseau horaire	Pays	Nom	Localisation	Site Web	Biographie
2	IslamTfna	islam_tfna@hotmail.fr	français	(utc+01:00)london	monde	larabi		http://	

Figure IV.7 : Exemple de table de profil Twitter

#### IV.4.6. Extraction de l'identité Google+

Notre interface de google+ présente comme suit :

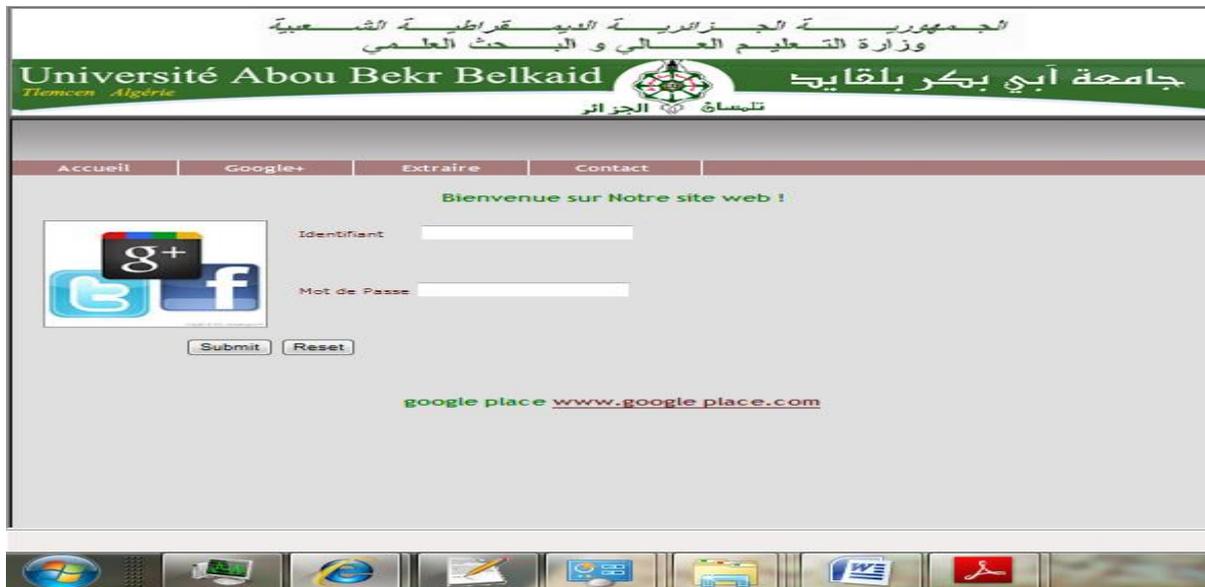
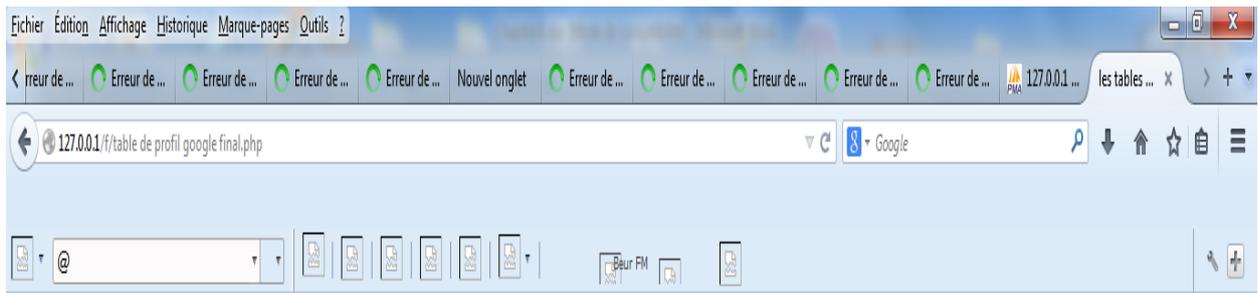


Figure IV.8 : L'interface d'extraction de l'identité Google+

#### Exemple

Voila un exemple de table de profil Google +



## table de profil google place

id_p	Adresse e-mail principale	E-mail de récupération de mot de passe	nom	Téléphone	Formation	Sexe	Anniversaire	Situation amoureuse	Autres noms	Autres profils	Liens	Domicile	Professionnel	Adresses	Profession	Compétences	Emplois	langue	
1	larabinorelislam@gmail.com	islam_tfna@hotmail.fr	islam asalam	669398494	etudiant	Homme	30 avril	celibataire	larabi					alger-tlemcen					ar

Figure IV.9 : Exemple de table de profil Google +

### IV.4.7. Intégration :

Notre interface d'intégration présente comme suit :

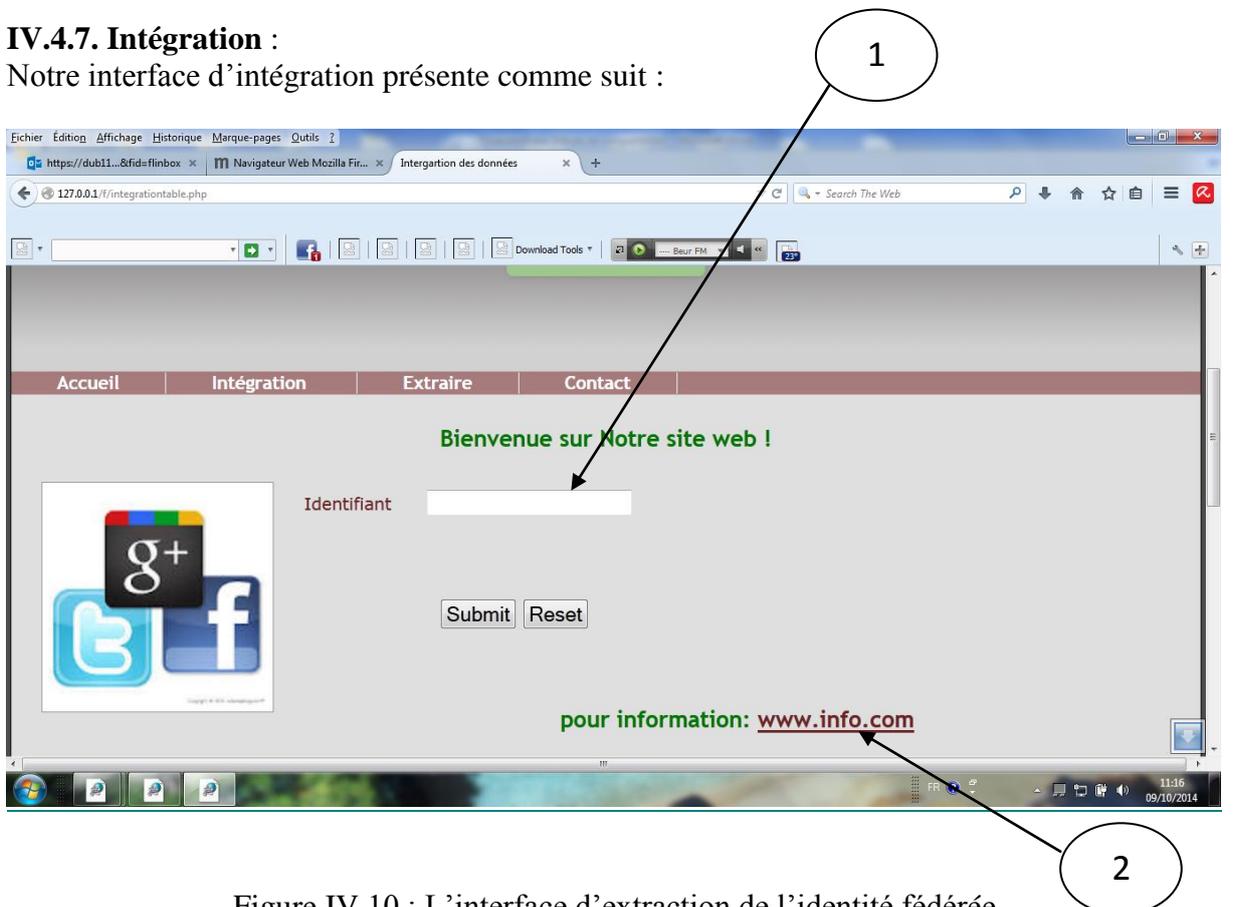


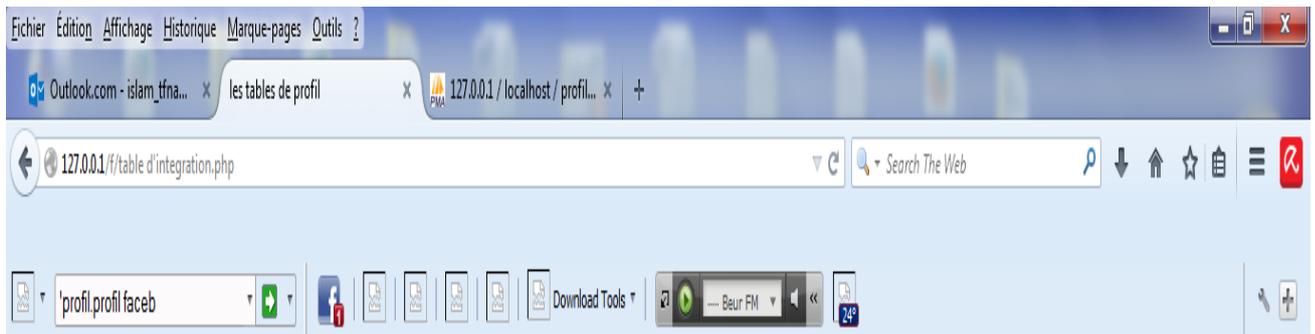
Figure IV.10 : L'interface d'extraction de l'identité fédérée

Explication :

- 1 : permet de remplir l'identifiant
- 2 : permet d'aller à une autre page d'information

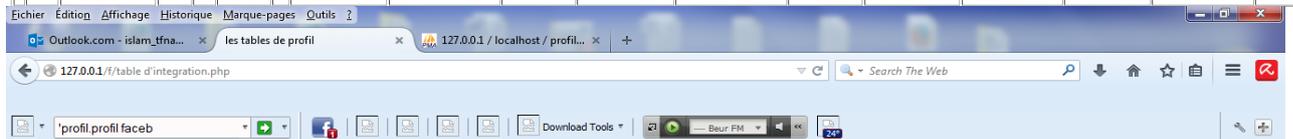
### Exemple schéma fédérée

Voila un exemple de schéma fédéré

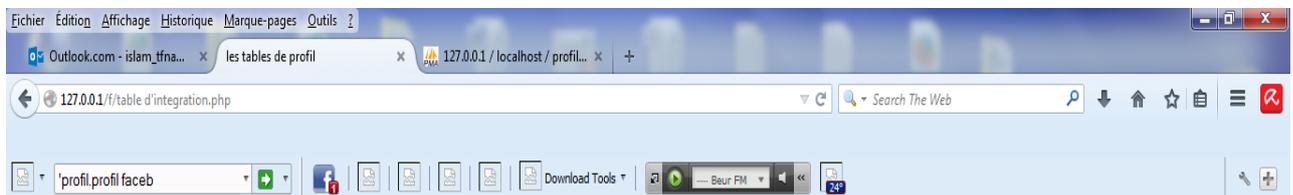


### table fédérer

id	Nom_f	Page Facebook	Email	Sexe	id_sexe	bio	education_school_id1	education_school_n1name	education_type1	education_school_id2	education_school_name2	education_type2	favorite_teams_id1	favorite_teams_name1	favorite_teams_id2	favorite_teams_name2	first_name
2	Isla Arabi	islam_tfna@hotmail.fr	Isla Arabi	male	2147483647		114649011879435	mofbi zakaria	High School	159873170744210	ABOU BAKR BELKAID	College	islam_tfna@hotmail.fr	683033455040975	Real Madrid Arabic	111091012309924	Nokat Marocain



hometown_id	hometown_name	last_name	link	location_id	location_name	name	timezone	updated_time	username	verified	work_employer_id	work_employer_name	Nom d'utilisateur	Adresse email	Langue	Fuseau horaire	Pays	Nom_t	Localis
Isla	105583976140814	Maghnia	Arabi	https://www.facebook.com/salam.apax	109589642400046	Tlemcen	Isla Arabi	1	2014-06-16T23:16:01+0000	islam.apax	1	2147483647	Student	IslamTfna	islam_tfna@hotmail.fr	français	(utc+01:00)london	monde	larabi



Adresse email	Langue	Fuseau horaire	Pays	Nom_t	Localisation	Site Web	Biographie	Adresse e-mail principale	E-mail de récupération de mot de passe	nom_g	Téléphone	Formation	Anniversaire	Situation amoureuse	Autres noms	Autres profils	Liens	Professionne
IslamTfna	islam_tfna@hotmail.fr	français	(utc+01:00)london	monde	larabi	http://		larabinoreislam@gmail.com	islam_tfna@hotmail.fr	islam asalam	669398494	etudiant	30 avril	cilbataire	larabi			

Figure IV.11 : Exemple d'identité fédérée

#### IV.4.8. Représentation de l'identité fédérée en Xml

```
<profil>
<tablefederée>
< colonne >
<colonne1>id </colonne 1>
<colonne2>Nom_f</colonne 2>
<colonne3>Page Facebook.</colonne 3>
<colonne4>Email1</colonne 4>
<colonne5>Sexe</colonne 5>
<colonne6>id_sexe</colonne 6>
<colonne7>bio</colonne 7>
<colonne8>education_school_id1</colonne 8>
<colonne9>.education_school_n1ame</colonne 9>
<colonne10>education_type1</colonne 10>
<colonne11>education_school_id2</colonne 11>
<colonne12>education_school_name2</colonne 12>
<colonne13>education_type2</colonne 13>
<colonne14>favorite_teams_id1</colonne 14>
<colonne15>favorite_teams_name1</colonne 15>
<colonne16>favorite_teams_id2</colonne 16>
<colonne17>favorite_teams_name2</colonne 17>
<colonne18>first_name</colonne 18>
<colonne19>hometown_id</colonne 19>
<colonne20>hometown_name</colonne 20>
<colonne21>last_name.</colonne 21>
<colonne22>link</colonne 22>
<colonne23>location_id</colonne 23>
<colonne24>location_name</colonne 24>
<colonne25>name</colonne 25>
<colonne26>timezone</colonne 26>
<colonne27>updated_time.</colonne 27>
<colonne28>username</colonne 28>
<colonne29>verified</colonne 29>
<colonne30>work_employer_id</colonne 30>
<colonne31>work_employer_name</colonne 31>
<colonne32>Nom d utilisateur </colonne 32>
<colonne33>Adresse email.</colonne 33>
<colonne34>Langue</colonne 34>
<colonne35>Fuseau horaire</colonne 35>
<colonne36>Pays</colonne 36>
<colonne37>Nom_t</colonne 37>
<colonne38>Localisation</colonne 38>
<colonne39>Site Web.</colonne 39>
<colonne40>Biographie</colonne 40>
<colonne41>Adresse e-mail principale</colonne 41>
<colonne42>E-mail de récupération de mot de passe.</colonne 42>
<colonne43>nom_g</colonne 43>
<colonne44>Téléphone</colonne 44>
```

```

<colonne45>Formation</colonne 45>
<colonne46>Anniversaire</colonne 46>
<colonne47>Situation amoureuse</colonne 47>
<colonne48>Autres noms.</colonne 48>
<colonne49>Autres profils.</colonne 49>
<colonne50>Liens</colonne 50>
<colonne51>Professionnel.</colonne 51>
<colonne52>Adresses</colonne 52>
<colonne53>Profession</colonne 53>
<colonne54>Compétences</colonne 54>
<colonne55>Emplois</colonne 55>
</ligne>
< /tablefederée>
</table>
</profil>

```

### **Conclusion**

Dans ce chapitre c'est la partie application, Nous avons utilisé 3 réseaux sociaux : Facebook, Twitter et Google+ pour la réalisation d'une identité social fédérée. Une approche d'intégration de données a été appliquée par l'utilisation d'une base de données lexicale Wordnet. Cette démarche peut être généralisée pour d'autre réseau social pour servir d'autres applications web pour une interaction intelligente avec les internautes. Dans ce projet on a eu les difficultés suivantes :

- Problème de permission et Interopérabilité : dans quelques plateformes de réseaux social ils ne donnent pas la permission à l'application web d'accéder aux informations des utilisateurs.
- validité des données : les données qui existe dans les réseaux social ne sont pas toujours crédible ou par fois manquantes qui influes sur la construction de l'identité et par conséquence la reconnaissance de l'utilisateur.

## Conclusion générale

Le but de ce projet c'est l'utilisation de l'intégration de données pour la construction d'un schéma fédérée pour une meilleurs réutilisation des informations sociaux dans d'autres applications et qui peut être généralisée pour d'autres réseaux sociaux.

Dans ce projet on a eu les difficultés suivantes :

-Problème de permission et Interopérabilité : dans quelques plateformes de réseaux social ils ne donnent pas la permission à l'application web d'accéder les informations des utilisateurs.

-validité des données : les données qui existe dans les réseaux social ne sont pas toujours crédible ou par fois manquantes qui influes sur la construction de l'identité et par conséquence la reconnaissance de l'utilisateur.

En perspectives :

Légaliser l'application pour avoir plus de permission

Intégrer cette application dans d'autre application.

Généraliser l'application par l'inclusion de d'autre réseaux.

### **Résumé :**

La Construction d'une identité sociale fédérée n'est pas une tâche facile si les différents intervenants (producteurs ou consommateurs d'information) ne s'entendent pas sur la sémantique des données. Il est donc très difficile de retrouver la sémantique d'un concept dans toutes les bases de données ou il peut être représenté

Dans ce sens, nous avons conçu un système d'intégration des données distribuée dans les réseaux sociaux (Facebook, Twitter et Google+). Le résultat d'intégration est un schéma fédéré représenté en XML pour qu'il puisse être réutilisé dans d'autre application web. Le système d'intégration de données utilise le réseau sémantique Wordnet pour la distinction de la redondance de l'information et le regroupement des attributs similaire.

### **Mots clé :**

Intégration de données, réseaux sociaux, XML, Schéma fédérée, WORDNET.

### **Abstract:**

Social federated identity is not easy work if the different interveners (Producers or information consumers do not agree on) do not agree on the semantics of data. It is so very difficult to find the semantics of a concept in all of data bases in which it can be represented.

In this sense, we conceive a data integration system which is distributed in the social networks (Facebook, Twitter and Google+) the result of the integration is a social federated schema represented in XML, to be re-used in other web applications. The developed data integration system uses the semantic network Wordnet for information redundancy distinction and similar attributes grouping.

### **Keyword:**

Data integration, Social networks, XML, federate schema, Wordnet

### **ملخص:**

إن تشييد هوية اجتماعية متألّفة ليس عملاً سهلاً لو أن المتدخلين (المنتجين أو المستهلكين للمعلومات) لا يتفقون على علم المعطيات إذن ليس من الصعب العثور على علم مفهوم في كل أسس المعطيات التي من الممكن أن يكون فيه ممثلاً.

-في هذه المعنى، تصورنا نظاماً لإدماج المعلومات الموزعة في الشبكات الاجتماعية (فايسبوك – تويتر و قوقل +) ان نتيجة الإدماج رسم متألّف متمثّل في XML حتى يستطيع أن يكون معداً لاستعمال في تطبيقات Web أخرى، يستخدم نظام الإدماج للمعطيات الشبكة Wordnet لتميّز تكرار المعلومة ولمقارنة المعطيات المتشابهة.

### **الكلمات المفتاحية:**

إدماج المعطيات ، الشبكات الاجتماعية ، رم تالفي، Wordnet مجرد.

## Bibliographies

- [1] [http://wwwbd.lip6.fr/index.php?module=documents&JAS\\_DocumentManager\\_op=downloadFile&JAS\\_File\\_id=119](http://wwwbd.lip6.fr/index.php?module=documents&JAS_DocumentManager_op=downloadFile&JAS_File_id=119)
- [2] [http://memsic.ccsd.cnrs.fr/docs/00/33/48/53/PDF/mem\\_00000130.pdf](http://memsic.ccsd.cnrs.fr/docs/00/33/48/53/PDF/mem_00000130.pdf)
- [3] [http://fr.wikipedia.org/wiki/Architecture\\_fc9C3A9rC3A9](http://fr.wikipedia.org/wiki/Architecture_fc9C3A9rC3A9)
- [4] <http://deptinfo.cnam.fr/Enseignement/CycleSpecialisation/ISA/ISCS/Chevance/Basesde donnees distribuees et federees.pdf>
- [5] <http://people.irisa.fr/Francois.Goasdoue/bib/Goasdoue-DEA-98.pdf>
- [6] C. Batini and M. Lenzerini, « A Methodologie for Data Schema Integration in the Entity Relationship Model », IEEE Transactions on Software Engineering, Volume SE-10, N°6, pages 650-664, Nov. 84.
- [7] C. Batini, M. Lenzerini and S.B. Navathe, « A Comparative Analysis of Methodologies for Database Schema Integration », ACM Computing Surveys, Volume 18, N°4, pages 323-364, Déc. 86.
- [8] U. Dayal and H.-Y. Hwang, « View Definition and Generalization for Database Integration in a Multidatabase System », IEEE Transactions on Software Engineering, Volume SE-10, N°6, pages 628-645, Nov. 84.
- [9] R. Hull, «Managing Semantic Heterogeneity in Databases : A Theoretical Perspective », 1997, <http://www-db.research.bell-labs.com/users/hull/pods97-tutorial.html>.
- [10] A. P. Sheth and J. A. Larson, « Federated Database Systems for managing Distributed, Heterogeneous, and Autonomous Databases », ACM Computing Surveys, Vol. 22, N°3, pages 183-236, Sept. 90.
- [11] V. Kashyap and Amit Sheth, « Semantics-based Information Brokering », Proceedings of the Third International and Knowledge Management (CIKM), Novembre 1994, <http://lsdis.cs.uga.edu/~amit/62-cikm-paper.ps>.
- [12] W. J. Labio, Y. Zhuge, J. L. Wiener, H. Gupta, H. Garcia-Molina, J. Widom, « The WHIPS Prototype for Data Warehouse Creation and Maintenance », Proceedings of the ACM Sigmod International Conference on Management of Data, pages 557-559, Tucson, Arizona, Mai 97.
- [13] J. Widom, « Research Problems in Data Warehousing », Proceedings of Fourth International and Knowledge management (CIKM'95), Baltimore, Maryland, Novembre 1995, <ftp://db.stanford.edu/pub/papers/warehouse-research.ps>.
- [14] M. Bonjour, G. Falquet, M. Leonard. *Integration de concepts et integration de bases de données*, Actes du <sup>x</sup>ème congrès INFORSID, Aix en provence 1994.
- [15] C. Parent, S. Spaccapietra. *Integration de bases de données. Panorama des problèmes et des approches*, Ecole polytechnique fédérale de Lausanne, ingénierie des systèmes d'informations Vol.4, N°3, 1996.
- [16]. <http://colibre.org/les-reseaux-sociaux-distribues/>
- [17]. <http://fr.wikipedia.org/wiki/Twitter>
- [18]. <http://colibre.org/les-reseaux-sociaux-libres-une-alternative/>
- [19]. <http://blog.polynet-online.fr/influence-reseaux-sociaux-ecommerce.html>
- [20]. <http://fr.wikipedia.org/wiki/OAuth>
- [21] R. Bourett "XML et les bases de données", <http://www.rpbouret.com>, 2003.

- [22] K. Staken, ‘‘Introduction to native XML database’’ <http://www.xml.com>, 2001.
- [23] M. Cyrenne, ‘‘When should you use a Native XML Database?’’, <http://www.edocmagazine.com/articlenew>, November/December 2002.
- [24] B. Verhaegen, ‘‘Requêtes OLAP sur une base de données XML native Académie Universitaire WALLONIE BRUXELLES, 2005
- [25] Heimbigner, D., McLeod, D.: A federated architecture for information management. *ACM transactions on Office Information Systems (TOIS)*, Vol. 3, No. 3, p.253-278, Juillet 1985
- [26] Sheth, A.P., Larson, J.: Federated database systems for managing heterogeneous, distributed and autonomous Databases. *ACM Computing Surveys*, Vol. 22, No. 3, 1990
- [27] Litwin, W., Abdelatif, A.: Multidatabase interoperability. *IEEE Computer*, Vol. 19, No. 12, 1986
- [28] Parent, P., Spaccapietra, S.: Intégration de bases de données: Panorama des problèmes et des approches. *Ingénierie des Systèmes d'Information*, Vol.4, No.3, 1996.
- [29] Manolescu, D. Florescu, D. Kossman : Answering XML queries over heterogeneous data sources, *BDA 2001*
- [30] G. Wiederhold: Mediation in information systems. *ACM Computing Surveys*, 27(2), 1995.

### Webographie

- (URL1): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL2): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL3): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL4): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL5): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL6): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL7): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL8): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL9): <http://www-poleia.lip6.fr/~doucet/CoursBDIA/Cours4.pdf>
- (URL10): [http://fr.wikipedia.org/wiki/Architecture\\_fédérée](http://fr.wikipedia.org/wiki/Architecture_fédérée)
- (URL11): [http://fr.wikipedia.org/wiki/Architecture\\_fédérée](http://fr.wikipedia.org/wiki/Architecture_fédérée)
- (URL12) : <http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb-20101206>
- (URL13): [http://fr.wikipedia.org/wiki/Architecture\\_fédérée](http://fr.wikipedia.org/wiki/Architecture_fédérée)