République Algérienne Démocratique et Populaires

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Djilali BOUNAAMA Khemis Miliana

Faculté des Sciences et de la Technologie

Département de Mathématiques et d'Informatique

Mémoire Présenté

Pour l'obtention de diplôme de

Master 2 en Informatique

Option : **« Ingénierie Du Logiciel »**

Titre :

Intégration De Données : Approche Semi-Automatique Pour La Mise En Correspondance De Schémas De Bases De Données Hétérogènes.

Réalisé par :

- Abdelkadir Regba

-Mahfoudh Ghalem

Devant le jury composé de:

president:F.Boudali

Encadreur : HanicheFaycel

Examinateur 1:D**.** Bahloul

Examinateur 2: H.Hachichi

Année Universitaire 2017/2018

Remerciements

Avant tout, je remercie Dieu le Tout Puissant pour m'avoir donné la force et le courage d'accomplir ce travail avec abnégation.

Mes vifs et sincères remerciements

S'adressent spécialement à,

M.faycel Haniche

Dont j'ai eu la chance de l'avoir comme Professeur, Encadreur et qui a bien voulu me confier ce travail riche d'expériences et me guider dans chaque étape de sa consécration. Vous m'avez toujours réservé un chaleureux accueil, malgré vos obligations et les contraintes professionnelles. Vos talents ainsi que vos compétences et votre sens du devoir m'ont marqué à jamais. Vos encouragements inlassables, votre amabilité, votre gentillesse et votre patience méritent toute notre attention. Veuillez trouver ici l'expression de notre estime et notre considération

Je remercie également,

Les membres du jury de nous avoir fait l'honneur de juger cette thèse. Veuillez accepter l'expression de ma vive gratitude

Enfin,

A toutes les personnes qui ont contribué de prés ou de loin, d'une manière directe ou indirecte à l'élaboration de ce travail de fin d'études.

Dédicaces

Je dédie ce mémoire

A mes chers parents ma mère et mon père

Pour leur patience, leur amour, leur soutien et leur

Encouragements

A mes frères.

A mes amies et mes camarades.

Sans oublier toute les professeurs que ce soit du

Primaire, du moyen, du secondaire ou deL'enseignement

supérieur

Abdelkadir

Dédicaces

Je dédie ce mémoire à :

Mes chers parents, qui m'ont soutenu,

encouragé pour que je puisse mener à bien mes

études, et qui attendu ce jour avec impatience.

A mes frères.

A mes amies de l'étudeet mes camarades.

A tous ceux qui ont contribué à la réalisation dece travai

Mahfoudh

Sommaire

listes des figure

Liste des tableaux

Résumé

l'intégration de différentes sources de données hétérogène a été posé depuis l'arrive de base de donnée , en effet, les nouvelles applicationsdéveloppéesde traitement de données font le plus souvent appel à des données déjà mémorisées dans l'ordinateur, que ce soit dans des fichiers, ou dans une bases de données indépendante , notre travail consiste à proposer une approche semi-automatique pour la Mise en correspondance de schémas de Bases de données d'un même domainenotre réalisation est proposé dans un contexte de système multi agents additionnel incluant un agent intégrateur de données qui doit posséder une capacité lui permettant de comprendre et couplé les attributs de la BD du Système Existant avec les attribues de sa propre BD, ainsi qu'une autre capacité lui permettant de faire le transfert de données nécessaire selon le besoin des autre agents et cela après une génération automatique de plan d'action et de requêtes nécessaires vers les deux BDs, l'approche proposée se base uniquement sur les information de schémas et combine des techniques linguistiques et structurelles, et propose l'exploitation des graphe de dépendances fonctionnelles et la notion de dépendances indirecte pour une détection plus efficaces de correspondances.

Mots clés : Base de données hétérogènes, processus d'intégration de données, schémas matching. Dépendance fonctionnelle, Agent-Intégrateur.

.

ملـــخـص

لقد ظهر مفهوم دمج قواعد البيانات الغير متجانسة منذ فترة مع تطور مفهوم قواعد البيانات و إنتشار إستعمالها في تطبيقات الإعلام الألي غالبا تعتمد التطبيقات الألية المطورة على إستخدام البيانات المخزنة في الحواسيب مثل الملفات او قواعد البيانات الشيئ الذي يتطلب فهم مخططاتها و تكييفها حسب هذه التطبيقات .

يتلخص مضمون هه المذكرة في إقتراح نهج نصف ألي يقوم بدمج مخططات قواعد البيانات الغير متجانسة من نفس مجال الدراسة

تم إقتراح هدا العمل في إطار أنضمة متعددة العملاء لعميل خاص بدمج المعطيات الدي يتطلب كفائة تسمح له بفهم و دمج خصائص قواعد البيانات للنضام القديم (المراد تحسينه) و مع خصائص قاعدة المعطيات الخاصة بنضام العملاء المظاف هدا العمل يتطلب أيضا كفائة تسمح له بتحويل البيانات من قاعدة إلى أخرى و هدا طبعا بعد التوليد الالي للمخططات العلمية و الطلبيات الضرورية بالتجاه القاعدتين

يعتمد النهج المقترح فقط على المعلومات الخاصة بالمخطط و يجمع بين التقنيات اللغوية و البنيوية و يقترح إستغلال شبكات التعلقات الوضيفية و مفهوم التعلق الغير مباشر و هدا من اجل الكشف الفعال للتزاوجات الممكنة .

الكلمات المفتاحية : قواعد البيانات,تطبيقات ,المخططات العلمية ,شبكات التعلقات

Abstract

The integration of different heterogeneous sources of data has been posed since the arrival of database concept, in fact, the new developed data processing applications often use data already stored in computers, whether in files, or in independent databases, our interest in this memory is to propose a semi-automatic approach for database schemas matching. The realization is proposed in a context of additional multi-agent system including a Data-Integrator-Agent that must have the capacity to understand and map the attributes of the Existing System DB with the attributes of its own DB, as well as another capability to transfer necessary data between the two DataBases, according to MAS requirements, after an automatic generation of action plan and necessary queries for both DBs. The proposed approach is based solely on schema information and combines linguistic and structural techniques, and proposes the exploitation of functional dependency graphs and the notion of indirect dependencies for more efficient correspondences detection.

Keywords: database**,** database integration, dependency, schemas Matching.

.

Introduction générale

Introduction générale

1.1 Contexte

Pour bien comprendre l'intégration de données et le développement qui a touché tous les domaine d'application ,car il n'existe aucune application informatique a été développé Sanbase de données ,c'est une évolutions technologiques qui permet de se développer et d'avoir des outils d'aide à la décision(la recherche ,l'insertion ,l'échange d'informations),pour être prendre une monde informatique décisionnelle, depuis le début de l'utilisation de la base de donnée l'amélioration de ce domaine est à chaque fois développé mais elle contient toujours des problème l'un des c'est problèmes si l'intégration des données avec des sources de données hétérogènes c'est un problème existe depuis l'arrive et l'adoption des bases de données.

L'intégration de données est un processus qui permet de combiner deux bases de donnée provenant de différentes applications d'un même domaine, ainsi que de combiner des sources de données hétérogènes et de les interroger via une seule interface de requête. L'intégration de données ne date pas d'aujourd'hui, elle est depuis l'apparition des bases de données.

Cette combinaison peut être effectuée de différentes façons et à différents niveaux de l'architecture du système car il existe dans le domaine d'intégration de donnée plusieurs méthode pour résoudre ses problèmes.
Le matching des schémas de données est une partie primordiale dans le processus d'intégration. Il s'agit d'une technique qui effectue la découverte de correspondances sémantiques entre les éléments et les attributs des schémas. Elle a été proposée dans plusieurs travaux de recherche tels que les travaux deRahm Erhard et Bernstein Philip[1] pour résoudre ce problème d'hétérogénéité. Ces travaux ont souvent abordé le matching dans un contexte où peu de schémas de données sont traités. Avec la prolifération des schémas de données sur la toile ou même au niveau des entreprises.

1.2 Organisation du mémoire

Partie une : chapitre 1

Nous parlerons dans ce chapitre sur le terme intégration de donnée, les domaines d'application et la problématique dans ce domaine.

Partie deux : chapitre 2

Nous aborderons dans cette partie sur les schémas matching d'unebase de données (data base schéma matching), la classification des approches existent et quelque exemple.

Partie trois : chapitre 3

Ce chapitre contient des descriptions de contexte de notre travail comme le processus du matching le langage et des logiciels utilisent dans la construction

Del'intégration des données avec des sources de données hétérogènes.

Enfin conclure d'une manière générale.

Chapitre I :
Intégration des données

Chapitre I : Intégration des données

1. Introduction

Le besoin de partager et de faire connecter des sources de données hétérogènes augmente de plus en plus, en particulier avec la croissance du nombre de sources de données, de nombreux travaux de recherche dans le domaine des systèmes d'information se sont focalisés sur le problème d'intégrations des données hétérogènes.

Ce chapitre fournit une description générales d'intégration des données sur les objectif est les domaines d'application et un aperçu plus détaillé sur des architectures d'intégrations des données.

2. Définition

L'intégration des données est un processus de combinaison des données appartenant à différentes sources et fournissant à l'utilisateur une vue unique de ces données[2]

Les Données:

Une donnée est une information numérique souvent codée, lisible par une machine, C'est la description élémentaire d'une réalité peut être (information, chiffre, date)**.** [3]

Intégration :

En informatique, le terme intégration désigne également la conception et la réalisation d'un système d'information intégré par la mise en relation de différents logiciels ou matériels existants. [3]

3. L'objectif de l'intégration des données

Aujourd'hui, les systèmes d'information et d'analyse au sein des organisations sont devenus de plus en plus complexes car le métier des organisations s'évolue de manière continue et la proposition de nouvelles application et/ou de de faire évoluer les applications existante devient une nécessité. D'où l'importance de coordonner les données manipulés par la différente application, généralement hétérogènes, ce nouveau souci est le domaine d'intégration des applications de l'entreprise dont l'intégration de données est une des solutions adoptées**.**

L'intégration de données permet de [3] :

- Piloter l'ensemble des bases de données interne et externe d'une organisation.
- Organiser les processus par lesquels des données informatiques sont collectées, regroupées dans un format unique.
- Assurer un accès rapide, simple, efficace aux données traitées.
- Améliorer la consultation et l'analyse des données effectuées par les collaborateurs et les consultants.
- Pour traiter de grandes quantités de données dans une même exécution
- Améliorer la prise de décision des décideurs.

4. Domaine d'applications

Les systèmes et techniques d'intégration de données sont utilisés largement dans plusieurs domaines, nous citons par exemple les domaines suivants [2]:

- **l'intégration des Applications de l'entreprise**

Aujourd'hui, Les Système d'Information (SI) d'une entreprise sont de plus en plus complexes car ils sont basés sur des logiciels et des sources de données hétérogènes, cette hétérogénéité a amené les directions informatiques à considérer la problématique d'intégration des données au sein des entreprises, pour répondre à la question importante suivante : comment permettre aux applications de se communiquent efficacement et sans erreurs? Par exemple: l'application des ventes a besoin de données présentes dans l'ERP(Enterprise Ressource Planning), et la gestion des commandes a besoin de données présentes dans le CRM(customer relationship management)D'où l'importance de coordonner ces données pour pouvoir suivre et analyser l'activité de l'entreprise ou pour faire communiquer différents départements ou services entre eux pour optimiser leurs processus métiers. [4]

- **Business to business**

Le commerce électroniques ou e-business (électronique business) pourrait être défini comme l'utilisation de moyens électroniques (particulièrement des technologies de l'information) pour réaliser des transactions commerciales (business). Il correspond donc au commerce électronique sur le net. Les transactions commerciales existaient depuis le début de web, même si celui-ci a permis une meilleure visibilité de ce domaine par le grand public. En effet, les partenaires commerciaux gèrent de plus en plus leur commerce en ligne.

Pour permettre l'échange de messages entre partenaires, les applications développées nécessitent de transformer les messages d'un format à un autre ; d'où la nécessité de la misse en correspondance des éléments de données, pour fournir un support pour la transformation de messages. Une tâche importante dans cette transformation est celle d'établir les règles detransformation. Comme dans l'intégration des schémas, il existe des conflits structurels et de noms car les schémas utilisent souvent des noms et types de données différents, différents intervalles de valeurs et différents groupes de champs. Aujourd'hui, les applications développées déterminent manuellement comment les formats de message sont reliés en transformant des programmes ou scripts. Cette approche manuelle pose un problème pour les environnements automatisés tels que le marché électronique. Le matching semi-automatique de schémas devrait réduire la quantité de travail manuel

- **Web sémantique**

L'intégrationdes moteurs de recherche pour la recherche d'une information implique l'utilisation de plusieurs sources de données, Il se focalise sur l'exploitation de la quantité importante de données sémantiques disponibles sur le Web. Les applications du web sémantique nécessitent de manipuler une hétérogénéité de données disponibles en ligne.Ces données décrivent différents domaines selon différents point de vue,D'où la nécessité de l'intégration des donnéespour la facilité de l'utilisation des données disponible sur le web

5. Problématiques liées à l'intégration de données

Plusieurs problèmes doivent être pris en compte pendant la conception de systèmes d'intégration de données, ces problèmes résultent de l'hétérogénéité des données. par exemple on a des difficulté pour relier des tableaux de différentes bases de données à cause des "Conflits" comme les Conflits de représentation, les conflits de nom,et les Conflits de contexte[5]c'est vrai que les données des sources sont structurellement indépendantes mais sont toujours supposées relever de domaines similaires c-`a-dire que les données relatives à un même sujet sont représentées différemment sur des bases de données différentes.

L'hétérogénéité des sources de données à intégrer a conduit à de nombreux problèmes qui peuvent se classer en deux catégories :

1-la mise en correspondances des schémas

2- l'évaluation des requêtes

5.1 Hétérogénéité des données

Le problème de l'hétérogénéité de données a longtemps été étudié dans la communauté des bases de données. En général, le problème de l'hétérogénéité peut concerner deux catégories[6]:

• **Hétérogénéité structurelle ou schématique**:

Elle provient quand les sources adoptent différents modèles de données, structures de données ou schémas, par exemple les modèles de bases de données relationnelles ou orientées objets. De nombreux travaux concernant ce type d'hétérogénéité ont été proposés dans les contextes des bases de données fédérées et des multi-bases de données.

• **Hétérogénéité sémantique** :

Elle est due aux conflits sémantiques dans les termes, les expressions, etc., qui sont adoptés par différents schémas de données mais exprimés de diverses manières. Autrement dit, elle est dûe aux différentes interprétations pour les objets du monde réel. En effet, les sources de données ont été conçues indépendamment par des concepteurs différents ayant des objectifs applicatifs différents. Chacun peut donc avoir un point de vue différent sur le même concept. L'interopérabilité sémantique de données présente un défi majeur dans le processus d'élaboration des systèmes d'intégration.

Exemple 1 :

Pour illustrer ce problème d'hétérogénéité, nous étudions l'exemple dans lafigure 1.

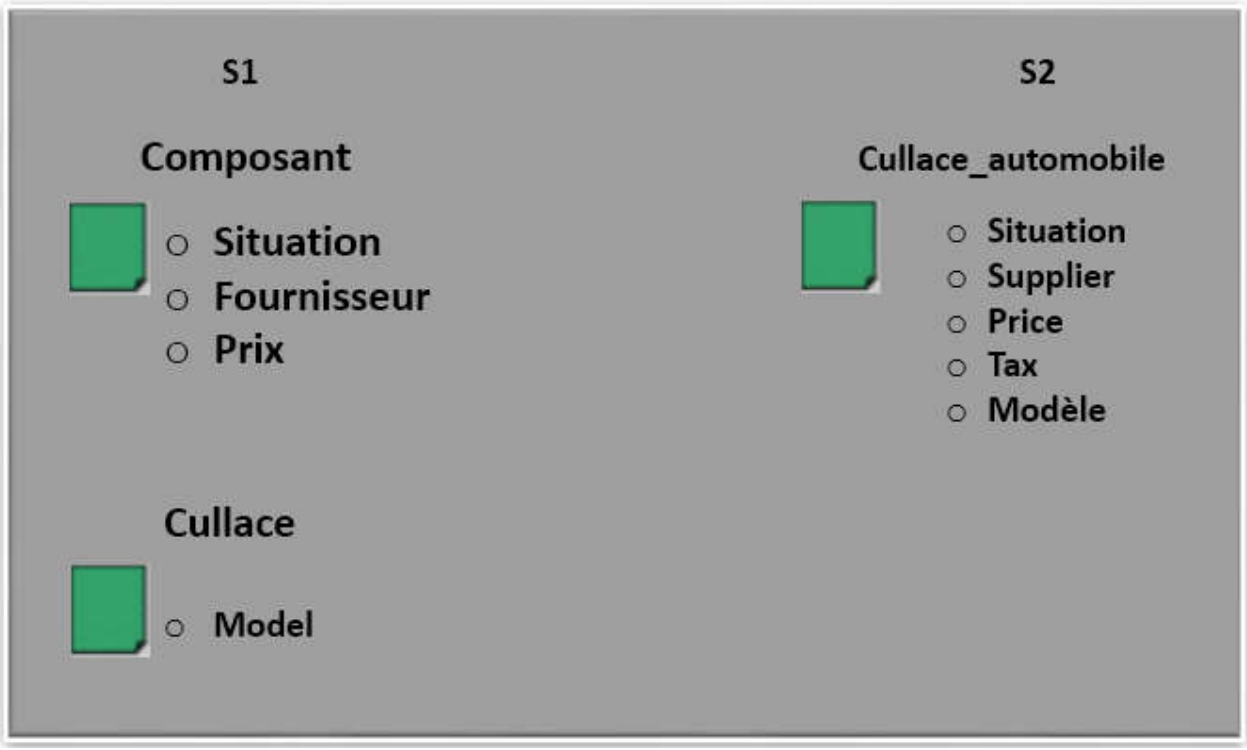


Figure 1 : Exemple de catalogues hétérogènes

Supposons qu'une entreprise de commerce pièce automobile vend des Culasse sur Internet. Les Culasse mis en vente sont fournis par des fournisseurs différents, où 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. Pour des raisons de simplicité, nous considérons deux catalogues S1 et S2 de deux fournisseurs.

Remarquons que les deux catalogues décrivent un Culasse différemment. Cette différence concerne le nombre de concepts utilisés pour définir chaque source, ainsi que l'aspect sémantique de chaque concept. LeTableau 1 décrit les différentes propriétés de S1 et S2.

	S1	S2
Classes	Composants : composant pièce Cullace : cullace	Cullace_automobile : Cullace
Propriétés	Situation (domaine : boolean) : neuf ou occasion	Situation (domaine : boolean) : disponible ou non
	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 cullace
		Tax (domaine : number) : la TVA d'un cullace
	Model (domaine : varchar) : model de type 1	Model (domaine : varchar) : modèle de type 2

Tableau 1: Sémantique de données de deux catalogues

Le problème qui se pose dans ce cas lorsque nous voulons faire interroger des bases de données hétérogènes est, d'une part l'identification de conflits entre les concepts dans les sources différentes qui ont des liens sémantiques, d'autre part, la résolution des différences entre les concepts sémantiquement liés.

Cette taxonomie des conflits sémantiques a été décomposée en trois catégories

(1) conflits de représentation.

(2) conflits de nom.

(3) conflits de contexte.

5.1.1 Conflits de représentation

Ces conflits se trouvent dans le cas où nous utilisons des propriétés différentes ou des schémas différents pour décrire le même concept. Cela signifie que les nombres de classes et propriétés représentant un concept *C* dans les sources ne sont pas égaux.

Exemple 2 :

Reprenons la Figure 1, où le fournisseur S1 utilise deux classes : Composants etCulasse et 4propriétés : situation, fournisseur, prix, et modèle pour décrire un Culasse. Tandis que le fournisseur S2 utilise une seule classe : Culasse _automobile et **5** propriétés : situation, supplier, price, tax, et modèle.

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 Culasse, tandis que S1 n'en utilise qu'une seule, à savoir prix.

5.1.2 Conflits de nom (termes)

Ces conflits se trouvent dans le cas où nous utilisons 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 Culasseest nommé parCulasse dans la S1, et par Culasse _autompobile dans la S2. La propriété Situation se trouve dans les deux sources, mais avec deux significations différentes (voir Tableau 1).

4.1.3 Conflits de contexte

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 Culasse ne s'applique que pour unCulasse neuf dans S1, mais peut-être pour tous les Culasse soit neufs ou d'occasions dans S2.

Par exemple, l'unité de mesure de la model dans la *S*1 est detype 1 tandis que celle dans la *S*2 est de type 2.

5.2 Evaluation de requêtes

Le traitement de requêtes est un mécanisme absolument obligatoire et important dans l'intégration des bases de données. Elle consiste à la réécriture d'une requêtebasée sur un schéma vers une requête adapté à un autre schéma pour être enfin exécuté selon le schémade la deuxième source de donnée. Dans l'architecture de l'entrepôt de donnée cela commence par une reformulation de celle-ci à partir du schéma global aux schémas à l'exportation des sources de données. Les algorithmes pour la reformulation des requêtes de chargement, dépendent de la manière dont la relation entre le schéma global et les schémas locaux a été définie (comme GAV ou LAV). Dans l'approche GAV, les requêtes sont reformulées simplement par le remplacement des éléments changeant. Dans l'approche LAV, la reformulation des requêtes est beaucoup plus compliquée, et connue comme étant la réponse aux requêtes formulées par l'utilisateur de vues, Il s'agit d'un problème NP-complet, et il ne peut être résolu facilement. [7]

6. Architectures d'intégration des données

Au cours des dernières années, trois architectures complémentaires de systèmes d'intégration d'informations se sont imposées : les entrepôts de données, les médiateurs et les systèmes pair-à-pair [8]**.**

6.1 Les entrepôts de données

Les entrepôts de données stockent localement une partie des données des sources distantes. L'avantage est d'avoir un temps de réponse faible puisque les requêtes posées sont évaluées localement. Toutefois, l'inconvénient majeur est qu'il faut trouver une bonne fréquence de rafraichissement des données de l'entrepôt afin qu'elles ne se périment pas. La Figure 2 représenté l'architecture de l'entrepôt

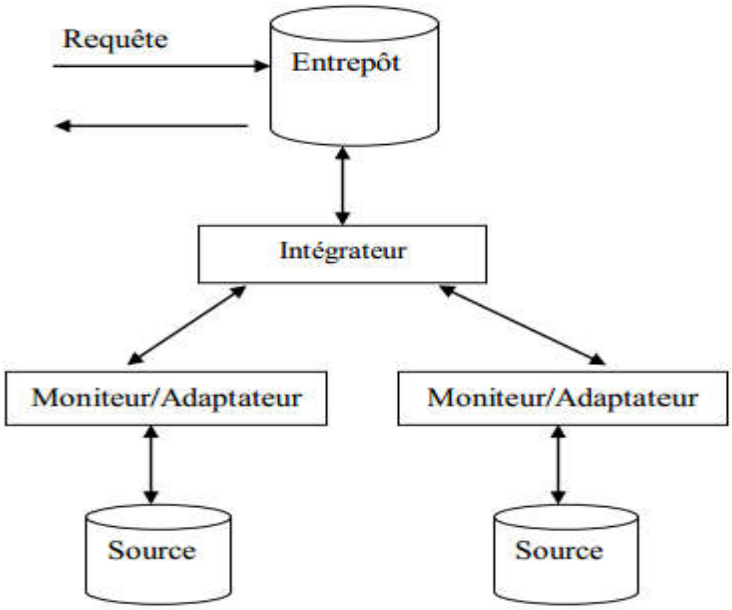


Figure 2: Architecture d'un entrepôt de données[9]

Cette architecture présentée par INMON et aL [9] comme une collection de données et d'informations intégrées, La démarche de cette architecture consiste à voir cette intégration comme la construction de bases de données réelles, etLes composants de cette architecture sont:

- **Les sources**

Les données de l'entrepôt sont extraites de diverses sources souvent réparties et hétérogènes, et doivent être transformées avant leur stockage dans l'entrepôt. Nous avons deux types de sources des données : internes et externes à l'organisation.

- **Le moniteur de source**

C'est un composant capable de détecter les mises à jours effectuées sur la source et de repérer les données à envoyer à l'entrepôt de données.

- **L'adaptateur de source**

Afin de préparer l'intégration des données dans l'entrepôt de données, il est nécessaire de convertir les données à envoyer dans un format plus ou moins commun. Pour ce faire, le moniteur doit être complété par un adaptateur capable de transformer les mises à jour, les questions et les réponses associées dans le modèle de l'entrepôt.

● **L'intégrateur**

Avant d'être déversées dans l'entrepôt, les données en provenance de sources multiples doivent être intégrées. Cette fonction est dévolue à l'intégrateur.

● **L'entrepôt de données**

Le support de destination des données est l'entrepôt de données. Il s'agit d'une base de données dont la structure dimensionnelle permet de faciliter le stockage et la disposition des informations afin de les analyser.

6.2 Les médiateurs

Contrairement aux entrepôts**,** les médiateurs ne stockent pas de données localement. Ceci les rend moins efficaces que les entrepôts lors de l'évaluation de requêtes, mais permet de garantir que les réponses sont calculées sur des données non périmées : les données des sources ,la figure 3 suivante représentel'architecture de médiation.

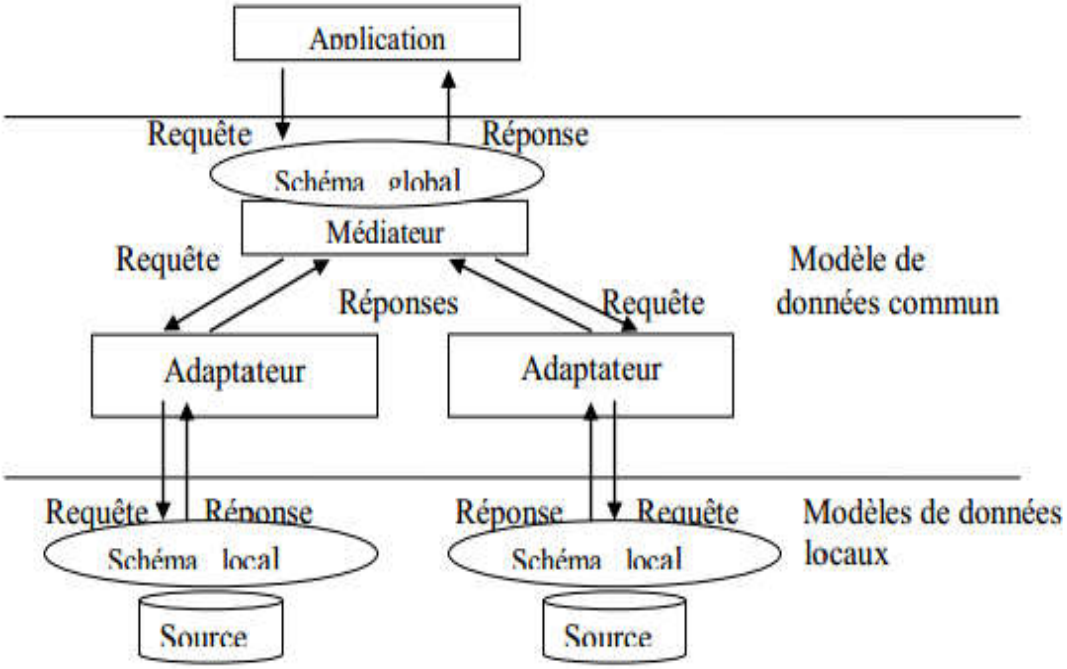


Figure 2: Architecture de médiation[10].

Cette architecture de médiation fut en 1992 par Gio Wiederhold [10]. Elle présente une nouvelle vision de l'architecture du traitement de l'information en entreprise, il tente de régler la

problématique de l'accès et de l'intégration de l'information en introduisant la notion de médiateur, Les composants essentiels d'un système de médiation sont :

Adaptateur (Wrapper):

Un adaptateur prend en compte les dimensions Distribution, Autonomie et Interopérabilité il permettre de Réalise la transformation entre le modèle de données dans lequel sont représentées les données de la source et le modèle choisi au niveau médiateur, et permet ainsi effectuer la transformation entre les expressions de requêtes.

Médiateur (Médiator)**:**

Le rôle d'un médiateur est de collecter, nettoyer et combiner les données attendues par le système de médiation et produites par les différents adaptateurs

7. Système d'intégration

LaFigure 4représenté par Xuan en 2006[11] pour comprendre comment le system d'intégrations des données est fonctionnée :

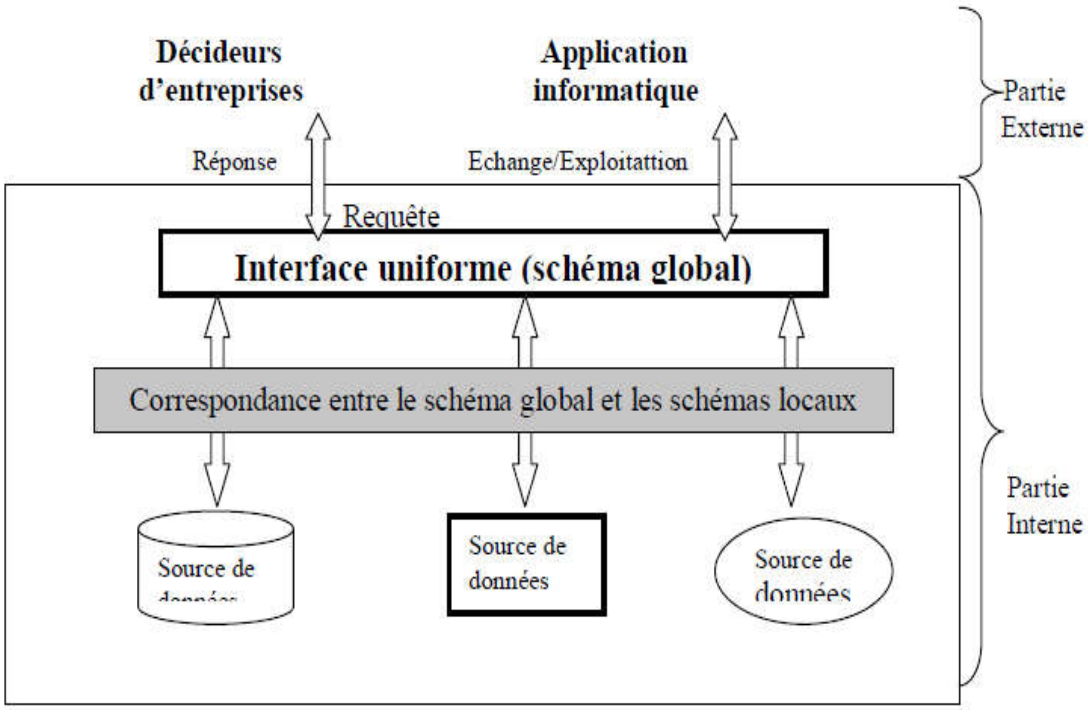


Figure 3 : Système d'intégration[11]

Dans les systèmes de médiation et les entrepôts de données, l'intégration de données consiste à éliminer d'abord les conflits entre les données causées par leur hétérogénéité et ensuite

les représenter dans un seul schéma cohérent. Tout système d'intégration doit fournir les solutions aux problèmes suivants :

(1) conflits de représentation.

(2) conflits de nom.

(3) conflits de contexte.

Généralement, un système d'intégration comporte deux parties principales une partie externe et une partie interne (voir la figure4).

• **La partie externe** correspond aux utilisateurs du système intégré comme, par exemple, les décideurs d'une entreprise, ou d'autres systèmes.
• **La partie interne** comprend d'une part les sources d'informations participant dans le processus d'intégration et d'autre part une interface permettant aux éléments de la partie externe d'accéder d'une manière automatique aux sources de données. Cette interface permettre de facilite l'accès à la base de donnée.

8. Nature du processus d'intégration

Les approches d'intégration de donnée sont classe selon le degré d'automaticité du processus d'intégration. Notons que cette automaticité concerne la résolution ou l'élimination des conflits sémantiques entre les sources durant le processus d'intégration. On distingue les approches manuelles, semi-automatiques et automatiques.

8.1 Les approches manuelles

Les premières approches proposées étaient des approches manuelles. Ces approches permettent d'automatiser l'intégration des données au niveau syntaxique. Les conflits sémantiques sont gérés manuellement et nécessitent la présence d'un expert humain pour interpréter la sémantique des données. Plusieurs systèmes ont été développés selon cette approche comme les systèmes multi-bases de données, Ces approches manuelles deviennent impraticables lorsque le nombre de sources de données à intégrer est important, ou lorsque les sources évoluent [12].

8.2 Approches semi-automatiques

Afin d'apporter plus d'automatisation dans la résolution des conflits sémantiques, plusieurs travaux se sont tournés vers les ontologies. Une ontologie permet de fournir la sémantique des concepts d'un domaine de manière formelle, Les approches semi-automatiquesreposent sur des ontologies linguistiques et permettent d'automatiser partiellement la gestion des conflits sémantiques. Les ontologies linguistiques traitent des termes, et non des concepts. Ceci peut créer des conflits de nom [13].

Dans notre travail nous nous intéressons à cette approche, l'approche semi-automatique car elle est fiable et qu'on doit y avoir des résultats d'intégration sans erreurs.

8.3 Approches automatiques

Les approches automatiques consistent à associer aux données des sources une ontologie conceptuelle qui en définit le sens. Une ontologie conceptuelle traite les concepts d'un domaine donné, la sémantique du domaine est ainsi spécifiée formellement à travers des concepts, leurs propriétés ainsi que les relations entre les concepts, la référence à une ontologie permet d'éliminer automatiquement les conflits sémantiques entre les sources [14].Il faut noter ici que l'approche purement automatique ne génère pas une intégration à 100% correcte.

9. Conclusion

Dans ce chapitre nous avons présenté une étude sur les systèmes d'intégration de données hétérogènes dans un contexte distribué pour bien comprendre le processus d'intégrations de données connaitre les différent problèmes et domaines d'applications .

Chapitre II
Schéma Matching

Chapitre II : Schéma Matching

1. INTRODUCTION

Dans le domaine de l'ingénierie logiciel, l'intégration de différentes sources de données devienne de plus en plus récurrente, en effet, la plupart des projets exploitent des sources de données externes du faite qu'elle contient des informations historiques et importantes, ce qui nécessite une compréhension de sa structure afin de les intégrer correctement avec le nouveau projet. Ce mécanisme peut être automatisé de manière à simplifier l'étude de données lors d'une conception logicielle. Ce chapitre est basé sur une étude détaillée de différentes approches existantes dans ce domaine.

L'objectif de ce chapitre est de présenter le contexte de notre travail en définissant les différents concepts et cas d'applications. Nous allons présenter le concept de Mise en correspondance de schémas de Bases de données (connue sous le terme schéma-matching) en général. Nous présentons également les techniques et approches existantes dans la littérature, et nous concluonsavec une conclusion.

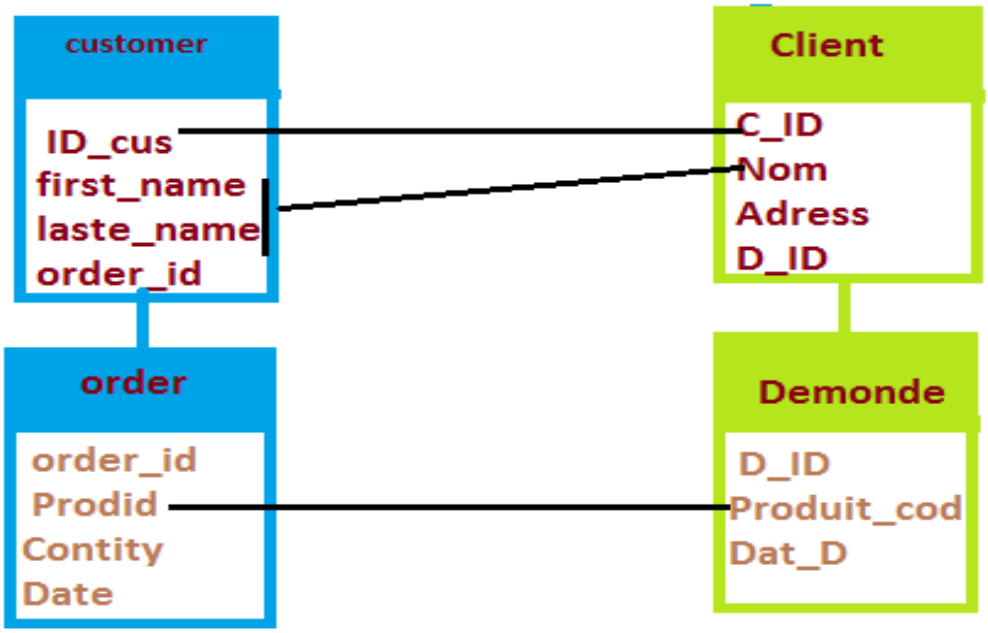
2. DataBase Schémas Matching

La mise en correspondance de schémas de Bases de donnéesa fait l'objet de recherche depuis un certain temps. Ce sujet est important dans des secteurs comme le commerce électronique, le web technologies, le marketing et le secteur de la santé [22**]** afin d'assurer une bonne intégration des données. Plusieurs études ont été menées pour résoudre les problèmes de mise en correspondance de schémas de base de données.

2.1Définitions

Le Matching des schémas est une technique qui effectue la découverte de correspondances sémantiques entre les éléments des schémas. C'est une opération qui prend par exemple des schémas de données et /ou du continue en entrée et retourne à la fin les valeurs de similarités sémantiques entre les éléments des schémas [2].

SchémaMatching est un processus qui prend deux schémas hétérogènes (par exemple S1 et S2 sur laFigure 5 comme entrée est produit en sortie un ensemble de mappages.



Schema S1 **Schema S2**

Figure 4: Simple schema matching demonstration

Cette figure représenté un exemple de résultat de schéma matching c-à-dire un mapping entre les élémentsde deux schéma S1 ,S2Le tableaux custemer de S1 contient quatre attributs ID_cus ,firste_name ,laste_name ,order_id ,est dans d'autre couté on a S2 contient le tableau client qui possède quatre attributs D_ID ,C_ID ,Nom ,Adress , les attributs des deux tableaux sont relier par des relation, l'attribut ID_cus est relié avec C_ID et le couple est relier avec l'attribut Nom de l'autre tableau.

3. Classification des approches de Schéma Matching

Les approches existantes ont été présenté par Rahm, E., Bernstein en (2001)[16],ils sontclassées en deux catégories: la première des catégories étant celle des approches basées sur des matchers simples (individuel), la deuxième étant celle des approches s'appuyant sur des matchers combinés, La figure 6 présente cette classification.

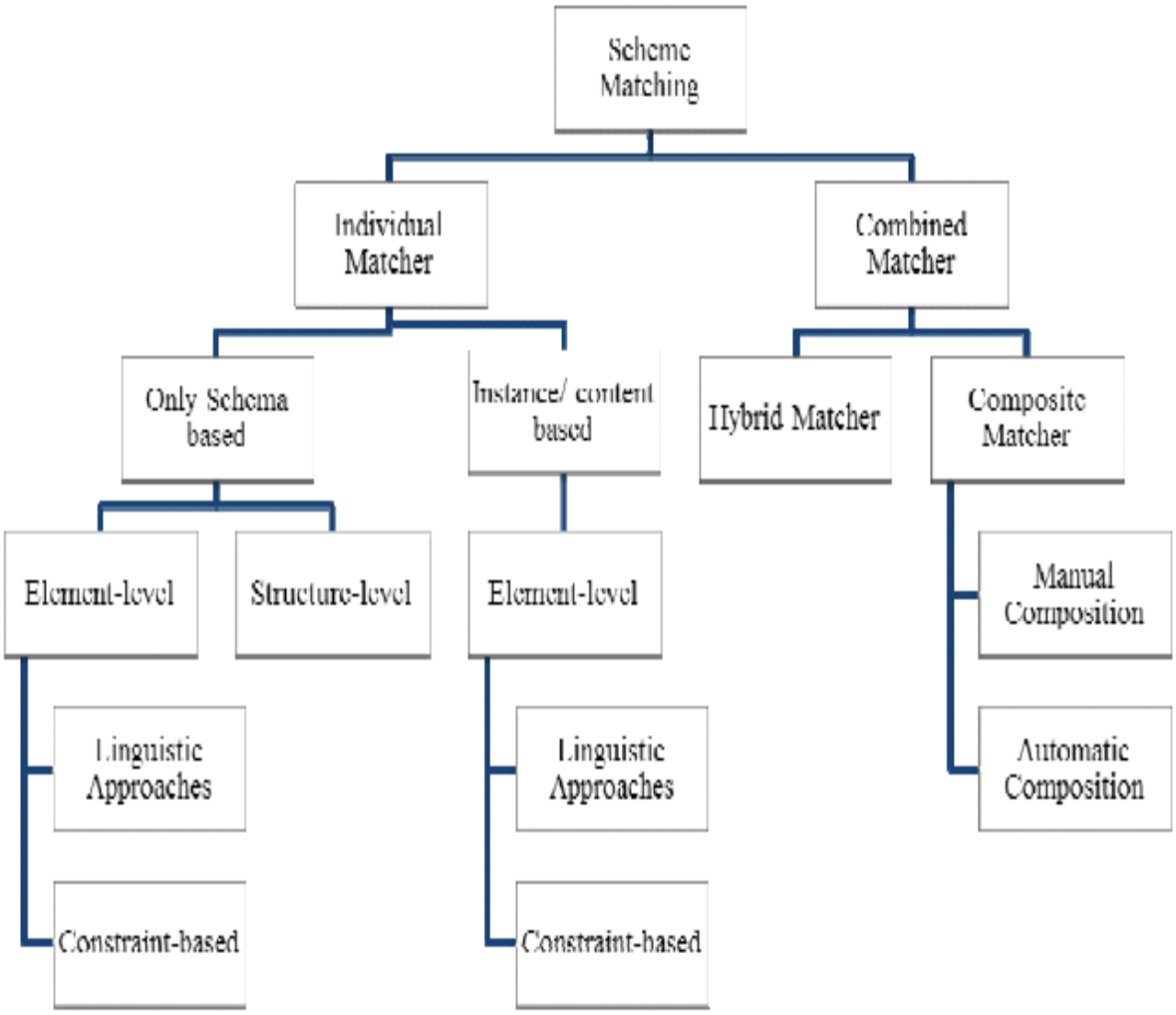


Figure 5:Schema Matching Approches

LaFigure 6 présente une classification des approche de mise en correspondance de schémas de base de données, cette classification sépare entre les approche qui se base sur les schéma uniquement et celle qui se base sur les instance (le contenu de base de données) uniquement, le niveau de l'élément et de la structure, entre les approche linguistique et celle basée sur les contraintes.

3.1Matchers individuels :

Cette catégorie inclut les matchers qui se base sur les schémas ou sur les instances, sur les éléments ou sur la structure seul, et sur des technique à base linguistique ou à base des règles.

Approche à base de Schéma contre instance:

Pour les approches basé sur les schémas, l'information au niveau du schéma est considérée, telles que les métadonnées, les noms d'éléments, les types de données et propriétés structurelles ou modèles, La méthode de similarité de texte sémantique (STS) détermine la similarité de deux textes en combinant la similarité des attributs entre eu, la similarité sémantique des attributs et similitude d'ordre des mots communs avec normalisation [17]

Par contre, dans les approches basées sur les instances, les données et le contenu des sources sont pris en considération uniquement. Ces approches utilisent des outils d'intelligence artificielle et de datamining [18].Dans [19]**,** Une examinassions des instances des attributs choisis et un calcul des valeurs de similarité entre eux (connue sous le nom de 'entropy-based distribution' EBD) sont effectués, en se basant sur un ensemble d'algorithme tels que N-gram et le nouvel algorithme TSim pour calculer EBD. et également K-medoid et normalised-Google-distance pour la segmentation et le clustering.

Eléments contre structure matcher :

L'action de matching peut considérerdes éléments simples tels que les attributs, ou des groupes d'éléments appelés structures.

Linguistique contre contrainte-based:

Les approches linguistique considère le nom et les descriptions textuelles des étiquettes de schéma ou des éléments, Différentes techniques, comme N-gram, EditDistance et SoundEX sont utilisés dans les approche linguistique [20].

Del'autre côté les approches se base sur des contraintes telles que les types de données, l'unicité et la clé, pour déduire les correspondances existantes.

3.2 Les approches combinés :

Un seul algorithme est utilisé par un comparateur individuel pour effectuer le processus de mise en correspondance. Pour les matcherscombinés, deux types d'appariement combinatoire peuvent être envisagés : (1) hybrides matcher qui prend en considération plusieurs critères pour effectuer le matching, et(2) les matchers composés, exécutent séparément plusieurs algorithmes de matching sur deux schémas et combine les résultats.

Différente matching approches ont été proposées par différents chercheurs. Cupid, développé par Jayant Madhavan , découvert les mappages entre les éléments des schémas en fonction de leurs noms, types de données, contraintes, et la structure du schéma en utilisant un plus large ensemble de techniques que les approches passées[21],Gillani S., Naeem, M., Habibullah, R., Qayyum ont fait une Taxonomie qui définie toutes les mesures de similarité sémantique possibles, de plus, a également proposé une approche qui exploite les relations sémantiques stocké dans l'ensemble de données DBpedia tout en utilisant un système de classement hybride pour déterrer la similitude entre les nœuds des deux graphs [22]

Hybride matchers

Combiner directement plusieurs approches d'appariement pour déterminer les candidats correspondants en fonction de plusieurs critères ou sources d'information. La plupart de ces techniques utilisent également des informations supplémentaires telles que des dictionnaires, des thésaurus et des informations de concordance ou d'incompatibilité fournies par l'utilisateur.

Autre classification :

Il existe d'autre classification des approches de schémas Matching. Pavel et zérome[23]ont proposé une classification qui est constitué de trois couches horizontales la Figure7 représente cette classification :

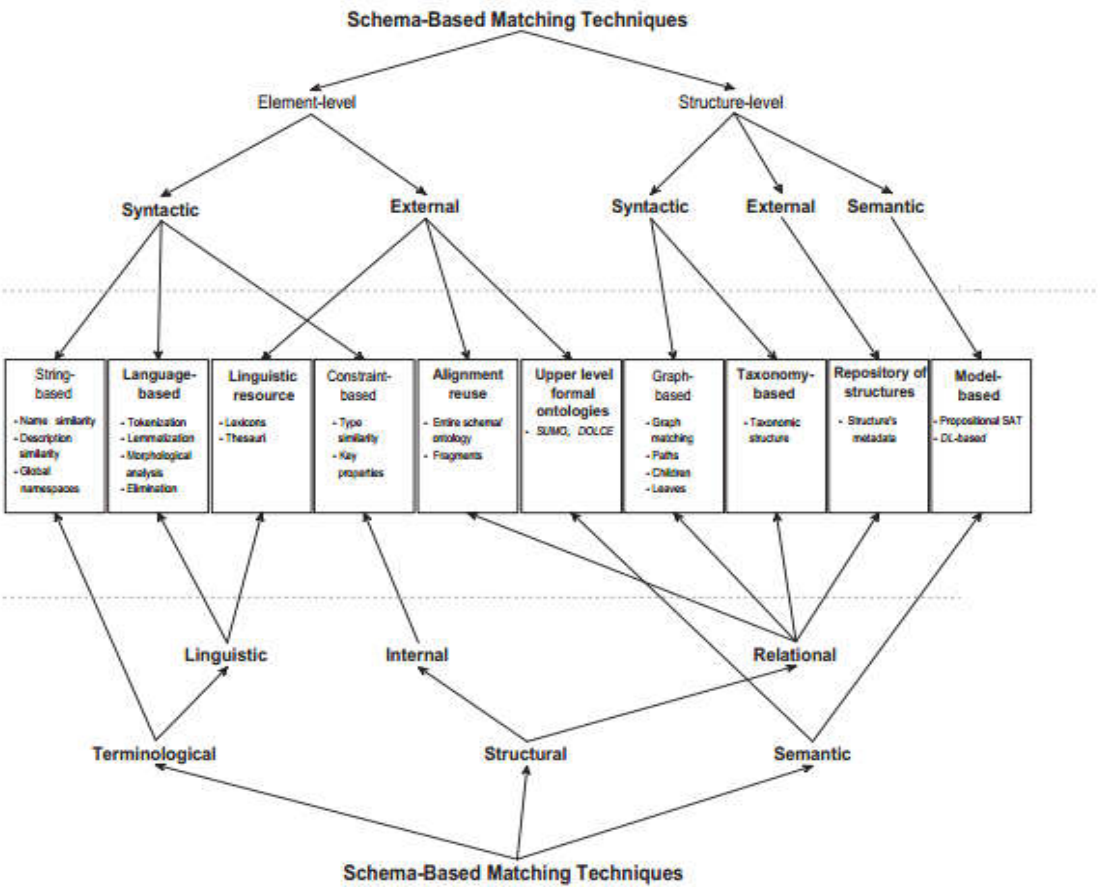


Figure 6: Classification des approches d'appariement par schéma élémentaire[23]

Couche 1

Elle classifie les techniques selon Le type d'entrée:

C'est le type d'entrée utilisé par les techniques du mapping. Il existe de nombreuxformalismes de représentation de données hiérarchiques disponibles. Une méthode demapping est généralement conçue pour fonctionner avec certains types de formalismes.

C'est la façon dans laquelle les techniques interprètent les informations en entrée, en distinguedeux niveaux :

Techniques basées sur les éléments

Ceux sont les techniques qui calculent des correspondances en analysant desentités ou des instances en isolation et ignorant leurs relations avec d'autres entités ou instances.

Techniques basées sur les structures

Ce sont celles qui calculent des correspondances en analysant comment les entités ou leursexemples apparaissent ensemble dans une structure.

Ces deux techniques se décomposent en trois niveaux:

-Niveau syntaxique,

-Niveau externe.

-Niveau sémantique.

Couche 2

Cette couche contient les techniques de bases utilisées pour établir descorrespondances entreles entités, chaque technique est portée sur une vue partielle d'ontologie, elles sont utilisables à la fois par les techniques de la couche 1 et la couche 2.

Couche 3

Elle classifie les techniques selon les types d'entrées :

Les techniques terminologiques :

Ces techniques reposent principalement sur des comparaisons de textes, elles s'appliquent surles noms, les commentaires et les propriétés des concepts afin de trouver ceux qui sontsimilaires, Il existe des méthodes de comparaisoncomme : Méthodes basées sur les chaines decaractères (texte) comme leur nom l'indique, ces méthodesconsidèrent l'entité comme une séquence de lettres. Etla Méthodes qui se basent sur les langagesles entités ici sont considérées comme untexte (des mots) dans un langage donnée, elles font appel généralement auxtechniques du traitement automatique du langage pour extraire les termesles plus représentatifs à partir d'un texte, elles utilisent généralement desressources externe tel que les dictionnaires.

Les techniques structurelles :

Se focalisent sur la comparaison des structures des entités à mapper, en fonction de la natureinterne ou externe des structures à comparer, on peut distinguer**:**

Méthodes structurelles internes, qui utilise la comparaison de structure interne d'une entité (ex : la comparaison des attributs, des noms, des types des attributs,..),et la Méthodes structurellesexternes, qui utilisent la comparaison externe en mettantpar exemple en jeux la disposition des entités dans leur hiérarchie, le voisinage.

Les techniques extensionnelles :

Elles se basent sur la comparaison des instances des entités des ontologies a mapper, en fonction de l'intersection des extensions de deux entités, on peut distinguer :Méthodes de comparaison d'extensions communes: elles sont utilisées lorsqu'il y a un ensemble des instances communes entre les ontologies**,**Méthodes d'identification des instances: elles sont appliquées quand il n'existepas un ensemble commun des instancescette méthodes est utilisables dans le cas où l'on sait que les instances sontidentiques (ex : lorsqu'on intègre deux ontologies de la même entreprise).

Les techniques sémantiques :

Elles sont très souvent des techniques qui se basent sur des modelés théoriques et se focalisent sur des techniques déductives exploitent très souvent la logique de description (test desubsumption) ou le calcul des prédicats on peut avoir deux méthodes dans cet technique:Méthodes basées sur des ontologies externes : si deux ontologies qui ne possèdent pas d'un terrain commun pour la comparaison vont être mappées, et la Méthodes déductives : cette méthodes né pas très performantes seules pourune tache inductive comme le mapping**.**

Autre classification des approches cette foissont classé par P.A. Bernstein[23] et dans cette classification Les approches de correspondance peuvent différer les unes des autres dans trois domaines, à savoir, les informations d'entrée, la façon dont les informations d'entrée sont traitées et les caractéristiques de la sortie correspondante au résultat. La Figure 8 représente cette classification.

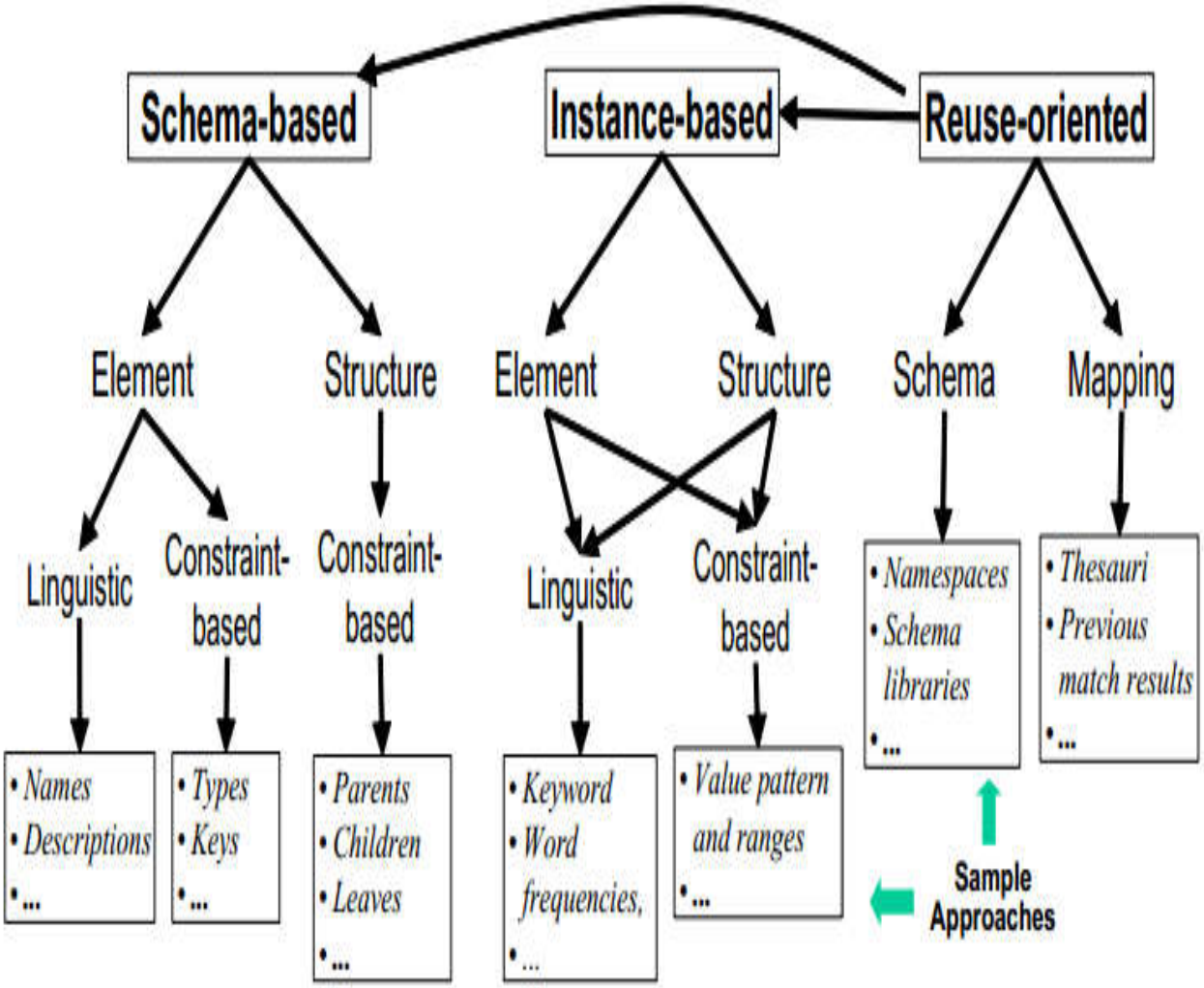


Figure 7: Classification des approches de correspondance[24]

4. Techniques de matching

Il existe un certain nombre d'algorithmes de matching qui cherchent à déterminer les correspondances entre les schémas. Ces algorithmes utilisent souvent des techniques appliquées aux éléments des schémas et des techniques structurelles. Ces techniquesont été classées par Shvaiko [24]

4.1 Techniques de matching d'éléments

Le matching d'éléments détermine les correspondances entre les éléments en analysant leurs labels (nom, prénom, sexe..etc.) Et en ignorant la relation structurelle avec les autres éléments. Ces techniques peuvent être décomposées en différentes catégories : méthodes terminologiques, linguistiques et basées sur les contraintes.

Techniques terminologiques :

Elles sont souvent utilisées afin de déterminer le matching des noms et de leurs descriptions. Ces techniques considèrent les chaînes de caractères comme des séquences de lettres de l'alphabet. Il existe un certain nombre de techniques terminologiques qui sont :

Préfixe:

Cette méthode prend en entrée deux chaînes de caractères et vérifie si la première chaîne est incluse au début de la deuxième chaîne de caractères (par exemple « int » et « integer »).

Suffixe:

Cette méthode prend en entrée deux chaînes decaractères et vérifie si la première chaîne est incluse à la fin de ladeuxième chaîne (par exemple « phone » et « téléphone »).

N-grammes:

Est une sous-séquence de n éléments construiteà partir d'une séquence donnée. En général, le n varie entre 1 et 5. Parexemple, les quatre-grammes (4) pour clientsont clie, lien, ient. Lamesure de similarité de quatre-grammes entre deux éléments *e* et *f* est définie comme suit : /ncliclie lien ient/

$$Sim_{tri}(e,f) = \frac{2 * |tri(e) \cap tri(f)|}{|tri(e) + tri(f)|}$$

Ainsi la distance entre client et nclientest de : 2*3/3+4=0.85.

Techniques de matching structurel :

Les techniques structurelles calculent le matching d'éléments en analysant comment les entités apparaissent ensemble dans la structure. Ces techniques considèrent soit la structure interne des entités (e.g les attributs et leurs types) et/ou la relation entre les entités (appelées aussi techniques relationnelles). Elles peuvent s'appuyer sur :

Une structure en graphe:

Les schémas de données en entrée du processus de matching (e.g schémas de base de données, des taxonomies ou ontologies)sont considérés comme des graphes contenant des nœuds

(ou sommets) et des relations (ou arcs). Souvent la comparaison de similarité entre une pairede nœuds (ou sommets) de deux schémas de données (schémas ouontologies) est basée sur l'analyse de leur position dans le graphe.L'hypothèse est que si deux nœuds (ou sommets) de deux schémas dedonnées sont similaires alors leurs voisins peuvent être aussi similaires. Par exemple SimilarityFlooding, proposé par Shavaiko,et,Jerome[25], se base sur cettehypothèse

Sur les taxonomies :

Il s'agit également de structures sous forme de graphesqui considèrent uniquement la relation de spécialisation. Les techniquestaxonomiques se basent sur l'hypothèse que si des nœuds possèdent desrelations de type is-aet sont souvent similaires alors leurs voisins peuventaussi être similaires.

5. Approches de matching à large échelle

Nous allons décrier deux approches de matching à large échelle: le matching deux à deux ou dit « pair-Wise » et le matching « holistique ». L'accent sera mis sur les stratégies et les techniques d'optimisation qui ont été utilisées et qui permettent d'améliorer le processus de matching et de réduire sa complexité dans un contexte dynamique et à large échelle. Nous présentons et comparons également les différents systèmes implémentant ces approches.

La différence entre les deux approches deux à deux (Pair-Wise) et holistique utilisées pour le matching à large échelle se situe au niveau des schémas de donnéesen entrée.

En effet, le matching deux à deux (Pair-Wise) permet de déterminer les correspondances uniquement entre deux schémas/ontologies volumineux à la fois (un schéma/ontologie source et une autre cible) alors que les approches holistiques peuvent être appliquées sur plusieurs schémas à la fois mais qui sont de petite taille. Nous allons détailler les différents travaux liés à ces approches en mettant l'accent sur les stratégies et les techniques utilisées pour améliorer la qualité des résultats, optimiser le temps d'exécution des matchers, réduire l'espace de recherche des matching candidats et assurer le passage à l'échelle.

5.1 Approches de Matching deux à deux (Pair-Wise)

L'approche classique de détermination du matching est l'approche « pair-Wise ». Elle est adoptée par la plupart des travaux de matching. Son principe est de trouver les correspondances entre paires d'éléments de deux schémas/ontologies [2], la figure 9 présenté l'approche de Matching deux à deux (pair-Wise).

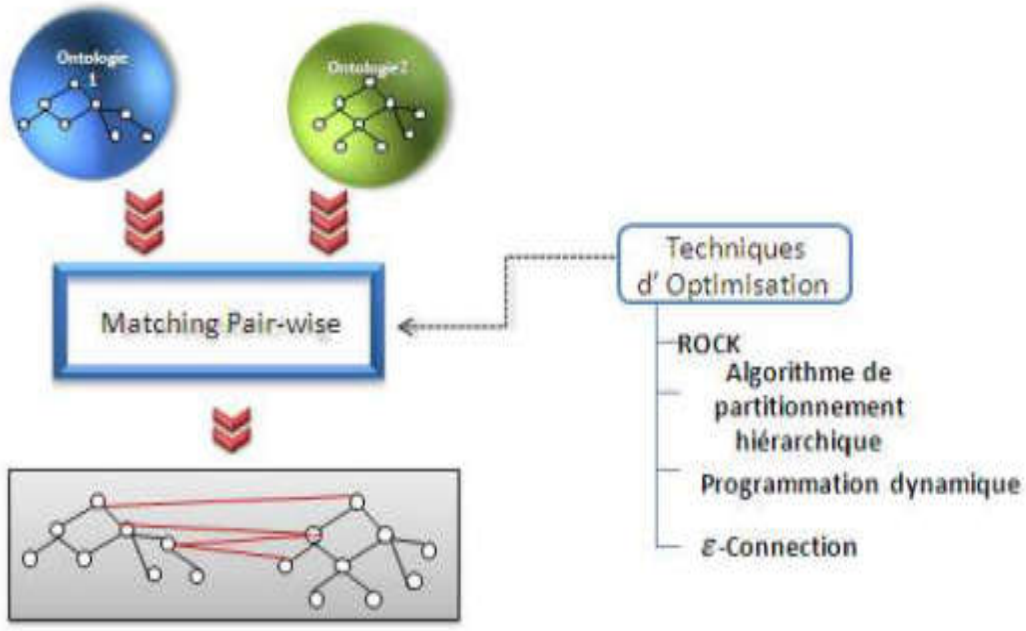


Figure 8: Approche de Matching deux à deux (pair-Wise)

5.2 Approches de Matching holistique

L'approche holistiquepermet de réaliser le matching de schémas multiples (par exemple il s'agit d'interfaces Web représentant des requêtes sur des bases de données du Web profond décrivant par exemple des réservations d'avion ou d'automobiles, etc.) pour trouver les attributs correspondants dans tous les schémas en une seule fois. Contrairement aux approches de matchingdeux à deux (pair-Wise), les approches holistiques prennent en entrée un ensemble d'interfaces web, la figure 10 déterminer Approches de Matching holistique.

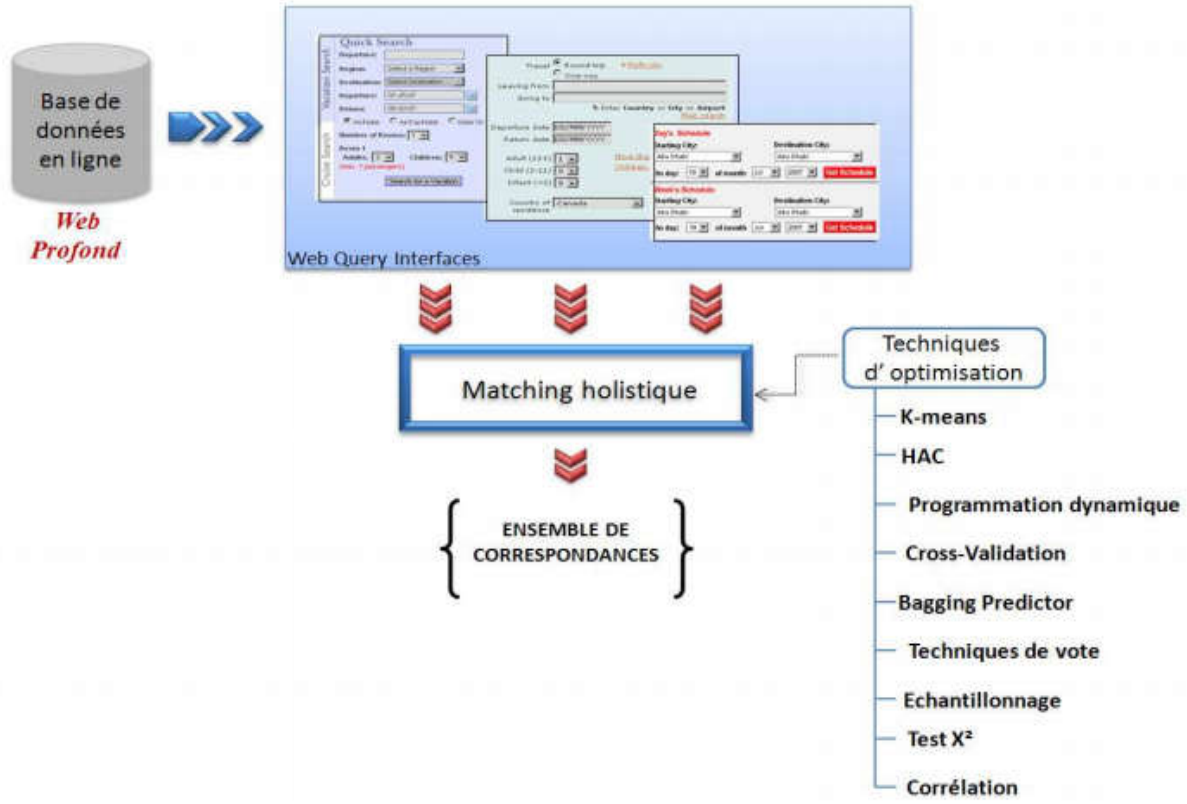


Figure 9: Approches de Matching holistique

Plusieurs travaux de matching de schémas à large échelle ont été proposés dans la littérature par [26], [27], [28], [29] et [30] utilisant autant d'approches différentes basées sur différentes stratégies et techniques d'optimisation et qui visent à améliorer la qualité et la performance du matching de deux schémas.

LeTableau 2illustreles différences entre les approches deux à deux (pair-wise) et holistique en se basant sur les avantages et inconvénients de leurs stratégies (tableaux

	Approches deux à deux (pair-wise)	Approches holistiques
Entrée	Deux schémas de données volumineux	Plusieurs interfaces web de petite taille
Avantages	Réduction de l'espace de recherche pour améliorer les performances.	Des résultats fiables qui montrent la pertinence des stratégies holistiques
Inconvénients	Risque de perte de bonnes correspondances suite à la décomposition. En effet, la plupart des stratégies deux à deux ne couvrent pas toutes les parties susceptibles d'avoir des correspondances	Aucune indication concernant le temps de traitement. Les algorithmes utilisés dans les stratégies holistiques possèdent pour la plupart une complexité exponentielle

Tableau 2: Comparaison entre les approches deux à deux (pair-wise) et holistique

Nous remarquons que les stratégies utilisées dans les approches deux à deux (Pair-Wise) ont comme inconvénient le risque de perte des bonnes correspondances contrairement aux stratégies holistiques qui offrent des valeurs élevées de fiabilité. Ceci s'explique par le fait que les approches deux à deux traitent des schémas volumineux contrairement aux stratégies holistiques.

Nous remarquons que les approches deux à deux ont pour avantage d'améliorer les performances en termes de temps d'exécution et ceci en limitant l'espace de recherche grâce aux stratégies qui décomposent les schémas de données.

6. Conclusion :

Nous avons présenté dans ce chapitre un état de l'art sur les travaux existants de schéma matching. Nous avons présentéplusieurs classifications des approches de schéma matchinget nous avons décrit les différentes stratégies et techniques qui ont été utilisées dans chacune des approches qui permettent de faire un matching d'élément, la plupart des approches existantes ont eu recours à des techniques permettant d'améliorer les performances et la fiabilité du matching.

Chapitre III
Contribution

Chapitre III : contrebutions

1. Introduction

Ce chapitre présente notre contribution au problème posé dans ce mémoire, pour la Mise en correspondance de schémas de Bases de données (connu sous le terme schémas-matching). Les schémas matching est un facteur pour assurer l'interopérabilité entre les systèmes hétérogènes dans plusieurs environnementscomme le Web sémantique est entreprises,ce qui permet un accès rapide simple est efficace aux données traitées et aide à la prise de décision efficace

Notre hypothèse s'appuie sur un modèle semi-automatique qui permettra de faire une intégration des données rapide est sans-erreurs

Ce chapitre est organisécomme suit : en premier lieu nous, présentons le contexte du travail qui parlera du cas pratique, nous présentons l'architecture du SMA à faire évolueret le Rôle del'agent intégrateur dedonnées, en deuxième lieu nous détaillonsl'approche utilisée pour le matching etnous terminerons avec une conclusion.

2. Contexte du travail

Ce travail aété proposé pour supporter l'intégration semi-automatique d'un système multi agent additionnelle conçu spécialement pour améliorer et évolué un système MRP(manifactiring resource palaninge) existant ,l'hypothèse se base sur le fait que le système existant (legacy system) possède ce propre base de données et le système additionnel possède aussi sa propre BD ,les deux base de données concernent le même domaine (manufacturing information système) mais ils ont été conçu séparément c.-à-d. qu'il ne possèdent pas le même schéma de base de données.

L'objectif de notre travail est de proposer une approche pour permettre à l'agent intégrateur de donnée,qui est un des agents du SMA additionnel, d'intégrer le système multi agent au système existant par le billé d'intégrations de données.Cette agent sera équipé par la capacité lui permettant de mettre en correspondance les deux schémas de BD et de tracer une Mapp pour la génération automatique de requêtes liées aux activités du système Multi Agent qui se résume par les actes de communication des messages échangésentre l'agent-Intégrateur et les autres agents du SMA.

2.1 Système multi agent (SMA)

Définition

Anne Nicole dit que le système multi-agents (SMA) est un système composé d'un ensemble d'agents (un processus, un robot, un être humain, etc.), situés dans un certain environnement et interagissant selon certaines relations [32].

2.2 Agent intégrateurs des données

Le rôle de l'agent intégrateur est premièrement de comprendre le schéma de la base de donnée de l'application existante (legacy-BD), effectué la mise correspondant du ce schéma avec celui de la base de donnée du system multi agent (SMA-BD) et deuxièmesdurant a vie l'agent est recommandé à faire une transformation et génération automatique de requête SQL vers le legacy-BD pour répondre auxbesoins des autres agent du SMA ,et d'assurer cette migration entre les deux base de données

Dans notre approche nous considèrent uniquement les informationsdes structures incluent dans le code SQL de définition des bases de données, nous nous n'utilisons pas les instances de donnée que contient la legacy-BD car la SMA-DB est incluent empty.

3. description de l'approche utilise pour la misse en correspondons desschémas

Nous allons présenter dans cette section l'approche d'intégration utilisée dans notre travail,on a pris en considérations lesdépendances fonctionnelles entreles attributs des deux schémas pour minimisé l'espace de recherche de correspondances ce qui permet de réduire la complexité du problème.

Dans notre approche on a prend deux schémas de données en entrée et retourne à la fin des valeurs de similarités entre les éléments des schémas. L'approche génère des relations de Matching entre les tableaux de la base de données , Il traite les élément (attribut) de chaque tableau , Il se base sur la similaritéde base entre les éléments individuels (attributs), en s'appuyant sur les descriptions textuelles de données à l'aide de schémas de base de donnée , et la similarité des éléments en comparant les voisinages d'une paire d'éléments. Les deux similarités sont utilisées pour calculer la similarité finale entre chaque paire d'éléments et cette

dernière permettre d'obtenir le résultat final de matching, la Figure 11 représente une phase de matching

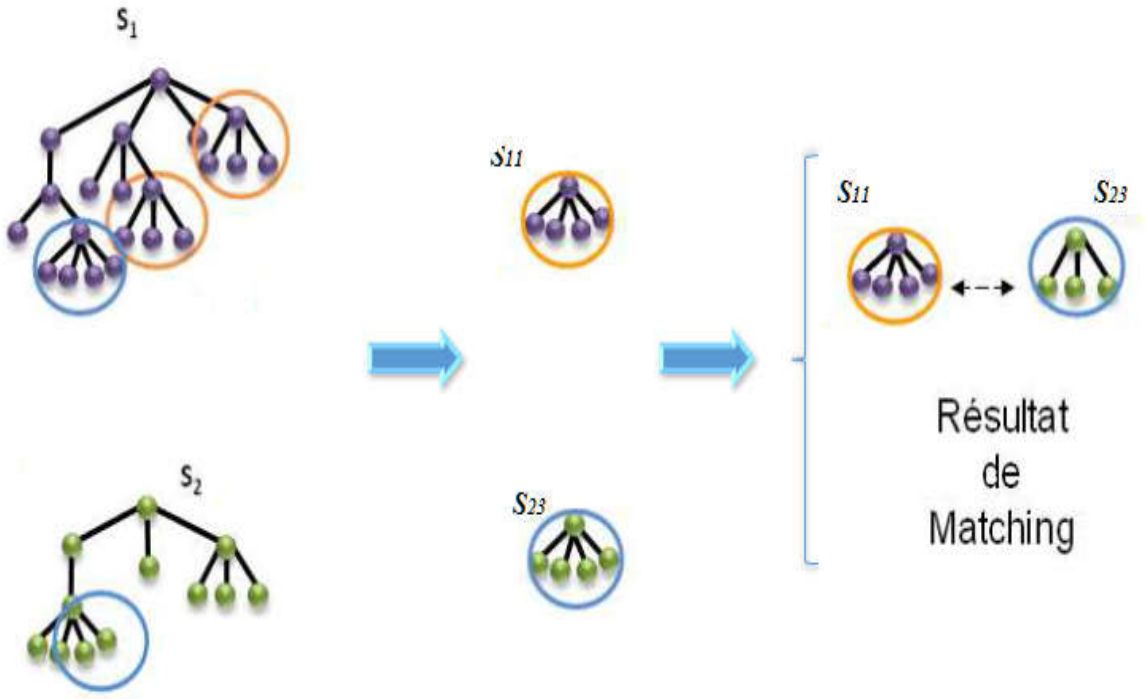


Figure 10: Phase de matching

On réalise donc le matching entre les tableaux les plus similaires pour obtenir les correspondances entre leurs éléments. Chaque groupe de sous structures similaires représente un problème de matching qui est résolu indépendamment. Le matching qu'on cherche à réaliser devrait déterminer la similarité entre les éléments. On a utilisé les similarités **terminologiques** et **linguistiques** entre les labels (synonymies). Le matching d'élément est important car les éléments référencés des schémas présentent plusieurs contextes, tels que l'administrateurpeut être représenté sou le nom de (Gestionnaire).

Pour déterminer le matching d'élément entre les tableaux, nous avons utilisé un dictionnaire et calcul de similarité des éléments pour la bonne correspondons de matching ente les tableaux, attributs.

3.1Position de l'approche dans la classification universel

La Figure12 représente le chemin de technique de matching d'élément que on a utilisépour la mise en correspondance entre les attributs des DBs.

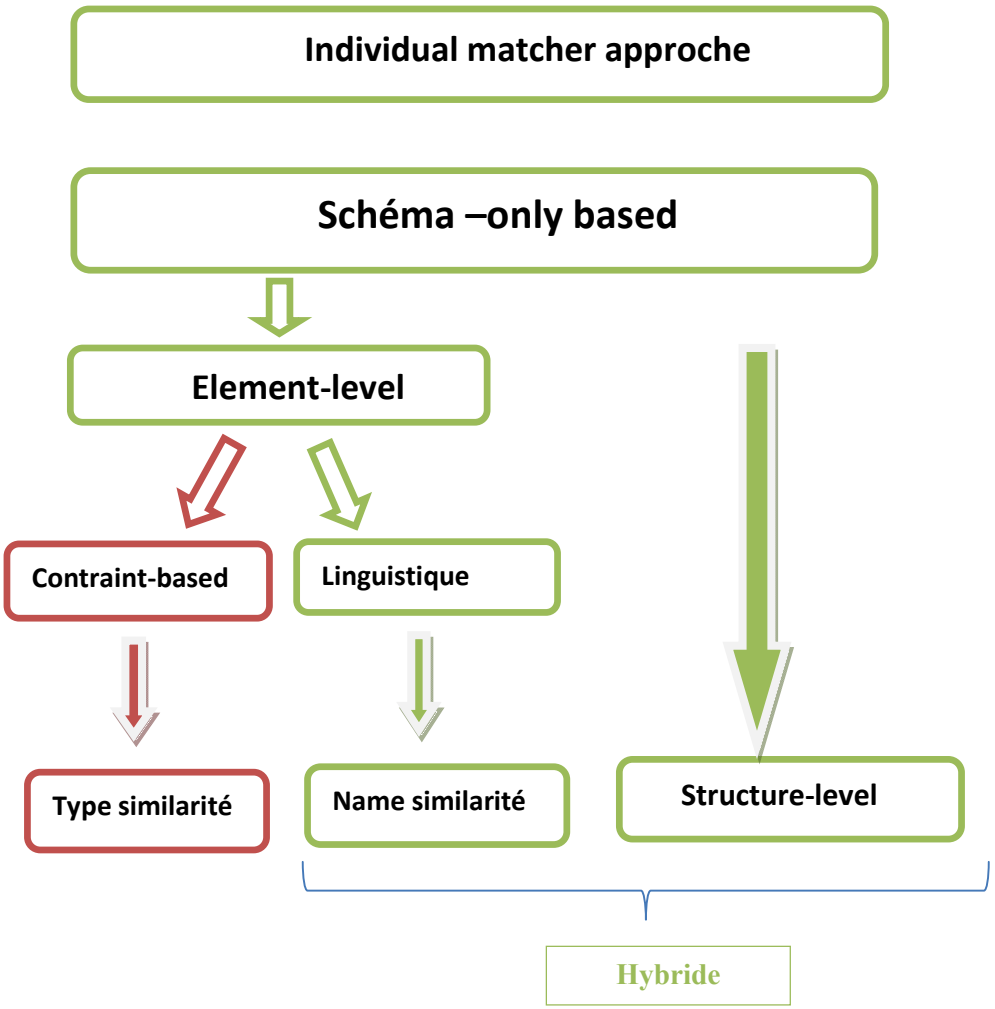


Figure 11: Situation de l'approche proposer dans la classification universelle

3.2 Problématiques liées à la réalisation

Plusieurs problèmes doivent être pris en compte pendant la conception de systèmes d'intégration de données ,dans notre étude d'application on résoudre les problèmes d'intégration entre legacy-BD et la deuxième base de donnée (local) permet ces problèmes on cétone les problème suivent :

Problèmes d'Hétérogénéité des données :

Ces conflits se trouvent dans le cas où nous utilisons des propriétés différentes (les attributs ou le nom des tableaux), les cas spécial dans la mise on correspondance des éléments sans représenté comme suite :

Le conflit de nom

Ces conflits se trouvent dans le cas où nous utilisons soit des noms différents pour le même concept ou propriété (synonyme), ou un nom identique pour représenter les attributs pour les deux schémas S1, S2

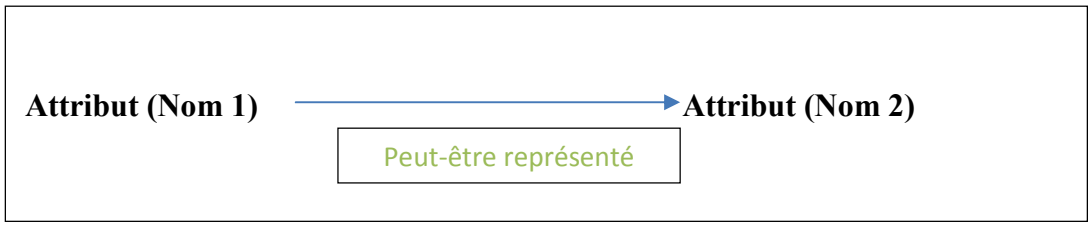


Figure 12: représente le problème de conflit de nom

Les problèmes de conflit de nom (1) sans des problèmes qu'il doit être résoudre d''une manier automatique par les algorithmes qui on a été créé et il sons pas besoin de les résoudre manuel.

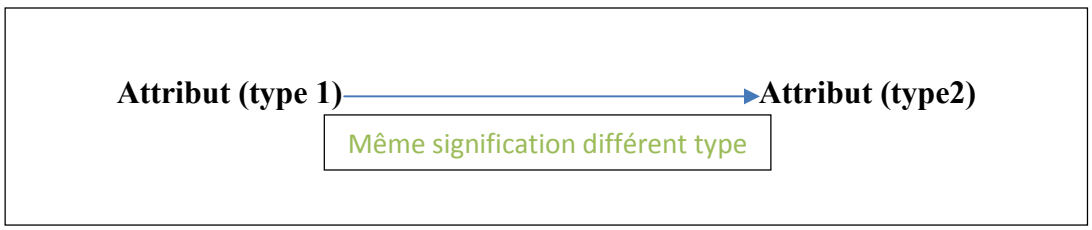


Figure 13: représente le problème de conflit de type

Les problèmes de conflits de type sont des problèmes qui ont trouvé lorsque on a deux attributs qui sont avoir la même signification, mais ils sont évalués dans différents type d'attribut, ce problème son résoudre après avoir consulté notre mappe (c'est l'étape de validation) d'une manier semi-automatique.

Conflit de représentations :

Ces conflits se trouvent dans le cas où nous représenté un tableau par plusieurs tableaux ou représenté un attribut par plusieurs attributs, c-à-dire schémas différents pour décrire le même concept.

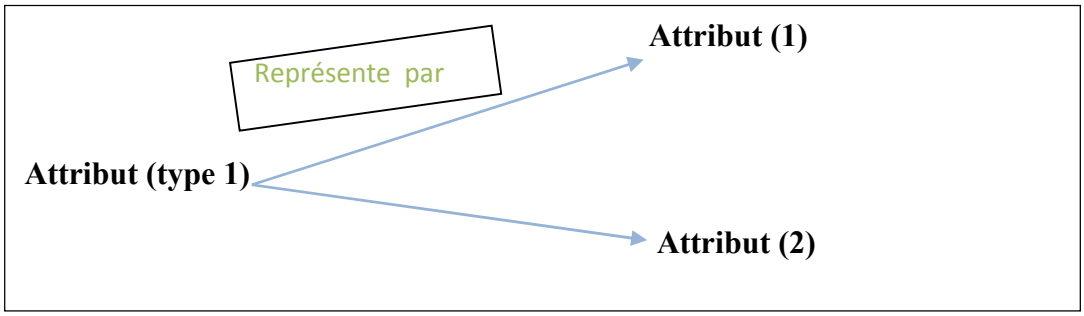


Figure 14: le problème de représentation des attributs

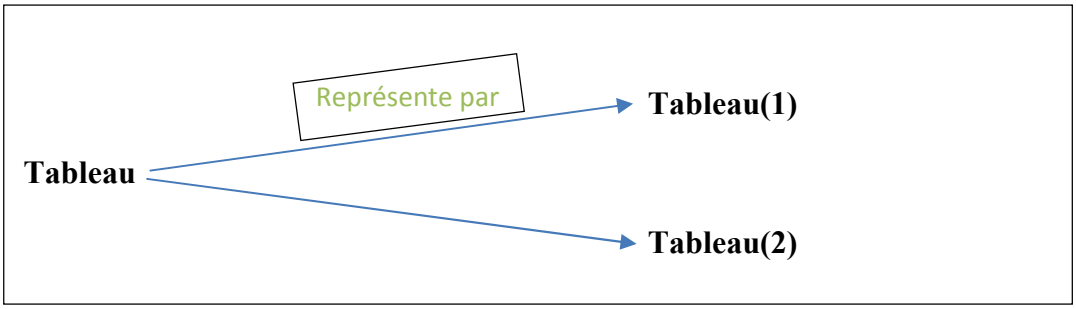


Figure 15: le problème de représentation des tableaux

Ce problème on le trouvait lorsque on a un tableau représenté par deux tableaux, on peut le résoudre le problème de représentation des tableaux d'une manier automatique à partir de notre approche ou d'une manière semi- automatique (manuel)

4. Processus de mise on correspondance

La Figure17 représenté les différentes étapes pour la préparation de la Mapp de correspondance des éléments des deux schémas, cette mappe sera utilisé comme étant la brique de base pour les transfère des données entre les deux bases de données avoir résoudre notre problème.

L'étape 1 : génération de graphe de dépendance fonctionnelle à partir de scripte SQL de définition de basse de données.

L'étape 2 : traitement itérativedes graphes de dépendance pour le couplage des éléments des deux schémas de base de données

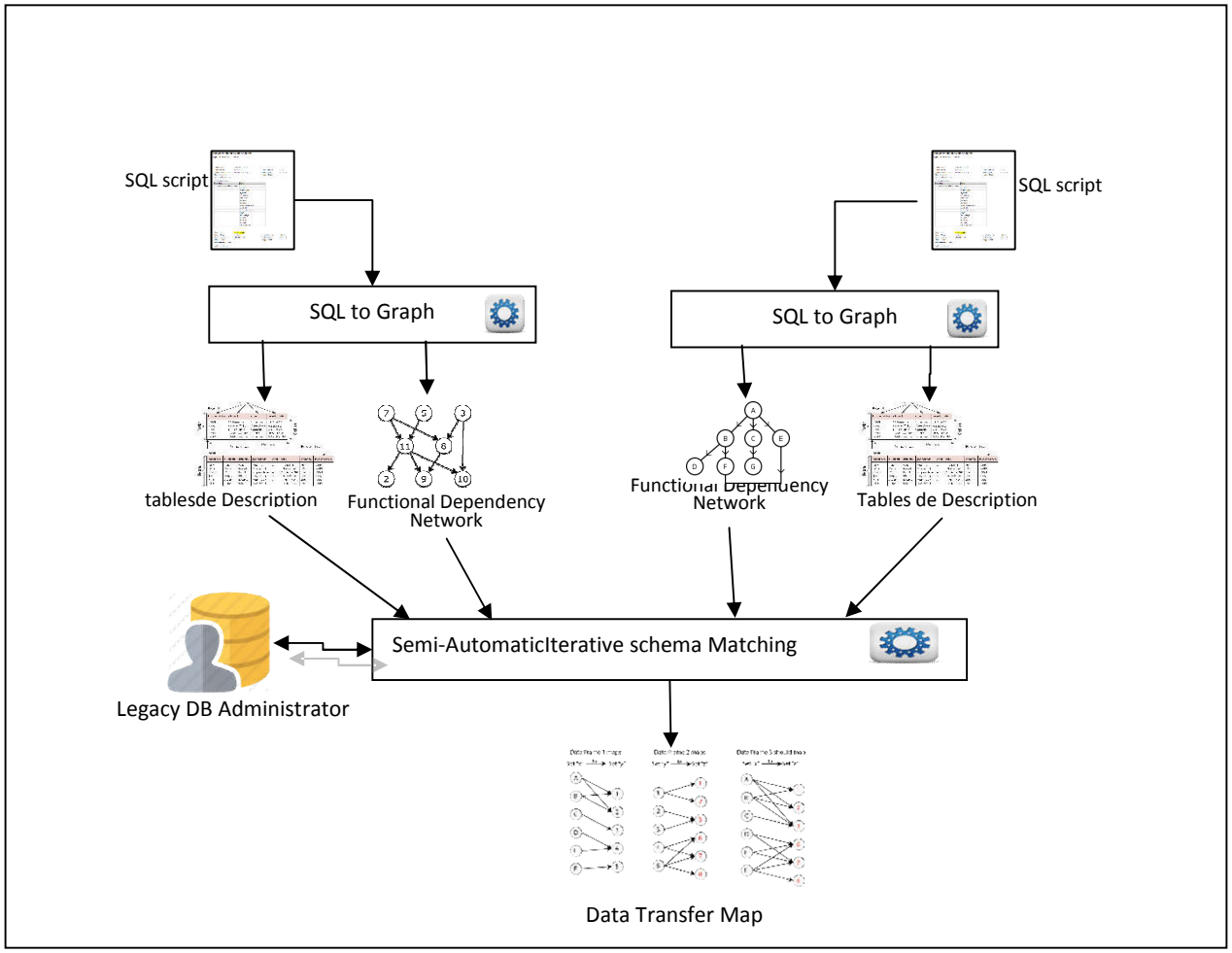


Figure 16: Matching des schémas

Cette méthode permet de déterminer la Correspondance de (Matching) uniquement entre deux schémas (unSchéma existant (legacy) et un autre schémadu système à ajouter)

L'approche se base surdeux étapes essentielles, laFigure 17présentée ce deux étape la premier étape et la Construction des réseaux de dépendances fonctionnelles et des tables de Description par une présentation graphique, la deuxième étape, et l'itérative Semi-Automatique de schéma Matching pour confirme la correspondance entre les schémas.

4.1 Etape 1 : SQL to Graph

Dans cette étape nous effectuons la Construction des réseaux de dépendances fonctionnelles et des tables de Descriptions correspondantes aux deux schémas de BD en se basant sur les scriptes SQL de définition des deux Bases de données.

Le réseau de DFs et les tables de descriptions générées seront utilisés par la suite Pourdétecter les liaisons entreles attributs et pour la construction de la Mapp de correspondance.

L'approcheconsidère que les deux schémas sont en Troisième Forme Normal (1FN, 2FN, 3FN).

La dépendance fonctionnelle :

Dansle domainedesbasesdedonnées, lesdépendancesfonctionnellessontlesliensexistantentredeuxpropriétésprésentesdanslabase de données.[33]

Soit R(U) une relation avec U ensemblesdeses attributs est x, y∈ U

On dit y dépond fonctionnellement de x (ou que x détermine y ou encore que y déterminé par x) et on notex $\xrightarrow{DF}$ y

Ssi La connaissance de la valeur de x Implique la connaissance d'une seule valeur de y

Le réseau de dépendances fonctionnelles capture tous les dépendances entre les attributs d'un domaine d'étude précis, il est utilisé initialement par les concepteurs de systèmes pour la modélisation de schéma de base donnée. Dans notre cas, nous faisons une réingénierie inverse c à d on génère le réseau de DFs à partir du schéma de la base de donnée décrit par le script SQL de définition. La Figure 18 présente un exemple de graphe de dépendances fonctionnelles.

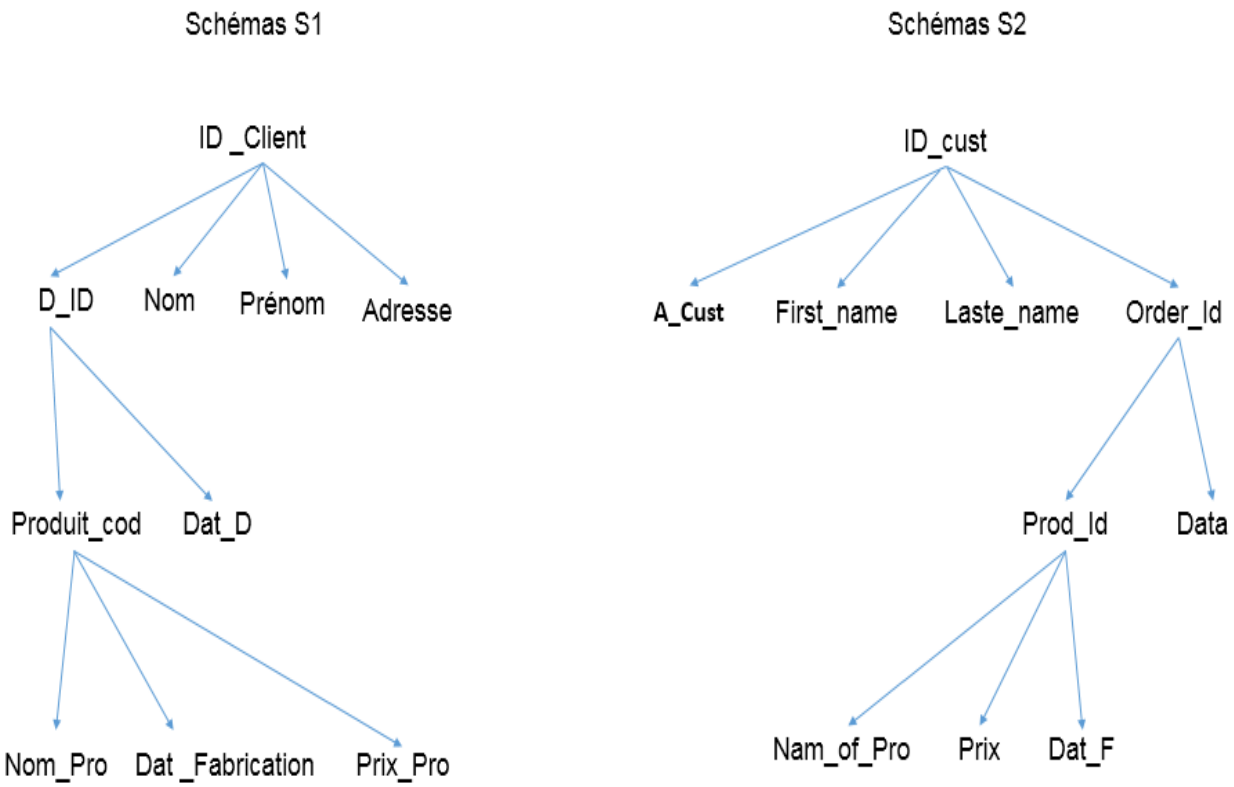


Figure 17: réseau de dépendances Fonctionnelles

4.2 L'étape 2 : Traitement itérative des graphes de dépendances fonctionnelles

Cette partie est la partie principale de l'approche, elle permet le couplage des éléments des deux schémas de base de données et la générationde la Mapp d'intégration qui sera utilisé par la suite dans le transfert de données.

Elle comporte les phases suivantes :

1- Phase de couplage des tableaux : cette phase permet de résoudre le problème d'intégration des tableaux de la base de données source et la deuxièmebase de données à l'aide de l'algorithme Match _Table qui permettre de couplet les pair des tableaux.
2- Phase de couplage des attributs utilise l'algorithme de Match_ Attribuesen se basant sur les dépendances directes pour faire le couplage entre les attributs et on utilisant un dictionnaire et comme une deuxième itération on utilisant la similarité entre les pair des attributs.
3- Phase de l'intégration pour les attributs en se basant sur les dépendances indirect (cette phase consiste à faire coupler les attribue non matché dans la deuxième phase et cela

après un traitement d'étendage du graphe de dépendances par ajout des dépendances indirecte).

On note qu'entre l'exécution de ces phases, le système propose à l'administrateur une interface de validation et de correction des correspondances détectées automatiquement.

4.2.1 Techniques Linguistique utilisées

La similarité linguistique :

La similarité entre deux entités représentées par des termes peut aussi être déduite en les analysants à l'aide des méthodes linguistiques. Les méthodes linguistiques permettent de déterminer la similarité entre deux entités, est-elle se base sur l'utilisation des ressources externes (dictionnaires, taxonomies. . .).

Les méthodes de la similarité linguistiques cherchent la forme canonique ou représentative d'un mot ou d'un terme (attribut) en exploitant ses variantes linguistiques, La similarité entre deux termes est mesurée en comparantes attribut. La recherche se faire à l'aide d'un dictionnaire.

Les similarités terminologiques :

Ces techniques et base principalement sur des comparaisons de textes, elles s'appliquent sur les noms (attribut), afin de trouver ceux qui sont similaires, cette techniques considèrent les chaînes de caractères comme des séquences de lettres de l'alphabet, Il existe un certain nombre de techniques terminologiques qui Il existe des méthodes de comparaison comme la méthode que l'on a utilisé :

la Méthodes qui basées sur les langages les entités ici sont considérées comme un texte (des mots) dans un langage donnée, elles font appel généralement aux techniques du traitement du langage pour extraire les termes les plus représentatifs à partir d'un texte, elles utilisent généralement des ressources externe tel que les dictionnaires , et pour déterminé les mot qui sont similaire ,et faire notre traitement sur la base de donne central (legacy) et la deuxième base de donnée on a créé un dictionnaire pour déterminé ce mottes et réalisé la Mapp des c'est deux base de données, le code suivant représenté la définition de notre dictionnaire utilise dont notre traitement de la réalisation de schémas.

```
Map <String, String[]> map = new HashMap <String, String[]> ();
```

Basé sur l'utilisation des ressources externes (dictionnaires, taxonomies. . .) : la similarité entre deux entités représentées par des termes est calculée à partir des liens sémantiques déjà existants dans les ressources externes

4.2.2 Présentation des algorithmes

Algorithme Match Table

Entrée :

- Liste1 des noms de tables de schéma1
- Liste2 des noms de tables de schéma2
- Dictionnaire de Matching
- La similarité entre pair des tableaux

Sortie :

- Mapp de correspondance de Tables (Mapp_tble)

Début

Pourtoute table_name1 dans Liste1

Pourtoutetable_name2 dans Liste2 et table_name2 non matché

Si check_ synonyme () et vrai faire

Marque liaison entre table_name1 et table_name2

Fin Si

Fin pour

Fin pour

Tant que existe entrée entré non traité faire

L :=recherche_max_sin ()

Couplé_table (MS[L,1],MS[L,2])

Suprême toute les entré de MS

Tant quetable_name1 =MS [L,1]

Suprême toute les entré de MS

Tant quetable_name2 =MS [L,2]

Fin tant que

Return Mapp_tble ;

Fin ;

Les fonctions

- La fonction check_ synonyme (table_name1, table_name2, dictionnaire).permettre de vérifier si table_name1 et table_name2 sont des synonyme dans le dictionnaire elle retourne vraie si table_name1 est synonyme de table_name2 sinon faux.
- La fonction calcul _sm (table_name1, table_name1) permettre de donner la similarité de 0 à 1 pour tout pair des tableaux.

L'algorithme Match_Table permettre de faire la mise on correspondance entre les tableaux, elle est prendre comme des entré la liste des noms de tables de schéma1 et la liste des noms de tables de schéma2, l'algorithme Match Table faire le traitement en deux étape la premier étape utilise le dictionnaire pour couplé les pair des tableaux et si il existe entrée entré non traité passé à la deuxième étape et utilise la similarités pour couple les outre tableaux ,lorsque l'algorithme Match Table et terminé elle retourne les paire des tableaux (Mapp_tble).

Algorithme Match_Attribues

Entrée :

——List des attributs de schémas1

—— List des attributs de schémas2

——Nombre_S1=nombre des attributs de schéma 1

——Nombre_S2=nombre des attributs de schéma 1

——Mapp_ table

——La similarités entre les attributs

Sortie :

——Mapp de correspondance des attributs (Mapp_attrubuts)

Début

Pour toute couple (Table1, Table2)dans la Map_Table

Pour tout Attribut_Att1∈ attribue(Table1)

Pourtout Attribut_ Att2∈ attribue(Table2)

Si check_ synonyme () et vrai faire

Marque liaison entre attrebut_Att1 et table_Att2

Fin Si

Fin pour

Fin pour

Tant que existe entrée entré non traité faire

H :=recherche_max_sin ()

Couplé_ table (MS[H,1],MS[H,2])

Suprême toute les entré de MS

Tant queattrebut_name1 =MS [H,1]

Suprême toute les entré de MS

Tant queattrebut_name2 =MS [H,2]

Fin tant que

Fin pour

Retourne Mapp _Att ;

Fin ;

L'algorithme Match _Attribut pour la dépondons direct permettre de faire le matching entre les attributs, elleprendre comme entrée la List des attributs de schémas 1 et la liste des attributs de loutre schémas, ,et prendre en considérations les dépendances Fonctionnelles directs entre les attributs et la Mapp des tableaux qui a été réalisé par l'algorithme de (Match Table) ,

l'algorithme Match_ Attribuesfaire le traitement en deux étape la premier étapeutilisé le dictionnaire pour faire le matching entre les attributs mi tant que elle existe entre non traite elle ce passéà la deuxièmeétapes et traite les pair attributs à l'aide de la similarités des attributs.

5.Processus de transfert de données et génération automatique des requête

La mise à jour et le transfert des donnéesentre les deux BD se fait selon le besoin et la logique de travail du système additionnel de renforcement, le scénario se fait selon le processus présenté dans la Figure 19 Quand un agent d'intégration reçoit un message ACL de type requête, il l'analyse et la traduit selon le contenue, il génère automatiquement les requête SQL nécessaire et un plan d'exécution en se basant sur la Mapp de Mise en correspondance, le plan d'exécution et de réassemblage des résultat sera après lancer en interrogeant les deux Base de Donnée, et en fin une réponse sera reformulé et renvoyé à l'agent demandeur.

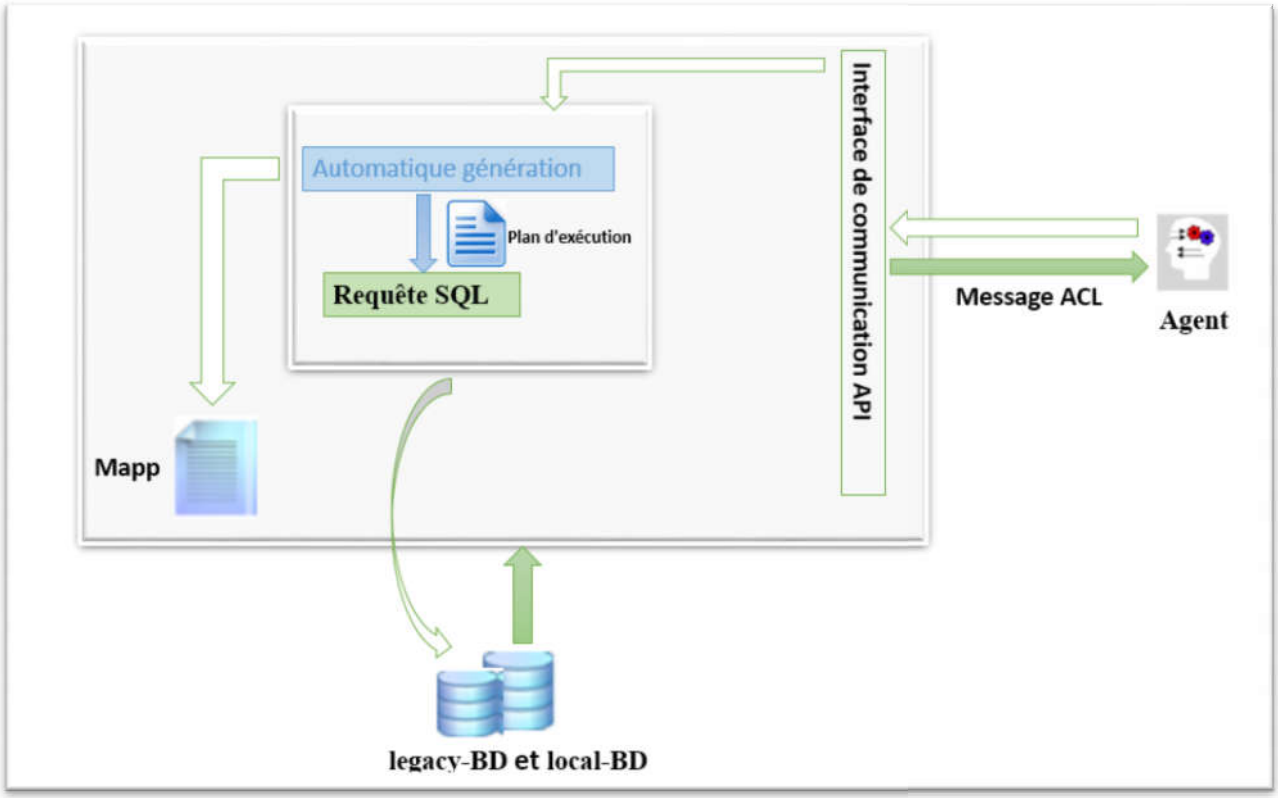


Figure 18: Processus de Transfert de données et/ou de Mise à jour

5.1 Analyse de la requête utilisateur :

L'utilisateur est invité à formuler ses requêtes en utilisant des langages formels pouvant être des langages de requête de BD, dans notre cas nous utilisent des langages de requête de BD (SQL) Ces requêtes doivent être analysées afin de déterminer leur type (simple ou bien mixte) et procéder à l'interrogation de la BD pour les satisfaire. Cette analyse se déroule en trois phases

assurant chacune un objectif bien déterminé :

-Première phasepermet de distinguer entre les mots réservés/clés (from, select, where, …) du langage formel utilisé et ceux introduits par l'utilisateur reflétant ses besoins réels (les données ou connaissances qu'il voulait extraire de la BD). Cette phase se base sur des analyseurs lexico-syntaxiques correspondant aux langages choisir

-Deuxième phaseaprès la séparation entre les mots réservés et ceux reflétant l'objectif de la recherche, il convient de déterminer la source de chacun de ces termes (BD centrale (legasy),et la deuxièmes BD)

-Troisième phasedétermine le rôle que les termes de la requête occupent aussi bien dans la requête que dans les sources de données : tables (concepts), tuples (instances), colonnes (propriétés) et transfère la requête utilisé par le system multi agent (SMA) à une requête mix pou accède ou deuxième base de données, etc.

Traitement des requêtes utilisateurs :

Cette étape s'intéresse au chargement des règles adéquates pour préciser le type de requêtes (simple, mixte) et le traitement nécessaire pour leur satisfaction qui est fortement lié à la nature de la requête. En effet, dans notre travail, nous traitons les requêtes de nature suivantes :

-Requête vide: requête bien formulée mais ayant comme résultat un ensemble vide d'enregistrements, autrement dit, aucune donnée de la BD correspond aux conditions spécifiées dans la requête.

Requêteinvalide: toute requête non conforme au langage d'expression. Il s'agit des requêtes incorrectes qui souffrent des problèmes syntaxiques ou structurels. Exemple, le cas d'une requête SQL qui demande d'afficher le contenu d'une colonne mis la requête contient une erreur de s'syntaxes.

Requête inadaptée: une requête mal formulée contenant des termes inadaptés ou erronés. Ces derniers doivent être remplacés par d'autres termes équivalents permettant de constituer une requête exécutable dans la BD.

Dans notre travail, Il est important de signaler que nous considérons les requêtes exprimées selon un langage standard : le langage SQL (Structured Query Language), le standard utilisé par les principaux SGBD.

5.2Traitement des requêtes SQL

Les requêtes SQL mixtes sont le résultat de la reformulation des requêtes inadaptées. Pour qu'une requête SQL inadaptée soit exécutable dans la BD, il faut que tous ses termes (après reformulation) fassent partie de la BD. Ceci nécessite : la reformulation des termes inadaptés en s'appuyant sur le mappe de la BD central (legasy) la et loutre base de donnée, Notons que la reformulation par synonymieest une technique d'expansion de requête qui consiste à remplacer un ensemble de termes par leurs synonymes extraits de dictionnaires

Exemple 1 : une requête SQL

En présenté un exemple de requête mixte qui nécessite à la fois l'accès à la base de donnée central (legasy) et à la 2eme BD. Il s'agit d'une requête SQL (select * from Client) constituée d'un seul terme (« Client ») appartenant à la base de donnée. Le système cherche l'existence de correspondants de ce terme apprêt avoir la mappe qui a été déterminé entre la base de donnée centrale (legasy) et la deuxième base de donnée afin de reformuler la requête pour qu'elle ne soit constituée que de termes appartenant à la BD (la recherche se base sur l'existence du nom de la table). Le système vérifie les termes synonymes du terme «Client» puis les remplacé.

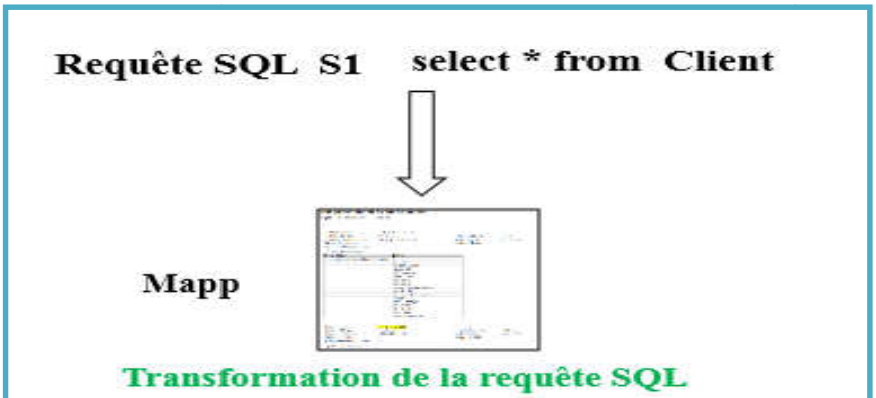


Figure 19: représenté la transformation de la requête SQL S1 à S2

6. Conclusion :

Nous avons présenté dans ce chapitre notre méthodologie de matchinghybride. Nous avons décrit les différentes phases de notre méthodologie qui sont la phase élément-level etstructurel-level. Cette méthodologie est appliquée sur deux schémas de base de données hétérogène, les analyse, détermine leurs similitudes et leur dépendance, les décompose et réalise le matching entre eux, et retourné une Mapp utilise pour faire le transfert des données.

Chapitre IV
Implémentation

Chapitre VI : Implémentation

1. Introduction

Après avoir achevé les étapes d'analyse et de conception du système, nous allons dans ce chapitre réaliser le système proprement dit. C'est la phase de réalisation de cette application pour l'intégration des données, qui utilise des technologies spécifiques. Nous allons commencer par la présentation des différents outils utilisés. Ensuite, la seconde partie concerne les principales interfaces graphiques.

2. Description de L'environnement de développement

Pour réaliser cette application, nous avons utilisé différents outils, Notre choix est justifié par ce qui suit:

2.1: le serveur

- **phpMyAdmin 3.3.8**

phpMyAdmin est une application web qui permet de gérer un serveur de bases de données MySQL. Dans un environnement multiutilisateur, cette interface écrite en PHP permet également de donner à un utilisateur un accès à ses propres bases de données[34].

La Figure 21 suivante montre une copie d'écran de la page d'accueil de PHPmyadmin. L'écran est divisé en deux parties. Sur la gauche on peut afficher toutes les bases de données et la partie droite présente l'ensemble des opérations disponibles en fonction du contexte.

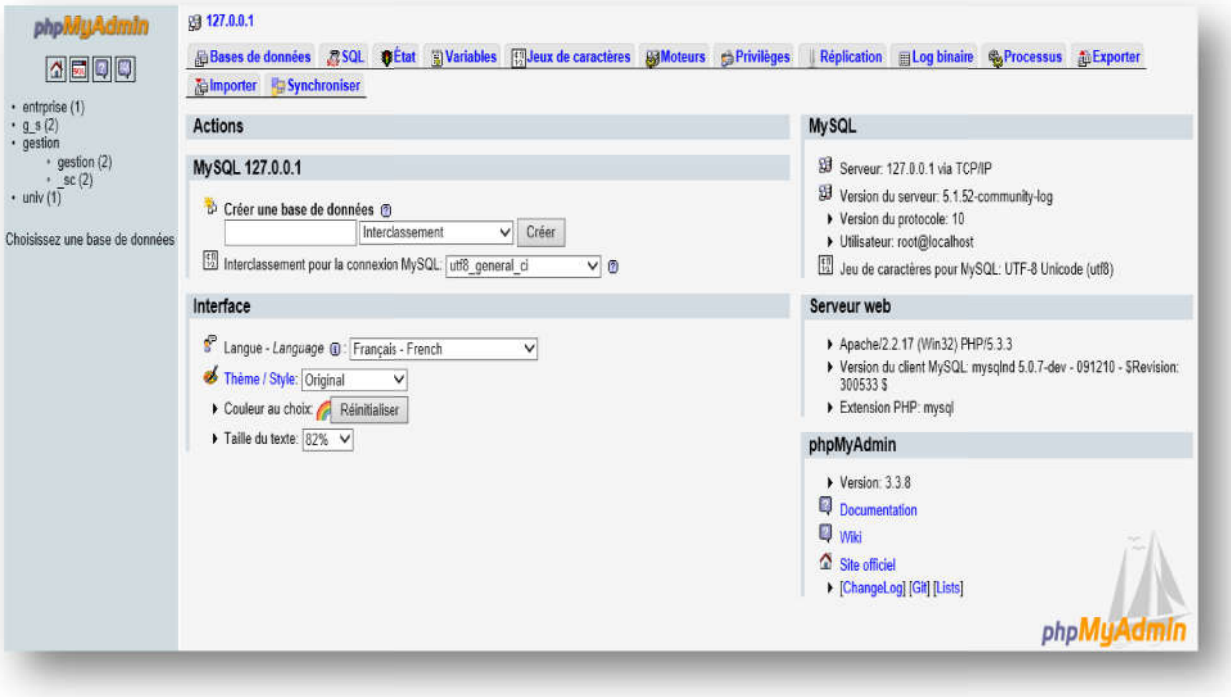
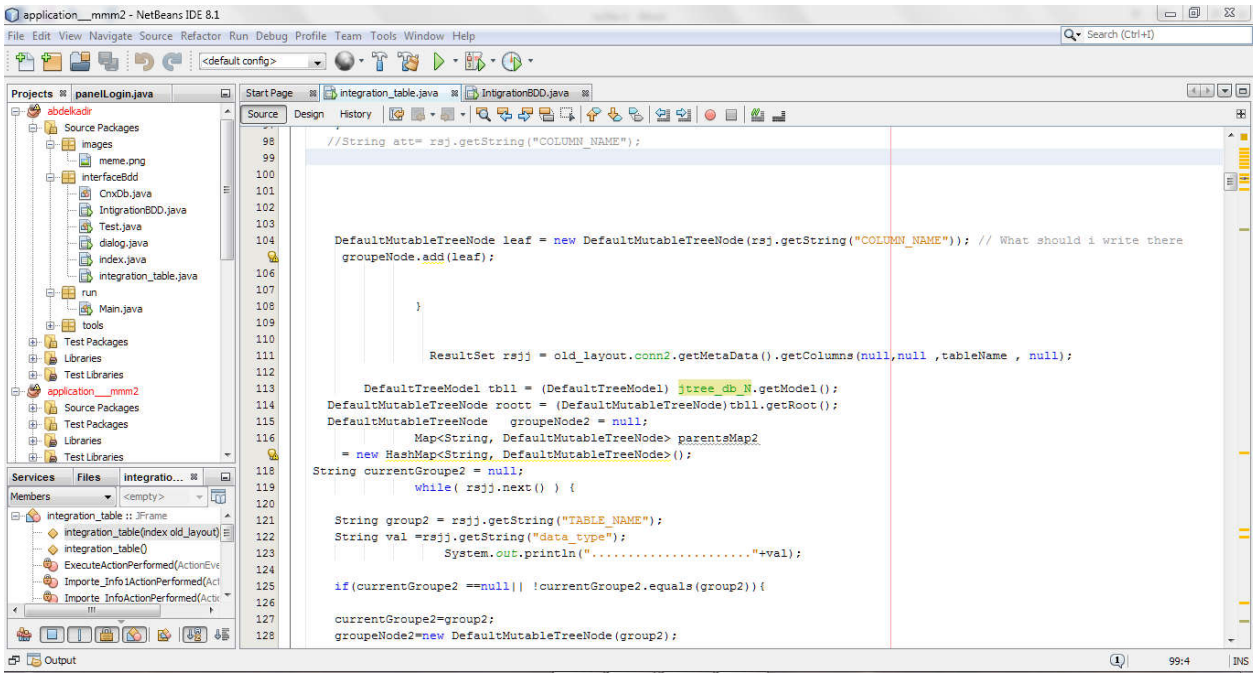


Figure 20:Interface phpMyAdmin

2.2 Les outils de développement

➢ NetBeans

NetBeans est un environnement de développement intégré (IDE) pour Java, placé en open source par Sun en juin 2000 sous licence CDDL (Common Développent and Distribution License). En plus de Java, NetBeans permet également de supporter différents autres langages, comme Python, C, C++, XML et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages web), NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X et Open VMS , La Figure 21 suivante montre une copie d'écran de la page d'accueil de netbeans[35].




```
//String att= rsj.getString("COLUMN_NAME");

DefaultMutableTreeNode leaf = new DefaultMutableTreeNode(rsj.getString("COLUMN_NAME")); // What should i write there
 groupeNode.add(leaf);

            }

              ResultSet rsjj = old_layout.conn2.getMetaData().getColumns(null,null ,tableName , null);

    DefaultTreeModel tbll = (DefaultTreeModel) jtree_db_N.getModel();
DefaultMutableTreeNode roott = (DefaultMutableTreeNode)tbll.getRoot();
DefaultMutableTreeNode   groupeNode2 = null;
            Map<String, DefaultMutableTreeNode> parentsMap2
 = new HashMap<String, DefaultMutableTreeNode>();
String currentGroupe2 = null;
            while( rsjj.next() ) {

 String group2 = rsjj.getString("TABLE_NAME");
 String val =rsjj.getString("data_type");
              System.out.println("....................."+val);

 if(currentGroupe2 ==null|| !currentGroupe2.equals(group2)){

 currentGroupe2=group2;
 groupeNode2=new DefaultMutableTreeNode(group2);
```

Figure 21: Interface NetBeans

3. Le langage de développement

3.1 Le langage de requête SQL

Pour communiquer avec une base de données, on a besoin de lui envoyer des commandes ou instructions appelées requêtes. Que ce soit pour la création, la suppression d'une table, la modification, l'insertion ou la sélection de données, le langage standard de requêtes est SQL.

SQL ou (Standard Query Language*)* est un langage permettant d'interroger les bases de données de manière simple. Il est doté d'une syntaxe particulière que l'on doit respecter pour que la communication avec la base se passe au mieux. Son succès est dû essentiellement à sa simplicité et au fait qu'il énonce des requêtes en laissant le SGBD responsable de la stratégie d'exécution.

SQL couvre les trois fonctions indispensables à la mise en œuvre et à l'exploitation de bases de données relationnelles :

- la création des données
- la manipulation des données
- le contrôle des données

A part le fait d'envoyer directement les requêtes SQL telles quelles au SGBD, le but ultime de l'utilisation de SQL sera aussi d'inclure ces requêtes SQL dans un programme écrit

Dans un autre langage. Ceci permet de coupler le SGBD à un langage informatique, donc à un programme. Tel est le cas de MySQL avec PHP.

3.2 Langage java

Java est un langage de programmation orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy (cofondateur de Sun Microsystems en 1982), présenté officiellement le 23 mai 1995 au SunWorld

La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes d'exploitation tels que Unix, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications, mais qui ont l'inconvénient d'être plus lourd à l'exécution (en mémoire et en temps processeur) à cause de sa machine virtuelle. Pour cela, divers plateformes et Framework associés visent à guider, sinon garantir, cette portabilité des applications développées en Java [36]

4. Description de notre application

L'accès à notre application d'intégration des données se fait via une page principale,Elle contient des interfaces faciles à manipuler. Notre but est de faciliter l'utilisation de l'application

Captures et interfaces

Nous présentons dans ce qui suit les interfaces principales de notre application pour la mise en correspondance de schémas de base de données hétérogène

4.1Partie authentification:

L'administrateur ou un utilisateur de la base de données se connecte par un Nom utilisateur et un mot de passe pour accéder à l'application

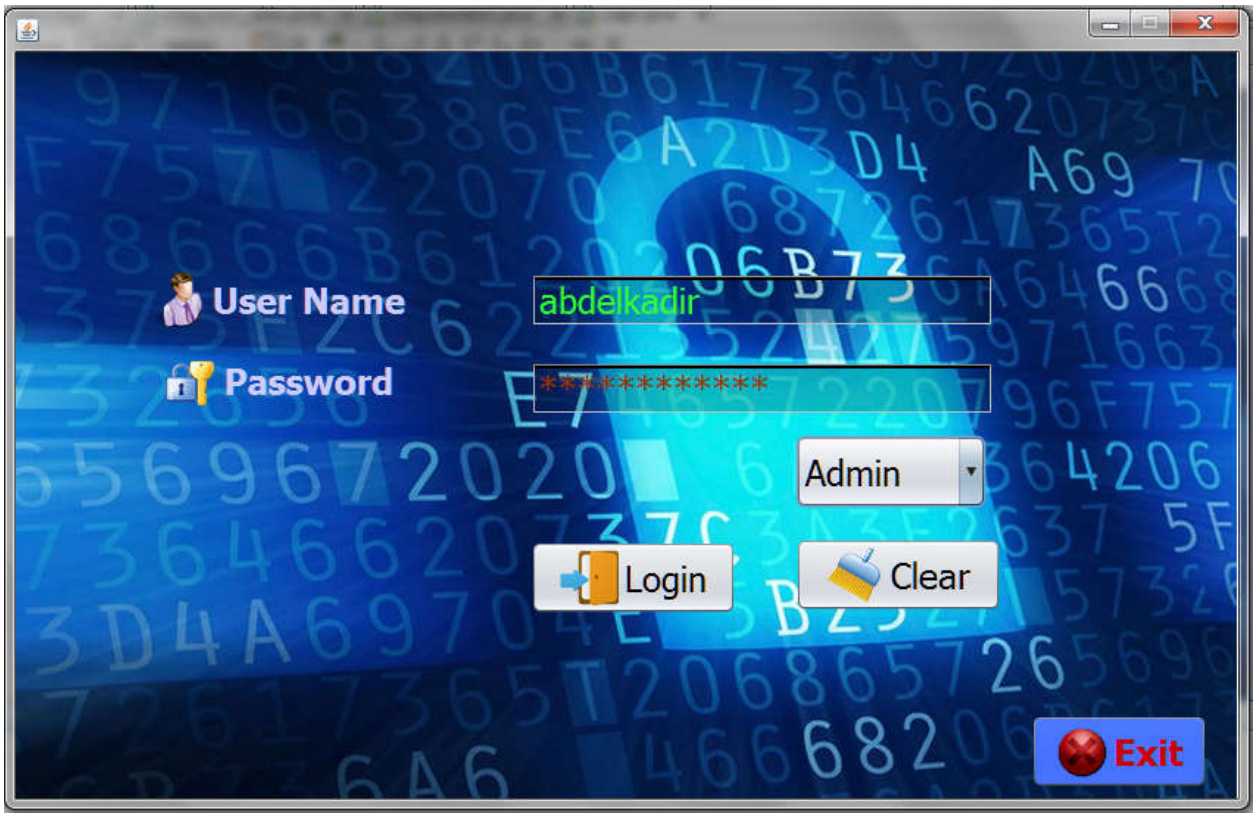


Figure 22: Formulaire d'authentification

4.2 Interface principale

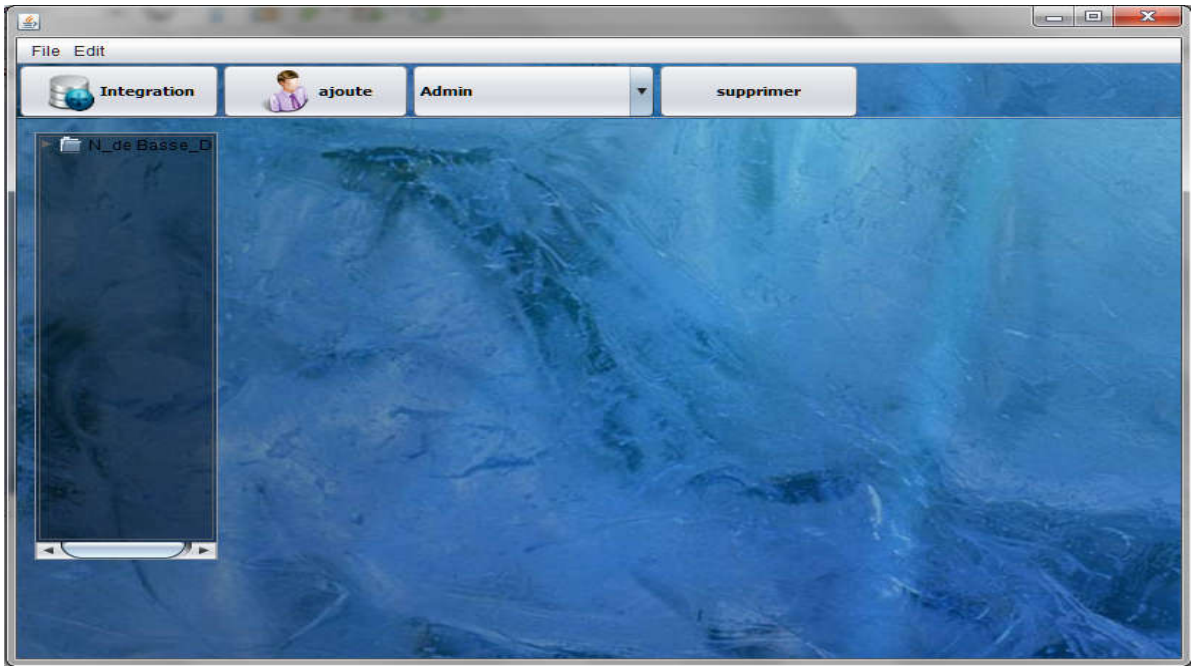


Figure 23:interface principale

Figure 24: Formulaire de connexion

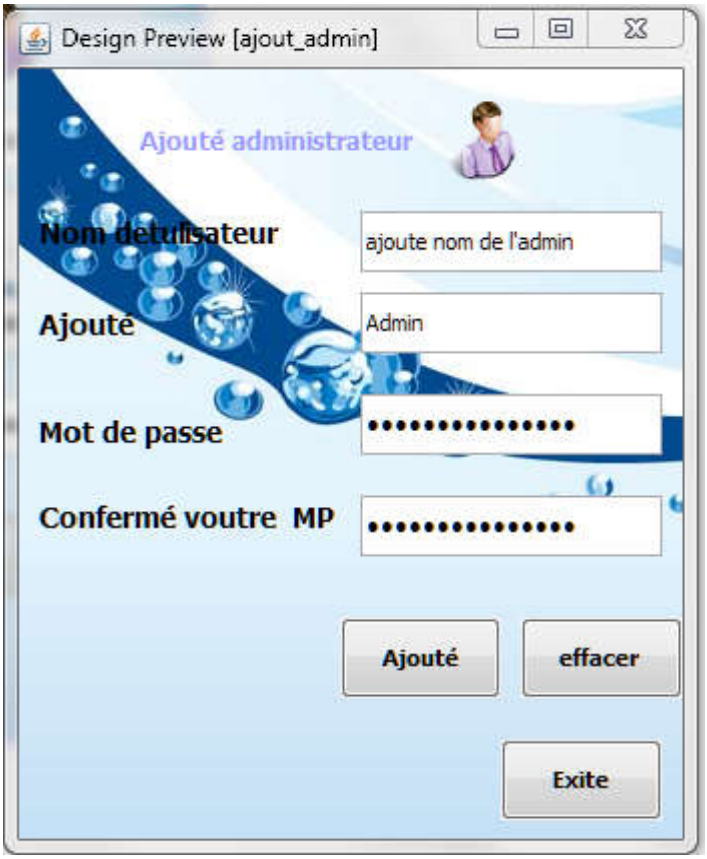


Figure 25: Formulaire pour ajouter un administrateur

4.3Manipulation des données

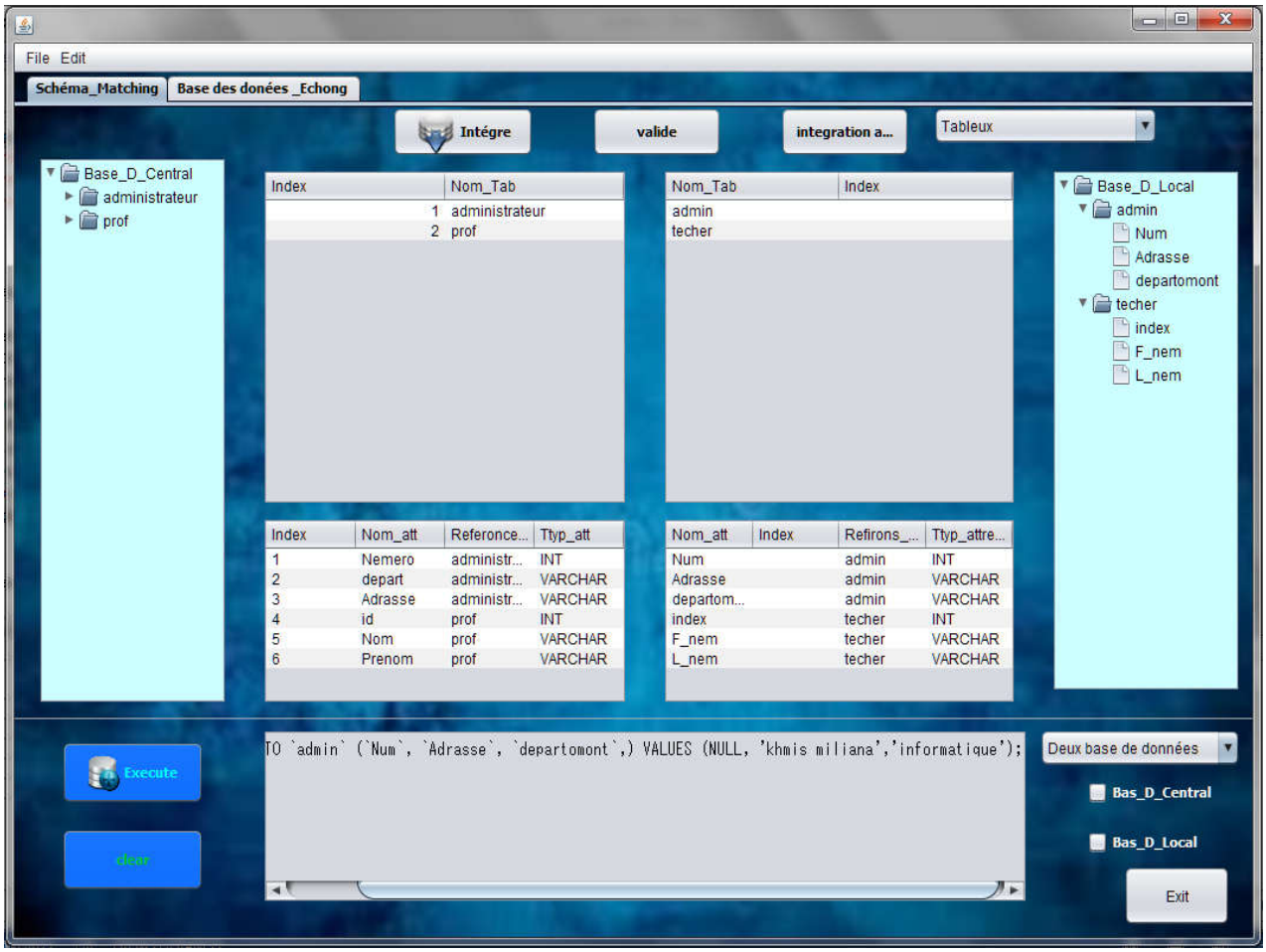


Figure 26: interface manipulation des données

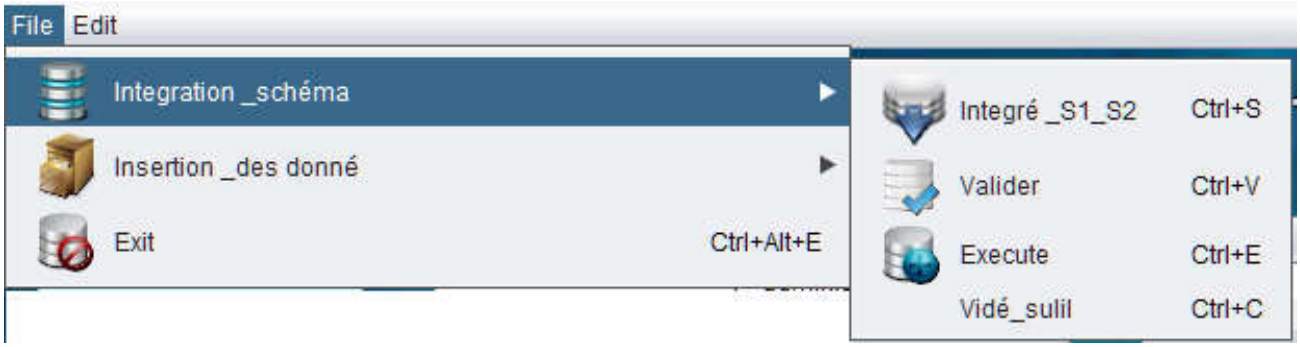


Figure 27: la barre de menu

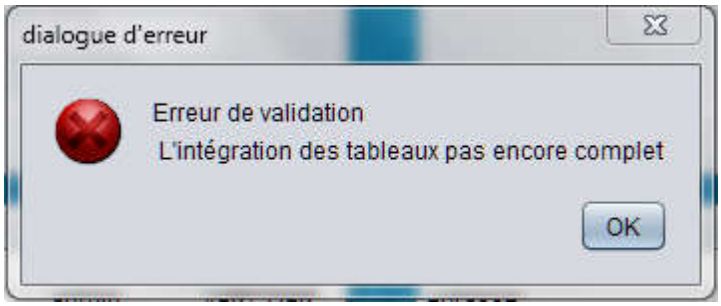


Figure 28: un dialogue d'erreur

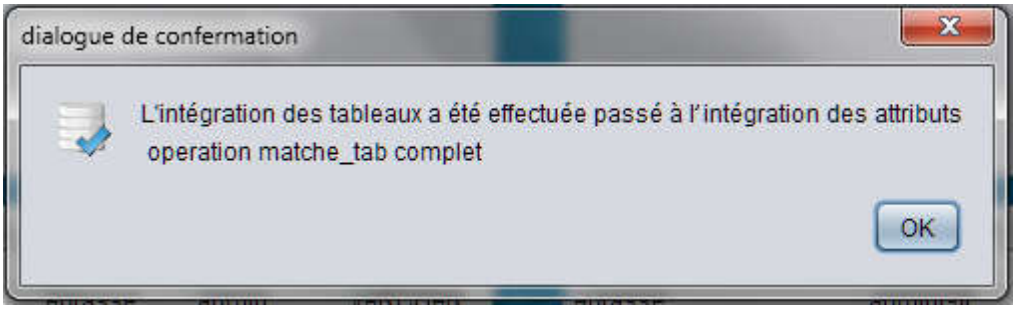


Figure 29: dialogue de confirmation pour l'intégration

4.4 Changement de type pour les attributs

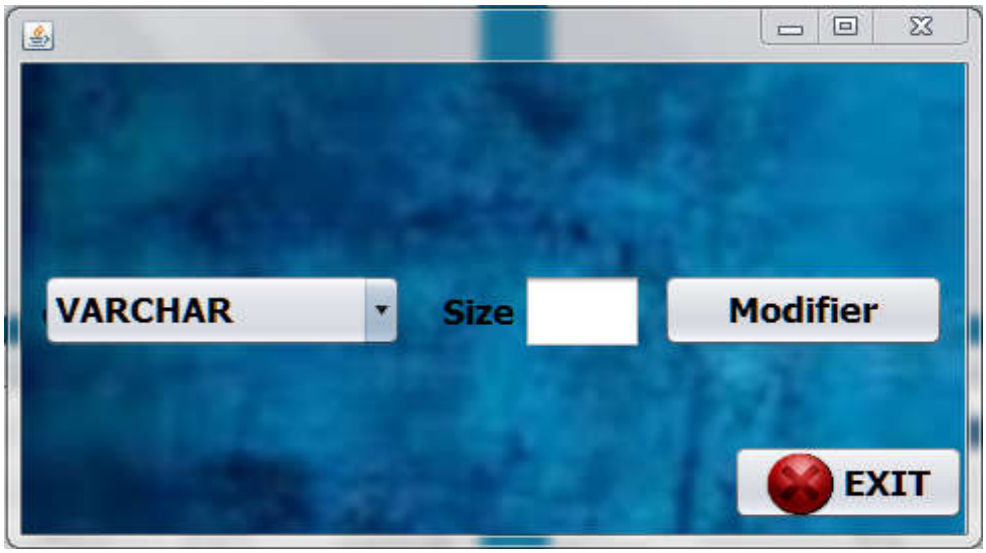


Figure 30: Changement de type attributs

4.5 Interface de transfert des données

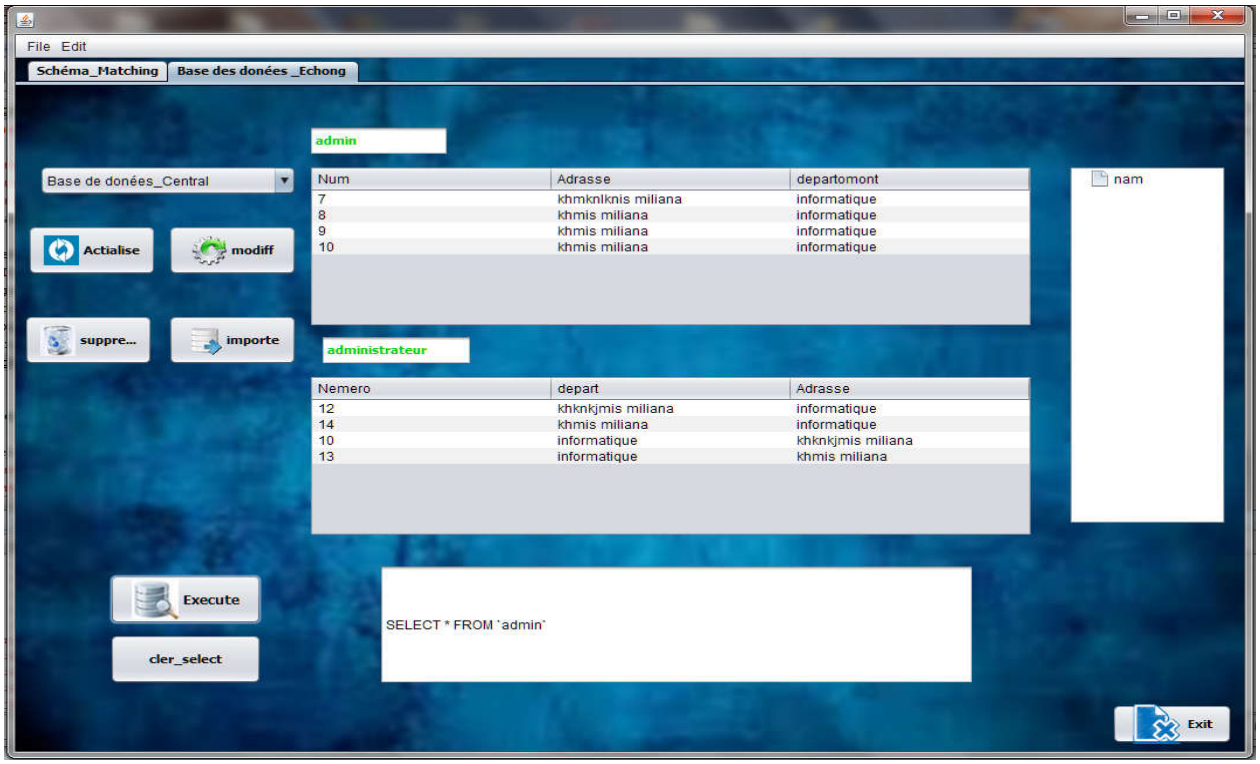


Figure 31: interface utilise pour faire des opérations des Transfer

5. Conclusion :

Nous avons montré d'une manière assez simple et clair les grandes lignes de notre implémentation pour comprendre comment l'application et exécuté. Quelques points restent encore à améliorer (par une API netbeans).

Conclusion Générale

Conclusion générale

Dans ce mémoire nous avant traité la problématique de l'intégration de bases de données hétérogènes, le butde notre travail était de combiner des sources de données hétérogènes pour permettre par la suite un échange de données efficace et sans erreurs.Pour résoudre ce problème, nous avant étudié les principaux approches d'intégrations de base de données, puis nous avant proposé une approche basée sur les dépondances fonctionnelles pour le matching des structures et des élements des deux BDs.

Notre choix des langages Java, SQL, on a utilisé le langage SQL pour le transir les bases de donnée via un graph et pour les interrogé, et pour le langage Java nous les utilise pour réalise notre system d'intégrations.

Pour réaliser notre système on a utilisé les schémas de la base de données pour bien comprendre notre BDD et applique l'approche qui on a proposé.

Notre approche est une approche hybridequi combine plusieurs techniques dans notre étude l'approche hybride se base sur la structure et les éléments(attributs) de la base de données

Pour appliquercette approche on a utilisé deux méthodes, la première méthode se basé sur un dictionnaire pour comparerune paire d'élément de la base de données et la deuxième méthode elle se base sur les dépondons fonctionnelles direct et indirect.

Références bibliographiques

Referencesbibliographiques

[1]Rahm Erhard, Bernstein Philip. A survey of approaches to automatic schema matching. International Journal on Very Large Data Bases, 2001, vol. 10, n°4, pp 334-350.
[2]http://liris.cnrs.fr/Documents/Liris-4535.pdf
[3]https://fr.wikiversity.org/wiki/Pentaho/Int%C3%A9gration_de_donn%C3%A9es
[4] Livre : « Enterprise Service Bus, theory in practice ». David A. Chappell. O'Reilly.

[5] C. Goh, S. Bressan, E. Madnick, and M. D. Siegel. Context interchange: New features and formalisms for the intelligent integration of information. ACM Transactions on Information Systems, 17(3) :270–293, 1999.

[6] A. Y. Levy, A. Rajaraman, and J. J. Ordille. The world wide web as a collection of views: Query processing in the information manifold. Proceedings of the International Workshop on Materialized Views : Techniques and Applications (VIEW'1996*)*, pages 43–55, June 1996.

[7] H. Garcia-Molina, J. Ullman, J. Widom. Database Systems: The Complete Book, Prentice hall, 2002.
[8] Nguyen Xuan D. Intégration de bases de données hétérogènes par articulation à priori d'ontologies : application aux catalogues de composants industriels.' Thèse de doctorat, ENSMA/Université de Poitiers, 2006.
[9] Inmon W.H, John Wiley & Sons. Using the Data Warehouse. 304 p.ISBN:0471059668.

[10] G. Wiederhold, "Mediators in the architecture of future information systems », IEEE Computer, vol.25, n°3, pp 38-49, 1992.
[11]Nguyen Xuan D. Intégration de bases de données hétérogènes par articulation à priori d'ontologies : application aux catalogues de composants industriels.' Thèse de doctorat, ENSMA/Université de Poitiers, 2006
[12] S. S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. D. Ullman, and J. Widom. The tsimmis project : Integration of heterogeneous information sources. Proceedings of the 10th Meeting of the Information Processing Society of Japan, pages 7–18, Mars 1994.

[13] D. Beneventano, S. Bergamaschi, S. Castano, A. Corni, R. Guidetti, G. Malvezzi, M. Melchiori, and M. Vincini. Information integration : The MOMIS project demonstration. In *The* VLDB Journal, pages 611–614, 2000.

[14] T. Gruber. A translation approach to portable ontology specification. Knowledge Acquisition, 5(2) :199–220, 1995.

[15]https://fr.wikipedia.org/wiki/schémas

[16] Rahm, E., Bernstein, P. A. (2001). A survey of approaches to automaticschemamatching, The VLDBJournal 10: 334–350.

[17] Islam, A., Inkpen, D. (2008). Semantictext similarityusing corpus-basedwordsimilarity

and string similarity, ACM Trans. Knowl. Discov. Data. 2, 2, Article 10.

[18]Kim B., Namkoong H., Lee D., Hyun S. J. (2011). A ClusteringBasedSchemaMatching

Scheme for ImprovingMatchingCorrectness of Web Service Interfaces, International

Conference on Services Computing, IEEE.

[19] Partyka, J., Khan, L., Thuraisingham, B. (2009). SemanticSchemaMatchingWithoutShared

Instances, International Conference on Semantic Computing, IEEE.

[20] Do H-H, Rahm E. (2002). COMA - A system for flexible combination of schemamatchingapproaches, Proceedings of the 28th VLDBConference, Hong Kong, China.

[21] Madhavan, J., Bernstein, P. A., Rahm, E. (2001). GenericSchemaMatchingwithCupid. In Apers, P. M. G., Atzeni, P., Ceri, S., Paraboschi, S., Ramamohanarao, K., and Snodgrass, R. T., editors, Proc. of the 27th International Conference on Very Large Data Bases (VLDB 2001), September 11-14, 2001, Roma, Italy, pages 49–58. Morgan Kaufmann.

[22] Gillani S., Naeem, M., Habibullah, R., Qayyum, A., (2013). SemanticSchemaMatchingUsingDbpedia, I.J. Intelligent Systems and Application.

[23] pavel. ontology matching. heidlberg (DE)springer-verlag,2007,341 p.

[24] P.A. Bernstein: A Survey of Approaches to Automatic Schema Matching. VLDB Journal, 10(4), 2001

[25]Shvaiko Pavel, Euzenat Jerome. A Survey of Schema-based Matching approaches. Journal on Data Semantics IV , 2005, vol.3730, pp 146-171

[26]Melnik Sergey, Hector Garcia-Molina and Rahm Erhard. Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching. **In:** 18th international Conference on Data Engineering, 26 February - 1 March 2002, San Jose, CA,USA: IEEE Computer Society, 2002, pp 117-128.

[27]Bernstein Philip, Melnik Sergey, Petropoulos, Michalis, and Quix, Christoph. Industrial-Strength Schema Matching. In ACM SIGMOD Record, 2004, vol. 33, n° 4, pp 38-43.

[28]Lu Jianguo, Wang Shengrui and Wang Ju. An experiment on the Matching and Reuse of XML Schemas. In: Lowe David and Gaedke Martin. Proceedings of the 5th International Conference on Web engineering, 27-29 july 2005, Sydney, Australia. Springer, 2005, p p273-284 (Lecture Notes in Computer Science, 3579).

[29]Rahm Erhard, Do Hong Hai, and Maβmann, Sabine. Matching Large XML Schemas. SIGMOD Record, 2004, vol. 33, n°4, pp 26-31.

[30]Wang Shengrui, Lu Jianguo and Wang Ju. Approximate Common Structures in XML Schema Matching. In: Fan Wenfei, Wu Zhaohui and Yang Jun. Proceedings of the 6th

[31]Yatskevich Mikalai, Giunchiglia Fausto and Avesani Paolo. A Large Scale Dataset for the Evaluation of Matching Systems. In: Posters of the 4th European Semantic Web Conference (ESWC), 2007.

[32] https://fr.wikipedia.org/wiki/Système multi agent

[33] https://fr.wikipedia.org/wiki/dépendance fonctionnelle

[34]https://www.projet-plume.org/fiche/phpmyadmin

[35]https://fr.wikipedia.org/wiki/NetBeans

[36] https://fr.wikipedia.org/wiki/Java_ (langage)