



République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**Université Abderahmane Mira de Béjaïa**

Faculté des Sciences Exactes

Département d'Informatique

*mémoire de Master en informatique*

*Option : Réseaux et systèmes distribués.*

***Thème : Propagation sémantique des requêtes  
dans une fédération de médiateurs***

Réalisé par :

*M<sup>elle</sup>* SACI Kahina

*M<sup>elle</sup>* YMMELI Fazia

Devant le jury composé de :

---

<b>Président</b>	<i>M<sup>r</sup></i> BAADACHE Abderrahmane	U. A/Mira Béjaïa.
<b>Examineur</b>	<i>M<sup>r</sup></i> ABBACHE Bournane	U. A/Mira Béjaïa.
<b>Examineur</b>	<i>M<sup>r</sup></i> DJEBARI Nabil	U. A/Mira Béjaïa.
<b>Promotrice</b>	<i>M<sup>elle</sup></i> GASMI Badrina	U. A/Mira Béjaïa.

---

Béjaïa, 2012

## *Remerciements*

”Ils t’interrogeront sur l’esprit ; dis : ”l’esprit est du domaine exclusif de mon seigneur et vous n’avez reçu de la science que fort peu de choses”.

Coran- le voyage nocturne 8 4(El Israa)

Nous tenons à remercier Le Grand Dieu de nous avoir donné la force et le courage pour poursuivre nos études, veuille t’il guider nos pas dans le droit chemin.

Nous remercions le président du jury et les examinateurs d’avoir accepter de juger notre travail.

Un merci particulier à nos sources de courage, nos parents, pour leur amour, leur sacrifices et leurs patiences.

Un énorme merci à nos amies pour leurs éternel soutien en particulier : Kahina RO, Nassima et ouidad.

## Dédicaces

Je dédie ce travail :

- ★ A mes très chers parents
- ★ A ma grand mère que Dieu la protège
- ★ A mes sœurs et mon frère
- ★ A toute ma famille
- ★ A tous mes proches et mes amis
- ★ A ma binôme " Kahina" et sa famille
- ★ A tous les étudiants de la promotion ReSyD 2012, en particulier : Samia, Meriem, Tinhinene, Khalida, Ghania, Lila et Rachida.

Fazia

## Dédicaces

Je dédie ce travail :

★ Tout d'abord à mes parents ; à ma chère mère qui m'a mis au monde et a veillé à mon bonheur ; à mon père qui a tout donné et sacrifié.

Je leurs éprouve ma profonde gratitude et je leurs promet d'être toujours à leurs cotés. Je vous aime et que Dieu vous garde pour nous tous. Amène.

★ A mes grands parents à qui je souhaite une longue vie.

★ A mes très chers frères que j'admire beaucoup ; Haroun et sa famille, Ali et sa famille.

★ A mes très chères sœurs ; Malika et sa famille, Karima et sa famille.

★ A mes oncles et mes tantes.

★ A mes cousins, et mes cousines.

★ A ma binôme et sa famille.

★ A tous mes amis et à toute ma promotion en particulier :

Samia, Aldjia, rachida, tinhinane, frouze, khalida, Mérième

Kahina

# Table des matières

<b>Table des Figures</b>	<b>vii</b>
0.1 Contexte et motivation . . . . .	2
0.2 Problématique et objectifs . . . . .	2
0.3 Plan du document . . . . .	3
<b>1 La médiation</b>	<b>4</b>
1.1 Système de médiation . . . . .	5
1.2 Evaluation de requête dans un système de médiation . . . . .	6
1.3 Approches de médiation . . . . .	7
1.3.1 Approche centralisée . . . . .	7
1.3.2 Approche distribuée . . . . .	7
1.4 Conclusion . . . . .	8
<b>2 Les systèmes Pair à pair</b>	<b>9</b>
2.1 Définitions . . . . .	10
2.1.1 Pair (Peer) . . . . .	10
2.1.2 Systèmes pair à pair (P2P) . . . . .	10
2.2 Caractéristiques des systèmes P2P . . . . .	10
2.3 Architectures des systèmes P2P . . . . .	12
2.3.1 Architecture centralisée . . . . .	12
2.3.2 Architecture distribuée . . . . .	13

2.3.3	Architecture hybride . . . . .	14
2.4	Classification . . . . .	15
2.4.1	Système P2P structuré . . . . .	15
2.4.2	Système P2P non structuré . . . . .	15
2.5	Routage dans les systèmes P2P . . . . .	16
2.5.1	Routage dan les systèmes P2P structurés . . . . .	16
2.5.2	Routage dans les systèmes P2P non structurés . . . . .	17
2.6	Exemples d’algorithmes de routage . . . . .	18
2.7	Conclusion . . . . .	18
<b>3</b>	<b>Routage sémantique dans les systèmes P2P</b>	<b>19</b>
3.1	Méthodes de routage sémantique . . . . .	20
3.1.1	Présentation de quelques algorithmes de routage sémantique . . . . .	21
3.2	Comparaison entre les algorithmes . . . . .	26
<b>4</b>	<b>Proposition</b>	<b>28</b>
4.1	Notions de base . . . . .	29
4.1.1	Notion de connaissance . . . . .	29
4.1.2	Graphes conceptuels . . . . .	29
4.1.3	Opérations de base sur les graphes conceptuels . . . . .	32
4.2	Notre proposition : Les graphes conceptuels pour le routage sémantique . . . . .	33
4.2.1	la base de connaissance d’un médiateur . . . . .	33
4.2.2	Méthode de routage . . . . .	34
4.2.3	Recherche sémantique de connaissances . . . . .	36
4.2.4	Contribution de la méthode . . . . .	58
4.3	Conclusion . . . . .	59
	<b>Bibliographie</b>	<b>61</b>

# Table des figures

1.1	Architecture générale d'un système de médiation. . . . .	5
2.1	Architecture P2P centralisée. . . . .	12
2.2	Architecture P2P distribuée. . . . .	13
2.3	Architecture P2P hybride. . . . .	14
3.1	Classification des méthodes de routage. . . . .	20
4.1	La représentation graphique d'un concept. . . . .	30
4.2	La représentation graphique d'une relation . . . . .	31
4.3	Base de connaissance d'un médiateur. . . . .	34
4.4	Résumé de la base de connaissances. . . . .	34
4.5	exemple d'architecture d'un réseau possédant trois pairs . . . . .	35
4.6	exemple de résumé global . . . . .	36
4.7	Représentation de la requête. . . . .	36
4.8	Support de concepts. . . . .	41
4.9	Support de relations. . . . .	41
4.10	Base de connaissances du pair1 . . . . .	43
4.11	La base de connaissances du pair P4 . . . . .	46
4.12	Le résumé local généré par le pair P1 avant la simplification . . . . .	47
4.13	Le résumé du pair P1 après simplification . . . . .	48
4.14	Le résumé local du pair P2 . . . . .	49

4.15 Le résumé local du pair P3 . . . . .	50
4.16 Le résumé local du pair P4 . . . . .	50
4.17 Architecture du système . . . . .	51
4.18 Résumé au niveau du pair P1. . . . .	52
4.19 Résumé au niveau du pair P2. . . . .	53
4.20 Résumé global. . . . .	55
4.21 Arbre des résumés. . . . .	55
4.22 La requête posée. . . . .	56
4.23 Résumé de la requête posée. . . . .	56



# Introduction générale

Le nombre croissant de données accessibles via des réseaux joue un rôle important dans la société d'information d'aujourd'hui. Donc, la difficulté d'évaluer l'information désirée est augmentée avec la croissance du volume de l'information disponible, ce qui engendre un manque d'outils suffisants pour aller vers une exploitation maximale de l'information disponible et éliminer le problème de surcharge d'information.

Le domaine de la recherche d'information (RI) est né pour automatiser la recherche d'information RI dans les bibliothèques. La première problématique en RI portait sur la coopération de connaissances hétérogènes qui a été et continue à être très étudiée sous un aspect particulier : l'intégration de sources hétérogènes, coopérant pour répondre à une requête de recherche envoyée par l'utilisateur, chaque source étant en mesure de fournir une partie des réponses ou encore des réponses partielles qui continue à être une problématique essentielle, notamment dans le cadre de la mise en correspondance d'ontologies, du fait du nombre croissant de sources d'informations disponibles via le Web.

Plusieurs approches ont apparu, pour faciliter la recherche d'information souhaitée est d'avoir la réponse aux différentes requêtes posées dont la première est celle du pair à pair (P2P) qui est considérée comme une base de solution très intéressante pour concevoir des communautés virtuelles de partage de données efficaces et capables de passer à l'échelle pour un très grand nombre de participants, permettant à la fois l'autonomie, l'hétérogénéité des participants et un fort contrôle des participants sur les données qu'ils proposent à la communauté. Une deuxième approche consiste à la construction de médiateur. Ce dernier joue le rôle d'interface de requêtes entre un utilisateur et des sources de données. La dernière approche est celle

d'entrepôt de données, elle consiste à la construction d'une base de données réelle en regroupant les informations pertinentes pour les applications considérées. Dans notre travail, on s'intéresse à l'approche pair-à-pair dans le contexte de fédération de médiateurs.

### 0.1 Contexte et motivation

Un système P2P représente un réseau abstrait dont les nœuds (pairs) sont équivalents en fonctionnalités, ils peuvent à la fois demander et fournir des services. Chaque pair établit une connexion avec un ensemble d'autres pairs et il n'y a pas de contrôle central ni hiérarchique de tout le système. Cette décentralisation entraîne un avantage primordial de P2P : l'agrégation des capacités des pairs pour augmenter la performance globale du système. Par ailleurs, un système P2P permet les entrées et sorties dynamiques des pairs. Grâce à ces propriétés, le paradigme P2P présente une bonne solution pour les calculs intensifs et à grande échelle. Ils sont utilisés dans plusieurs domaines comme : le partage de fichiers, le partage des capacités de calcul et l'échange de messages instantanés. Grâce à leurs caractéristiques avantageuses, de nouveaux domaines se dirigent vers l'utilisation de ces systèmes dans de nouvelles applications.

### 0.2 Problématique et objectifs

Dans ce mémoire, nous nous intéressons au problème de routage sémantique des requêtes dans une fédération de médiateurs. Notre objectif est de trouver une méthode pour localiser les pairs (médiateurs) pertinents pour répondre à une requête donnée dans le but de minimiser le nombre de messages échangés lors de la satisfaction de la requête. Pour répondre à cet objectif, nous nous sommes retrouvés face à plusieurs questions auxquelles il faut répondre :

- Quel langage utilisé pour la représentation des connaissances ?
- Comment localiser les pairs les plus pertinents à une requête donnée ?
- Quel algorithme de propagation pourra être efficace pour le routage des requêtes posées ?

Pour cela, nous avons proposé une méthode de routage sémantique basée essentiellement sur le contenu en utilisant le formalisme des graphes conceptuels.

## 0.3 Plan du document

Le reste de notre document est organisé en quatre chapitres dont les trois premiers sont consacré à l'état de l'art :

Dans le premier chapitre, nous introduisons la notion de médiateurs et leur fédération.

Dans le deuxième chapitre, nous présentons l'essentiel des systèmes pair à pair (P2P).

Dans le troisième chapitre, nous nous focalisons sur le routage sémantique des requêtes dans les systèmes P2P en citant ses différentes approches ainsi que quelques algorithmes de routage.

Le quatrième chapitre est consacré à notre proposition, dans ce dernier nous donnons une brève introduction au formalisme des GCs ainsi que l'utilisation que nous faisons de ce formalisme pour résoudre notre problème à savoir ; le routage sémantique des requêtes dans une fédération de médiateurs.

Nous concluons ce papier en rappelant les contributions de notre travail ainsi, nous proposons quelques perspectives qui peuvent être liées à l'amélioration de notre proposition.

# 1

## La médiation

### Introduction

Le développement des réseaux et la masse énorme des informations disponibles sur ces réseaux étaient derrière la naissance de la médiation durant les années quatre-vingts, cette approche de médiation était proposée comme une solution pour donner accès à des nombreuses sources d'informations qui peuvent être distribuées et hétérogènes. Alors, la médiation consiste à rendre le plus transparent possible l'échange requête-réponse avec le système médiateur, et donc elle présente l'intérêt de pouvoir construire un système d'interrogation de sources de données sans toucher aux données qui restent stockées dans leurs sources d'origine.

## 1.1 Système de médiation

La notion de médiateur a été initialement proposée par Wiederhold Gio [GIO]. Il définit un médiateur comme suit : "A mediator is a software module that exploits encoded knowledge about some sets or Subsets of data to create information for a higher layer of applications".

Dans un contexte de médiation, les ressources pré-existantes, avec leurs données mais aussi leurs propres métadonnées, et le but d'un médiateur est de réconcilier certaines de ces métadonnées de façon à donner l'illusion à un utilisateur qu'il interroge un système d'information homogène [LIBOUREL]. La médiation consiste à rendre le plus transparent possible l'échange requête-réponse avec le système médiateur. L'utilisateur n'a qu'à se préoccuper de formuler la requête de son choix (navigation ou interrogation), le système doit lui retourner un résultat intelligible. La médiation présente donc l'intérêt de pouvoir construire un système d'interrogation de sources de données sans toucher aux données qui restent stockées dans leurs sources d'origine [RAYNAUD].

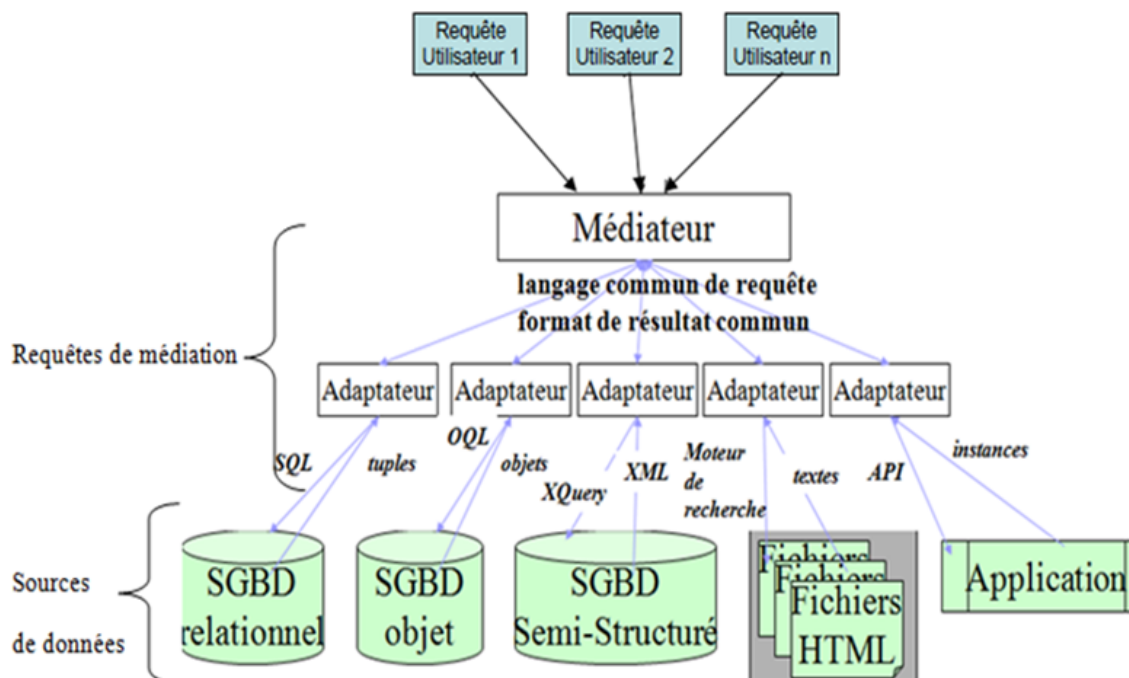


FIGURE 1.1 – Architecture générale d'un système de médiation.

### 1.1.1 Les composants d'un système de médiation :

les composants essentiels d'un système de médiation sont : le schéma global appelé schéma de médiation, les mappings du schéma global avec les sources, les fonctions de réécriture de requêtes et les fonctions de composition des résultats [KOSTADINOV]. Les mappings du schéma global avec les sources sont des requêtes, appelées requêtes de médiation.

## 1.2 Evaluation de requête dans un système de médiation

Le médiateur présente des vues intégrées des sources de données. Ainsi une requête formulée au médiateur est posée indépendamment des localisations des différentes données intervenant pour calculer le résultat. Cela introduit trois difficultés [GARDARIN] :

- la décomposition d'une requête : il s'agit à partir d'une requête posée sur une vue intégrée, de localiser les données intervenant dans sa résolution, de produire des sous requêtes spécifiques à chacune des sources, d'ordonner ces sous requêtes et éventuellement d'introduire des opérateurs au niveau du composant de médiation afin de compléter cet ensemble de sous requêtes. La localisation des sous requêtes nécessite des structures spécifiques de gestion de métadonnées.
- la recomposition des résultats : une fois les sous-requêtes soumises à chacune des sources, il s'agit de savoir recomposer les différents résultats entre eux. Les résultats de chacune des sous requêtes peuvent éventuellement faire l'objet d'un traitement additionnel soit parce que l'évaluateur de sous requête n'a pas la capacité nécessaire pour traiter entièrement cette dernière, soit parce que les sous requêtes comportent des dépendances entre elles.
- l'optimisation du traitement : Le médiateur a rarement une vision sur la façon dont les sous requêtes sont traitées au niveau des sources (placement des données, type de stockage, indexation, stratégie d'évaluation).

L'interrogation effective des sources se fait par des adaptateurs (ou " wrappers ") qui constituent une interface d'accès aux différentes sources. Ces adaptateurs traduisent les sous requêtes exprimées dans le langage de requête spécifique de chaque source. Les résultats sont ensuite renvoyés au médiateur qui se charge de les intégrer avant de les présenter à l'utilisateur.

## 1.3 Approches de médiation

On distingue deux approches de médiation : approche centralisée et approche décentralisée.

### 1.3.1 Approche centralisée

Un médiateur est spécifique à un domaine d'application donné. Il comprend un schéma global, ou ontologie, dont le rôle est central. Il fournit un vocabulaire structuré servant de support à l'expression des requêtes. Par ailleurs, il établit une connexion entre les différentes sources accessibles. Dans l'approche de médiation centralisée, l'intégration d'information est fondée sur l'exploitation de vues abstraites décrivant de façon homogène et uniforme le contenu des sources d'information dans les termes de l'ontologie. Ces vues sont définies par des modules spéciaux appelés des adaptateurs. Les sources d'informations pertinentes pour répondre à une requête, sont calculées par réécriture de la requête en termes de vues [GASMI ].

### 1.3.2 Approche distribuée

Pour les nouvelles applications à titre d'exemple l'échange de fichiers multimédia (musique, vidéos, etc.) entre internautes, l'approche centralisée n'est pas adaptée car l'accès à un tel volume de données offre des performances trop faibles [ROUSSET]. Les médiateurs peuvent intégrer efficacement les données de quelques grandes entreprises, mais pas celles de millions d'utilisateurs. Pour ces besoins nouveaux, une médiation distribuée est en passe d'émerger avec l'apparition des systèmes pair-à-pair. Pour permettre le partage de données hétérogènes distribuées à large échelle, la construction d'un réseau pair-à-pair de médiateurs interopérables est nécessaire. Les médiateurs sont une solution pour traiter l'hétérogénéité des données et le Pair-à-pair est celle pour traiter la large échelle.

#### 1.3.2.1. Définition de la fédération de médiateurs

La fédération de médiateurs désigne un groupe de médiateur qui accepte d'être considérés comme une entité unique par d'autre médiateur quand ceux-ci demandent un service à la fédération. En d'autres termes un importateur peut envoyer une demande à un médiateur de la

fédération et celle-ci peut être transmise à n'importe quel ou à chaque médiateur de la fédération [DONG].

### 1.3.2.2 Propagation de requête dans une fédération de médiateur

Une fédération est vue comme un système distribué dynamique dans lequel un médiateur peut coopérer avec des médiateurs différents pour des requêtes différentes. Ainsi, un médiateur peut propager une requête ou une partie de la requête qui n'est pas satisfaite localement à un autre (ou plusieurs) médiateur (s) en se fondant sur deux critères [DONG] :

1. Critère sémantique : en confrontant la requête au contexte qu'il gère, le médiateur identifie les médiateurs les plus appropriés pour satisfaire la requête.
2. Critère de proximité : la requête est propagée vers des médiateurs déjà identifiés. Ces deux types de critères peuvent être combinés dans un seul algorithme de propagation (algorithme hybride).

## 1.4 Conclusion

Dans ce chapitre, nous avons présenté l'une des principales technologies de partage de données et gestion de données distribuées qui est la médiation. Cette dernière doit servir d'intermédiaire de coopération non seulement entre les utilisateurs et le système, mais aussi réciproquement. Nous avons pu donner les différentes approches de médiation, l'approche centralisée et l'approche distribué.



# 2

## Les systèmes Pair à pair

### Introduction

Depuis quelques années, le paradigme P2P<sup>1</sup> attire, de plus en plus, l'attention de plusieurs chercheurs en Informatique. Ce paradigme est apparu comme une solution aux quelques limites du paradigme Client/ Serveur. En évitant des goulets d'étranglement susceptibles et en offrant une bonne tolérance aux pannes, le paradigme P2P est approprié aux environnements distribués à grande échelle dans lesquels les nœuds peuvent partager leurs ressources d'une façon autonome et décentralisée.

Dans ce chapitre, nous introduisons brièvement les systèmes pair-à-pair. Nous présentons quelques définitions qui sont utiles pour la compréhension des systèmes P2P, puis, nous citons certains caractéristiques spécifiques au domaine pair à pair. Par la suite, nous passons à la

---

1. peer to peer

classification des systèmes P2P.

## 2.1 Définitions

### 2.1.1 Pair (Peer)

Un pair est un nœud sur un réseau, il peut communiquer avec n'importe quel pair du réseau et il est caractérisé par un identifiant (PID : Peer Identifier). Son rôle est à la fois [BAKHTOUCI] :

**Client** : une partie du pair permet de demander des services au réseau P2P.

**Serveur** : une partie du pair permet d'offrir des services au réseau P2P.

**Routeur** : une partie du pair permet de router les demandes de service des utilisateurs.

### 2.1.2 Systèmes pair à pair (P2P)

Ces systèmes sont définis comme des systèmes répartis consistants des nœuds interconnectés l'un aux autres et capables de s'organiser selon des topologies de réseau afin de partager leurs ressources (contenu, cycles de CPU, stockage et bande passante). Ils sont tolérants aux pannes de leurs nœuds et capables d'accueillir des informations concernant d'autres nœuds, tout en maintenant une connectivité et une performance acceptable, sans exiger l'intermediation d'un serveur global ou l'utilisation d'une autorité centralisée [THEOTOKIS].

En général, un système P2P est composé d'un protocole permettant la communication entre les pairs, des algorithmes trouvant les ressources et les applications se situent au dessus de l'environnement distribué. Grâce à l'échange direct entre les pairs, la technologie P2P permet un partage optimal des ressources et des services tels que les informations, les fichiers, les processus de traitement et les espaces de stockage [ISMAIL].

## 2.2 Caractéristiques des systèmes P2P

Les systèmes P2P possèdent des caractéristiques avantageuses par rapport aux autres systèmes basés sur le paradigme Client/ Serveur. Chaque système P2P essaie de posséder un grand

nombre de ces caractéristiques [AL KING].

- **Grande échelle** : Il s'agit de faire coopérer un grand nombre de nœuds (jusqu'à des milliers ou des millions) pour partager leurs ressources tout en maintenant une bonne performance des systèmes, cela signifie qu'un système P2P doit offrir des méthodes bien adaptées avec un environnement dans lequel il y a un grand volume de données à partager, un nombre important de messages à échanger entre un grand nombre de nœuds partageant leurs ressources via un réseau largement distribué.
- **Autonomie des nœuds** : Chaque nœud gère ses ressources d'une façon autonome. Il décide quelle partie de ses données à partager. Il peut se connecter ou/et se déconnecter à n'importe quel moment. un nœud possède également l'autonomie de gérer sa puissance de calcul et sa capacité de stockage.
- **Auto-Organisation** : Les systèmes P2P ne nécessitent pas l'administration et la maintenance nécessaire est prise en compte par les algorithmes sous-jacents. Ces réseaux autorisent l'arrivée de nouveaux pairs ou le départ de pairs courants tout en adaptant les connections réseaux en conséquence.
- **Communication Symétrique** : Par opposition au modèle client/serveur dans lequel chaque nœud joue le rôle de client ou de serveur, dans un réseau P2P chaque nœud peut se comporter à la fois comme client et serveur.
- **Contrôle Distribué** : Les systèmes P2P ne disposent pas de point de contrôle central. Le contrôle est distribué entre les pairs de telle sorte qu'aucun pair ne devient un goulot d'étranglement. En cas d'indisponibilité d'un pair, sa part de contrôle est prise en charge par les autres nœuds.
- **Hétérogénéité** : A cause de l'autonomie de nœuds possédants des architectures matérielles et/ou logicielles hétérogènes, les systèmes P2P doivent posséder des techniques convenables pour résoudre les problèmes liés à l'hétérogénéité de ressources.

## 2.3 Architectures des systèmes P2P

Les systèmes P2P peuvent être classés en trois grandes familles que nous présentons dans cette partie selon le niveau de centralisation qui se traduit par une distribution des tâches à accomplir plus ou moins importante. Nous décrivons ainsi le modèle P2P centralisé, distribué et hybride [KTARI].

### 2.3.1 Architecture centralisée

Le modèle centralisé est à la limite du modèle P2P, car il repose sur un serveur dédié qui centralise et maintient l'ensemble des connaissances des pairs, les ressources étant toujours hébergées sur les pairs. Dans ce modèle, le serveur est une entité de localisation et ne contient aucune ressource. Seule la récupération d'objets est décentralisée [KTARI].

L'intérêt de ce modèle est de permettre une recherche exhaustive, une localisation rapide des ressources et une gestion simple grâce à l'entité centrale. L'inconvénient majeur est que le fonctionnement du système repose uniquement sur cette entité car elle doit être capable de supporter un grand nombre de requêtes et d'effectuer beaucoup de recherches [KTARI].

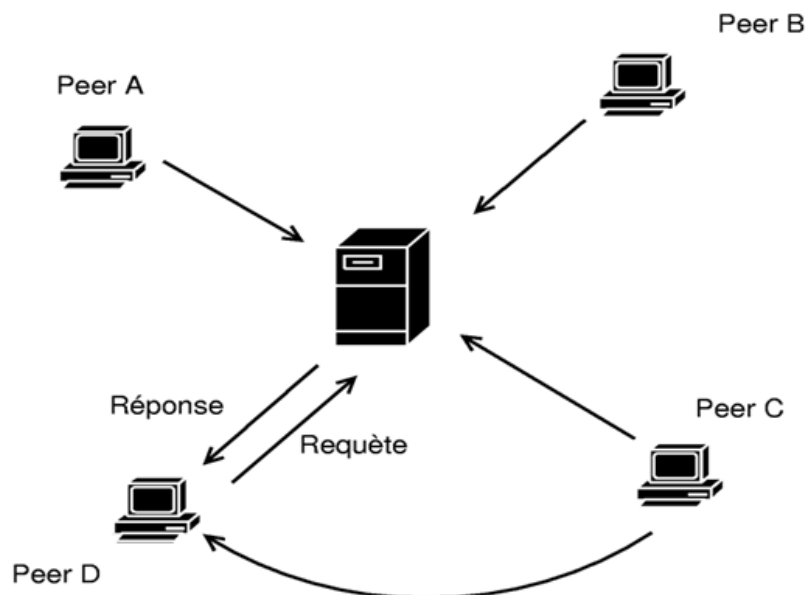


FIGURE 2.1 – Architecture P2P centralisée.

### 2.3.2 Architecture distribuée

Les systèmes P2P distribués essayent de répartir la totalité des fonctions du système entre les pairs à savoir : la recherche, le routage et la récupération des objets dans le réseau logique. Les pairs doivent s'organiser pour former une architecture dynamique efficace les connectant entre eux [KTARI]. Contrairement aux systèmes centralisés, où il suffisait de se connecter au serveur pour avoir accès aux informations, pour avoir accès à une information en décentralisé il faut :

1. Apprendre la topologie du réseau sur lequel le client est connecté.
2. Rechercher l'information sur tous les nœuds.
3. Recevoir une réponse d'un nœud répondant aux critères [BUDAN].

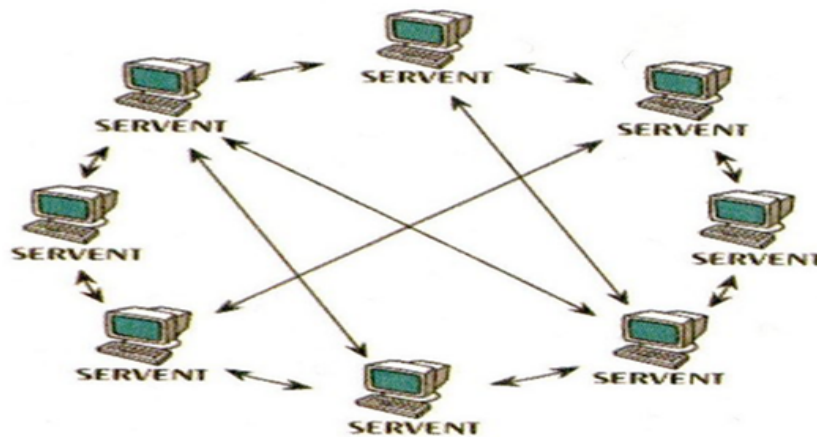


FIGURE 2.2 – Architecture P2P distribuée.

### 2.3.3 Architecture hybride

Le système hybride combine à la fois le centralisé et le distribué [CARON]. Ce modèle repose sur des interconnexions des super nœuds sur le niveau haut de la hiérarchie, suivant le modèle distribué. Chaque nœud feuille se rattache à un ou plusieurs super nœuds. Un super nœud gère un ensemble de nœuds feuilles. Les objets partagés par une feuille sont enregistrés sur le super nœud responsable de cette feuille. Lorsqu'une feuille recherche un objet, elle envoie sa requête à son super nœud. Celui-ci effectue alors la recherche parmi les objets des nœuds qui lui sont rattachés, voire les super nœuds voisins si besoin [KTARI].

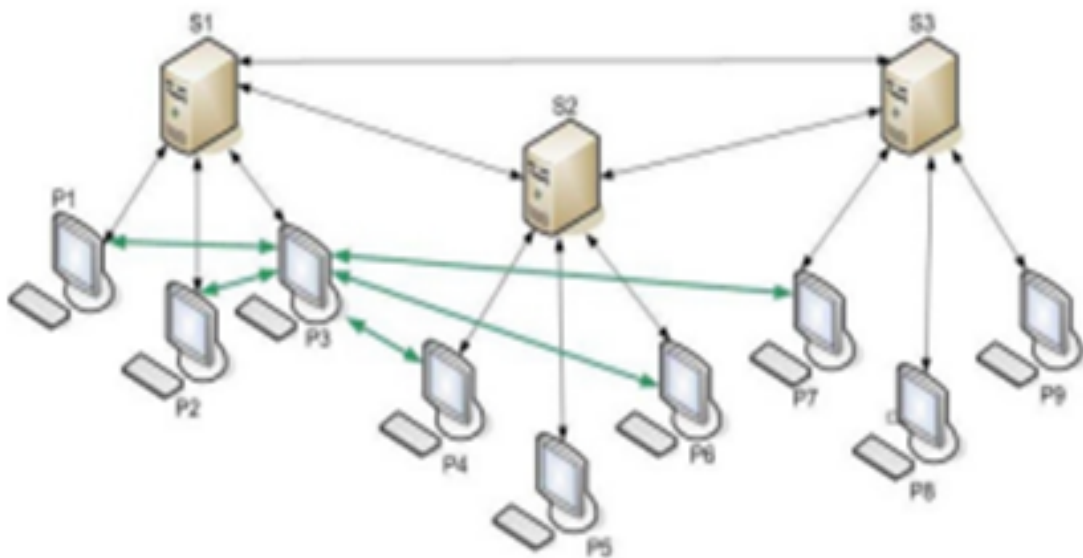


FIGURE 2.3 – Architecture P2P hybride.

## 2.4 Classification

Selon le degré de structuration, les systèmes P2P se répartissent en deux catégories ; les systèmes pair à pair structurés et les systèmes pair à pair non structurés [LE].

### 2.4.1 Système P2P structuré

Ils sont dis structurés, car au dessus du réseau physique sous-jacent les nœuds sont reliés par un réseau recouvrant construit sous certaines contraintes, répondant à plusieurs propriétés et connectant les pairs selon une structure particulière donnée (Anneau Chord, espace de coordonnées cartésiennes CAN, hypercube. . .). La gestion des ressources quant à elle s'appuie souvent sur l'utilisation d'une table de hachage distribuée. Le principe des tables de hachage distribuées consiste à utiliser une fonction de hachage pour faire correspondre à chaque ressource une valeur de hachage sous forme d'une chaîne de bits de longueur fixée.

### 2.4.2 Système P2P non structuré

Les systèmes pair-à-pair non structurés constituent une catégorie importante dans le domaine du P2P. Ils admettent que tous les pairs soient égaux, il n'ya pas de pairs spéciaux pour organiser le système. Dans ce type de système, l'emplacement des fichiers est indépendant de la structure du réseau, il n'ya pas une structure explicite à maintenir [AMAD].

Cette catégorie est caractérisée par une grande flexibilité, un pair peut rejoindre ou quitter le système sans influencer sur sa structure globale. Cependant, le manque de structure explicite rend le processus de la recherche et de la localisation plus complexe [DJAGHLOUL].

Dans ce système, les messages échangés sont transmis en utilisant une connexion directe entre deux pairs et dans le cas contraire, le message parcourt un chemin et passe le long des autres pairs qui jouent le rôle des ponts.

La recherche des services dans un tel réseau se fait généralement selon la technique d'inondation " flooding ou épidémique ".

## 2.5 Routage dans les systèmes P2P

Le routage est une méthode d'acheminement des informations à la bonne destination à travers un réseau de connexion donné. Le problème de routage consiste à déterminer un acheminement optimal des paquets à travers le réseau au sens d'un certain critère de performance [LEMLOUMA]. Et donc le routage de requêtes a la responsabilité de trouver les pairs pertinents pour satisfaire une requête. Une fois que cette dernière est envoyée aux pairs pertinents, elle est exécutée sur ces pairs et les réponses sont renvoyées au pair initiateur de cette requête [DEDZOE]. Les méthodes de routage des requêtes dépendent beaucoup des topologies des systèmes P2P. nous consacrons la suite pour détailler les méthodes de routage dans les différentes structures des systèmes P2P.

### 2.5.1 Routage dan les systèmes P2P structurés

Les systèmes structurés prennent très souvent la forme de tables de hachage distribuées (distributed hash table : DHT). Ces tables permettent d'associer une clé à tout objet à indexer dans le système P2P. Les pairs sont responsables de données qui leurs sont attribuées suivant leurs position dans le réseau [VENTRESQUE]. Dans une DHT, les pairs et les objets sont identifiés dans un même espace de nommage. Chaque nœud est responsable d'une partie des clés dans le système et chaque objet est stocké dans le nœud dont l'identifiant est le plus proche de sa clé selon la métrique de distance utilisée.

- Lorsqu'un nœud met à disposition un objet, il doit d'abord l'annoncer/ publier au système. La publication consiste à calculer la clé associée à l'objet, puis à envoyer un message au nœud responsable de la clé à travers le réseau logique [ABDENNADHER].
- Lorsqu'un nœud cherche un objet associé à une clé, il cherche directement la clé correspondante. Quand le nœud responsable est trouvé, ce dernier répond au nœud source et envoie les éventuelles informations relatives à l'objet.
- Pour assurer le routage, chaque nœud dans la DHT maintient régulièrement les adresses de quelques nœuds voisins, choisis selon la structure. Lors de l'envoi d'un message dans le réseau, cette structure permet de se rapprocher du destinataire à chaque saut.



Le routage dans une DHT permet à chaque nœud recevant un message de décider localement à quel voisin faire suivre le message. Le routage des requêtes est très efficace : trouver le pair responsable d'un objet se fait en  $O(\log n)$  décisions de routage,  $n$  étant le nombre de pairs dans le réseau. Mais les requêtes sont généralement limitées à la recherche par clé [CERQUEUS], [LE].

### 2.5.2 Routage dans les systèmes P2P non structurés

Le routage des requêtes dans les systèmes pair à pair peut être classé en deux catégories [YEFERNY] : les systèmes de routage par inondation et les systèmes de routage utilisant la sémantique. Dans cette section, nous intéressons au routage par inondation. Le routage sémantique sera présenté en détails dans le chapitre suivant.

#### ✓ 2.5.2.1 Routage par inondation :

Ce type de routage est utilisé sur les réseaux pair-à-pair non-structuré. Un nœud désirant localiser une ressource demande à ses voisins s'ils connaissent cette ressource. À leur tour, ces voisins demandent à leurs voisins si ils ont connaissance de cette ressource et ce, jusqu' à une profondeur fixée par le système [BUSNEL]. Lorsque la ressource est trouvée, son adresse est propagée le long du chemin inverse. La méthode par inondation utilise certains dispositifs tels que :

- champ de vie de la requête TTL (Time To Live) [BUSNEL] : qui indique le nombre de fois que l'on peut retransmettre le message et il est décrémenté à chaque passage par un nœud.
- mémorisation des requêtes [BUSNEL] : une mémorisation des dernières requêtes acheminées par chacun des nœuds permettent d'éviter une circulation éternelle des requêtes.

Cette méthode possède des avantages ainsi que des inconvénients, les avantages nous pouvons les résumer dans (i) l'administration simple où l'ajout et la suppression des nœuds est facile au sein du réseau, (ii) la sûreté d'où on assure que la requête soit reçue par tous les pairs, (iii) cette méthode ne nécessite pas des mises à jour en ce qui concerne le contenu des pairs. Et pour les inconvénients nous pouvons les résumer dans (i) la Surcharge du réseau qui entraîne une latence élevée pour la satisfaction des requêtes, (ii) le manque de fiabilité, (iii) la complexité exponentielle de la recherche.

## 2.6 Exemples d'algorithmes de routage

1/**BFS** : Dans sa version de base, la recherche par parcours en profondeur (breath-first search, BFS) inonde le réseau P2P à une distance TTL (Time-To-Live) de l'initiateur de la requête. Chaque pair recevant la requête décroît la valeur du TTL, exécute la requête et la transmet à tous ses voisins (si  $TTL > 1$ ). Cela permet donc d'atteindre tous les pairs situés à une distance TTL de l'initiateur [CERQUEUS].

2/**Random walks** : L'algorithme de marche au hasard (random walks) génère  $k$  "marcheurs" qui partent de différents voisins du pair initiateur de la recherche et parcourent le réseau. A chaque fois qu'un marcheur atteint un pair, ce dernier exécute la requête et interroge le pair initiateur pour savoir si la condition de terminaison est remplie. Si elle n'est pas remplie, le marcheur se dirige vers l'un des voisins du pair courant. Il est aussi possible de n'interroger que régulièrement le pair d'origine. L'efficacité de cet algorithme est assez bonne au sens où le nombre total de messages ( $k * TTL$  dans le pire des cas) n'est pas dépendant de la topologie[CERQUEUS].

L'avantage majeur de cet algorithme est la réduction significative des messages qu'il réalise. Cependant, ses principaux inconvénients sont[DEDZOE] :

- sa grande variation de performance.
- ne garantit pas l'obtention de toutes les réponses recherchées même si elles existent.
- ne permet aucun apprentissage du réseau.

## 2.7 Conclusion

Les systèmes P2P représentent un domaine de recherche très actif grâce à leurs actuelles utilisations ainsi leurs futures applications.

Dans ce chapitre, nous avons introduit les concepts des systèmes pair-à-pair, représenté leurs caractéristiques principales, leurs classifications en terme d'architecture puis en terme de structurations. Ensuite, nous avons passé à un point très important qui est le routage en le spécifiant dans les différents types de ce système, où nous avons remarqué l'importance d'introduire la sémantique dans le processus de routage qui sera le point à traiter dans le chapitre suivant.

# 3

## Routage sémantique dans les systèmes P2P

### Introduction

Les systèmes utilisant l'approche par inondation, souffrent d'efficacité et d'évolutivité limitée. Les stratégies de routage intelligentes sont essentielles dans des tels arrangements, de sorte que les requêtes deviennent seulement routées à un sous-ensemble de pairs choisi capables de répondre aux requêtes entières (ou à des parties des requêtes). Le problème de routage efficace des requêtes a été étudié par de nombreux chercheurs, puisque sa solution affecte le mécanisme global de recherche du réseau P2P. D'ailleurs, afin de gagner une récupération rapide des données sans une grande consommation de la bande passante, des moyens efficaces pour le traitement et le routage des requêtes vers un ensemble de pairs bien choisi est nécessaire.

Dans ce qui suit, nous allons voir les différentes méthodes de routage avec exemple d'algorithme pour chaque approche ainsi qu'une comparaison entre ces algorithmes.

### 3.1 Méthodes de routage sémantique

Bien qu'étant encore au stade de la recherche, la sémantique dans le processus de routage dans les systèmes P2P nous semblait intéressante pour avoir des réponses exactes à nos questions. Il ne s'agit plus ici de s'intéresser à une architecture précise, puisque la notion de P2P sémantique peut s'appliquer à toutes les architectures. Le principe est d'ajouter de l'information dynamique aux tables de routage. L'information ajoutée peut être sur le contenu des nœuds, sur les requêtes ou sur les utilisateurs comme elle peut même être un mélange des trois [DOUCET]. Cette information peut être généraliste, ou bien au contraire très spécifique, en fonction des besoins.

Dans la littérature, plusieurs travaux de recherche ont essayé d'améliorer la méthode classique de routage des requêtes, qui les propage à un ensemble de pairs choisi d'une manière aléatoire, en introduisant la sémantique dans le processus de propagation de requêtes d'une manière à avoir un routage guidé. [TEMPICH], [DEFUDE]. La méthode de routage sémantique possède plusieurs approches bien citées dans la figure 3.1 :

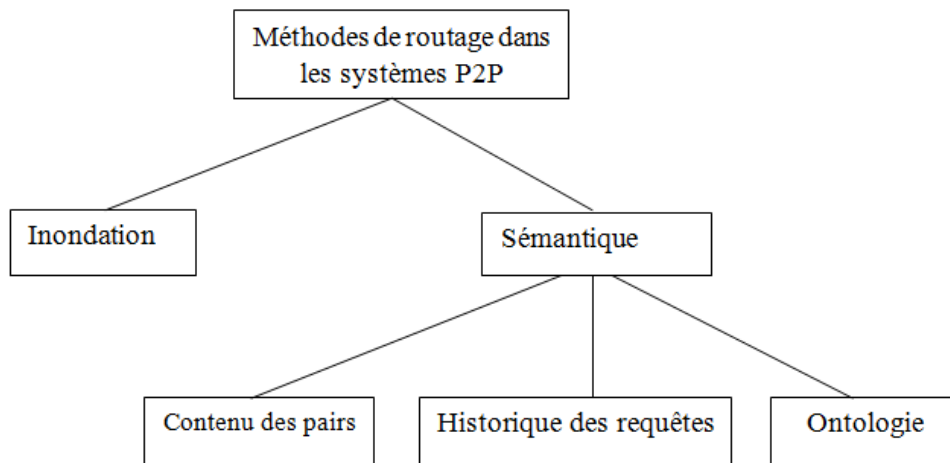


FIGURE 3.1 – Classification des méthodes de routage.

1. **Par contenu** : Une première approche de cette méthode est appelée système de routage par contenu, où le contenu de chaque pair est résumé dans une étiquette de contenu et un routeur fait suivre les requêtes de recherche selon les méta-informations des pairs qui se sont

enregistrées chez lui.

**2.Par Historique** : La deuxième approche est celle basée sur l'historique des requêtes déjà traitées. Dans cette approche, les pairs sauvegardent des informations sur les requêtes déjà passées pour faciliter et réduire le temps de réponse une fois cette requête est répétée. Dans [RAGHAVAN] les auteurs représentent un profil de l'historique de recherche servant comme une base de données des requêtes soumises précédemment ainsi que leurs résultats associés.

**3.Par Ontologie** : Une autre approche de cette méthode est celle appelée système de routage par ontologie. Une ontologie est un ensemble structuré permettant de donner un sens aux informations. Elle peut être vue comme un modèle conceptuel d'un domaine particulier, qui décrit les concepts de ce domaine et les relations entre ces concepts.

L'ontologie est généralement considérée comme une base de connaissances. Les ontologies permettent la modélisation d'informations, elles jouent un rôle de référence pour décrire la sémantique des informations à partager.

L'utilisation d'ontologies permet la représentation formelle des connaissances à l'aide des modèles basés sur des logiques (logiques de représentation, logique de description)[DOBSON].

### 3.1.1 Présentation de quelques algorithmes de routage sémantique

**1/ Psearch[TANG]** : C'est une méthode basée sur le contenu des pairs. Son principe est de distribuer les indices des documents décrivant le contenu des pairs selon l'implémentation *LSI* (*Latent Semantic Indexing*)<sup>1</sup> du modèle vectoriel à travers le réseau P2P. Elle propose que les pairs stockent les documents en gardant leur autonomie de stockage, mais l'indexation de ces documents est stockée dans une table de hachage distribuée. Le but de cette indexation est de permettre de calculer quels documents (resp ; chapitres, pairs...) qui peuvent répondre à une requête donnée.

Les requêtes sont propagées récursivement au voisin dont le contenu du vecteur sémantique est le plus proche de celui de la requête, jusqu'à arriver au pair qui maintient la portion d'espace à laquelle appartient la requête.

---

1. est une technique qui tend à implanter partiellement la recherche sémantique ou orienter les concepts

**2/ INGA**<sup>2</sup> [LOSER] : C'est une méthode hybride ; utilise le contenu des pairs ainsi l'historique des requêtes déjà passée , elle propose une stratégie de sélection de raccourcis en mesure d'identifier leurs pairs du groupe de manière efficace. Un raccourci est un lien unidirectionnel entre deux pairs. Un pair peut être lié à n'importe quel pair distant tant qu'il connaît son *PID* (identifiant). INGA distingue deux types de raccourcis :

1. Raccourci du fournisseur de contenu : il est crée et mis à jour à chaque fois qu'un pair reçoit une réponse à sa requête d'un pair à distance.
2. Raccourci de recommandation : est créé à la reception de la requête.

INGA analyse les raccourcis stockés localement pour déterminer les pairs distants à qui envoyer la requête. Si aucun raccourci ne correspond exactement à la requête, INGA procède au calcul de la similarité entre les raccourcis et la requête.

**3/ LPS**<sup>3</sup> [YEFERNY] : Présente un algorithme de routage par apprentissage des requêtes basé sur le profil<sup>4</sup> des utilisateurs (comportement implicite) qui est déduit à partir d'un historique de requêtes stockées dans un fichier *log*.

Pour le routage des requêtes, les auteurs se sont basés sur deux contextes qui sont des projections sur le fichier *log*. La première projection représente le lien entre les requêtes passées et les termes associés. La seconde représente le lien entre les requêtes passées et les pairs positifs, et donc les résultats de ces projections sont respectivement les termes des requêtes et les pairs ayant répondu à ces requêtes. Un algorithme de génération de *concepts formels*<sup>5</sup> est par la suite appliqué pour générer deux ensembles de concepts, le premier ensemble aura la forme (ensemble d'identifiants de requêtes, ensemble des termes) et le deuxième (ensemble d'identifiants de requêtes, ensemble des pairs), ces deux ensembles là sont utilisés pour la construction d'une base de connaissances pour l'algorithme de sélection des pairs.

---

2. Interested-Base Node Grouping Algorithm

3. LearningPeersSelection

4. représente des corrélations sémantiques entre les requêtes passées et les pairs positifs construit selon la méthode des concepts formels

5. Définit une relation binaire reliant les objets à leurs attributs

4/ **REMINDIN**<sup>6</sup> [TEMPICH] : C'est une méthode basée sur l'historique, les auteurs proposent un système pair-à-pair qui exploite différentes descriptions sémantique représentées comme des déclarations RDF<sup>7</sup> et organisées d'une manière rendant les connaissances locales suffisantes pour localiser les pairs pertinents, en effet, cette méthode est inspirée des métaphores sociales qui permet d'améliorer la recherche d'information et ainsi les réponses aux requêtes.

Cette méthode consiste à observer les pairs ayant répondu avec succès aux requêtes, à mémoriser cette information va être utilisée par la suite pour sélectionner les pairs pertinents pour des demandes futures. REMINDIN maintient une table de routage au niveau de chaque pair.

5/ **ExSI2D** [VENTRESQUE] : C'est une méthode basée sur l'historique des requêtes ainsi l'ontologie. Elle utilise une expansion particulière, appelée expansion structurante sous forme des DSE<sup>8</sup> pour expliquer les concepts prenants part à une requête émise par un participant. Cela lui permet de préciser les dimensions de sa requête sans modifier la requête elle-même. EXSI2D offre aussi la possibilité au fournisseur d'information d'interpréter l'expansion structurante dans sa propre ontologie.

pour le routage, les auteurs proposent une expansion de requête où chaque propagation de concept est mémorisée dans des vecteurs séparés : des dimensions sémantiquement enrichies. L'ensemble de ces vecteurs constitue l'expansion structurante de requête. La requête initiale n'étant pas modifiée, elle peut être utilisée lors du processus de classement par pertinence des documents. Pour le routage des requêtes cette méthode utilise la fonction de similarité entre des concepts appartenant à une ontologie.

6/ **Méthode de reformulation**[AZZAZ] : C'est une méthode hybride basée sur l'historique des requêtes ainsi sur l'ontologie, cette méthode permet d'éviter la redondance des chemins empruntés par une requête et de ne la faire propager que vers les pairs pertinents. Les auteurs supposent que les pairs partagent la même ontologie, et maintiennent une expertise et

---

6. Routing Enabled by Memorizing INformation about Distributed INformation

7. est un modèle, associé à une syntaxe, dont le but est de permettre à une communauté d'utilisateurs de partager les mêmes méta données pour des ressources partagées.

8. Dimension Sémantiquement Enrichie

également faire intervenir leur comportement passé (historique des requêtes). Le pair émetteur de la requête forme une vision plus au moins globale à propos du déroulement de sa requête.

A la réception d'une requête, une reformulation de cette dernière en terme d'ontologie global est effectuée en détectant les sous concepts, supposant que Path requêtes passées  $\leftarrow$  vide. En collaboration avec les expertises des autres pairs, les degrés de similarité par rapport au concept en question est calculée.

**7/ PlanetP** [CUENCA] représente le contenu de chaque pair de manière compacte à l'aide d'un filtre de Bloom décrivant les termes des documents stockés par celui-ci. Ces filtres de Bloom sont distribués dans le réseau en utilisant un algorithme de propagation appelé "algorithme de gossiping". Cette diffusion permet à chaque pair de construire une table de routage. Et donc Chaque pair possède un index global constitué d'une liste d'autres pairs, associés chacun à leur filtre de Bloom [CHIKY]. Un nœud qui reçoit une requête commence d'abord par faire une recherche dans son index local. S'il ne peut pas honorer la requête, il calcule les rangs des pairs de son index global et il propage la requête aux pairs de plus grand rang [SICARD].

A la réception d'une requête le nœud commence à construire le Filtre de Bloom correspondant à cette requête puis il compare le Filtre de Bloom de la requête aux signatures de pair présentes dans la table de routage. Cela permet de cibler les pairs les plus susceptibles de répondre à la requête, et donc de réduire le nombre de messages utilisés pour propager les requêtes en calculant la mesure de pertinence d'un pair pour une requête qui est donnée par la somme des mesures IPF<sup>9</sup> des termes de la requête apparaissant dans le corpus du pair. Cette mesure donne un rang (Ri) à chaque pair, ce qui permet de les ordonner par ordre de pertinence à la requête.

Dans[CHIKY08] les auteurs proposent une approche améliorant PlanetP en tenant compte de l'historique des requêtes traitées.

**8/ Routage basée sur les résumés**[ALEM] : Dans ce travail, les auteurs proposent une méthode basée sur les résumés. Cette méthode propose de construire un index de routage local et un index global (distribué) décrivant les données des différentes sources. L'idée de base

---

9. Inverse Peer Frequency



consiste à parcourir l'index local, puis l'index global des résumés représenté sous forme d'un arbre en cas de non satisfaction de la requête afin de trouver les pairs distants pouvant la satisfaire.

L'index local est construit en se servant du modèle SaintEtiQ<sup>10</sup>[Raschia].

### ✓ Procédure de construction de l'index global

- 1) Tous les nœuds feuilles envoient une copie des informations sur leurs données (synthétisées sous forme d'une hiérarchie de résumés) vers ses voisins dans le réseau de P2P.
- 2) Tous les pairs qui ont reçu des messages depuis leurs voisins sauf un (appelé N), fusionnent leurs informations locales (disponibles sous forme de résumé) avec toutes les informations reçues des différents voisins à l'exception du nœud N. Puis, ils envoient le résumé fusionné à N.
- 3) L'étape 2 est répétée itérativement jusqu'à ce que le nœud central reçoive des informations de tous ces voisins.
- 4) Le nœud central produit un résumé global (après la fusion de tous les résumés de ses voisins avec son résumé local). Ce résumé décrit toutes les données disponibles dans le réseau P2P, il sert comme un index sur les données du réseau.
- 5) La dernière étape consiste à faire l'opération inverse (le résumé global est diffusé dans le même chemin mais de haut en bas), et l'algorithme se termine quand les nœuds feuilles reçoivent l'index global de routage .

A la réception d'une requête, l'index (hiérarchie globale) est exploré. Pour chaque nœud  $z$  du résumé qui correspond exactement à la requête,  $Pz$  est ajouté à l'ensemble des pairs pertinents pour la requête  $PQ$ . Puis, la requête est propagée vers l'ensemble des pairs contenus dans  $PQ$ . Chaque pair qui a reçu le message, doit vérifier la satisfaction de la requête par rapport à ses données stockées localement et ainsi identifier les meilleures réponses. Ensuite, il renvoie la réponse vers l'initiateur de requête. Lorsque le pair initiateur reçoit tous les résultats des pairs de l'ensemble  $PQ$ , il filtre les résultats et renvoie seulement les réponses les plus satisfaisantes.

---

10. un modèle structuré pour les résumés de données. Il prend en entrée deux types d'informations : les données à résumer les données relatives au domaine

## 3.2 Comparaison entre les algorithmes

Le tableau ci dessous compare les différentes méthodes étudiées selon différents critères, à savoir :

**Information utilisée :** informations utilisées pour définir la sémantique.

**Représentation de l'information sémantique :** structures de données utilisées pour stocker les informations.

**Passage à l'échelle :** par passage à l'échelle, nous voulons savoir si la méthode permet de supporter un grand nombre de pairs (i.e + +), un nombre moyen (i.e + -) ou peu (i.e - -).

**Maintenance :** fréquence des mises à jour des informations sémantiques. La fréquence peut être importante (i.e + +) ou plus au moins importante (i.e + -).

**Espace de stockage utilisé :** espace mémoire utilisé pour stocker les informations sémantiques.

**Partage de connaissances :** coopération des pairs pour construire des connaissances. La coopération peut être forte (i.e + +) ou sans coopération (i.e - -).

?? indiquent que l'information n'est pas spécifiée dans les documents traités.

Méthodes Critères	Information utilisée	Représentation de l'information sémantique	Passage à l'échelle	Espace de stockage utilisé	Partage de connaissances	Maintenance
MAY	Historique des requêtes	Table	++	--	--	++
PlanetP	Contenu des pairs	Filtre de Bloom	+ -	++	++	+ -
REMINDIN	Historique des requêtes + Ontologie	Ensemble de déclarations RDF	++	++	--	++
Psearch	Contenu des pairs	Vecteurs	+ -	++	++	++
INGA	Contenu des pairs	Ensemble de déclarations RDF	++	++	--	++
ExSI <sup>2</sup> D	Historique des requêtes + Ontologie	vecteurs	++	--	++	??
Résumés	index	résumés	--	++	++	??
LPS	Historique	??	++	--	--	++

## Conclusion

Dans ce chapitre, nous avons illustré les trois approches de routage sémantique à savoir ; le contenu des pairs, l'historique, et l'ontologie .Quelques algorithmes de routage sémantique des requêtes ont été évoqués,ces derniers ont été comparés selon des critères qu'ils nous semblent intéressants : l'approche utilisée, la representation de l'information sémantique, le passage à l'échelle,l'espace de stockage utilisé, le partage de connaissances et la maintenance.

Dans le chapitre prochain, nous présentons notre méthode de routage sémantique inspirée de la méthode de routage basée sur les résumés [ALEM] et adaptée au contexte de la recherche de connaissances dans une federation de médiateurs.

# 4

## Proposition

### Introduction

Dans notre proposition, nous nous intéressons aux formalismes basés sur les graphes. Par formalismes basés sur les graphes nous entendons les formalismes dans lesquels non seulement la représentation est faite par les graphes mais aussi les raisonnements sont faits par les opérations sur les graphes. Nous nous inspirons aussi de la méthode de routage sémantique nommée "Routage basée sur les résumés" [ALEM] (voir la section 3.3) utilisée à l'origine dans les domaines des bases de données distribuées. Nous commençons ce chapitre par une présentation de quelques notions de base sur le formalisme des GCs nécessaire à la compréhension du reste de ce mémoire.

## 4.1 Notions de base

### 4.1.1 Notion de connaissance

La représentation des connaissances dans un ordinateur est un système formel qui vise la codification de la connaissance humaine à des fins de traitement automatique (GERBE).

Les connaissances sont aussi définies dans [DICO] comme étant des objets, concepts et relations qui sont supposés exister dans un certain domaine d'intérêt.

#### ✓ Représentation des connaissances

La représentation des connaissances désigne un ensemble d'outils et de technologies destinés d'une part à représenter et d'autre part à organiser le savoir humain pour l'utiliser et le partager [CHAMPAVERE]. Il existe plusieurs langages de représentation de connaissances, parmi ces langages nous citons : la logique de frames, les logiques de description [DL], les GCs [CG], etc.

### 4.1.2 Graphes conceptuels

#### 4.1.2.1 Définition

Les graphes conceptuels ont été introduits par John-F Sowa [Sowa 84]. Ils ont été conçus pour représenter la sémantique du langage naturel [CHAWK]. C'est un modèle de représentation des connaissances de la famille des réseaux sémantiques. Il présente un double intérêt théorique : il repose d'une part sur la théorie des graphes, permettant d'utiliser des opérations de graphes pour le raisonnement [Chein et Mugnier, 1992] ; il s'appuie d'autre part sur une interprétation en logique du premier ordre [Roberts, 1992]. Autrement dit, le modèle des Graphes Conceptuels (GC) est un formalisme de représentation des connaissances fondé sur la définition de concepts et de relations entre concepts. En général, un graphe conceptuel possède deux sortes de nœuds :

- ✓ Les nœuds concepts qui représentent des entités, des attributs, des états, des événements, etc.
- ✓ Les nœuds relations conceptuelles qui symbolisent les liens qui existent entre deux concepts.

- Formellement

Un graphe conceptuel est défini comme étant un graphe [MARTIN] :

- orienté.

- fini : tout graphe dans une mémoire d'un ordinateur ne peut avoir qu'un nombre fini de nœuds.
- bipartie : il ne possède que deux sortes de nœuds : les concepts et les relations conceptuelles où chaque arc reliant une sorte de nœud à l'autre sorte de nœud.
- et non nécessairement connexe : dans un graphe connexe, si deux parties n'étaient pas connectées entre elles on aurait deux graphes.

### ✓ 4.1.2.2 Les éléments du modèle

#### a. Les concepts

D'après [LYON] un concept est : toute idée, toute pensée, ou toute construction mentale au moyen de laquelle l'esprit appréhende les choses ou parvient à les reconnaître .[RASTIER] évoque la notion de concept en ces termes :

- le concept est une forme de la pensée humaine qui permet de dégager les caractères généraux essentiels des choses et des phénomènes de la réalité objective ou plus simple, une représentation mentale, générale et abstraite d'un objet. Ce concept-là, philosophique et logique, est posé sans aucun rapport nécessaire avec les langues ni avec les systèmes de signes.
- le concept est universel de représentation qui appartient au langage.[CHAWK]

Les concepts sont représentés graphiquement par des rectangles (voir figure 4.1) ou entre crochets [Concept] en utilisant la notation linéaire.

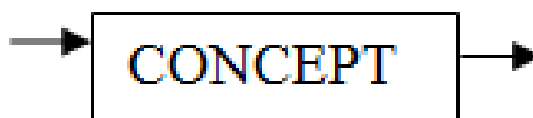


FIGURE 4.1 – La représentation graphique d'un concept.

**b. Relation conceptuelle**

Une relation conceptuelle représente une association entre un ou plusieurs concepts. Elle est composée d'un libellé de relation qui classe la relation, et d'arcs qui lient la relation aux concepts associés [GERBE00]. Les relations sont représentées graphiquement par des ellipses (voir figure 4.2) ou entre parenthèses (RELATION CONCEPTUELLE) en utilisant la représentation linéaire. **c. Arc**

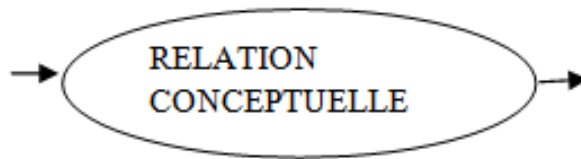


FIGURE 4.2 – La représentation graphique d'une relation

Les arcs sont les éléments qui permettent de lier les concepts aux relations. Un arc est dit appartenir à la relation et est dit attaché au concept. Les arcs d'un graphe conceptuel sont orientés, et on distingue les arcs orientés vers la relation, les arcs source, de ceux orientés vers le concept, les arcs destinations [GERBE00].

**d. support**

Un support  $S$  est un quintuplet  $(TC, TR, \partial, I, \mathfrak{S})$  tel que [THOMOPOULOS] :

- **TC** : La hiérarchie des concepts Un modèle sémantique permet de généraliser ou au contraire de spécialiser. C'est le rôle de la hiérarchie des concepts. Elle classe par groupes les différents concepts du plus particulier au plus général [CHAWK]. La hiérarchie des concepts est composée de diverses familles de concepts du même type. Les concepts sont ordonnés selon leur degré de généralité par une relation d'ordre partiel notée " $\leq$ ". Les termes de sur-type et de sous-type sont employés pour désigner la position respective de deux concepts dans la hiérarchie : Quelque soit  $t$  et  $s$  deux types de concepts Si  $t \leq s$  alors :

- $t$  est un sous-type de  $s$ .
- $s$  est un sur-type de  $t$ .

Tout type de concept est à la fois sous-type et sur-type de lui même : Quelque soit  $t$  un type

de concepts  $t \leq t$ .

- **TR** : Hiérarchie de relations : ensemble des types de relations, est un ensemble partitionné en sous-ensembles de types de relations de même arité.  $TR = TR_{i1}U\dots\dots\dots U TR_{ij}$  où  $TR_{ij}$  est l'ensemble des types de relation d'arité  $ij$ ,  $ij \neq 0$ . Tout  $TR_{ij}$  est ordonné par une relation " sorte de " $\leq$ " [GENEST].

- $\partial$  : Signature d'une relation : est une application, appelée signature, qui à tout type de relation associe le type de concept maximal de chacun de ses arguments[THOMOPOULOS].

La signature d'une relation représente une contrainte sur les types de concepts qui peuvent être reliés par ses arcs [GEORGET].

- **I** : ensemble de marqueurs ou référents qui représentent les entités du domaine que l'on cherche à modéliser.

- $\mathfrak{S}$  :Relation de conformité : est une application de  $I$  dans  $TC \perp$ , qui à tout marqueur individuel  $m$  associe un type de concept  $t$  : on dit que  $m$  est une instance de  $t$ .

### 4.1.3 Opérations de base sur les graphes conceptuels

Nous présentons ici les opérations et définitions de base pour la dérivation et la structuration des graphes conceptuels. La dérivation permet de créer un nouveau graphe à partir de graphes existants. Soient  $w$  le graphe dérivé de  $u$  et  $v$  ( $u$  et  $v$  pouvant être identiques). Il y a quatre opérations de dérivation [LECLERE] :

- **Restriction de concept** : Un concept de  $u$  peut être restreint en remplaçant son type par un type plus spécialisé (sous-type) ou en remplaçant un concept générique représentant une entité anonyme par un concept individuel représentant une entité identifiée.

- **Restriction de Relation** : Le type d'une relation de  $u$  peut être remplacé par un sous type.

- **Jointure** : Si un concept  $c$  de  $u$  est identique à un concept  $d$  de  $v$ ,  $w$  est le graphe résultant de l'union de  $u$  et  $v$  en supprimant  $d$  et en reliant les arcs de  $d$  à  $c$ .

- **Projection** : La projection est une opération qui permet de calculer si un graphe est plus spécialisé qu'un autre.

- **La normalisation** : C'est l'opération qui rend un graphe sous forme normale [MUGNIER]. Un graphe sous forme normale respecte une structure ou les référents sont uniques : Elle est



obtenue en fusionnant les nœuds concepts ayant le même marqueur individuel. Ainsi, lorsqu'on utilise la forme normale, il ne peut exister deux appellations différentes pour référer à une même instance d'un certain concept dans un même graphe [DIBIE].

- **La somme disjointe de deux graphes conceptuels** : La somme disjointe de deux graphes est le graphe obtenu en mettant deux copies des deux graphes d'origines en juxtaposition [GASMI].
- **La Simplification** : consiste à supprimer des informations redondantes dans un graphe [ES-PINASSE].

## 4.2 Notre proposition : Les graphes conceptuels pour le routage sémantique

Notre proposition est inspirée d'une méthode originale nommée "Routage basée sur les résumés" [ALEM](voir section 3.3) utilisée à l'origine dans le domaine des bases de données distribuées, que nous avons adapté à la recherche de connaissance dans une fédération de médiateurs. Dans ce qui suit, nous présentons la base de connaissances des médiateurs, puis la méthode routage sémantique dans la fédération.

### 4.2.1 la base de connaissance d'un médiateur

Nous supposons qu'un médiateur gère les connaissances d'un domaine particulier, ces connaissances sont représentées par le formalisme de représentation de connaissances des graphes conceptuels.

- **Exemple** : dans un domaine des gestion d'entreprise on pourra avoir le graphe qui représente les connaissances suivantes (voir figure 4.3) :

-Un employé nommé 1 habite dans une ville proche Lkseur, il possède un salaire élevé de 60000 et il possède un grade élevé chef de projet.

-Un employé nommé 2 habite dans une ville proche Béjaia, il possède un salaire élevé de 50000 et il possède un grade élevé chef de d'équipe.

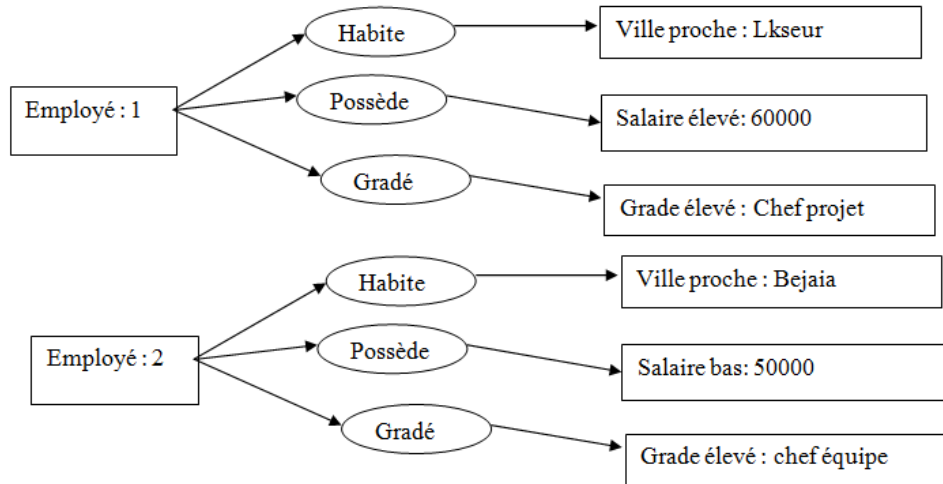


FIGURE 4.3 – Base de connaissance d’un médiateur.

### 4.2.2 Méthode de routage

Notre proposition est basée sur le contenu des pairs. Pour avoir une vue global du contenu de médiateur nous basons sur la construction de deux types de résumés : résumé local et resume global.

a) **Résumé local** : Le résumé local d’un pair  $i$  est noté  $Z_i$  et représenté sous forme d’un graphe conceptuel. Il est construit suivant ces étapes :

1. Éliminer tous les référents du GC de la base de connaissances pour avoir une information plus générale.
2. Appliquer l’operation de normalisation pour éliminer les redondances.

**Exemple** : le résumé local de la base de connaissances(figure 4.3) est donné dans (figure 4.4)

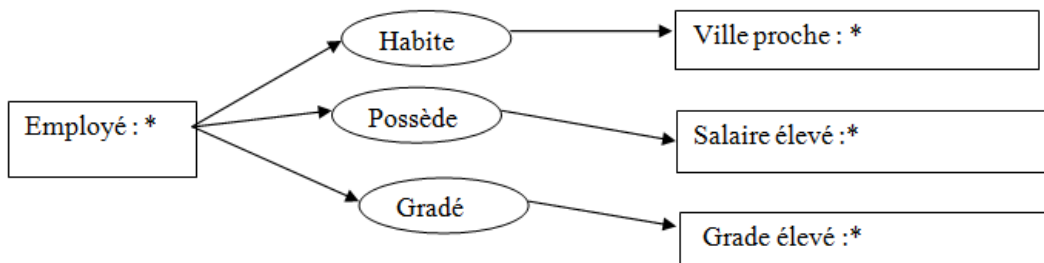


FIGURE 4.4 – Résumé de la base de connaissances.

b) **Résumé global** : Il est noté ZG, il englobe tous les résumés locaux des pairs existants dans un réseau. Il est représenté sous forme d'un arbre et construit sur un nœud central recevant les résumés des différents pairs existants suivant ces étapes :

1. Les nœuds possédants un seul voisin envoient leurs résumés locaux.
2. Chaque nœud  $P_i$  recevant le résumé local  $Z_j$  de son voisin  $S_j$  construit un autre résumé en utilisant  $Z_i$  et  $Z_j$  suivant ces étapes :
  - Somme disjointe de  $Z_i$  et  $Z_j$ .
  - Normalisation du graphe résultant.
3. L'étape 2 est répétée itérativement jusqu'à ce que le nœud central reçoive des informations de tous ces voisins.
4. Le nœud central produit un résumé global sous forme d'un arbre.
5. La dernière étape consiste à faire une diffusion du résumé global construit.

### ✓ Exemple de construction de l'index global

Soit un réseau contenant 3 pairs : P1, P2 et P3 formants l'architecture montrée dans la figure 4.5 :

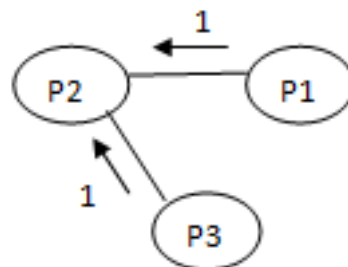


FIGURE 4.5 – exemple d'architecture d'un réseau possédant trois pairs

Donc, suivant cette architecture les pairs P1 et P3 envoient leurs résumés locaux au pair P2. On obtient comme résumé global l'arbre montré dans la figure suivante :

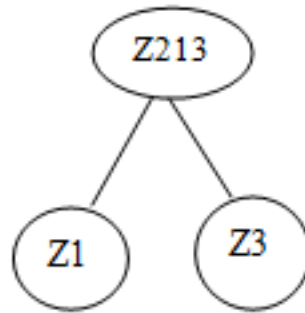


FIGURE 4.6 – exemple de résumé global

### 4.2.3 Recherche sémantique de connaissances

Dans notre cas, la recherche est faite en exploitant les résumés : local ; si la requête peut être satisfaite localement. Global sinon (si la requête peut être satisfaite par un pair distant).

#### ✓ Représentation de la requête

La requête est représentée sous forme d'un graphe conceptuel.

#### Exemple

Une requête de recherche des employés qui habite à Béjaïa peut être représentée comme illustrée dans la figure (4.7) :

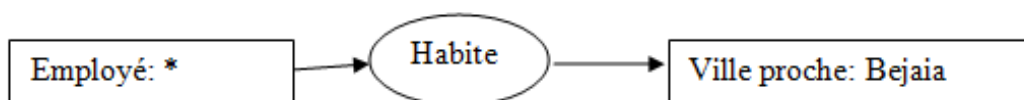


FIGURE 4.7 – Représentation de la requête.

#### ✓ Traitement de la requête

Deux types de satisfaction de requête sont possibles : la satisfaction locale et la satisfaction distribuée.

a) **Satisfaction locale** : un pair recevant la requête essaye de la satisfaire localement, suivant ces étapes :

- Construire le résumé de la requête suivant les mêmes étapes décrites dans (section 4.2)

- Projection du résumé de la requête sur le résumé local.
- S'il y a une projection alors :
  - \*Projection de la requête sur la base.
  - \*Si une projection est trouvée alors rendre la réponse.
- sinon essayer de satisfaire la requête en se servant des autres pairs de la fédération.

**b)Satisfaction distribuée** : pour la satisfaction distribuée de la requête, il faut exploiter le résumé global afin de localiser les pairs pertinents pour répondre à la requête. Les étapes de a satisfaction distribuée sont données ci-après :

- projection du résumé de la requête sur la racine de l'arbre.
- s'il y a une projection alors projeter sur les fils.

✓ **Algorithme de propagation** : les étapes suivies sont décrites formellement par :

**Algorithme** Routage Semantique

**Entrées** :  $Q$  : requête posée ;

$Z_{local}$ : résumé local d'un pair  $P_i$  ;

$Z_{global}$ : résumé global de tous les résumés des pairs du réseau ;

**Sortie**: *Pairs pertinents* : ensemble des pairs pouvant répondre à une requête posée.

**Var** : trouve= faux ;

**debut**

Génération de résumé pour  $Q$  ( $ZQ$ ) ;

*Traitement local* ( $Q$ ,  $Z_{local}$ , *trouve*) ;

**Si** (*trouve* = faux) **alors**

*Pairs pertinents* := *SelectionPairs pertinents* ( $ZQ$ ,  $Z_{global}$ );

**Si** (*Pairs pertinents*  $\neq \emptyset$ ) **alors**

*Propager*  $Q$  au *Pairs pertinents*;

**Sinon**

*Echec* ;

**Finsi** ;

**Finsi** ;

**Fin.**

**Procédure** *Traitement local* ( $Q$ ,  $Z_{local}$ , *trouve*) ;

**Entrée:**  $BC$  : Base de connaissances locale ;

**Sortie:**  $List_{projection}$ : Liste des réponses satisfaisantes la requête ;

**debut**

$List_{projection} :=$  projection de  $ZQ$  sur  $Z_{local}$  ;

**Si** ( $List_{projection} \neq \emptyset$ ) **alors**

$List_{projection} =$  projection de  $Q$  sur  $BC$  ;

$Trouve =$  vrai ;

**Sinon**

$Trouve =$  faux ;

**Finsi** ;

**Fin**

**Fonction** *SelectionPairs pertinents* (*ZQ*, *Z*)

**Entrée** : *Z* : racine de l'index global ;

*ZQ* : le résumé de la requête;

*Zfils* : nœud dans l'index global ;

**Résultat** : *PQ* /\* la liste des pairs pertinents pour Q \*/

**Var** : *PQ* :=  $\emptyset$  /\* la liste des pairs est vide \*/

*Listprojection* =  $\emptyset$  Liste des réponses satisfaisantes la requête ;

**debut**

*Listprojection* := projection de *ZQ* sur *Z*;

**Si** (*Listprojection*  $\neq \emptyset$ ) **alors**

**Si** *Z* est une feuille **Alors**

*PQ* := *PQ*+*PZ*; // *PZ* : L'identifiant d'un pair qui a généré *Z*;

**Sinon**

**Pour** chaque nœud fils *Zfils* de *Z* **faire**

*PQ* := *PQ*+ *SelectionPairs pertinents* (*ZQ*, *Zfils*) ;

**Fin pour** ;

**Fin si** ;

**Fin si** ;

**Fin.**

### ✓ Scénario d'exécution

Nous donnons ici un exemple détaillé de déroulement de notre algorithme de routage proposé sur un système contenant quatre médiateurs (pairs)

- Le support



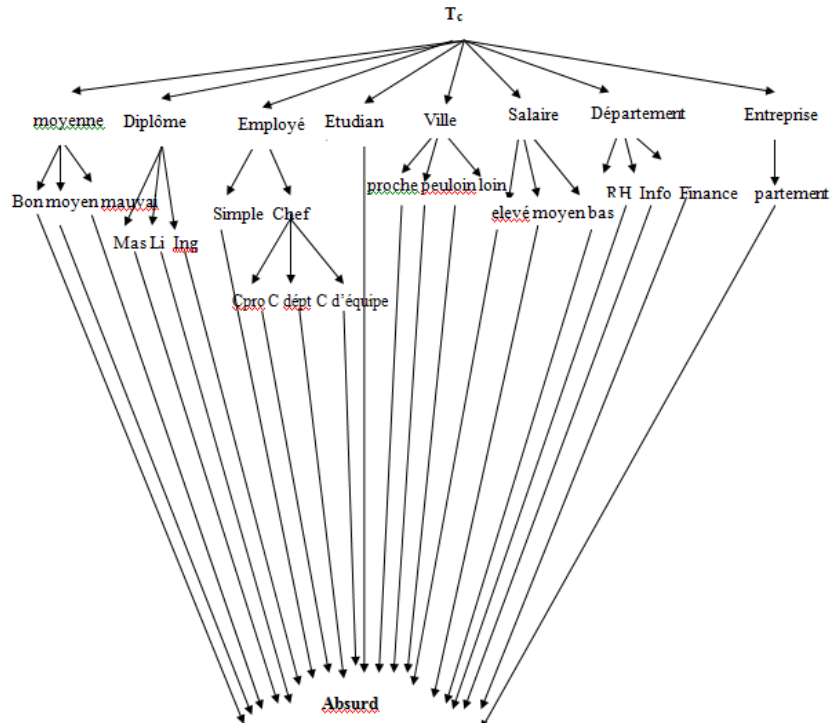


FIGURE 4.8 – Support de concepts.

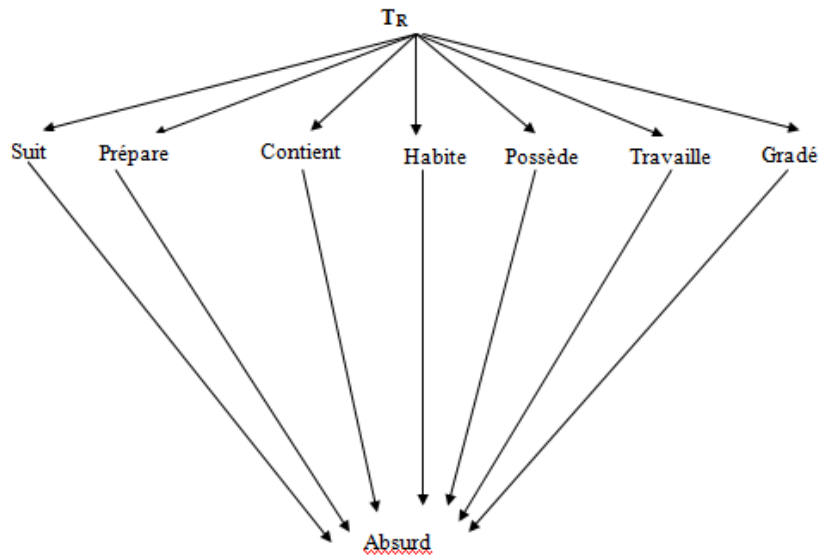


FIGURE 4.9 – Support de relations.

1/ **Les référents** On remarque que les catégories sont les concepts feuilles :

Ville loin (Tazmalt, Akbou, .....).

Ville Un peu loin (Sidi Aich, Amizour,.....).

Ville Proche (oued ghir, lekseur).

Salaire Elevé [50000,...].

Salaire Moyen [35000,50000].

Salaire Bas [15000,35000].

Note Bonne [15,20].

Note Moyenne [8,14].

Note Mauvaise [0,8].

Grade élevé (Chef département, Chef projet, Chef D'équipe).

Grade simple (simple).

### 2/ Quelques signatures

Habite (Employé, Ville).

Suit (Employé, Informatique).

Possède (Employé, Salaire).

Grader (Employé, grade).

Habite (Étudiant, Ville).

Suit (Étudiant, Informatique).

Possède (Étudiant, Moyenne).

• **Hypothèse** : on considère que les bases de connaissances sont normalisées au départ par l'application de l'opération de normalisation (voir la section 4.2).

### 3/ Les bases de connaissances de chaque pair

#### Pair P1

Le pair1 contient dans sa base de connaissances quatres employés tel que :

l'employé 1 habite à Lkseur, il suit le département informatique, il possède un salaire de 54000 et il possède un grade de chef de projet.

l'employé 2 habite à Bejaia, il suit le département informatique et il possède un salaire de 60000.

l'employé 3 suit le département informatique et il possède un salaire de 25000.

l'employé 4 habite à Oued ghir, il suit le département informatique et il possède un grade de chef de projet.

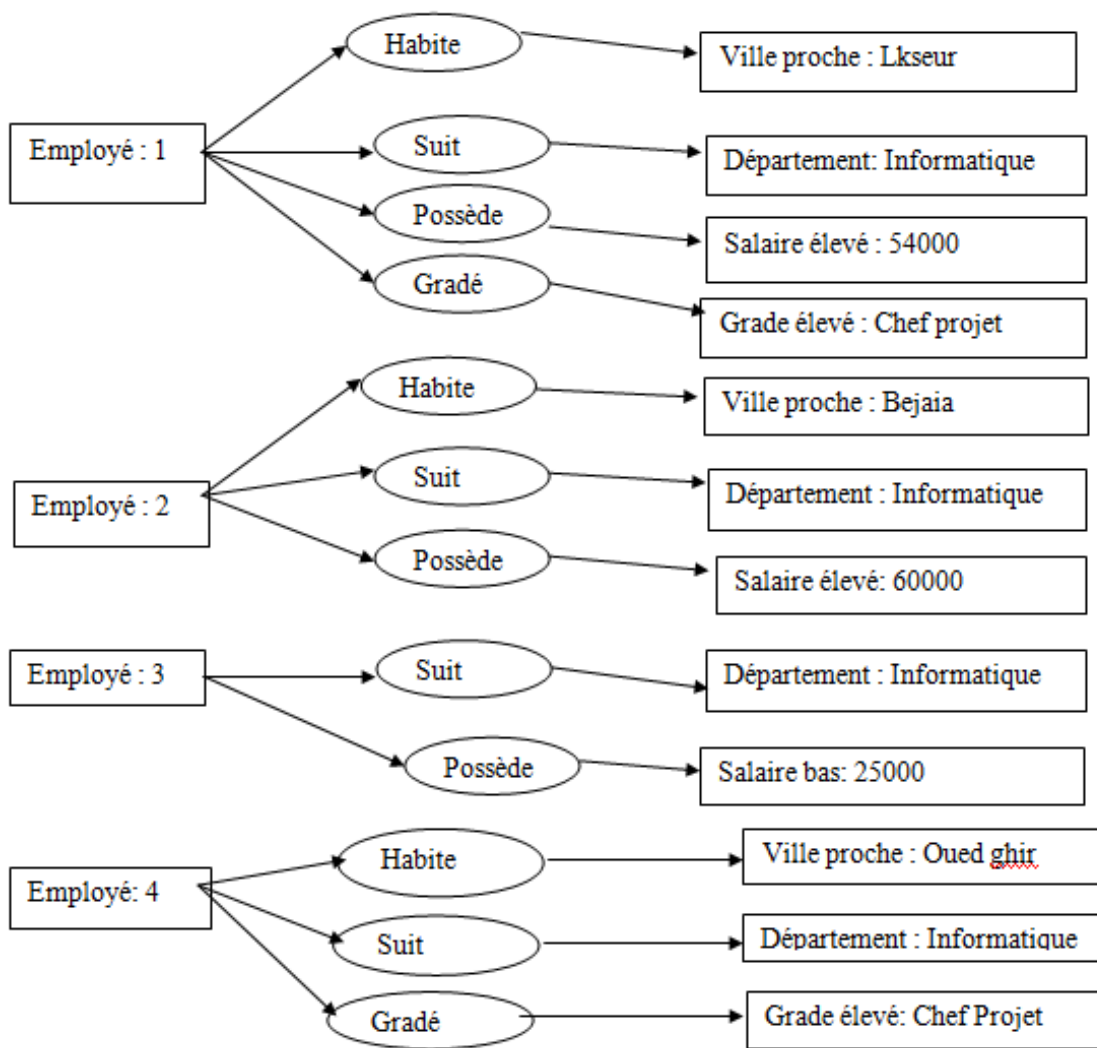


FIGURE 4.10 – Base de connaissances du pair1

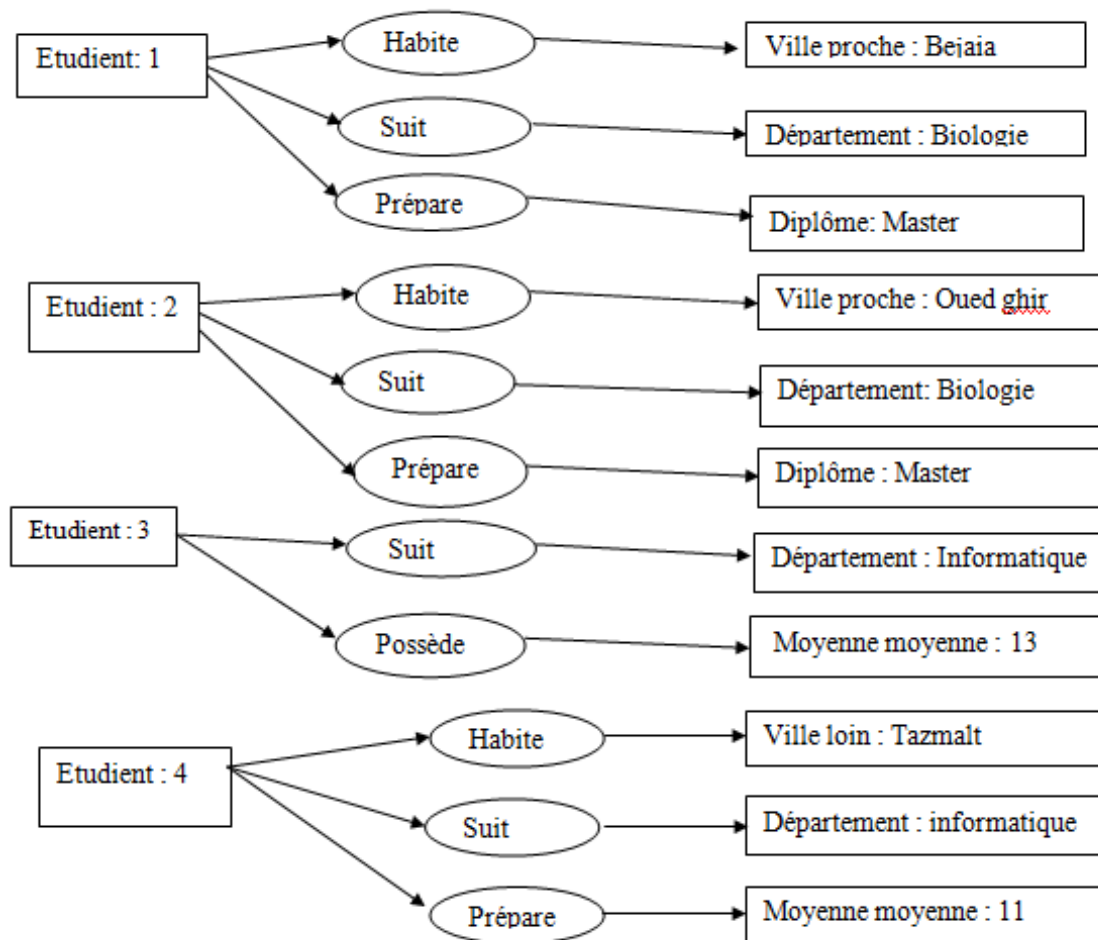
### Pair P2

Le pair2 contient dans sa base de connaissances quatres employés tel que :

l'étudiant 1 habite à Béjaia, il suit le département Biologie, il prépare le diplôme de master.

l'étudiant 2 habite à oued ghir, il suit le département Biologie et il prépare le diplôme de master. l'étudiant 3 suit le département informatique et il possède une moyenne qui est égale à 13

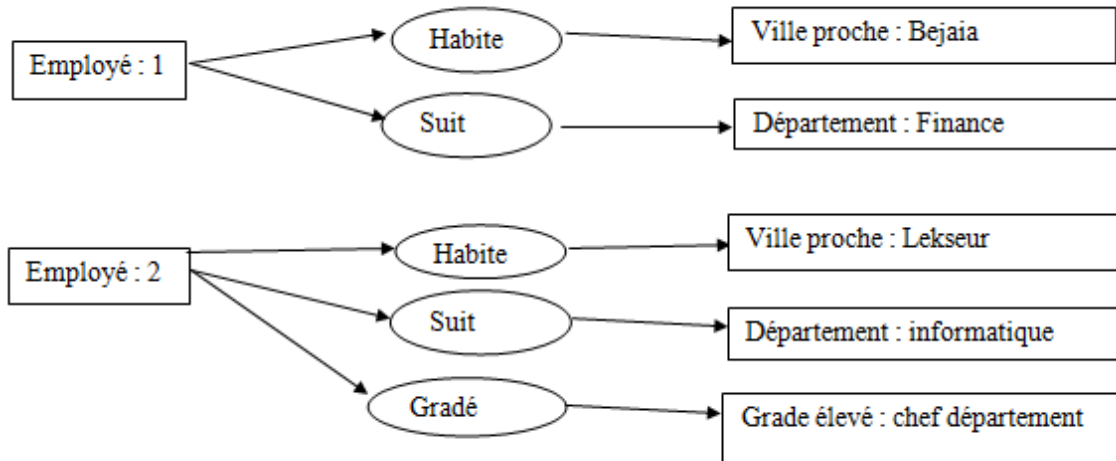
l'étudiant 4 habite à Tazmalt, il suit le département informatique et il possède une moyenne qui est égale à 11



**Pair P3**

l'employé 2 habite à Lkseur, il suit le département informatique et il possède un grade de chef de département.

l'employé 1 habite à Bejaia et il suit le département finance.



**Pair P4**

l'employé 1 habite à Ouedghir, il suit le département finance, il possède un salaire de 50000 et il possède un grade d'un simple employé.

l'employé 2 suit le département finance, il possède un salaire de 56000 et il possède un grade d'un simple employé. l'étudiant 1 suit le département Biologie, il prépare le diplôme de ingénieur et il possède une moyenne qui est égale à 15.

l'étudiant 2 suit le département Biologie, il prépare le diplôme de ingénieur et il possède une moyenne qui est égale à 16.

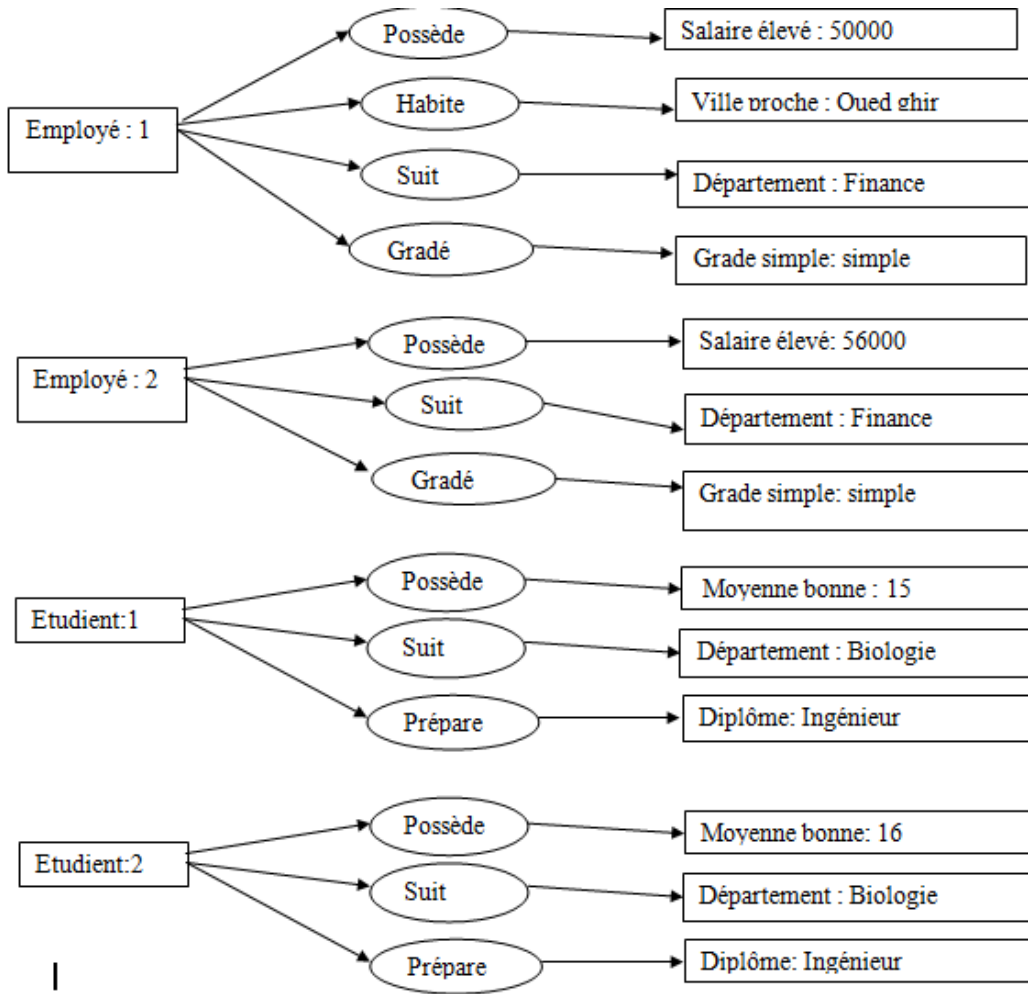


FIGURE 4.11 – La base de connaissances du pair P4

4/ Génération d'un résumé local

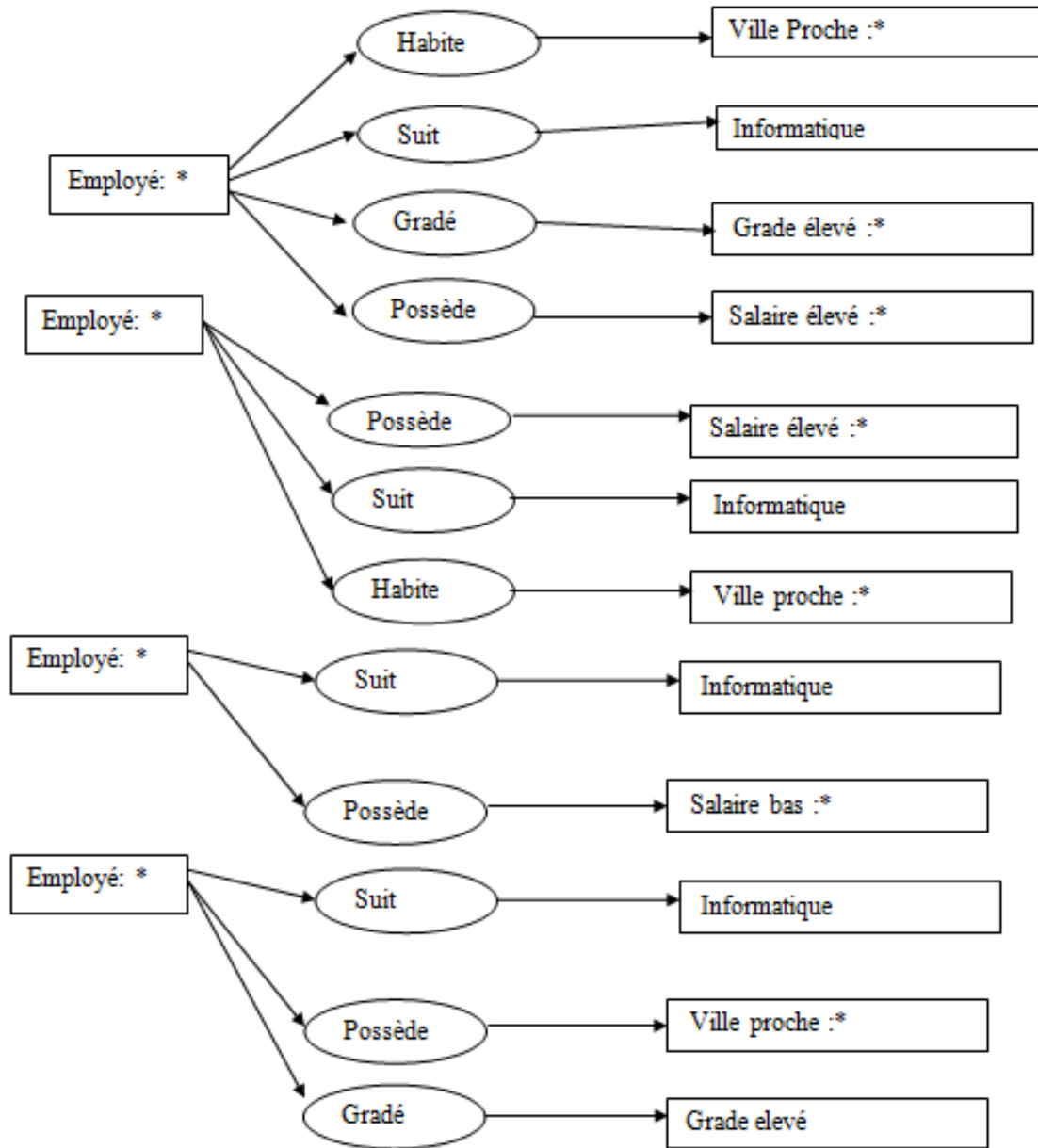


FIGURE 4.12 – Le résumé local généré par le pair P1 avant la simplification

Le graphe obtenu contient des redondances, donc les mêmes informations sont répétées plusieurs fois, pour les éliminer et pour avoir un résultat plus correct on applique une deuxième règle appelé " simplification". On aura comme résultat :

Pair P1

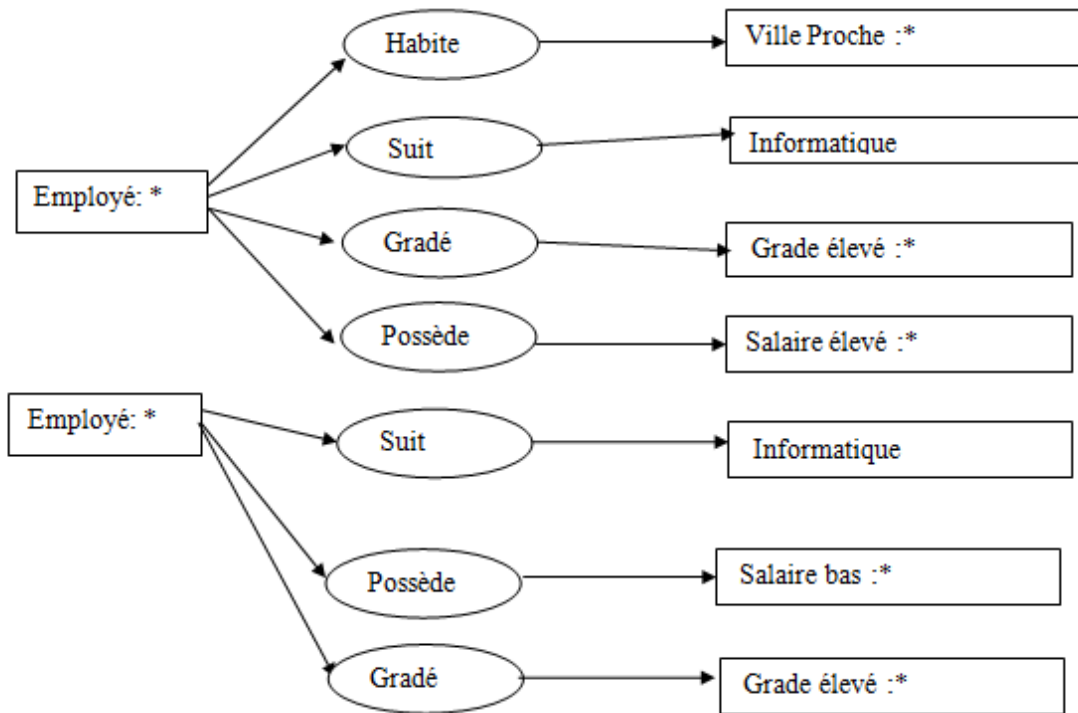


FIGURE 4.13 – Le résumé du pair P1 après simplification



La même règle est appliquée sur la base de connaissance de pair P2, on obtient :

**Pair P2**

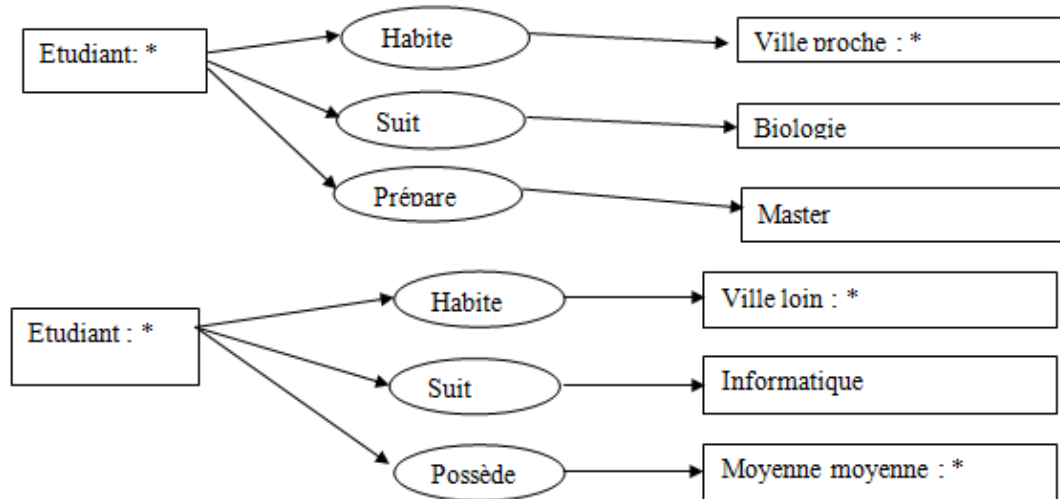


FIGURE 4.14 – Le résumé local du pair P2

**Pair P3** : pour générer le résumé de ce pair, on a besoin d'appliquer une autre règle en plus de celle de simplification, cette opération est celle de la somme disjointe de deux graphes conceptuels. Le résultat final obtenu après application des deux règles est représenté dans la figure 4.15.

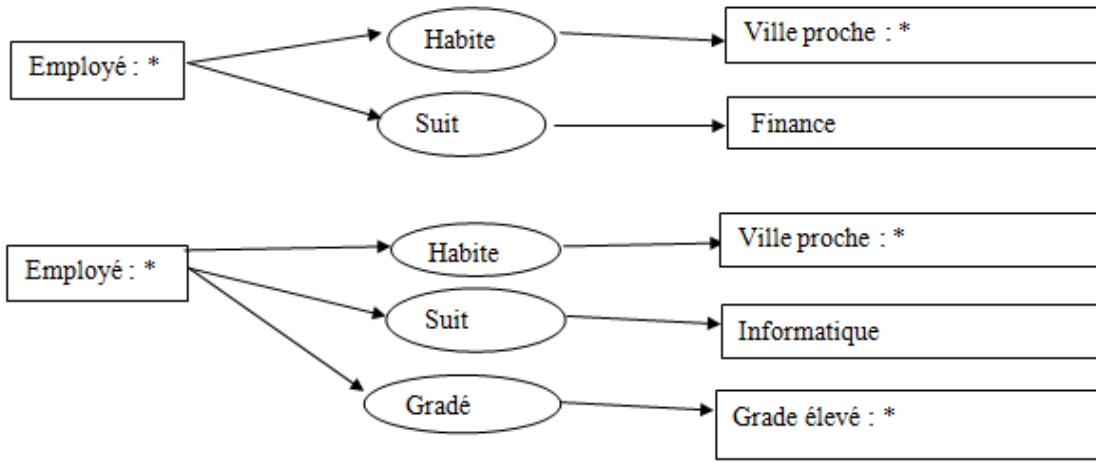


FIGURE 4.15 – Le résumé local du pair P3

**Pair P4** : les mêmes règles sont appliquées sur les graphes du pair P4, on aura comme résultat :

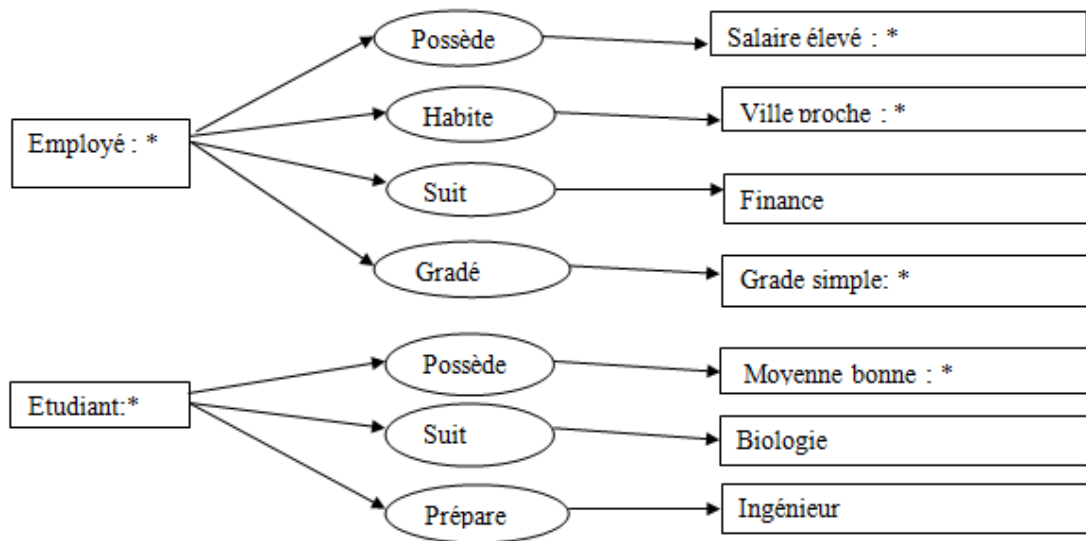


FIGURE 4.16 – Le résumé local du pair P4

5/ Génération d'un index global

nous appliquons les deux règles (Somme disjointe de deux graphes, et la Simplification) définies précédemment pour construire un résumé global (l'index global) contenant des résumés sur toutes les données existantes sur le réseau. Si on prend un réseau qui possède l'architecture suivante :

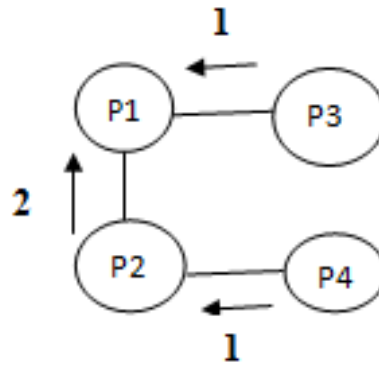


FIGURE 4.17 – Architecture du système

P3 envoie son résumé local à P1 qui va construire un autre résumé à partir de son index local et en utilisant l'index qui lui est envoyé par P3. Nous obtenons un nouveau résumé Z31 au niveau de P1. Le résumé Z31 construit après l'application de la règle de simplification est :

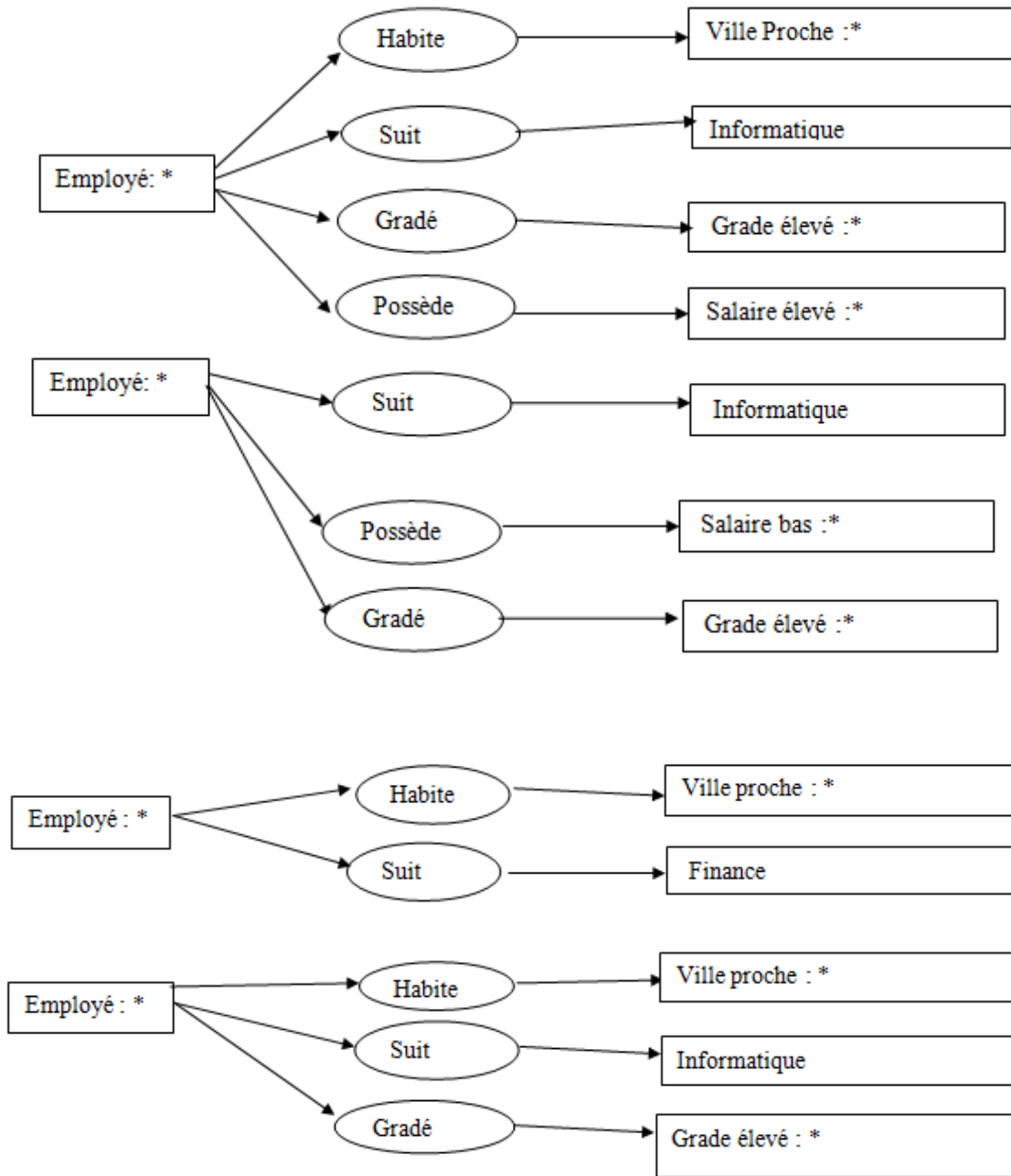


FIGURE 4.18 – Résumé au niveau du pair P1.

P4 envoie son index local au pair P2. En utilisant le même principe expliqué et suivi par le pair P1, ce pair construit le résumé Z42. Nous obtenons le graphe de la figure 4.19 :

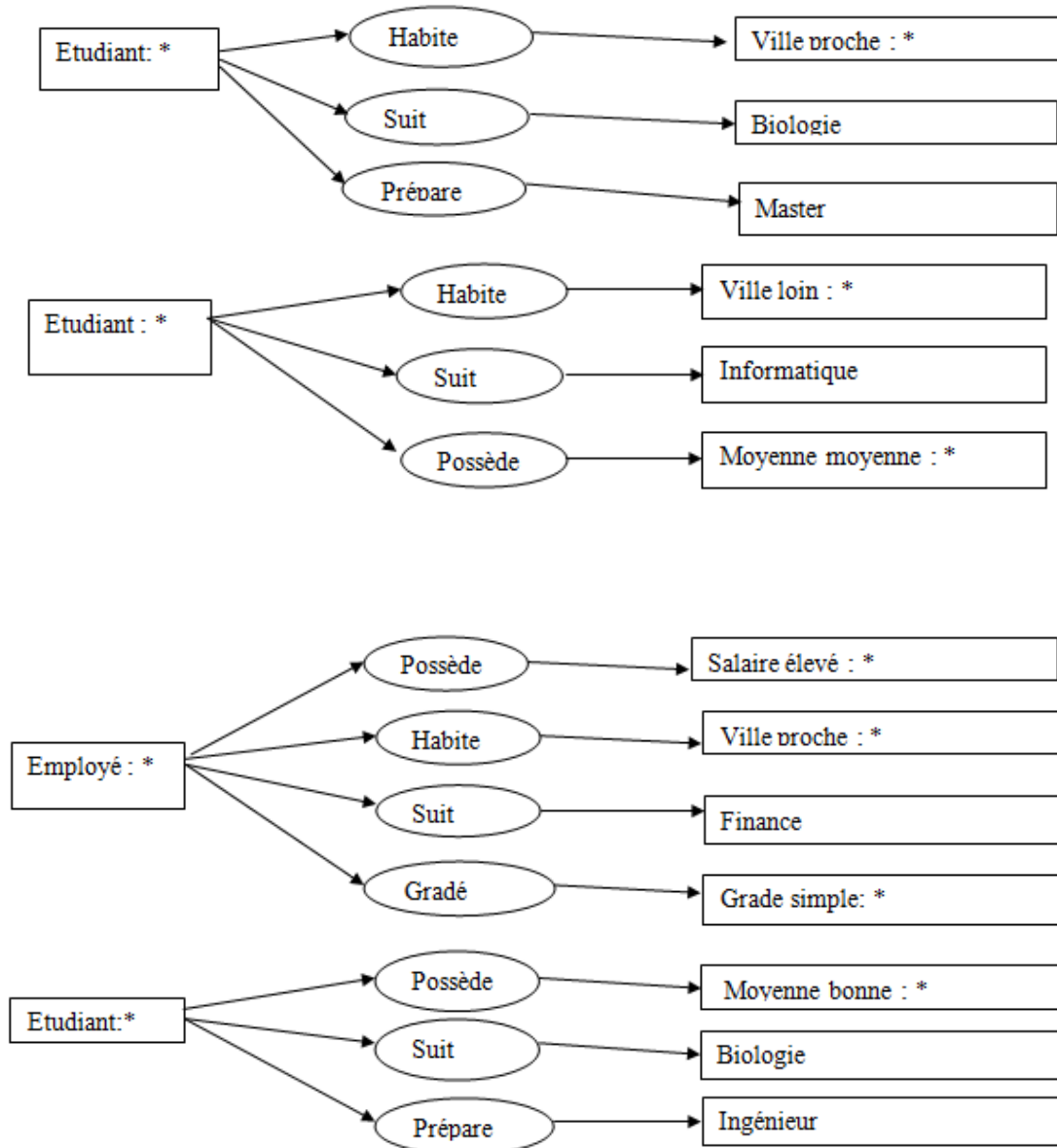
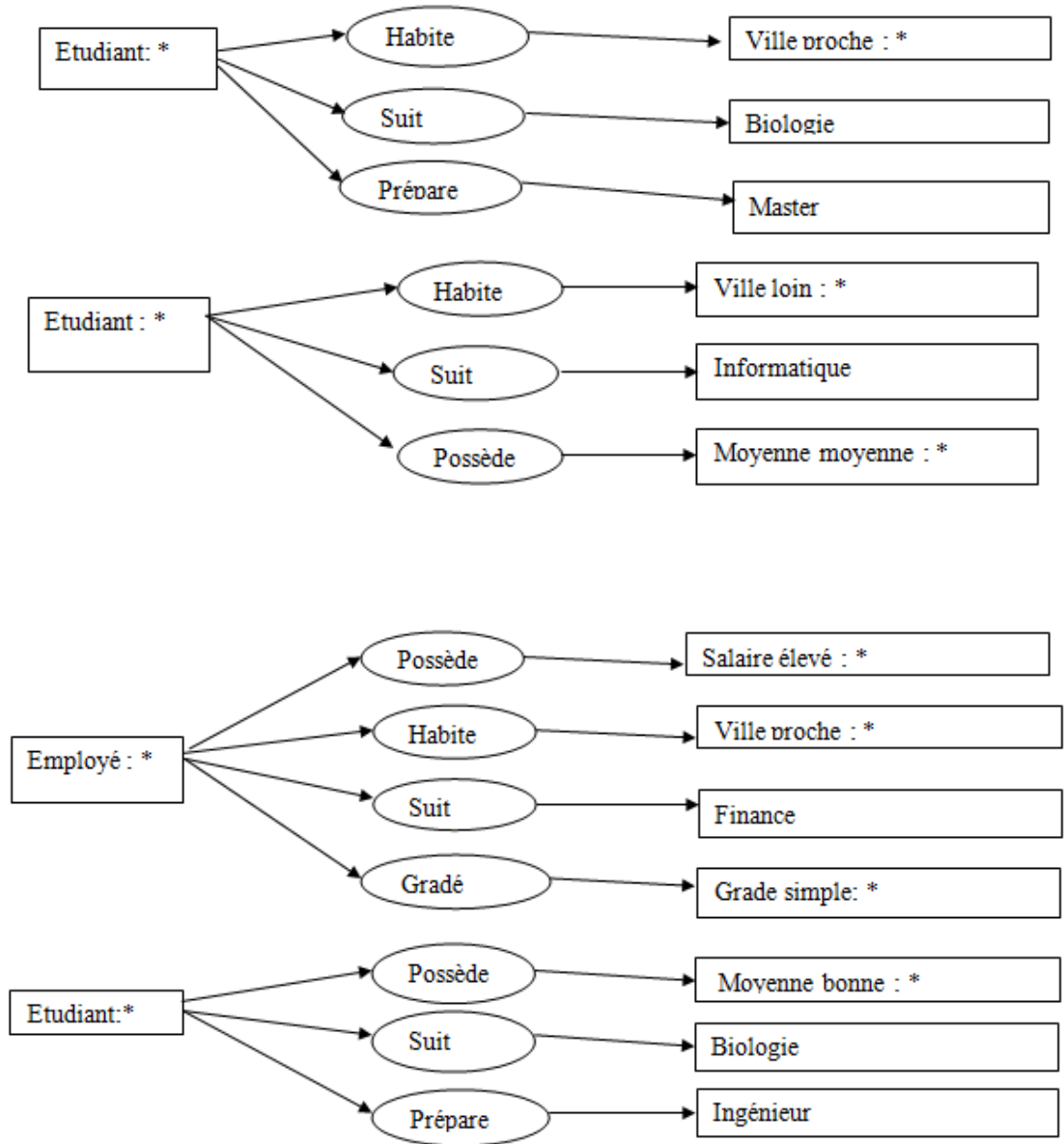


FIGURE 4.19 – Résumé au niveau du pair P2.

Le résumé Z42 généré va être envoyé par le pair P2 au pair P1 et en utilisant le résumé Z13 déjà construit par ce pair, il va construire le résumé Z1342 de la même façon. Donc on obtient après application des deux règles définies précédemment (Somme disjointe et simplification) l'index global représenté dans la figure 4.20 :



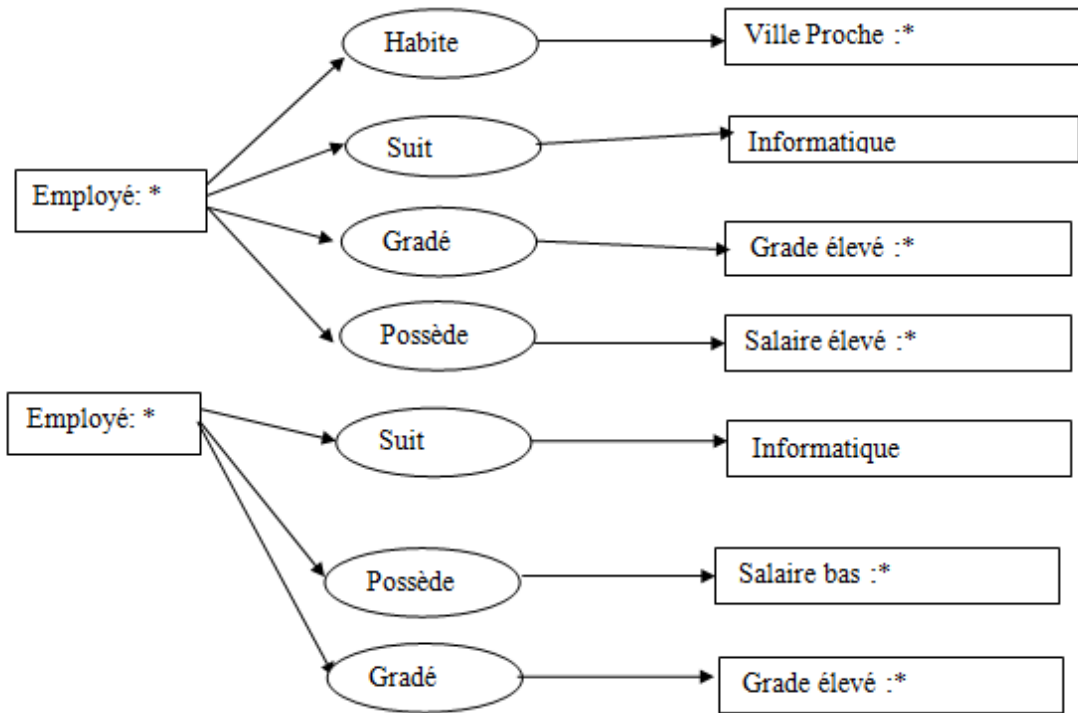


FIGURE 4.20 – Résumé global.

Les résumés construits forment un arbre comme il est représenté ci-dessous :

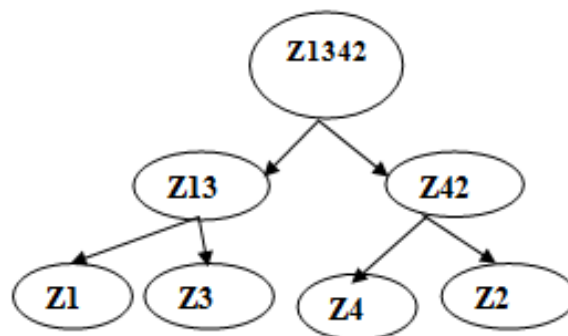


FIGURE 4.21 – Arbre des résumés.

Après construction et obtention de résumé global, le pair P1 va le diffuser à tous les autres pairs. Donc, on aura à la fin, chaque site possède son index local et un index global. Ce qui leur permet de définir le site ou les sites la satisfont.

### 6/ Traitement d'une requête :

Soit la requête posée sur le pair P2 : **la liste des employés qui habitent à Bejaia possédant un salaire de 50000.**

Le processus à suivre par le pair P2 :

#### a. Représentation de la requête sous forme d'un graphe conceptuel :

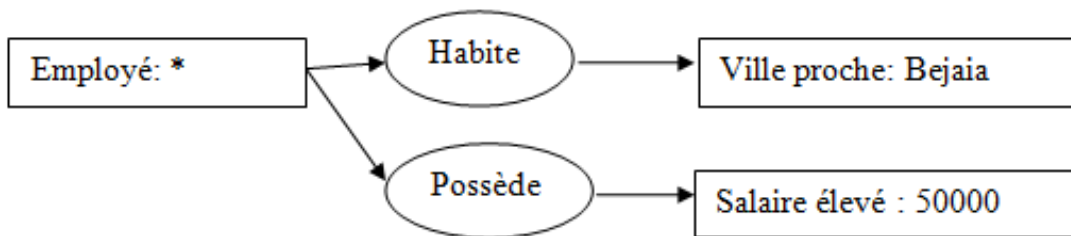


FIGURE 4.22 – La requête posée.

#### b. Construction d'un résumé

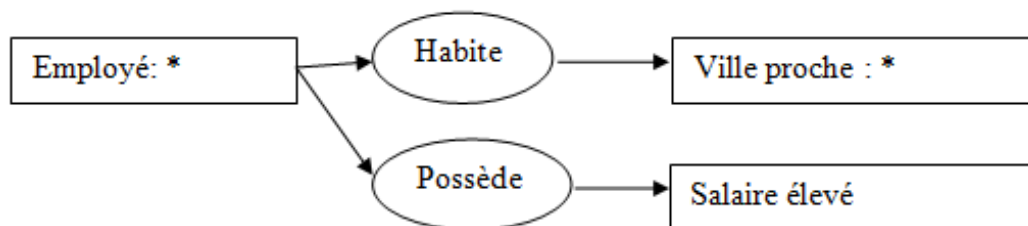


FIGURE 4.23 – Résumé de la requête posée.

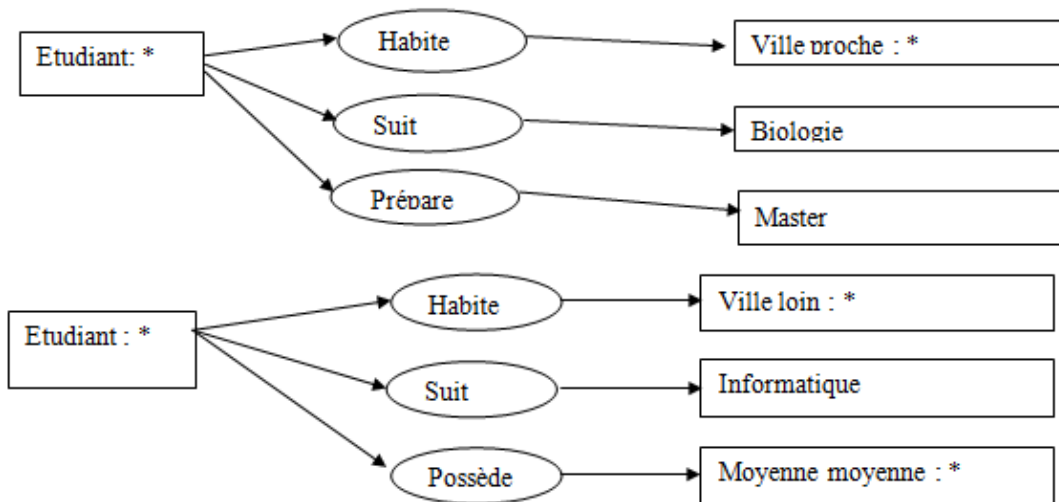
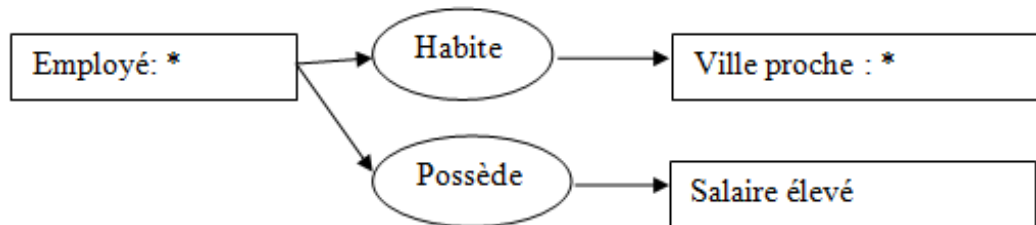
### 7/ Comparaison du résumé local du pair et celui de la requête

Après normalisation de résumé de la requête pour éviter et supprimer toutes les redondances,



## La proposition

on utilise l'opération de projection de résumé de la requête sur le résumé local du pair sur lequel la requête est posée. Donc les relations sont projetées sur les relations et les concepts sont projetés sur les concepts. On obtient :



Le résumé local du pair2 ne satisfait pas la requête. Alors, nous passons à une autre étape :

### 8/Comparaison du résumé global du réseau et celui de la requête

le résumé global contient les IDs des pairs qu'il résume. Ce qui facilite au pair P2 de déterminer les pairs les plus pertinents en parcourant l'arbre d'index global.

Dans cet exemple, en parcourant l'arbre des résumés, l'index global Z1342 va savoir que la réponse à sa requête existe dans le résumé Z42 en projetant sur le noeud fils, donc, il va continuer le parcours d'arbre suivant la branche possédant Z42 ainsi de suite jusqu'à ce qu'il arrive à la feuille possédant le même résumé que celui de la requête, il récupère l'IDs du pair contenant ce résumé, il lui envoie la requête. les noeuds recevant la requête vont projeter cette dernière sur leurs base de connaissance pour rendre la réponse si elle existe.

#### 4.2.4 Contribution de la méthode

✓ **Utilisation des graphes** : Nous avons choisi le formalisme des Graphes Conceptuels, car :

1. comme le souligne [CHEIN], ce formalisme partage les avantages des modèles de type réseau sémantique (i.e. les modèles utilisant des graphes étiquetés) qui sont d'un usage relativement intuitif (notamment grâce à leur aspect graphique) et qui ont de bonnes capacités descriptives (les étiquettes peuvent être complexes et des graphes peuvent être emboîtés dans d'autres graphes) ;
2. les "Graphes Conceptuels" (GC) ont une interprétation consistante et complète en logique du premier ordre ;
3. une relation de spécialisation peut être calculée sur les GC. Les spécialisations d'un GC "requête" peuvent ainsi être retrouvées dans la BC. La recherche de connaissances peut ainsi se faire de manière abstraite et "par le contenu" (i.e. en donnant une partie de la connaissance au lieu d'une référence exacte) comme le préconisent [ACKER] pour faciliter la recherche de connaissances et la production d'explications.

✓ **Résumé** : L'utilisation des résumés nous permet de réduire le nombre d'information circulant sur le réseau.

✓ **Contenu** : L'utilisation de l'approche par contenu est la meilleure façon pour la localisation des pairs pertinents.

✓ **La sémantique** : nous avons utilisé la sémantique pour améliorer considérablement la localisation des données. Donc, l'utilisation du critère de la sémantique sur le contenu des pairs nous permet de router plus efficacement les requêtes.

### 4.3 Conclusion

Dans ce chapitre, nous avons présenté une méthode de routage hybride basée sur le contenu des pairs. La représentation de l'information sémantique est faite par le formalisme de représentation de connaissances des graphes conceptuels. Cette méthode profite des avantages de l'utilisation de contenu des pairs tels qu'elle réduit le nombre de pairs interrogés pour une requête donnée ce qui diminue de son tour le nombre de messages circulant dans le système entier.

# Conclusion et perspectives

La masse énorme d'informations disponibles sur des sources de données distribuées et hétérogènes, nécessite un plan de recherche et de sélection, de plus en plus performants, pour permettre à l'utilisateur de localiser et d'extraire précisément l'information désirée d'une manière simple et efficace.

Afin de contribuer à la proposition des solutions à ce problèmes de recherche et de sélection, notre travail consistait à proposer une méthode de routage dans une fédération de médiateurs en intégrant l'aspect sémantique dans le processus de routage concernant le contenu des pairs. L'information sémantique dans notre proposition est représentée par un formalisme de représentation des connaissances des graphes conceptuels pour leurs capacité de description. L'utilisation de ce formalisme nous a permet d'extraire les catégories de connaissances directement à partir du support en se retrouvant sur ses concepts feuilles, sans passer par une technique de catégorisation.

L'autre point positif de cette méthode est l'utilisation des résumés qui nous donnent une vue sur le contenu de la base de connaissance qui nous a permet de réduire le coût de la recherche, i.e. au lieu de projeter la requête sur le graphe de la base de connaissance, cette opération se fait seulement sur un graphe réduit de son résumé ce qui nous permet de savoir si la requête peut être satisfaite par cette base de connaissances. La représentation de l'indexe global construit à partir des résumés des autres pairs sous forme d'un arbre présente aussi un avantage du fait que la recherche dans un arbre est de complexité logarithmique.

Plusieurs autres extensions de ce travail sont possibles. En terme de perspectives possibles, nous envisageons de :

- Implémenter l'algorithme proposé et simuler les résultats afin de pouvoir le comparer aux résultats des autres méthodes proposées.
- Étendre cette approche pour retourner les k-meilleures réponses dans la fédération.
- Utiliser une structure permettant à chaque médiateur d'indiquer le nombre de résultats existant dans sa base de connaissance pour chaque résumé généré, cela permet d'utiliser une recherche probabiliste sur le nombre de résultat sur chaque pair pertinent.
- Enfin, et en terme d'hétérogénéité, nous n'avons traité dans ce travail que la fédération des médiateurs homogènes qui adoptent tous le même langage de représentation de connaissances or que les médiateurs sont très souvent hétérogènes, i.e. ils utilisent des langages différents pour représenter les connaissances, il est alors très intéressant d'adapter notre approche pour fonctionner dans un environnement hétérogène.

# Bibliographie

[ABDENNADHER] : *Fatma ABDENNADHER*. " Traitement de la Forte Dynamicit  dans un Syst me Publier/Souscrire Bas  DHT ", M moire de Master de l'Universit  de Sfax. Juillet 2011.

[ACKER] : *Liane ACKER et Bruce PORTER*. "Extracting viewpoints from knowledge bases". Proceedings of the Twelfth National Conference on Artificial Intelligence. pp. 547-552. AAAI Press/MIT Press, Seattle, USA. Mars1994.

[AL KING] : Raddad AL KING. " Localisation de sources de donn es et optimisation de requ tes r parties en environnement pair- -pair " Th se de doctorat de l'Universit  Toulouse III - Paul Sabatier, Mai 2010.

[ALEM 09] : Abdelkader ALEM " Vers une Approche bas e sur les R sum s Distribu s pour le traitement des Requ tes Flexibles dans un Syst me P2P ", Programme des doctoriales STIC' 09, M'Sila, Alg rie, D cembre 2009.

[AMAD] : Mourad AMAD, " D couverte et localisation des services en mode P2P ". Th se de Magist re de l'Universit  Abderrahmane Mira - Bejaia, 2005.

[AZZAZ] : Amina AZZAZ, Malki MIMOUN, Ladjel BELLATRECHE, Youcef BENMIMOUN. " Une Approche Ontologique d'Int gration de Sources de Donn es dans un Environnement de Pair   Pair ". Proceedings of the 2nd Conf rence Internationale sur l'informatique et ses Applications (CIIA'09) Saida, Algeria, May 2009.

[BAKHTOUCHI] : BAKHTOUCHI Abdelghani. " Etude et proposition d'une architecture de m diation entre sources de donn es h t rog nes " M moire de Magister   l'Institut National de formation en Informatique (INI) Oued Smar, Alger, 2006.

[BUSNEL] : Yann BUSNEL. " Prise en compte de la proximité sémantique dans la restructuration de réseau logique pair à pair ", Mémoire du Master d'Ecole Normale Supérieure de Cachan - Bretagne, Juillet 2006.

[CERQUEUS] : Thomas CERQUEUS. " Mise en place de la méthode ExSI2D dans un système P2P hétérogène", Memoire de Master 2 ALMA - Nantes, Septembre 2009.

[CHAMPAVERE] : Jérôme CHAMPAVERE " de la représentation de connaissance au Web sémantique ", Disponible sur <http://www.scribd.com/doc/39489963/null>. (dernière consultation : février 2012).

[CHAWK] : Mohamad CHAWK " Modèle de graphes conceptuels et représentation sémantique du langage naturel ", Rapport de recherche, ERSICO, Université Jean Moulin, Lyon 3, Mars 2000.

[CHEIN] : Michel CHEIN " présentation des activités sur les graphes conceptuels simples " In Actes des 4èmes journées du LIPN, Villetaneuse, Paris, septembre 1994, pp 81-97.

[CHIKY] : Raja CHIKY, Bruno DEFUDE, Georges HEBRAIL. " Définition et diffusion de signatures sémantiques dans les systèmes pair-à-pair", Actes des journées Extraction et Gestion des Connaissances, Lille, France, Janvier 2006.

[CHIKY 08] : Raja CHIKY, Abbas SLIMANI. " Routage des requêtes optimisé par apprentissage dans une architecture P2P sémantique ". Actes des journées Extraction et Gestion des Connaissances (EGC'2008), Namur, Belgique, Janvier 2008.

[DICO] : Computing Dictionary, 1996 Accessible at <http://wombat.doc.ic.ac.uk/> (dernier consultation : Juin 2012).

[CUENCA] : Francisco Matias CUENCA-ACUNA, Christopher PEERY, Richard P. MARTIN, Thu D. NGUYEN. " PlanetP : Using Gossiping to Build Content Addressable Peer-to-Peer Information Sharing Communities". 12th International Symposium on High-Performance Distributed Computing (HPDC-12 2003), Seattle, WA, USA, pages 236-249, June 2003.

[DANG] : Frédéric DANG NGOG, " Moteur de recherche personnalisé et décentralisé ". Thèse de doctorat de l'université Paris 11, Orsay. Janvier 2006.

[DEDZOE] : Kokou DEDZOE. " Traitement de Requêtes Top-k dans les Communautés Virtuelles P2P de Partage de Données" Thèse de doctorat à l'université de Henri Poincaré - Nantes, Novembre 2011.

[DEFUDE] : Bruno DEFUDE, " Organisation et routage sémantiques dans les systèmes pair-à-pair ", Actes du XXVème Congrès INFORSID, Perros-Guirec, France, Mai 2007.

[DIBIE] : Julliete DIBIE-BARTHELEMY, Ollivier HAEMMERLE, Eric SALVAT "a semantic validation of conceptual graphs". Knowledge Based Systems, 19 (7) : 498-510, Février 2006.

[DJAGHLOUL] : Younes DJAGHLOUL, Zizette BOUFAIDA, " Intégration des ressources Web dans un environnement P2P, basé sur les ontologies et la gestion de la confiance ". Thèse de doctorat de l'Université Mentouri - Constantine, 2007.

[DOBSON] : Simon DOBSON, Lorcan COYLE, Paddy NIXON, "Hybridising Events and Knowledge as a Basis for Building Autonomic Systems", Journal of Trusted and Autonomic Computing, December 2006.

[DONG] : Cheng DONG. " Gestion et découverte de compétences dans des environnements hétérogène ", Thèse de doctorat à l'université de Henri Poincaré - Nancy1, Octobre 2008.

[DOUCET] : Anne DOUCET, Genoveva Vargas SOLAR "Médiation de données : solutions et problèmes ouverts", Actes des 2ème assises nationales du GDR I3, Nancy, France, décembre 2002.

[ESPINASSE] : Bernard ESPINASSE " Représentation des connaissances : Introduction aux Graphes Conceptuels ". accessible par <http://www.lsis.org/epinasseb/Supports/MR-TC/GraphesConceptuels-oct08-4p.pdf> (dernière consultation : Avril 2012).

[GARDARIN] : Georges GARDARIN, "Intégration de données hétérogènes distribuées" Cours N°8. Accessible par <http://www.informatik.uni-trier.de/ley/db/indices/a-tree/g/Gardarin:Geor> (dernière consultation : Juin 2012).

[GASMI] : Badrina GASMI "gestion distribuée de compétences ". Mémoire de Magistère de l'Université Abderrahmane Mira - Bejaia, 2007.

[GEORGET] : Laurent GEORGET " Modélisation et exploitation de traces de simulation d'agents émotionnels ". Mémoire Master, Ecole Polytechnique de l'Université de Nantes, 2005.

[GERBE] : Olivier GERBE " Introduction au formalisme des graphes conceptuels " Cours N° 1. Disponible sur : <http://zonecours.hec.ca/documents/H2006-1-687571.introductioncgs.pdf>. (dernière consultation : Mars 2012).

[GERBE 00] : Olivier GERBE " Un modèle uniforme pour la modélisation et la méta-



modélisation d'une mémoire d'entreprise ". Thèse doctorat de l'université de Montréal, Janvier 2000.

[GIO] : Wiederhold GIO. "Mediators in the Architecture of Future Information Systems" IEEE Computer, 25(3) :38-49, March 1992.

[HUGO] : St-Louis HUGO, "Création d'une mémoire collective ". Thèse doctorat de l'Université de Laval, Québec, Juin 2008.

[HUON] : Alexandre HUON " Routage dans les Réseaux Pair à Pair " Mémoire Master2 d'Ecole Nationale Supérieure d'Informatique de Caen, Juin 2006.

[INFO] : "InfoWebMaster". Disponible sur : <http://glossaire.infowebmaster.fr/index-/>. (dernière consultation : Mai 2012).

[ISMAIL] : Anis ISMAIL. " Communautés dans les réseaux sémantiques pairs-à-pairs" Thèse de doctorat de l'université de la méditerranée Aix-Marseille II, Juillet 2010.

[KALOGERAKI] : Vana KALOGERAKI, Dimitrios GUNOPULO, Dimitrios YAZTI, "A local search Mechanism for peer- to- peer Networks". In Proceedings of the 2002 ACM CIKM International Conference on Information and Knowledge Management, McLean, VA, USA, pages 300-307, November 2002.

[KTARI] Salma KTARI. "Interconnexion et routage dans les systèmes pair à pair" Thèse de doctorat de l'École Nationale Supérieure des Télécommunications - Paris, Décembre 2009.

[KOSTADINOV] : Dimitre KOSTADINOV, Verónica PERALTA, Assia SOUKANE, Xiaohui XUE. "Système adaptatif d'aide à la génération de requêtes de médiation", 20<sup>ème</sup> journée Bases de Données Avancées (BDA), Montpellier, France, Octobre 2004.

[LE] : Viet Dung LE " Équilibrage de Charge dans un Réseau Pair-à-Pair Structuré : une Méthode Dynamique Intégrée ". Thèse de doctorat de l'Université de Montréal, 2006.

[LECLERE] : Michel LECLERE " Raisonner avec des définitions de types dans le mod ?le des graphes conceptuels ", Revue d'Intelligence Artificielle, 12(2) :243-278. Hermès. Juillet 1998.

[LIBOUREL] : Thérèse LIBOUREL, " Médiation via les métadonnées ", Disponible sur : [www.lirmm.fr](http://www.lirmm.fr). (Dernière consultation : septembre 2009).

[LOSER] : Alexander LOSER, Steffen STAAB, Christoph TEMPICH, "Semantic Social Overlay Networks", IEEE Journal on Selected Areas in Communications, 25(1), Janvier 2007.

[MARTIN] : Philippe MARTIN " Exploitation de graphes conceptuels et de documents

structurés et hypertextes pour l'acquisition de connaissances et la recherche d'informations " Thèse de doctorat de l'université Sophia Antipolis - Nice, Octobre 1999.

[MUGNIER] : Marie-Laure MUGNIER. " Contributions algorithmique pour les graphes d'héritage et les graphes conceptuels " Thèse de doctorat de l'Université Montpellier II, France, Octobre 1993.

[OSPINA] : Victoria eugenia OSPINA, Alain-Jérôme FOUGERES " aide à la conception collaborative un système de médiation pour l'usage de micro-outils logiciels ", In Cognito, Cahiers Romains de Sciences Cognitives (3), pp. 89- 122, Mars 2009.

[RAGHAVAN] : Vijay V. RAGHAVAN, Hayri SEVER "On the reuse of past optimal queries". Proceeding SIGIR '95 Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, Pages 344 - 350 New York, NY, USA, July 1995.

[RASTIER] : François RASTIER : "sémantique et recherches cognitives ", Presses Universitaires de France., France, 1991.

[REYNAUD] : Chantal REYNAUD, Gloria GIRALDO. " Médiation de services sur le Web ". Dans actes des Journée Francophones de la Toile (JFT), pages 59-68, Tours France, Juillet 2003.

[ROBERTS] Don D. ROBERTS. "The existential Graphs of Charles S. Peirce". Wiredweird Publishing Company, 1973.

[ROUSSET] : François GOASDOUE, Marie-Christine ROUSSET " Intégration d'Informations par Médiation ". Actes du XXIIIème Congrès INFORSID, Grenoble, France, Mai 2005.

[SICARD] : Jérôme SICARD. " Contributions à la Modélisation, Évaluation et Conception de Systèmes de Recherche d'Information en Pair-à-Pair ", Thèse de doctorat de Télécom SudParis, Novembre 2010.

[SOWA 84] : John Florian SOWA "Conceptual structures : information processing in mind and machine" Addison-Wesley Publishing Company, USA, 1984.

[TANG] : Chunqiang TANG, Zhichen XU, Sandhya DWARKADAS, "Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks", In SIGCOMM '03 : Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications, pages 175-186, New York, NY, USA. ACM, January 2003.

[TEMPICH] : Christoph TEMPICH, Steffen STAAB, Adrian WRANIK, " Semantic Query Routing in Peer-to-Peer Networks based on Social Metaphors ", 13th International World Wide Web Conference, , New York City, NY, USA, May 2004.

[THEOTOKIS] : Stephanos THEOTOKIS et Diomidis SPINELLIS "A Survey of Peer-to-Peer Content Distribution Technologies", ACM Computing Surveys, Vol.36, pp. 335-347, December 2004.

[THOMOPOULOS] : Rallou THOMOPOULOS " Représentation et interrogation élargie de données imprécises et faiblement structurées ", Thèse doctorat de l'Institut national agronomique Paris-Grignon. Décembre 2003.

[VENTRESQUE] : Anthony VENTRESQUE " Espaces vectoriels sémantiques : enrichissement et interprétation de requêtes dans un système d'information distribué et hétérogène ", Thèse de doctorat de l'université de Nantes. Septembre 2008.

[Wiki] : " Wikipedia ". Disponible sur : [http //www.wikipedia.org/](http://www.wikipedia.org/). (dernière consultation : fevrier 2012).

[YEFERNY] : Toufik YEFERNY, Khedija AROUR, Yahya SLIMANI. " Routage sémantique des requêtes dans les systèmes pair-à-pair ", Actes de la Sixième Conférence Francophone en Recherche d'Information et Applications, Hyères, France, Mai 2009.

## Résumé

La recherche d'information est un problème crucial dans différentes types d'application et plus particulièrement dans les applications distribuées et coopératives, où l'ensemble des informations sont répartis sur plusieurs pairs qui peuvent être des médiateurs fédérés ou coopératifs. Dans ce cas, satisfaire une requête de recherche d'information peut être soit non structuré (propagation aléatoire des requêtes dans la fédération), soit structuré (propagation des requêtes selon une structure d'organisation des médiateurs).

Dans ce mémoire, nous nous intéressons au deuxième type de recherche dont l'efficacité peut être améliorée en introduisant de la sémantique dans le processus de propagation (routage) des requêtes dans la fédération. Cette sémantique est généralement construite à partir du contenu des médiateurs mais peut également faire intervenir leur comportement passé.

**Mots clés** : Système pair-à-pair, Médiation, Routage sémantique, Graphes conceptuels.

## Abstract

The search for information is a crucial problem in different types of application and more particularly in the distributed and co-operative applications, or the whole of information are distributed on several pairs which can be federate or co-operative mediators. In this case, to satisfy a request of search for information can be either not structured (random propagation of the requests in the federation), or structured (propagation of the requests according to a structure of organization of the mediators).

In this memory, we are interested in the second type of research whose effectiveness can be improved by introducing semantics into the process of propagation (routing) of the requests in the federation. This semantics is generally built starting from the contents of the mediators but can also utilize their last behavior.

**Key words** : System peer to peer, Mediation, Semantic routing, Graph conceptual.