

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider Biskra



Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département d'Informatique

N° d'ordre :.....

Série :.....

THESE

Présentée pour obtenir le grade de
DOCTORAT EN SCIENCES EN INFORMATIQUE

Par

REZEG khaled

THEME

Découverte des services dans les systèmes d'information
géographiques répartis sous réseau ad-hoc

Soutenue le : 18/01/2011

Devant le jury composé de :

Pr DJEDDI Nouredine, Professeur à l'Université de Biskra

Président

Pr LASKRI Mohamed Tayeb, Professeur à l'Université d'Annaba

Rapporteur

Pr KOUDIL Mouloud, Professeur à l'Ecole Nationale Supérieure

Examineur

d'Informatique Alger

Dr KAZAR Okba, Maître de Conférences à l'Université de Biskra

Examineur

Dr BOUKARAM Abdallah, Maître de Conférences à l'Université de Sétif

Examineur

Année Universitaire 2010/2011

Remerciements

Mes sincères remerciements et ma profonde gratitude à Monsieur LASKRI Mohamed Tayeb Professeur des universités à l'Université d'Annaba, pour avoir accepté de diriger cette recherche, pour m'avoir suivi pendant cette période riche par ses orientations scientifiques, ses précieux conseils et sa disponibilité malgré ses occupations.

Mes vifs remerciements à Madame Sylvie SERVIGNE Maître de conférences à LIRIS de Lyon pour son co-encadrement de mes travaux de recherches durant mon séjour scientifiques à l'INSA de Lyon pour ses efforts qui n'ont pas été épargnés en termes de disponibilité, de précieux conseils et sa vive collaboration et enfin je la remercie pour ses encouragements.

Mes remerciements à Monsieur DJEDDI Nouredine, Professeur des universités à l'Université de BISKRA pour son acceptation de présider le jury. Egalement à Monsieur KOUDIL Mouloud Professeur des universités à l'Ecole Nationale Supérieure en Informatique Alger, à Monsieur KAZAR Okba Maître de conférences à l'université de BISKRA, et à Monsieur BOUKARAM Abdallah Maître de conférences à l'université de Sétif d'avoir accepté de juger ce travail et de faire partie du jury.

Mes remerciements vont de même à tous les enseignants du département d'informatique pour leurs encouragements. Plus particulièrement, je tiens à exprimer ma reconnaissance à Monsieur ZERARKA Mohamed Faouzi. Je le remercie pour ses aides dès le début de ma thèse, sa disponibilité et sa générosité pour faire avancer mes recherches. Je remercie aussi les membres de LIRIS de Lyon pour leurs soutiens et encouragements.

Enfin, je tiens à remercier tous les enseignants qui ont participé dans ma formation, tous mes amis et tous les gens qui je n'ai pas cités pour leurs aides et encouragements.

Dédicaces

À mon père...

À ma chère mère ...

*Sans eux je n'aurais jamais été ce que je suis, ni atteindre le stade que j'ai atteint
aujourd'hui.*

*Aux deux êtres les plus chers au monde, qui ont été - pour moi- l'exemple du respect,
du devoir consciencieux ; exemple qui a su me guider dans mes études.*

*Pour tout l'amour, la patience qu'ils ont su m'offrir et la fierté qu'ils éprouvent
envers moi.*

*À ma femme qui m'a encouragé à y aller de l'avant. Elle était la lumière étincelle
qui se fortifie jour après jour et qui animait mon chemin avec plein d'émotions,
d'enthousiasme, du bonheur et d'ambitions.*

À mes enfants Mohamed slimane, Takia et Mohamed Rafik

À mon frère et sa famille

À mes sœurs et leurs familles

À mon oncle et mes tentes et leurs familles.

Je dédie ce travail

Résumé

Les systèmes d'information géographiques actuels donnent accès à un grand nombre de sources hétérogènes et distribuées. La multiplication des ressources et l'accroissement des données nécessitent l'utilisation des services Web qui apparaissent comme solution pour assurer l'interopérabilité entre les différentes ressources. Ils sont capables de collaborer et de partager le contrôle des processus de données et d'informations entre les applications sur différentes plateformes. L'exploitation des données géographiques à travers les services Web fait apparaître les Services Web Géographiques (Geospatial Web Services). Un nombre croissant de services Web géographiques (SWG) sont conçus pour interroger, faire interopérer des informations spatiales sur le Web. Les SWG sont en train de changer la manière dont les informations spatiales, les systèmes et les applications sont conçus, développés et déployés. Au fur et à mesure que les SWG se multiplient, la difficulté de la découverte de service s'accroît. La maîtrise de cette phase nécessite la personnalisation de la sélection des services web géographiques désirés par l'utilisateur. Notre travail concerne particulièrement la découverte des services web géographiques répartis en se basant sur la norme ISO 19119 et les critères de qualité.

Une approche SMA à base de métadonnées pour la découverte sémantique des SWG repose sur la coopération d'un ensemble d'agents utilisant les capacités sémantiques des services web géographiques à travers les catalogues basés sur les métadonnées selon la norme ISO19119. Les agents de notre architecture sont conformes aux propriétés générales des agents (Portabilité, Autonomie, Stabilité, Persistance, Flexibilité, Sociabilité). L'idée de base de notre approche est de classer les SWG selon les fonctionnalités en exploitant des modèles de métadonnées basés sur la norme ISO19119. Une optimisation de l'approche précédente est effectuée en faisant appel à la qualité et la personnalisation. La qualité joue un rôle primordial pour satisfaire les besoins explicites ou implicites de l'utilisateur et la personnalisation facilite l'expression du besoin utilisateur et rend l'information sélectionnée intelligible à l'usager et exploitable. La concrétisation de l'approche de découverte sémantique des SWG à base de qualité est assurée à travers deux modèles de métadonnées basés sur la qualité: le premier concerne les services web géographique et sera adapté aux classes de services web géographiques dans l'objectif de la construction du catalogue de métadonnées. Le deuxième concerne le profil de l'utilisateur qui sera utilisé pour enrichir et compléter les requêtes de l'utilisateur. La correspondance entre les requêtes utilisateurs et les annonces des SWG publiés est assurée à travers l'utilisation de l'algorithme de « matching » LARKS. Cet algorithme sera adapté en appliquant les trois premiers filtres (contexte, profil, similarité) sur les critères de qualités figurant dans la requête et les descriptions des classes de SWG disponibles dans le catalogue de métadonnées. Cette solution nous permet d'avoir un résultat

de type «relaxed match». Au niveau registre UDDI, l'application de l'algorithme de « matching » est basée sur le mécanisme d'interrogation du registre UDDI en faisant appel aux différentes pages (pages blanches, pages jaunes, pages vertes). Une liste de SWG classée selon l'ordre de correspondance (aspect qualité, aspect fonctionnel) représente le résultat de la découverte

La mise en œuvre de cette approche est assurée par un prototype QGWSSD (Quality Geospatial Web Service Semantic Discovery), en exploitant un ensemble d'outils (Protégé, XML/XML Schemas, PHP,...) en prenant comme domaine d'application le domaine du tourisme. Ce prototype est une application Web adaptée d'une architecture Web classique "n-tiers". Il est basé sur l'exploitation des métadonnées et l'algorithme de «matching». Pour rendre ces métadonnées persistantes, nous avons choisi le couple XML/XML Schema, basé sur le principe des langages à balises de la famille SGML. Un ensemble d'interfaces est mis à la disposition de l'utilisateur (Fournisseur/client / Utilisateur), il est réalisé en utilisant le PHP qui assure une simplicité d'interfaçage avec des bases de données. Cet ensemble contient :

1. des interfaces d'acquisition de connaissance du profil utilisateur;
2. des interfaces de publication des SWG;
3. des interfaces de découverte des SWG.

Mots Clés :

Catalogue, Découverte sémantique, ISO 19119, Métadonnées, Ontologie, Profil utilisateur, Qualité, Service Web géographique, Systèmes multi-agents.

Abstract

The actual geographic information systems today provide access to a large number of heterogeneous and distributed sources. The multiplication of resources and increased data requires the use of Web services that appear as a solution to ensure interoperability between different resources. They are able to collaborate and share control of process data and information between applications on different platforms.

The use of geographic data through Web services shows the geospatial Web Services (GWS). A growing number of GWS designed to interoperate spatial information over the network have emerged. GWS are changing the way in which spatial information systems and applications are designed, developed, and deployed. When GWS is increasing, the difficulty of the service discovery increases too. The service discovery represents one of the interactions between Web service components (supplier, customer and UDDI). It is unfortunately that it is based only on the syntax. The semantic aspect necessary to satisfy the user is absent. The use of semantic Web technology in the Web service discovery phase gives rise to the Web service semantic discovery. The latter is based on semantic reasoning. It enhances the accuracy of search results compared to traditional techniques of Web service discovery, the additional matching accuracy in terms of computing power. In this paper we are interested to the distributed GWS semantic discovery based on geographical metadata held in geo-catalogs. The metadata used in our case is designed according to the ISO 19119 standard reinforced by the quality criteria.

The suggested approach is Geospatial Web services semantic discovery using metadata and multi-agents system. It is based on the co-operation of a set of agents using the semantic capacities of the geospatial Web service through the catalogues based on the metadata according to standard ISO19119. The agents of our architecture take in account the principal properties of the agent technology (Portability, Autonomy, Stability, Persistence, Flexibility, and Sociability). The based idea of our approach is to classify to geospatial Web services by functionalities by using metadata of ISO 19119 standards.

Optimization of the recent approach is done by using quality and personalization. Quality management plays a vital role to satisfy the user needs (explicit or implicit). Personalization aims to facilitate the expression of user needs and make information intelligible to the selected user and exploitable. Our discovery approach based on semantic metadata organized in catalogs, proposed in the publication phase of GWS two levels of description. The first level concerns the GWS classes catalog based on the ISO19119 standards and quality criteria; the second level concerns the UDDI registry through the WSDL. In the discovery phase, this solution allows to restrict the search space and increase the number of relevant services through the application of the algorithm of

"matching". In the metadata level, the matching algorithm "LARKS" is adjusted by applying the first three filters (context, profile, similarity) on the quality criteria contained in the request and description of classes available in GWS metadata catalog. This solution enables us to have a result like "relaxed match". In the UDDI registry level, the application of "Matching" algorithm is based on the interrogation mechanism of the UDDI registry using the different pages (White pages, yellow pages, green pages). A list of GWS classified by order of correspondence (quality aspect, functional aspect) is the result of the discovery.

For the implementation of our approach, we propose the tourism field as application domain; we use a prototype ontology which includes only the part of the terminology field of tourism. We develop a Web application, an adaptation of a traditional Web architecture. Our implementation is based on the use of metadata and the algorithm of "matching". To make these persistent metadata, we chose the pair XML / XML Schema, based on the principle of markup languages SGML family. A set of interface is available to the user (Provider /Customer / User), the interface is made using PHP that provides ease of interfacing with databases. These interfaces are:

- 1. Interfaces for acquiring knowledge of the user profile;*
- 2. GWS publication interfaces;*
- 3. GWS discovery interfaces.*

Keywords

Catalog, Semantic discovery, ISO 19119, Metadata, Ontology, User profile, quality, Geospatial Web Service, Multi-agents systems.

Sommaire

Introduction générale	01
Chapitre 1: La découverte sémantique des services Web géographiques.....	05
1.1 Les Systèmes d'Information Géographiques (SIG)	05
1.1.1 L'information géographique	05
1.1.2 Les Systèmes d'Information Géographiques (SIG)	06
1.1.3 Les fonctionnalités du SIG.....	08
1.1.4 Les systèmes de gestion de base de données géographiques (SGBDG).....	08
1.2 Les services Web	09
1.2.1 Architecture Orientée Service (AOS)	09
1.2.2 Les Services Web	09
1.2.2.1 Architecture des services Web.....	10
1.2.2.1.1 WSDL: Web Service Description Language.....	11
1.2.2.1.2 UDDI: Universal Description, Discovery and Integration	12
1.2.2.1.3 SOAP: Simple Object Access Protocol.....	15
1.2.2.2 Les opérations des services Web.....	15
1.2.2.2.1 La description	15
1.2.2.2.2 La publication	15
1.2.2.2.3 La découverte	16
1.2.2.2.4 L'invocation	16
1.2.2.2.5 La composition	16
1.3 Les services Web géographiques	16
1.3.1 Définition du service Web géographique (SWG)	17
1.3.2 La sémantique des Services Web Géographique	17
1.3.2.1 Classification des SWG selon point de vue technique.....	18
1.3.2.2 Classification des SWG selon la norme ISO	19
1.3.2.3 Les métadonnées des SWG selon la norme ISO 19119	19
1.4 La découverte sémantique des services Web géographiques.....	23
1.4.1 La découverte	23
1.4.2 La découverte sémantique	24
1.4.2.1 Les travaux relatifs	24
1.4.3 La découverte sémantique des services Web géographiques	26

1.4.4 Les algorithmes de « matching »	27
1.4.5 Les travaux relatifs	27
1.5 Conclusion	28
Chapitre 2 : La qualité des Services Web Géographiques	30
2.1 Qualité de service	30
2.1.1 La qualité du service traditionnel	30
2.1.2 Qualité de service électronique	31
2.1.3 Qualité de services Web	33
2.1.3.1 Classification 1 : Selon A Mani et al [MAN02]	33
2.1.3.2 Classification 2: Selon R Shuping [SHU03].....	34
2.1.3.3 Classification 3 : Selon L Médini [MED05]	36
2.1.4 Qualité de services de réseaux.....	37
2.2 Les indicateurs de qualité	38
2.3 La qualité des services Web géographiques.....	40
2.3.1 Qualité des informations géographiques	40
2.3.1.1 Normalisation de la qualité des informations géographiques.....	42
2.3.1.1.1 La norme ISO 19113 : Les critères de qualité	42
2.3.1.1.2 La norme ISO 19114 : Mesurer les critères de qualité	43
2.3.1.1.4 La norme ISO19115:Les métadonnées des informations géographiques.....	44
2.3.1.1.5 La norme ISO 19138 (Projet préliminaire de spécifications techniques)	48
2.3.2 La qualité d'usage	48
2.3.4 Les travaux relatifs	48
2.4 Conclusion	49
Chapitre 3 : Profil utilisateur	50
3.1 Système de recherche d'information (SRI)	50
3.1.1 Les fondements de la recherche d'information	51
3.1.2 Les types du système de recherche d'information (SRI).....	53
3.1.2.1 Le Système de Recherche d'Information classique	53
3.1.2.2 Le Système de Recherche d'Information adaptatif	53
3.1.2.3 Le Système de Recherche d'Information personnalisé	54
3.2 Profil Utilisateur	54

3.2.1	Approches de représentation du profil utilisateur	55
3.2.2	Méta modèle de profil utilisateur	55
3.2.2.1	Domaine d'intérêt	56
3.2.2.2	Données personnelles	57
3.2.2.3	Qualité	58
3.2.2.4	Données de livraison	59
3.2.2.5	Données de sécurité	60
3.2.2.6	Dimensions et sous-dimensions du modèle de profil utilisateur	61
3.2.3	Méta modèle de contexte	62
3.2.4	Méta modèle de préférences	63
3.2.5	Relations entre le profil, le contexte et les préférences	64
3.2.6	Gestion de profils	65
3.2.6.1	Instanciation des modèles	66
3.3	Détail de la dimension qualité	66
3.3.1	Facteurs sur le contenu des données	67
3.3.1.1	Fraîcheur	67
3.3.1.2	Exactitude	67
3.3.2	Facteurs sur les processus	68
3.3.2.1	Qualité de performance	69
3.3.2.2	Qualité liée au contexte d'exécution	69
3.3.2.3	Qualité de sécurité	70
3.3.3	Facteurs sur le contenant des données	70
3.3.3.1	Popularité	70
3.4	Requête utilisateur	71
3.4.1	Classification de requêtes selon A. YAN	71
3.4.1.1	Requêtes de recherche de traitements	71
3.4.1.2	Requêtes pour connaître les traitements	71
3.4.1.3	Requêtes de l'utilisation des traitements	72
3.4.2	Requête géographique	72
3.4.3	Forme canonique des requêtes géographiques	73
3.5	Conclusion	74

Chapitre 4 : Une approche de la découverte sémantique des SWG basée Agent	75
4.1 Concepts exploités.....	75
4.1.1 Les services Web sémantiques	75
4.1.1.1 Les ontologies	76
4.1.1.1.1 Les composants d'une ontologie	76
4.1.1.1.2 Méthodologie de l'ingénierie ontologique	77
4.1.1.1.3 Les types d'ontologies	78
4.1.1.1.3.1 Les ontologie Géographique	78
4.1.1.1.3.2 Les ontologie du domaine de tourisme	79
4.1.2 Les géocatalogues	81
4.1.3 Les systèmes multi-agents (SMA)	83
4.1.3.1 Qu'est ce qu'un agent ?.....	83
4.1.3.2 Les systèmes multi-agents (SMA)	84
4.1.3.2.1 Utilité des systèmes multi-agents	84
4.1.3.2.2 Couplage SMA/SIG	85
4.1.3.2.3 Couplage SMA/SW	85
4.1.3.2.3 Couplage SMA/SWG.....	86
4.2 Modèle proposé	86
4.2.1 Architecture générale.....	86
4.2.2 Architecture et fonctionnement des composants	87
4.2.2.1 Profil utilisateur	87
4.2.2.2 Base de connaissances.....	89
4.2.2.2.1 Catalogue des métadonnées des SWG	89
4.2.2.2.2 La table d'index	91
4.2.2.2.3 Le registre UDDI	92
4.2.3 Agents de l'architecture	92
4.2.3.1 Agent interface client	92
4.2.3.2 Agent interface fournisseur	93
4.2.3.3 Agent découverte	94
4.2.3.4 Agent publication.....	97
4.2.3.5 Agent source.....	97
4.2.4 Protocoles de communication entre agents	98

4.2.4.1 Protocole de publication de SWG.....	98
4.2.4.2 Protocole de découverte de SWG	100
4.3 Conclusion	101
Chapitre 5:Optimisation de la découverte sémantique des SWG à base de qualité	103
5.1 Le langage de modélisation UML	103
5.1.1 Diagrammes de classes	104
5.1.1.1 Les classificateurs	105
5.1.1.2 Les relations	106
5.2 Le modèle proposé	107
5.2.1 Aspect architectural et fonctionnel	107
5.2.1.1 Sous système de publication des SWG	107
5.2.1.2 Sous système de découverte des SWG.....	108
5.2.1.3 Sous système de profil utilisateur	109
5.2.1.4 Tableau de correspondance fonctionnelle entre les deux approches proposées	109
5.2.2 Description du modèle proposé	110
5.2.2.1 Éléments nécessaires à notre modèle	111
5.2.2.1.1 Le catalogue des métadonnées	111
5.2.2.1.2 Le service	111
5.2.2.1.3 Le domaine d'application	111
5.2.2.1.4 Le profil utilisateur	111
5.2.2.2 Vue d'ensemble du modèle proposé	112
5.2.2.2.1 Le package catalogue	112
5.2.2.2.2 Le package service	116
5.2.2.2.3 Le package domaine d'application	117
5.2.2.2.4 Le package profil utilisateur	118
5.2.2.2.5 Relations entre les packages	123
5.3 Représentation formelle de la qualité	126
5.3.1 Qualité des catégories des SWG.....	126
5.3.2 Qualité de la requête utilisateur.....	126
5.3.3 Méthodes d'évaluation de la qualité.....	127
5.3.3.1 La méthode SAW (Simple Additive Weighting).....	128
5.4 Algorithme de « matching ».....	129

5.4.1 L'algorithme de « matching » LARKS.....	129
5.4.2 Pourquoi LARKS	132
5.4.3 Notions Utilisées dans l'algorithme de « matching » de notre approche	133
5.4.3.1 Filtrage multidimensionnel selon les critères de qualité.....	133
5.4.3.2 La distance sémantique entre deux concepts dans une ontologie...	134
5.5.3.2.1 La mesure de Resnik.....	135
5.5.3.2.2 La mesure de Jiang-Conrath.....	135
5.5.3.2.3 La mesure de Hirst-St.Onge.....	135
5.5.3.2.4 La mesure de Wu-Palmer.....	135
5.4.4 Mesure sémantique utilisée dans le filtre de profil.....	136
5.4.5 Algorithme de «matching ».....	136
5.4.5.1 Algorithme de «matching» au niveau catalogue des catégories des SWG	137
5.4.5.2 Algorithme de « matching » niveau registre UDDI.....	139
5.5 Conclusion.....	139
Chapitre 6 : Prototype de découverte sémantique des SWG à base de Qualité	141
6.1 Outils de développement.....	141
6.1.1 XML/XML Schémas.....	142
6.1.2 OWL.....	142
6.1.2.1 Protégé.....	143
6.1.3 PHP.....	144
6.1.4 MySQL.....	145
6.2 Ontologies exploités.....	145
6.2.1 Ontologie de la qualité des SWG.....	146
6.2.2 Ontologie du domaine du tourisme.....	148
6.3 Description du prototype.....	148
6.3.1 Architecture de l'application.....	148
6.3.2 Les composants du QGWSSD.....	149
6.3.2.1 Le sous système de profil utilisateur.....	150
6.3.2.1.1 Interface groupe utilisateurs.....	150
6.3.2.1.2 Interface des données personnelles d'utilisateur.....	151
6.3.2.2 Le sous système de publication.....	152

6.3.2.2.1 Interface de description des catégories des SWG.....	153
6.3.2.2.2 Interface de description des SWG.....	154
6.3.2.3 Le sous système de découverte.....	155
6.3.2.3.1 Connexion d'utilisateur.....	155
6.3.2.3.2 Construction de la requête utilisateur.....	155
6.3.2.3.3 Recherche et classement des SWG.....	157
6.3.2.3.3.1 Analyse de la description WSDL du service web.	159
6.4 Conclusion.....	160
Conclusion générale.....	162
Références bibliographiques.....	165

Liste des figures

Figure 01: Les composantes d'un SIG	08
Figure 02: Architecture des services Web.....	10
Figure 03 : Structure fondamentale d'un document WSDL.....	11
Figure 04 : Relation entre les éléments d'un document WSDL.....	12
Figure 05 : Relation entre WSDL et UDDI.....	14
Figure 06: Structures de données "noyau" d'UDDI	14
Figure 07: Structure d'un message SOAP	15
Figure 08: Les classes principales ISO 19119 pour la description d'un service	21
Figure 09: Roue de DEMING	39
Figure 10: Packages des métadonnées géographiques norme ISO 19115	45
Figure 11: Diagramme décrivant la sémantique de la norme ISO 19115	46
Figure 12: Processus en U de la recherche d'informations	52
Figure 13: Méta-modèle du profil utilisateur	56
Figure 14 : Méta modèle de la dimension domaine d'intérêt	57
Figure 15 : Méta modèle de la dimension données personnelles.....	58
Figure 16 : Méta modèle de la dimension Qualité	59
Figure 17 : Méta modèle de la dimension données de livraison	60
Figure 18 : Méta modèle de la dimension sécurité.....	61
Figure 19 : Dimensions et sous-dimensions du modèle de profil	62
Figure 20 : Méta modèle du Contexte	63
Figure 21 : Méta modèle de préférences	64
Figure 22 : Associations entre profil, contexte et préférences	65
Figure 23 : Niveau d'instanciation	66
Figure 24: Exemple de requête spatiale.....	73
Figure 25: Différentes acceptations du terme "ontologie"	76
Figure 26 : Processus de construction d'ontologie.....	78
Figure 27 : Aperçu de l'ontologie « tourisme »	81
Figure 28: Modèle sémantique basé agent pour la découverte des SWG.....	87
Figure 29 : Mécanisme d'adaptation de requête utilisateur	88
Figure 30 : Architecture de la base de connaissances	89
Figure 31: Diagramme de classes du service Web géographique	90

Figure 32: architecture d'agent interface client	93
Figure 33: architecture d'agent interface fournisseur	93
Figure 34 : Fonction Englobe	95
Figure 35 : Procédure de degré de « matching »	95
Figure 36: Fonction d'affectation de valeur de « Matching »	96
Figure 37: Architecture d'agent découverte.....	96
Figure 38: Architecture d'agent publication.....	97
Figure 39: Architecture d'agent source	98
Figure 40 : Diagramme de séquences de publication de services Web géographique....	100
Figure 41 : Diagramme de séquences de découverte de services Web géographique....	101
Figure 42 : Extrait simplifié du méta modèle pour les diagrammes de classes UML 2	105
Figure 43 : Représentations graphiques plus ou moins détaillées d'une classe UML ...	105
Figure 44 : Architecture générale du modèle proposé	107
Figure 45 : Vue d'ensemble du QGWSR-Model	112
Figure 46 : Diagramme de modèle de description de catégorie des SWG.....	113
Figure 47 : Classes et associations du package Service	117
Figure 48 : Classes et associations du package Domaine d'Application	118
Figure 49 : Diagramme simplifié des classes de package profil utilisateur	118
Figure 50 : Diagramme des classes de profil	119
Figure 51 : Diagramme des classes de la requête utilisateur	123
Figure 52 : Diagramme de classes représentant les relations entre les packages	125
Figure 53 Filtrage d'information en utilisant des facteurs qualité.....	134
Figure 54: Les relations conceptuelles.....	136
Figure 55 : Algorithme de « matching » au niveau catalogue des catégories des SWG.	130
Figure 56 : interface du logiciel protégé 3.4.....	144
Figure 57 : Le graphe d'une partie de l'ontologie qualite.OWL.....	146
Figure 58 : La description OWL de l'ontologie qualité SWG Qualite.OWL.....	147
Figure 59 : Une prise d'écran de l'ontologie TOURISME à travers protégé 3.4.4.....	148
Figure 60 : Architecture générale de l'application proposée.....	149
Figure 61 : Interface groupe utilisateurs.....	150
Figure 62 : Interface données personnelles utilisateur.....	151
Figure 63 : Extrait du fichier XML Schema de la classe utilisateur (XML Spy 2005)	152
Figure 64 : Interface de publication des catégories des SWG.....	153

Figure 65 : Extrait du fichier XML Schema de la classe catégorie du SWG (XML Spy 2005)	154
Figure 66 : interface de publication des SWG.....	154
Figure 67 : Interface de connexion d'utilisateur.....	155
Figure 68 : Interface de construction de requête utilisateur.....	156
Figure 69 : Extrait du fichier XML Schema de la requête utilisateur (XML Spy 2005)	157
Figure 70 : Interface des résultats de « matching » niveau catalogue.....	158
Figure 71 : Interface des résultats de « matching » niveau Registre UDDI.....	159
Figure 72: Processus d'extraction d'une description WSDL vers les instances d'une base de connaissances AROM selon le modèle de représentation de services...	160

Liste des tableaux

Tableau 01: Parallèle entre les SGBD relationnels classiques et les SGBD spatiaux..	09
Tableau 02 Description d'un SWG selon les points de vue ISO/IEC 10746.....	21
Tableau 03: Les éléments des métadonnées représentant les SWG.....	22
Tableau 04 : Tableau comparatif des solutions de la découverte sémantique des SW	26
Tableau 05 Métriques de fraîcheur	67
Tableau 06. Métriques d'exactitude	68
Tableau 07 Métriques de popularité	70
Tableau 08: Les éléments des métadonnées représentant les classes des SWG	91
Tableau 09 : Multiplicités syntaxiquement correctes pour les propriétés en UML.....	106
Tableau 10: Tableau de correspondance fonctionnelle entre les deux approches proposées.....	110
Tableau 11 : Tableau de métriques de fraîcheur	115
Tableau 12: Tableau des métriques de popularité	116
Tableau 13 : Tableau de description des composants du formalisme de la requête ...	122
Tableau 14: Comparaison des méthodes d'analyses multicritères.....	128
Tableau 15: Matrice de décision de la méthode SAW.....	128
Tableau 16: Tableau de correspondance entre LARKS et QGWSR-Model.....	132

Introduction générale

Introduction générale

L'information géographique est devenue très onéreuse et importante dans notre vie quotidienne. Son importance réside dans son existence dans la plupart des activités humaines à travers leurs localisations dans l'espace géographique, elle est aussi très utile pour affiner certaines requêtes. Il n'est pas étonnant que la plus grande partie des documents présents sur le Web contiennent des références spatiales à travers l'utilisation de noms de lieux ou d'adresses. Les informations géographiques sont exploitées à travers des traitements informatiques principalement réunis dans les Systèmes d'Information Géographiques (SIG).

Les SIG actuels donnent accès à un grand nombre de sources hétérogènes et distribuées. La multiplication des ressources et l'accroissement des données nécessitent l'utilisation des services Web qui apparaissent comme solution pour assurer l'interopérabilité entre les différentes ressources. Ils sont capables de collaborer et de partager le contrôle des processus de données et d'informations entre les applications sur différentes plates-formes. L'exploitation des données géographiques à travers les services Web fait apparaître les Services Web Géographiques (Geospatial Web Services). Ce domaine est en pleine évolution, il est à la convergence de trois domaines de recherches: les systèmes d'information géographiques, les services Web et le Web sémantiques. Les services sont implantés par les fournisseurs, qui mettent à disposition les descriptions de services sous forme de fichiers. Ces descriptions sont centralisées et stockées dans des annuaires (publication). Les applications clientes envoient des requêtes aux annuaires pour sélectionner les services (découverte). Elles téléchargent ensuite les descriptions des services sélectionnés, et les invoquent directement (invocation). Un nombre croissant de Services Web Géographiques (SWG) sont conçus pour interroger et faire interopérer des informations spatiales sur le Web. Les SWG sont en train de changer la manière dont les informations spatiales, les systèmes et les applications sont conçus, développés et déployés [ZHA07]. Au fur et à mesure que les SWG se multiplient, la difficulté de la découverte de service s'accroît. La maîtrise de cette phase nécessite la personnalisation de la sélection des services Web géographiques désirés par l'utilisateur.

La découverte représente l'une des opérations d'interaction entre les éléments de service Web (le fournisseur, le client et l'annuaire UDDI) [BEN08, BEN04]. Elle est définie comme suit :

La découverte : Les services sont conçus afin d'être sélectionnés via des mécanismes de recherche. La découverte est permise par la description préalable des services et leur publication au sein d'un registre. Elle est effectuée en deux phases [LOP08]:

1. La recherche: elle peut se faire via les registres dans lesquels les fournisseurs ont publié leurs services Web ;

2. La sélection: le client doit ensuite sélectionner parmi la collection de services Web issus de l'étape de recherche, le service Web qui convient le mieux à ses attentes.

Il s'avère malheureusement que la découverte des services Web est basée seulement sur la syntaxe. L'aspect sémantique indispensable pour satisfaire l'utilisateur y est absent [FRI05]. La capacité de représenter la sémantique géospatiale est d'une grande importance lors de la construction des applications géospatiales pour le Web. Les services Web sémantiques fournissent des solutions de technologies intelligentes pour des services d'annotation, de découverte, de composition et de l'invocation dans les environnements distribués. Le déploiement de cette technologie dans des applications géospatiales Web a un potentiel pour améliorer la découverte, la recherche et l'intégration de l'information géographique, ainsi que sa réutilisation dans d'autres contextes que celui d'origine [DUM07]. L'utilisation de la technologie Web sémantique dans la phase de découverte des services Web donne naissance à la découverte sémantique des services Web. Cette dernière est basée sur le raisonnement sémantique. Elle ajoute à la précision des résultats de recherche traditionnels par rapport aux techniques de découverte de service Web, la précision supplémentaire de « matching » en termes de puissance de calcul [FRI05].

Cette thèse a pour objectif d'optimiser la découverte sémantique des services Web géographiques en réduisant l'espace de recherche et augmentant le nombre de services pertinents. Pour cela nous avons exploité des catalogues de métadonnées. Ces dernières sont basées sur la norme ISO 19119 renforcées par les critères de qualité.

Notre contribution répond à notre problématique de la découverte sémantique des services Web géographiques à travers deux propositions :

- **Une approche de découverte sémantique des SWG basée sur les métadonnées à travers les SMA :** Notre approche fait appel à quatre domaines en pleine évolution: les SMA, les SIG, le Web sémantique et les services Web. Ce couplage permet de profiter pleinement de la puissance des SMA dans la modélisation, la richesse des SIG comme outils de traitement de données géographiques, la prise en charge de l'aspect sémantique par le Web sémantique et l'interopérabilité par les services Web grâce à ses standards (WSDL, SOAP, UDDI). L'approche proposée repose sur la coopération d'un ensemble d'agents utilisant les capacités sémantiques des services Web géographiques à travers les catalogues basés sur les métadonnées selon la norme ISO19119. Les agents de notre architecture sont conformes aux propriétés générales des agents (Portabilité, Autonomie, Stabilité, Persistance, Flexibilité, Sociabilité).
- **Une approche d'optimisation de la découverte sémantique des SWG basée qualité :** Cette approche représente une optimisation de l'approche précédente en faisant appel à la qualité et la personnalisation. La qualité joue un rôle primordial pour satisfaire les besoins

explicites ou implicites de l'utilisateur [TRI07, GUT07, BET05] et la personnalisation facilite l'expression du besoin utilisateur et rend l'information sélectionnée intelligible à l'utilisateur et exploitable [KOS7, BOU04]. Dans l'objectif de profiter pleinement de la qualité, la personnalisation et des normes assurant l'interopérabilité telle que l'ISO 19119 qui propose une taxonomie des services web géographiques [YUQ08], nous proposons notre approche qui consiste à renforcer les métadonnées exploitées dans [REZ09] par les critères de qualité aux différents niveaux (données, service et usage).

- **Un prototype QGWSSD (Quality Geospatial Web Service Semantic Discovery):** Le QGWSSD est le prototype de validation de notre travail, pour lequel nous avons choisi le domaine de tourisme comme champ d'application. Ce prototype est une application Web en adaptant une architecture Web classique "n-tiers". Il est basé sur l'exploitation des métadonnées et l'algorithme de «matching». Pour rendre ces métadonnées persistantes, nous avons choisi le couple XML/XML Schema, basé sur le principe des langages à balises de la famille SGML. L'application de l'algorithme de «matching» est appliquée à deux niveaux :
 - Au niveau métadonnées, l'algorithme de «matching» LARKS [SYC02] adapté en appliquant les trois premiers filtres (contexte, profil, similarité) sur les critères de qualité figurant dans la requête et les descriptions des classes de SWG disponible dans le catalogue de métadonnées.
 - Au niveau registre UDDI, l'application de cet algorithme [PAO02] est basée sur le mécanisme d'interrogation du registre UDDI en faisant appel aux différentes pages (pages blanches, pages jaunes, pages vertes).

Une liste de SWG classée selon l'ordre de correspondance (aspect qualité, aspect fonctionnel) représente le résultat de la découverte.

Un ensemble d'interfaces est mis à la disposition de l'utilisateur (Fournisseur/client / Utilisateur), il est réalisé en utilisant le PHP qui assure une simplicité d'interfaçage avec des bases de données. Cet ensemble contient:

- des interfaces d'acquisition de connaissance du profil utilisateur;
- des interfaces de publication des SWG;
- des interfaces de découverte des SWG.

Organisation du document

Ce document est organisé en deux parties, décrivant respectivement l'état de l'art et notre contribution. La première partie du document contient trois chapitres :

- le premier chapitre est consacré à la présentation de la découverte sémantique des services Web géographiques en faisant un survol des systèmes d'information géographiques, des services Web, des services Web géographiques ainsi que l'aspect sémantiques des SWG et la norme ISO 19119. La découverte sémantique des services Web géographiques et les différents travaux existant dans ce domaine représentent la deuxième partie.
- le deuxième chapitre décrit la qualité des services Web géographiques à travers la présentation de la notion de qualité à travers ses différents niveaux (qualité de données, qualité de services et qualité d'usage) exploités dans la phase de découverte ainsi que les travaux relatifs à ce domaine.
- le troisième chapitre propose un état de l'art sur le profil utilisateur en spécifiant le profil utilisateur, sa définition, ses composantes et ses différents méta-modèles, ainsi qu'une présentation détaillée de la dimension qualité du profil utilisateur et la requête géographique.

La deuxième partie est consacrée à notre contribution, elle présente les deux approches proposées ainsi que la mise en œuvre de la deuxième approche. Elle est composée de trois chapitres:

- le quatrième chapitre présente l'approche de découverte sémantique des services Web géographiques basée sur les métadonnées à travers les SMA. Dans ce chapitre nous faisons en premier lieu un survol des concepts exploités dans notre travail à l'image des ontologies et les SMA ainsi qu'une présentation du couplage SMA/SIG, le couplage SMA/SW et le couplage SMA/SWG. En deuxième lieu nous présentons l'approche proposée en indiquant l'architecture du modèle suivie de l'aspect architectural et comportemental de chaque composant.
- le cinquième chapitre fait l'objet de la description de l'optimisation de l'approche précédente en se basant sur l'aspect qualité en présentant l'aspect architectural et fonctionnel, le modèle de services Web Géographiques et les critères de qualités exploités ainsi que la requête utilisateur.
- le sixième chapitre met en évidence l'approche de découverte sémantique des services Web géographiques dans le domaine du tourisme. Dans lequel nous spécifions l'architecture du modèle, les outils techniques exploités ainsi qu'une description détaillée du prototype en précisant ses fonctionnalités et ses interfaces.

Nous avons conclu notre manuscrit par une conclusion générale qui discute les apports de notre travail, ainsi que les perspectives de recherche envisagées.

*Chapitre 1: La découverte sémantique
des services Web géographiques*

Chapitre 1: La découverte sémantique des services Web géographiques

Le couplage des systèmes d'information géographique (SIG) et les services Web (SW) fait apparaître les services Web géographique (SWG). L'enrichissement sémantique des SWG permet de profiter pleinement de leur capacité assurée par la richesse des SIG comme outils de traitement de données géographiques, la prise en charge de l'aspect sémantique par le Web sémantique et l'interopérabilité par les services Web grâce à ses standards (WSDL, SOAP, UDDI).

Ce chapitre est organisé de la façon suivante : la section 1 est consacrée à la présentation des Systèmes d'Information Géographique (SIG). Ensuite, la section 2 fait l'objet de l'étude des services Web en indiquant leur architecture, les standards de cette technologie ainsi que les opérations appliquées sur les SW. La section 3 décrit les services Web géographiques à travers la définition d'un SWG, la sémantique des SWG ainsi que la norme ISO 19119. La section 4 est réservée à la découverte sémantique des services en présentant la découverte, la découverte sémantique et la découverte sémantiques des SWG ainsi que les algorithmes de « matching » et certains travaux relatifs.

1.1 Les Systèmes d'Information Géographiques (SIG)

L'objectif de cette section est de donner un aperçu des principales fonctionnalités des systèmes d'information géographique (SIG). Pour cela, nous commençons par donner des définitions sur l'information géographique et les SIG. Ensuite, nous décrivons brièvement leurs différentes fonctionnalités ainsi que le système de gestion de base de données géographiques.

1.1.1 L'information géographique

L'information géographique peut être vue comme une abstraction de la surface terrestre. Elle représente des objets géographiques de différentes natures (bâtiments, routes, rivières, etc.) dans un espace, de manière à pouvoir les localiser. Elle est définie par le CNIG comme l'« information qui est reliée à une localisation sur la Terre, exprimée par rapport à un système de référence » [CNIG 94 cité dans HUB02], et par Denègre et Salgé [SAL96] comme un ensemble reliant :

- une information relative à un objet ou un phénomène du monde terrestre, décrit plus ou moins complètement par sa nature, son aspect et ses attributs (p. ex. un bâtiment décrit par sa hauteur, son nombre d'étages, sa fonction, etc.) ; cette description peut inclure des relations avec d'autres objets ou phénomènes (p. ex. ce bâtiment appartient à telle commune, etc.) ;
- et sa localisation sur la surface terrestre, décrite dans un système de référence explicite (p. ex. système de coordonnées ou adresse postale).

À cette définition peut être ajoutée la composante temporelle : on précise alors que la localisation de l'information est valable à un moment donné [YAN06].

L'apparition de l'information géographique sur le Web est très intéressante pour la gratuité de certaines données, pour l'accès direct à des cartes et à des fonctionnalités, pour la visualisation de données en temps réel ou pour répondre à des questions d'ordre géographique. Cette apparition est assurée à travers une multitude de sites Web à composante géographique dont le potentiel d'utilisateurs est très important avec des profils très différents par leurs compétences, leurs langues ou leurs cultures. Ces sites peuvent être classés comme étant des *sites archives* pour référencer d'autres sites Web géographiques, des *sites catalogues* pour accéder à des grandes quantités d'informations géographiques, des *sites vitrines* pour montrer le savoir-faire en matière d'information géographique ou des *sites de services* pour utiliser des applications géographiques liées principalement à la cartographie [HUB02].

1.1.2 Les Systèmes d'Information Géographiques (SIG)

Un SIG est un ensemble de logiciels qui permettent de saisir, de stocker, de visualiser de l'information géographique numérique. La norme EdiGéo définit un SIG comme étant un "système pour saisir, stocker, vérifier, intégrer, manipuler, analyser et visualiser des données qui sont référencées spatialement à la terre. Il comprend en principe une base de données localisées et les logiciels applicatifs appropriés." [HUB02].

Il est défini aussi comme un environnement conçu pour l'analyse et la modélisation de la distribution spatiale de phénomènes. Il se compose d'une base de données géographique (BDG), d'une boîte d'outils contenant des procédures d'analyse, de gestion, de saisie et de représentation ainsi que d'une interface utilisateur.

Le but d'un SIG est d'offrir un système capable de répondre aux diverses questions de base sur les informations géographiques [HUB02] :

- **Où ?** : Cette question permet de mettre en évidence la répartition spatiale d'un objet géographique. Où se trouve cette intersection?
- **Quoi ? Qui ? Quel(le) ?** : Cette question renseigne sur les objets géographiques présents dans une zone donnée. Quelle est le nombre de voies de cette route?
- **Comment ?** : Ici, une référence est faite à l'analyse spatiale. Comment limiter les accidents?
- **Quand ?** : Cette question informe sur l'aspect temporel des objets géographiques. Quand cette autoroute a été mise en service?
- **Et si ?** : Cette dernière question permet de faire des prévisions ou des simulations sur l'avenir. Si le niveau de l'Orne augmente de 3 mètres, quelles seront les villes touchées?

Il est possible de distinguer trois types de SIG [YAN06] :

- Les SIG généralistes bureautiques ont pour vocation essentielle l'import de données externes et leur analyse pour donner des cartes à insérer dans des rapports ou des

présentations.

- Les SIG généralistes de gestion disposent des mêmes capacités que les SIG bureautiques, sont fréquemment moins conviviaux, mais disposent d'outils de modélisation beaucoup plus puissants, qui vont imposer des contraintes à la saisie et donc assurer une certaine qualité des données. Ces SIG vont également disposer de capacités client/serveur permettant à plusieurs personnes en réseau de travailler sur la même base de données à partir de postes informatiques distants. Ils disposent d'outils de développement pour s'adapter à tout type d'application.
- Les SIG métiers sont, dès le départ, très spécialisés, destinés à des métiers particuliers. Leur champ d'application est réduit mais ils sont souvent les seuls ou les meilleurs dans leur domaine.

Un système d'information géographique se divise généralement en cinq principaux éléments qui sont :

- *Matériel*: le développement de l'informatique graphique a permis d'utiliser les SIG dans une grande variété d'ordinateurs connectés en réseau ou autonomes.
- *Logiciel*: plusieurs fonctionnalités sont offertes par les logiciels SIG.
- *Données*: les SIG manipulent des données géographiques et des données tabulaires associées pouvant être constituées en interne, ou acquises depuis les producteurs de données. Elles sont de type:
 - *Données factuelles*: un SIG doit permettre la manipulation des données de nature alphanumérique liées aux concepts géographiques, des noms de lieux, des valeurs numériques comme la population, les surfaces etc.
 - *Données géométrique*: liées à la spécialisation des concepts dans un univers réel à deux ou trois dimensions, et sa représentation devient tributaire d'un système de coordonnées (référentiel) et de son environnement (relation spatiale), grâce à des primitives de représentation des informations géométriques.
- *Utilisateurs*: Les SIG s'adressent à une très grande communauté d'utilisateur depuis ceux qui créent et maintiennent les systèmes jusqu'aux personnes qui utilisent quotidiennement la dimension géographique. Avec l'avènement des SIG sur Internet, la communauté des utilisateurs de SIG s'agrandit de façon importante chaque jour. Il est raisonnable de penser à une brève échéance nous serons tous à des niveaux différents d'utilisateurs de SIG.
- *Méthodes*: Le succès de la mise en application d'un SIG est fonction de la bonne conception du plan de sa mise en œuvre selon des méthodes et des pratiques propres à chaque organisation.



Figure 01: Les composantes d'un SIG

1.1.3 Les fonctionnalités du SIG

La classification de [SAL96] se décompose en cinq grands domaines de fonctionnalités SIG : les "5A".

- l'**A**cquisition des données : ensemble de fonctions permettant de saisir des données géographiques suivant différentes sources;
- l'**A**rchivage des données : stockage de l'ensemble des données dans des bases de données aptes à contenir des données géo référencées et sémantiques;
- l'**A**bstraction des données : modélisation des données géographiques au niveau des bases de données et de l'échange des données;
- l'**A**nalyse des données : analyse fournissant un ensemble de fonctions telles que le croisement, la transformation, la manipulation pour pouvoir ressortir de nouvelles informations;
- l'**A**ffichage des données : interface entre l'utilisateur et le cœur du système pour présenter des graphismes, des données et informations relatives à l'information géographique.

1.1.4 Les systèmes de gestion de base de données géographiques (SGBD Géographique)

Les données géographiques sont stockées dans des bases de données géographiques (BDG). Une BDG est un ensemble d'objets géographiques organisé de manière à être manipulé dans un SIG. La gestion et l'accès à une base de données (BDG) est assuré par un SGBD géographiques. Ce dernier est un SGBD, il offre un type de donnée spatial dans son modèle de données et son langage de requêtes et il implémente ce type et ses opérateurs, fournissant au moins l'indexation spatiale et des algorithmes efficaces pour la jointure spatiale [ZEI00].

Les SGBD Géographiques peuvent être vus comme une extension des SGBD relationnels comme indiqué dans le tableau suivant :

	SGBD relationnel classique	SGBD spatial
Données	Entier, Réel, Texte, ...	Plus complexes: Point, Ligne, Région ...
Prédicats et Calculs	Tests : =, >, ... Calculs : +, /, ... et fonctions simples	Prédicats et calculs géom. et topologiques: Tests : intersecte, adjacent à, ... Fonctions géom. : intersection, surface...
Manipulation	Opérateurs de l'algèbre : Sélection, Projection, Jointure... Agrégats : Count, Sum, Avg...	Manipulation par thème ou inter-thèmes Sélection et jointure sur critère spatial Agrégats : fusion d'objets adjacents
Liens entre Objets	Par clés de jointures	Liens spatiaux (souvent) implicites
Méthodes d'accès	Index B-tree, hachage	Index R-tree, quad-tree, etc.

Tableau 01: Parallèle entre les SGBD relationnels classiques et les SGBD spatiaux [ZEI00]

1.2 Les services Web

Les applications logicielles d'entreprises deviennent de plus en plus distribuées, complexes et coûteuses en efforts de gestion. Pour remédier à ces insuffisances, il faut développer de nouveaux concepts technologiques contribuant au développement d'applications plus étendues et plus complexes sur un marché en mutation constante. C'est dans ce contexte que les concepts de la programmation orientée objet, l'approche composant et l'architecture orientée services sont apparues [BEN09].

1.2.1 Architecture Orientée Service (AOS)

L'AOS est une approche architecturale permettant la création des systèmes basés sur une collection de services développés dans les différents langages de programmation, hébergés sur différentes plates formes avec divers modèles de sécurité et processus métiers [BAR03]. Chaque service représente une unité autonome de traitement et de gestion de données, communiquant avec son environnement à l'aide des messages [BOO04]. L'AOS est axée autour de trois concepts fondamentaux, à savoir, le fournisseur des services, le client des services et l'annuaire de publication. Le fournisseur permet l'accès à son service à travers une interface [OSU02]. La technologie des services Web représente la technologie la plus utilisée pour migrer vers l'architecture AOS.

1.2.2 Les services Web

Selon [OSU02], un service Web est une agrégation de fonctionnalités publiées pour être utilisées. Il utilise Internet comme conduit pour réaliser une tâche. Il est semblable à un processus

métier virtuel qui définit des interactions au niveau application.

Selon [W3C04], un service Web est un système logiciel conçu pour supporter les interactions entre applications à travers le réseau. Les services Web offrent un moyen standard d'interopérabilité entre différentes applications qui s'exécutent sur une variété de machines/plates formes. Ils sont caractérisés par leur grande interopérabilité et extensibilité, ainsi qu'une description interprétable/compréhensible automatiquement par la machine grâce au standard XML. Ils peuvent être combinés d'une façon faiblement couplée afin de réaliser des opérations complexes. Les programmes offrant des services simples, peuvent interagir afin de mettre en place des services sophistiqués avec des valeurs ajoutés.

Dans notre travail nous nous intéressons à des services Web élémentaires sémantiques dans le domaine géographiques. Un service Web élémentaire est un service isolé, pouvant servir de services de base à une composition de services Web. Selon sa description, un service élémentaire peut être classique ou sémantique [LOP08].

- **Service Web élémentaire classique.** Un service Web est dit classique si sa description n'intègre que son (ou ses) profil(s) fonctionnel(s) (la description syntaxique des méthodes proposées, des paramètres de ces méthodes et des protocoles à utiliser).
- **Service Web élémentaire sémantique.** Un service Web est dit sémantique si sa description intègre, en plus de la description fonctionnelle, une dimension sémantique. Cette dimension englobe toute description complémentaire (du fournisseur, des objectifs du service Web, etc.) de la description fonctionnelle qui permet d'ajouter de l'information à la description classique du service Web.

1.2.2.1 Architecture des services Web

Les services sont implantés par les fournisseurs, qui mettent à disposition les descriptions de services sous forme de fichiers. Ces descriptions sont centralisées et stockées dans des annuaires. La notion d'annuaire est comparable aux annuaires téléphoniques. Les applications clientes envoient des requêtes aux annuaires pour sélectionner les services. Elles téléchargent ensuite les descriptions des services sélectionnés, et les invoquent directement [MRI07]. Ces mécanismes sont illustrés par la figure 02.

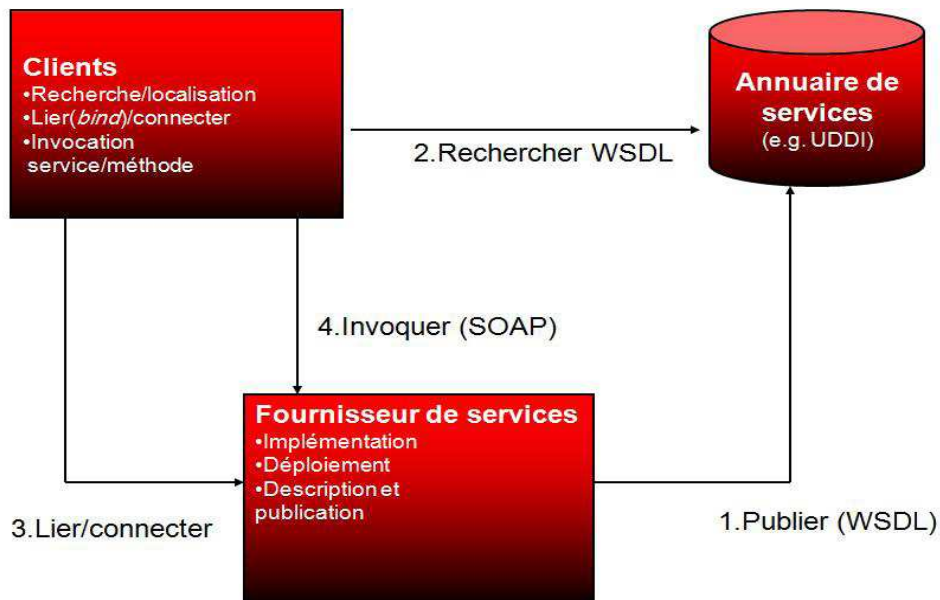


Figure 02 : Architecture des services Web

Pour garantir l'interopérabilité des trois opérations précédentes (publication, recherche et liaison/invocation), des standards ont été élaborés pour chaque type d'interaction.

1.2.2.1.1 WSDL: Web Service Description Language

Le standard de description des services Web proposé par le W3C est le langage WSDL (Web Service Description Language). WSDL, basé sur XML, permet de décrire le service Web, en précisant les méthodes disponibles, les formats de messages d'entrée et de sortie, et comment y accéder. Un document WSDL décrit un service Web en deux parties : une partie abstraite et une concrète. C'est à dire, la séparation de "quelle" fonctionnalité est fournie de "comment" et "ou" celle-ci est offerte. La figure 03 présente la structure fondamentale d'un document WSDL.

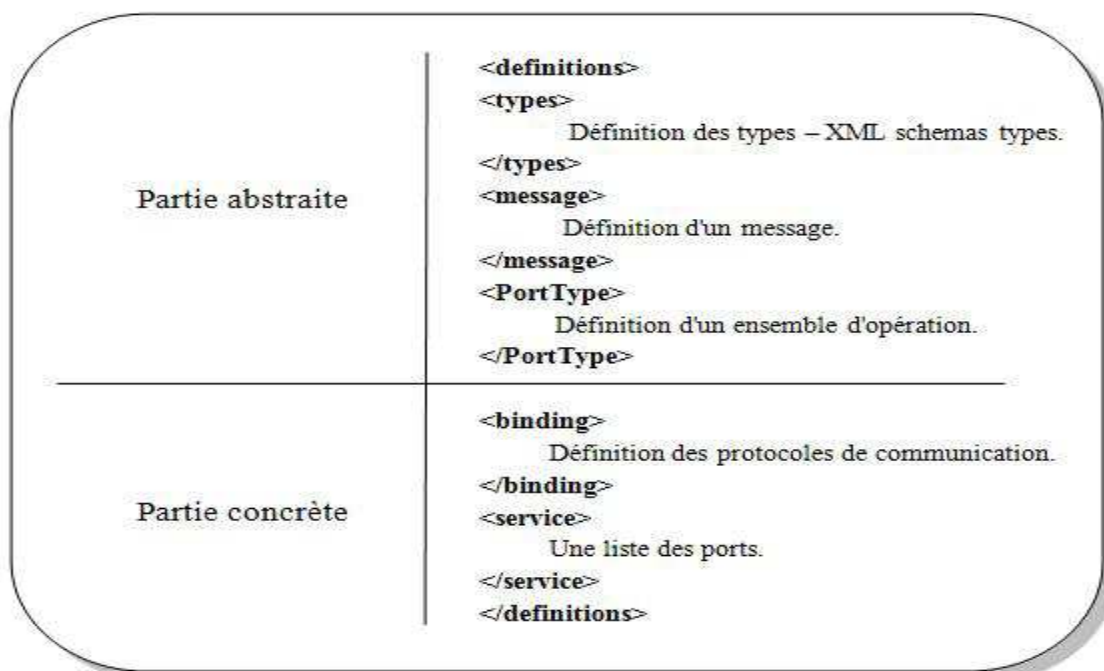


Figure 03 : Structure fondamentale d'un document WSDL

Pour la définition des services un document WSDL utilise les éléments suivants [CHR01]:

- *Types*: ils définissent des types de données échangées;
- *Message*: Il définit d'une manière abstraite des données transmises;
- *Opération*: elle décrit d'une manière abstraite les actions supportées par le service;
- *Port type*: il représente un ensemble d'opération correspondant chacune à un message entrant ou sortant;
- *Binding (Rattachement)*: est un protocole de communication et un format de données échangées pour un port;
- *Port*: est une adresse d'accès au service;
- *Service*: il regroupe un ensemble de port.

La relation entre les composants d'un document WSDL est illustrée dans la figure 04. Chaque document WSDL définit un service comme une collection de points finaux ou ports. Chaque port est associé à une liaison spécifique qui définit la manière avec laquelle les messages seront échangés. Chaque liaison établit une correspondance entre un protocole et un type de port. Ce dernier se compose d'une ou plusieurs opérations qui représentent une définition abstraite des capacités fonctionnelles du service. Chaque opération est définie en fonction des messages échangés au cours de son invocation (inputs/outputs) [MEL 04].

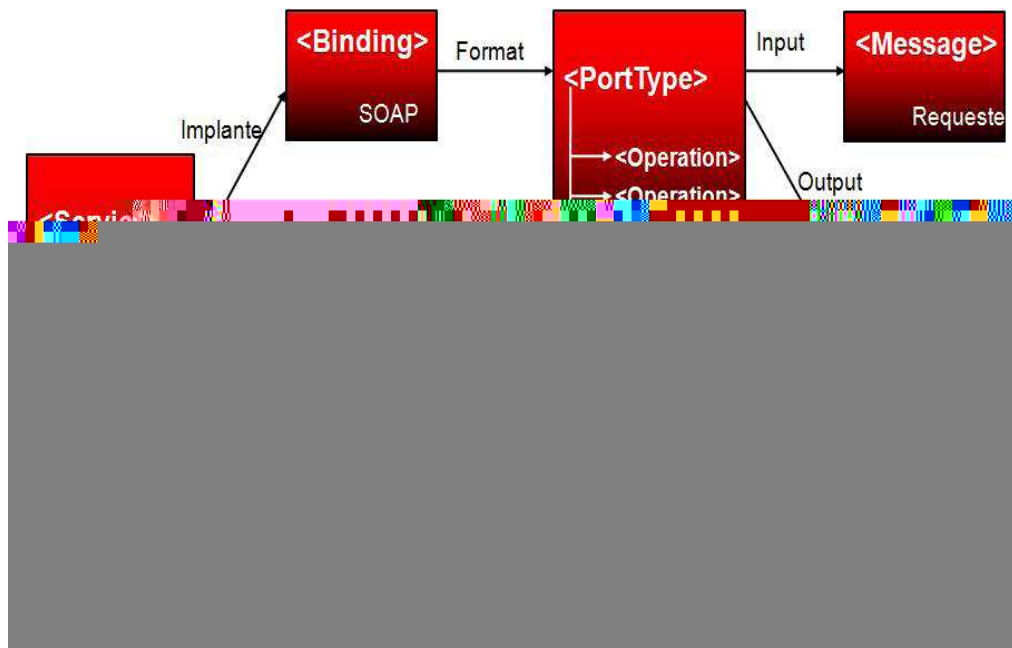


Figure 04 : Relation entre les éléments d'un document WSDL

1.2.2.1.2 UDDI: Universal Description, Discovery and Integration

Une fois le service Web décrit par son fournisseur, pour qu'il puisse être utilisé (et réutilisé) par le plus grand nombre de clients, il doit être publié dans un registre public. Nous présentons le registre UDDI qui a un ensemble de spécifications pour la mise en œuvre de registres pour les

services Web, couramment utilisé dans l'implémentation de registres internes aux organisations. Il définit comme un annuaire de service accessible sur le Web décrivant les sociétés et les services proposés, aussi bien d'un point de vue fonctionnel que technique, permettant aux utilisateurs de découvrir les services disponibles. Cet annuaire définit comment réagir avec un service et permet une immatriculation mondiale. Il peut être implanté au niveau mondial sur Internet ou au niveau d'entreprise sur intranet. Parmi le panel de registres accessibles via le Web (tels que *Web Service List*), nous citons *XMethods* et *RemoteMethods* (détaillés dans [LOP08]).

Le protocole UDDI est une plateforme destinée à stocker les descriptions des services Web disponibles, à la manière d'un annuaire de style "Pages Jaunes". Des recherches sur les services peuvent être effectuées à l'aide d'un système de mots-clés fournis par les organismes proposant les services. UDDI propose également un système de "Pages Blanches" (adresses, numéros de téléphone, identifiants...) permettant d'obtenir les coordonnées de ces organismes. Un troisième service, les "Pages Vertes", permet d'obtenir des informations techniques détaillées à propos des services et permettent de décrire comment interagir avec les services en pointant par la suite vers une description WSDL [BAG03].

Comme le montre la figure 05, une description UDDI doit contenir quatre catégories de données [OAS02] :

- *BusinessEntity* (*entité d'affaires*): information sur le fournisseur du service.
- *BusinessService* (*service d'affaires*): information décrivant une famille particulière de services techniques.
- *BindingTemplate* (*modèle de rattachement*): information technique sur les points d'accès aux services et leur implémentation.
- *TModel*: (modèle technique) description des spécifications des services.

Figure 05: Structures de données “noyau” d’UDDI (extrait de [OAS02])

Le modèle de description UDDI semble très orienté “commerce” (ventes, prestations), donc a priori éloigné des besoins de consultations de traitements géographiques [YAN06]. Les annuaires UDDI existent sous une forme XML. Le modèle de description UDDI est défini dans un schéma XML. Des API sont disponibles pour interroger les annuaires UDDI et y référencer des services Web [OAS02]. Le protocole d'utilisation de l'UDDI contient trois fonctions de base [BEN09]:

- *publish* pour enregistrer un nouveau service;
- *find* pour interroger l'annuaire;
- *bind* pour effectuer la connexion entre l'application cliente et le service.

La relation entre le WSDL et l’UDDI est assurée par les correspondances (interface du service (WSDL) -> tModel (UDDI) et implémentation du service (WSDL) -> Business Service (UDDI)) comme illustré dans la figure 06.

Figure 06 : Relation entre WSDL et UDDI

1.2.2.1.3 SOAP: Simple Object Access Protocol

SOAP (Simple Object Access Protocol) [SOA07] est un protocole léger destiné à l'échange d'informations structurées dans un environnement distribué. Il utilise des technologies XML pour définir une structure de messages pouvant être échangés sur divers protocoles sous-jacents (p.ex. HTTP, SMTP). Cette structure a été conçue pour être indépendante de tout modèle de programmation et autre sémantique spécifique d'implémentation. Un message SOAP est une transmission unidirectionnelle entre des nœuds SOAP (d'un émetteur vers un récepteur SOAP). Un message SOAP contient deux sous-éléments à l'intérieur de l'élément *Envelope* externe: un élément *Header* (en-tête) et un élément *Body* (corps) (voir figure 07). Les contenus de ces éléments sont définis par l'application et ne font pas partie de la spécification de SOAP.

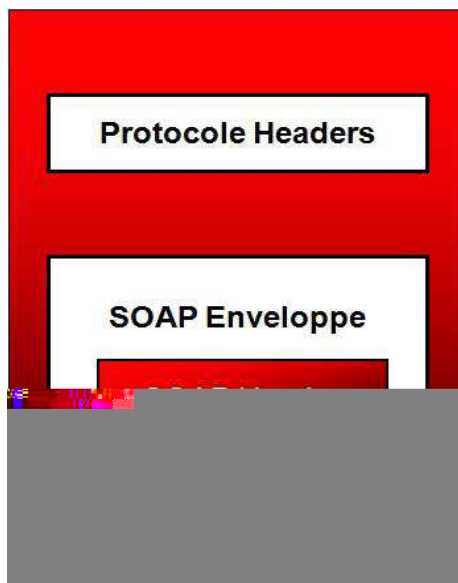


Figure 07 : Structure d'un message SOAP

L'objectif du protocole SOAP est de permettre l'invocation des services Web, qui offrent des opérations métier sur des ports de connexion. Couplé à WSDL, le SOAP fournit à l'application cliente des outils nécessaires pour invoquer des services distants en lui donnant l'illusion qu'ils sont locaux.

1.2.2.2 Les opérations des services Web

Un service Web peut subir les opérations suivantes [BAR03]:

1.2.2.2.1 La description

Afin que le service Web déployé puisse être visible, dans le but d'être utilisé (et réutilisé), le fournisseur doit le décrire à l'aide du standard WSDL. Cependant, les informations fonctionnelles du service Web contenues dans une description WSDL (nom et paramètres des méthodes proposées).

1.2.2.2.2 La publication

La recherche des services en vue de les utiliser (et réutiliser) repose sur leur publication

préalable dans un registre. Les différents composants de la publication faite par UDDI sont des documents XML manipulant de l'information à propos du fournisseur (*Business Entity*), le service lui-même (*Business Service*), les accès au service (*Binding Template*), le type de service (*tModel*) et des relations entre deux parties (*Publisher Assertion*) (Figure 05).

1.2.2.2.3 La découverte

La découverte est un point clé nécessaire à la réutilisation des services élémentaires dans une AOS. Les services sont conçus afin d'être sélectionnés via des mécanismes de recherche. La recherche et la sélection dans UDDI reposent sur la publication préalablement décrite du service et de son fournisseur. Le futur client peut connaître par l'intermédiaire de l'UDDI : les fournisseurs d'un service, les services proposés par un fournisseur donné, les moyens d'invoquer un service. Pour apporter aux clients la réponse à ces questions, UDDI organise l'ensemble des informations qu'il contient en trois parties (les pages blanches, les pages jaunes et les pages vertes), spécifiées en XML.

1.2.2.2.4 L'invocation

Lorsque le client connaît au préalable le service Web qu'il veut invoquer (par l'intermédiaire des étapes précédentes telles que la recherche et la sélection), son invocation peut être réalisée via le protocole de communication standard SOAP.

2.2.2.5 La composition

Des collections de services peuvent être coordonnées et assemblées afin de former une composition de services. Cette possibilité de construire de nouveaux systèmes à partir de services existants constitue l'un des avantages de l'AOS.

1.3 Les services Web géographiques

L'exploitation traditionnelle de l'information géographique rencontre pas mal de problèmes comme cité dans [ZHA07] à savoir :

- les informations géographiques sont livrées dans des fichiers de base non consommables, regroupés selon le format et la taille du fournisseur, ce qui conduit à une surcharge du réseau;
- le processus d'obtention des informations géographiques est intensif. Avec l'approche traditionnelle l'utilisateur doit passer par un ensemble de processus qui ne peuvent pas être automatisés par traitement des lots (batch-processing plans). Ces processus sont présentés comme suit :
 1. identification des sources de données et leurs catalogues;
 2. recherche de façon interactive dans l'ensemble de données à plusieurs tentatives dans l'objectif de restreindre le jeu de données souhaité;
 3. téléchargement des données attendues;

4. transformation et importation des données vers le système exploité par l'utilisateur;
 5. le traitement des données et présentation des résultats.
- l'information géographique n'est pas interopérable. Les informations sont stockées dans un format propriétaire au vendeur, et un traitement spécifique doit être pris en charge avant son utilisation dans le système utilisateur.
 - la difficulté de partager la valeur ajoutée de l'information géographique.

Les services Web peuvent réduire considérablement le volume des données et des ressources informatiques nécessaires à la fin du côté utilisateur [YUE06]. Ils représentent une solution des problèmes cités en profitant de ces capacités mentionnés comme suit [ZHA07]:

1. modularité pour le partage des fonctions spécifiques;
2. interfaces standards pour l'interopérabilité des applications;
3. capacité de la prise en charge des demandes des utilisateurs en matière d'informations géographiques;
4. disponibilité des fonctionnalités pour des opérations faiblement couplés sur l'ensemble de données.

1.3.1 Définition du service Web géographique

Le mariage de services Web avec des informations géographiques conduit à des services Web géographiques. En règle générale, un service Web géographique est défini comme une application modulaire Web qui fournit des services sur les données, les informations ou les connaissances géographiques. Elle se réfère à l'utilisation des technologies des services Web pour gérer, analyser et distribuer l'information géographique. En outre, un service Web géographique peut être trié et fouillé par ses caractéristiques géographiques, comme la localisation, la superficie, le voisinage et d'autres caractéristiques du territoire. Les interfaces de certains services Web géographiques ont été normalisées. Les acteurs les plus importants pour les efforts de normalisation sont l'ISO/TC211 et de l'Open Geospatial Consortium (OGC) » [ZHA07].

1.3.2 La sémantique des services Web géographique

Un des grands principes du Web sémantique est qu'il est nécessaire d'associer aux ressources du Web des informations exploitables par des agents logiciels afin de favoriser l'exploitation de ces ressources. Deux termes principaux sont utilisés dans la littérature afin de décrire les informations associées à des ressources : métadonnées et annotations [PRI03]. Les métadonnées, qui sont « des ensembles de données structurées pour décrire, expliquer, localiser les ressources et en faciliter la recherche, l'usage et la gestion » [NIS04]. Dans le domaine de l'information géographique, les métadonnées sont définies comme suit : « ensemble d'informations décrivant une donnée, une série de données (lot de données) ou plus globalement une ressource. Elles se composent d'éléments relatifs à l'identification, la représentation spatiale, la qualité, le

contenu, les modalités d'accès et de diffusion, etc». En fonction de l'exigence attendue, différents niveaux de précision sont généralement distingués [CNI07]:

- les métadonnées de découverte qui permettent d'identifier les données,
- les métadonnées d'exploration qui caractérisent avec plus de détails les données disponibles et permettent d'estimer leur utilité dans le cadre d'un projet,
- les métadonnées d'exploitation qui apportent l'ensemble des informations nécessaires pour intégrer les données dans un SIG et les utiliser.

Les métadonnées des SWG peuvent être classées en trois niveaux: syntaxique, structurelle et sémantique [ZHA07]:

- Métadonnées syntaxique: décrivent les sources de base et des informations simples sur la paternité des SWG, tels que les données de format de fichier, date de création, des sources, taille, et les auteurs.
- Métadonnées structurelles: décrivent la structure des données géographiques et les fonctions des SWG comme les schémas XML, les types d'organisation des données (vecteur ou raster) et des types de fonctionnels (conversion, manipulation).
- Métadonnées sémantique: définissent le contexte ou le domaine des données géographiques et de SWG, comme le type de données thématiques (par exemple, des glissements de terrain dans les géosciences, la distribution de la population en géographie économique) et de types de domaines fonctionnels (par exemple, le calcul d'indices de végétation en matière de télédétection, l'estimation de la température de surface des terres dans le climat).

La sémantique des services Web géographiques est assurée à travers des classifications sémantiques. Parmi ces classifications nous présentons celles proposées dans [FOR05, PER02, YUQ08] comme suit :

1.3.2.1 Classification des Services Web Géographiques selon point de vue technique

Ce classement propose les catégories suivantes [FOR05]:

- Viewer (visionneuse) : accès en lecture aux géodonnées;
- Web Catalog Service (WCAS resp. WRS) : il permet la recherche de données selon des critères techniques, temporels et spatiaux (métadonnées);
- Web Map Service (WMS) : obtention de cartes sous la forme d'une image ou d'informations attributaires concernant des objets géographiques;
- Web Feature Service (WFS) : accès en lecture et en écriture à des géodonnées (vectorielles);
- Web Coverage Service (WCS) : il permet l'accès à des données tramées (raster)(sous forme brute);

- Web Gazetteer Service (WFS-G) : il établit la liaison entre des géoréférences indirectes (adresses, noms de lieux, unités administratives) et des géoréférences directes (coordonnées);
- Web Coordinate Transformation Service (WCTS) : les géodonnées sont transmises sous forme GML puis transformées dans un système de référence spatiale indiqué;
- Web Terrain Service : il génère des « vues » à la demande, sur des données en 2,5D ou en 3D;
- Applications spécifiques : routage par exemple ou solutions spécifiques à une branche donnée.

1.3.2.2 Classification des Services Web Géographiques selon la norme ISO :

La prénorme ISO (text for IS 19119 Geographic information – Services) classe les différents services de manière encore plus détaillée et distingue six classes de services [FOR05, PER02, YUQ08]:

- Human interaction services : services destinés au contrôle et à la gestion d'interfaces utilisateur, de graphiques, de données multimédia, etc (Catalogue viewer, Geographic viewer, Geographic feature editor)
- Model/Information management services : services de gestion du développement, de traitement et de mémorisation de métadonnées, de concepts et de jeux de données (Feature access service, Map access service, Coverage access service, Catalogue service.);
- Workflow/Task services : services d'assistance aux utilisateurs dans le cadre de tâches et d'activités spécifiques (Chain definition service, Workflow enactment service);
- Processing services : Un service de traitement Web donne accès à des calculs et à des modèles qui traitent des données à référence spatiale.
- Communication services : services de codage / décodage et de transfert de données au sein de réseaux de télécommunication;
- System management services : services de contrôle de composants de systèmes, d'applications et de réseaux (exemple : gestion de comptes d'utilisateurs).

1.3.3 Les métadonnées des services Web géographiques selon la norme ISO 19119

Les métadonnées exploitées dans notre travail sont basées sur la norme ISO 19119 qui représente l'une des normes du modèle de référence du comité technique de l'ISO/TC 211 relatif à l'information géographique. Ce modèle propose d'utiliser les standards des technologies de l'information en l'appliquant à l'information géographique. Il assure plusieurs avantages comme indiqué dans [SER06]. La norme ISO 19119 offre un cadre de développement de logiciel permettant aux utilisateurs l'accès et le traitement des données géographiques provenant de sources

diverses à travers une interface générique au sein d'un environnement ouvert [PER02].

L'architecture des services géographiques spécifiés dans la norme ISO 19119 a été développée pour répondre aux objectifs suivants [PER02] :

- fournir un cadre abstrait pour permettre un développement coordonné de services spécifiques,
- fournir des services interopérables à travers des interfaces standards,
- soutenir le développement de catalogue de services à travers les métadonnées,
- assurer la séparation en les données et les services,
- permettre l'utilisation les services d'un fournisseur sur les données d'un autre fournisseur;
- définir un cadre abstrait, qui peut être mis en œuvre de multiples façons.

Les deux parties principales de la norme ISO 19119 sont constituées par les diagrammes de classes et par le catalogue des objets (Data Dictionary). L'ensemble des éléments de descriptions ISO 19119 est assez succinct (figure 08) [YAN06];

La structure d'une description d'un service comprend 3 classes principales :

- description générale du service SV Service Identification;
- description des opérations SV OperationMetadata;
- description des entrées et sorties du service MD DataIdentification.

Le diagramme de classe de la figure 08 montre les classes en question, ainsi que SV Parameter et SV OperationMetadata. La classe MD DataIdentification décrit les données ; elle appartient à ISO 19115.

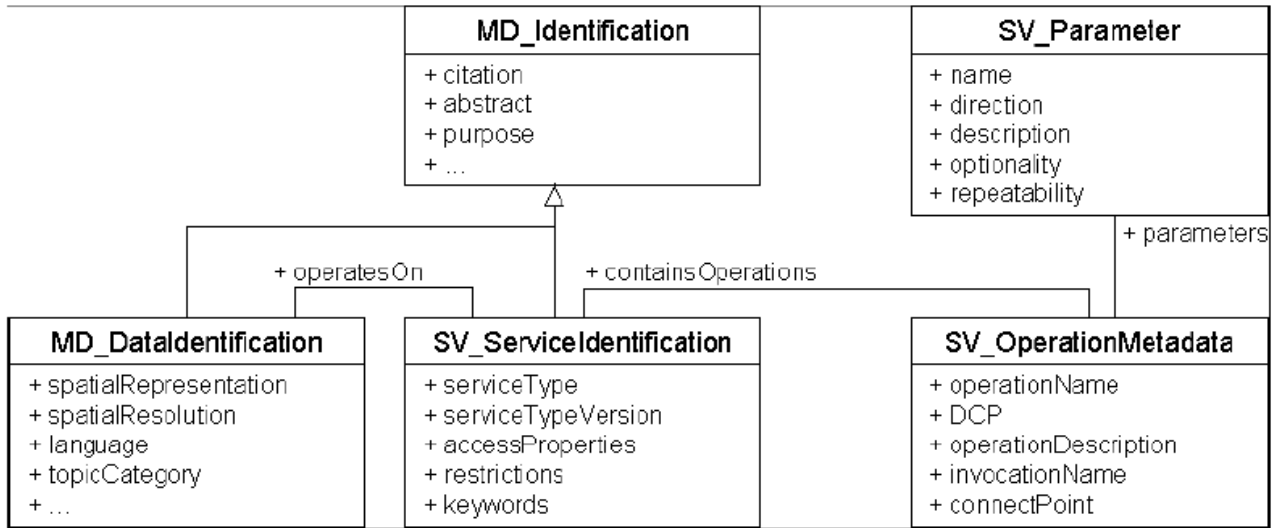


Figure 08 : Les classes principales ISO 19119 pour la description d’un service (d’après [ISO01])

Comme le montre le tableau 02, ISO 19119 prévoit que les consultations des descriptions de service puissent s’effectuer selon quatre points de vue différents. Ces points de vue correspondent à ceux définis par ISO/IEC 10746 (Reference Model–Open Distributed Processin) [YAN06, PER02].

Point de vue informatique	Chaînage des services : Définir des services avec des interfaces réutilisables : Service Metadata, Service/Data coupling Service Chaining
point de vue informationnels	Interopérabilité sémantique : Une taxinomie de services géographiques est définie : – éditeurs et visualisation (viewers) de données destinés aux humains, – simples fournisseurs de données (vecteur, raster ou texte), – services réalisant des traitements (spatiaux, thématiques, temporels, etc.), ... – services utiles à la communication (p.ex. encodage).
Point de vue ingénieur	Distribution des services : Cette classification proposée par l’OGC des services par catégorie recoupe celle du point de vue “information”, mais du point de vue de la mise en œuvre : services avec IHM ou non, entre services effectuant des traitements ou délivrant des informations, etc.
Point de vue technologique	Spécification des services et de leur plateforme : protocoles pour l’interroprabilité des services (DCP – Distributed Computing Platform).

Tableau 02 Description d’un service Web géographique selon les points de vue ISO/IEC 10746[YAN06]

Selon la norme ISO 19119 les champs indiqués dans le tableau 03 sont obligatoires pour définir les métadonnées d’un service Web géographiques et peuvent êtres complétées en faisant appel au AMFprofil défini par l’OGC : « OpenGIS Catalogue Services Specification 2.0.2 - ISO

Metadata Application Profile (1.0.0) » disponible sur le site de l'OGC [GEO07].

N°	Désignation du champ	Utilisation
01	Identifiant de la métadonnée	Numéro unique permettant de différencier chaque métadonnée par adhérent au sein du Géocatalogue
02	Titre du service	Libellé du service, s'affiche dans les tableaux de résultats,...
03	Langue de la métadonnée	Par défaut, la langue est « FRA ». La valeur « fre » demandée par INSPIRE (ISO 629-2/B) est également gérée par le Géocatalogue.
04	Date de création / mise à jour de la métadonnée	Date à laquelle la fiche a été créée ou actualisée.
05	Jeu de caractères des métadonnées	Jeu de caractères du fichier contenant les métadonnées
06	Résumé sur le service	Texte décrivant en quelques lignes le contenu du service. Cette information s'affiche dans le tableau de résultats.
07	Catégorie des données / services	Principale(s) thématique(s) de la donnée selon une typologie définie par la norme 19115
08	Niveau de sécurité des services	Information permettant d'indiquer si le service est accessible avec ou sans contrainte
09	Date de création ou de publication de la donnée	Information afin de dater les données disponibles dans le Géocatalogue
10	Point de contact sur la métadonnée	Nom d'un organisme (ou d'une personne) assurant la fonction de contact pour la métadonnée.
11	Niveau hiérarchique du service	Fixé à un service
12	Type de service	Décrit le type de service et permet de gérer les regroupements par service. Dans le Géocatalogue, les types suivants sont interprétés : Site Web / Services cartographiques (WMS, WFS, WMS-C, WPS) / Service de téléchargement
13	Paramètre des services	Contient au minimum l'adresse Internet vers le service + informations techniques du service

Tableau 03: Les éléments des métadonnées représentant les Services Web Géographiques [GEO07]

1.4 La découverte sémantique des services Web géographique

Cette section est réservée à la présentation de la découverte, découverte sémantique et la découverte sémantique des SWG.

1.4.1 La découverte

Les services sont conçus afin d'être sélectionnés via des mécanismes de recherche. La découverte est permise par la description préalable des services et leur publication au sein d'un registre. Elle est effectuée en deux phases [LOP08]:

- La recherche: Elle peut se faire via les registres dans lesquels les fournisseurs ont publiés leurs services Web;
- La sélection: Le client doit ensuite sélectionner parmi la collection de services Web issus de l'étape de recherche, le service Web qui convient le mieux à ses attentes.

La recherche et la sélection dans UDDI reposent sur la publication préalablement décrite du service et de son fournisseur. Le futur client peut connaître par l'intermédiaire de l'UDDI : les fournisseurs d'un service, les services proposés par un fournisseur donné, les moyens d'invoquer un service. Pour apporter aux clients la réponse à ces questions, UDDI organise l'ensemble des informations qu'il contient en trois parties, spécifiées en XML. Chacune d'elles peut être utilisée pour faire une recherche via UDDI. Ces parties sont les suivantes [LOP08] :

- *Les pages blanches (White Paper)*: Ce composant permet de connaître les informations à propos de l'organisation proposant le service. Cette description contient toutes les informations jugées pertinentes pour identifier l'organisation (telles que son nom, son adresse physique). Le futur client du service retrouve dans les pages blanches les informations que le fournisseur a renseignées dans l'élément Business Entity lors de la publication.
- *Les pages jaunes (Yellow Paper)*: Les pages jaunes de l'UDDI détaillent la description de l'organisation faite dans les pages blanches en répertorient les services proposés. Dans cette section, sont décrits : la catégorie de l'entreprise, le secteur d'activité dans lequel exerce l'entreprise, les services offerts par cette organisation, le type de services et les conventions d'utilisation – prix, qualité de service, etc. Cette description repose sur les classifications standard de l'industrie nord américaine (NAICS28 et UNSPSC29). La description des services contenue dans les pages jaunes est non technique et renseignée par les fournisseurs eux mêmes.
- *Les pages vertes (Green Paper)*: Les pages vertes comportent les informations techniques liées aux services Web et basées sur leur description WSDL.

1.4.2 La découverte sémantique

Il s'avère malheureusement que la découverte des services Web est basée seulement sur la syntaxe. L'aspect sémantique indispensable pour satisfaire l'utilisateur y est absent [FRI05]. L'utilisation de la technologie Web sémantique dans la phase de la découverte des services Web donne naissance à la découverte sémantique des services Web. Cette dernière est basée sur le raisonnement sémantique. Elle ajoute à la précision des résultats de recherche traditionnels par rapport aux techniques de découverte de service Web, la précision supplémentaire de « matching » en termes de puissance de calcul [FRI05].

La découverte sémantique des services Web est définie par A. Friesen et M. Altenhofen dans [FRI05] comme suit : « Semantic discovery of Web Services means semantic reasoning over a knowledge base where a goal describes the required web service capability as input. Semantic discovery adds accuracy to the search results in comparison to traditional Web Service discovery techniques, which are based on syntactical searches over keywords contained in the Web service descriptions. The additional accuracy of a match is expensive in terms of required computational power. The expensiveness of semantic matchmaking has several aspects influencing the design of the Semantic Web Services infrastructure in different ways ».

1.4.2.1 Les travaux relatifs

Plusieurs travaux sur la découverte sémantique des services Web ont été proposés, entre autres nous citons certains concernent la description sémantique des services, d'autres se sont focalisés sur les langages d'interrogation et l'intégration de la description sémantique.

- OWL-S (Ontology Web Language Service) [OWL04] définit une ontologie de services Web, basée sur le langage OWL. Une instance de cette classe est décrite par trois concepts à savoir :

- ServiceProfile : qu'est-ce que le service fournit aux agents (ici des programmes clients) qui l'invoquent ? Qu'est-ce qu'il attend d'eux ?
- ServiceModel : comment fonctionne-t-il ? Quel est son modèle d'exécution ?
- ServiceGrounding : comment y accéder ? (descripteurs WSDL).

- SAWDL (Semantic Annotations for Web Services Description Language)[SEM07] est une extension des langages de définition WSDL et XML schema, et est une recommandation du W3C pour l'annotation sémantique des services Web. SAWSDL ne désigne pas un langage particulier décrivant le modèle sémantique. Il emploie les mécanismes d'extension WSDL2.0 et supporte WSDL1.1. Les spécifications SAWSDL se concentrent sur l'annotation sémantique de la définition abstraite d'un service et n'adressent pas l'annotation de l'implémentation du service.

- XMethods : Le site XMethods (<http://xmethods.com/>) proposé par l'organisation du même nom, est un registre de services Web hébergé sur Internet. Ce site permet de publier des services Web et de faire la recherche de services préalablement publiés. La publication des services sur le site XMethods repose sur une publication UDDI. Le fournisseur doit tout d'abord être inscrit auprès du site XMethods. Ensuite, le fournisseur doit décrire un document tModel et un document Business Service selon les spécifications UDDI et une extension spécifique au site XMethods.
- Benna et al [BEN08] présentent une approche qui utilise le SAWDL ce qui permet de ne pas dépendre d'un langage particulier du modèle de représentation sémantique. Aussi les mises en correspondance entre la requête et les services Web et les compositions inter-service sont connues au préalable, dans une couche décrite en XML et appelée base de liens sémantiques, ce qui réduit le temps de recherche lors de la phase découverte. Un autre avantage de cette solution est l'utilisation de XQUERY pour répondre à la requête utilisateur. Elle retourne, sous la forme d'un document XML, aussi bien les fonctionnalités du service Web que les informations relatives à sa localisation et à son invocation.
- Naveen et al [SRI04] propose une approche hybride utilisant l'UDDI renforcé par l'OWL-S. L'UDDI fournit un site Web d'un registre des services Web, mais son manque de capacité de représentation explicite et sa syntaxe de recherche pour autant obtenir des résultats qui sont grossiers dans la nature. L'approche est fondé sur le mécanisme de découverte sur OWL-S. Ce dernier (OWL-S) permet de décrire les services Web sémantique en termes de capacités d'offre et d'effectuer l'inférence logique de faire correspondre les capacités demandées avec les capacités offertes. Cette approche OWL-S/UDDI combine le meilleur des deux technologies.
- Dans [BAA05], les auteurs proposent une représentation de service Web particulière formalisée à l'aide d'une logique de description. Les auteurs définissent un formalisme avec lequel on peut raisonner sur les descriptions de services Web, et ainsi rechercher les services par le biais d'inférences. Ce formalisme repose sur deux notions : les conditions et les interprétations.
- Dans [MAS07], El-Masri et al proposent une approche de découverte de service Web dont l'objectif principal est de concevoir un système intelligent qui a le potentiel de l'examen des propriétés de la qualité du service Web d'une manière ouverte et transparente, et en permettant aux clients de sélectionner le meilleur service Web disponible.

Le tableau suivant présente une synthèse des travaux relatifs à la découverte sémantique des services Web.

		Aspects fonctionnels	Fourni -sreur	Description de la tâche du service		Description de l'utilisation du service	
				Catégorie du service	Objectif du service	Qualité de service	Pré- Post conditions
Langage de description de services	WSDL	+	-	-	-	-	-
	OWL-S	+	+	-	+	-	+
	SAWSDL	+	-	-	+	-	-
Registres accessibles via le Web	UDDI	+	+	+	-	-	-
	XMethods	+	+	-	+	-	-
	QWSDataset [MAS07]	+	-	-	-	+	-
Solution de publication , recherche et sélection	Naven et al [SRI04]	+	+	-	+	-	+
	Benna et al [BEN08]	+	+	+	+	-	-
	Aproche basée sur une logique de descriptions [BAA05]	+	-	-	++	-	+

Légende : + prise en compte / - Nom prise en compte

Tableau 04 : Tableau comparatif des solutions de la découverte sémantique des SW.

1.4.3 La découverte sémantique des services Web géographiques

Le couplage de l'OGC et le Web Sémantique a donné naissance au « *Semantic Geospatial Web* » [EGE02]. Cette discipline assure la capacité de représenter la sémantique géospatiale, qui est d'une grande importance lors de la construction des applications géospatiales pour le Web. Les services Web sémantiques fournissent des solutions de technologies intelligentes pour des services d'annotation, de la découverte, la composition et de l'invocation dans les environnements distribués. Déploiement de cette technologie dans des applications Web géographiques a le potentiel pour améliorer la découverte, la recherche et l'intégration de l'information géographique, ainsi que sa réutilisation dans d'autres contextes que celui d'origine [DUM07]. Cette nouvelle tendance vise à assurer la recherche Web et la récupération de données et les services géographiques non seulement fondée sur des informations syntaxiques ou lexicales, comme beaucoup d'approches actuelles, mais sur les informations sémantiques précisées par les métadonnées, les nomenclatures et geo-ontologies [BRE07]. La recherche est assurée à travers l'application des algorithmes de « matching ».

1.4.4 Les algorithmes de « matching »

L'algorithme de «matching» compare toutes les annonces à la requête, pour fournir comme résultat les annonces qui sont proches de la requête. Proche peut prendre plusieurs sens. En effet, le souhait est que notre algorithme de «matching» soit capable de reconnaître non seulement les services identiques mais d'autres services utilisables pour satisfaire la requête,...[PAO02]

Ces différents niveaux (exact match, plug in match, relaxed match) ne sont pas équivalents. Ils sont appelés « degré de correspondance » et ils varient d'après les mécanismes. L'algorithme permet généralement au client de fixer le degré de correspondance désiré entre sa requête et les annonces.

L'algorithme de «matching» doit permettre une recherche [PAO02]:

- précise: les résultats renvoyés par l'algorithme doivent correspondre aux définitions des degrés de correspondance;
- efficace: le temps de réponse doit être le plus court possible.
- Efficace: l'ensemble des résultats ne doit pas être trop grand. Mieux vaut un petit ensemble avec un degré de correspondance élevé.

L'algorithme doit réduire le nombre de faux positifs et de faux négatifs. De plus, il serait intéressant d'encourager les acteurs à faire une description honnête de leur service. Autrement dit, si la description n'est pas honnête, le service n'est pas ou mal repris dans les résultats lui correspondant.

Parmi les algorithmes de «matching» exploités dans la recherche des services nous citons [PAO02]:

- UDDI: le premier but d'UDDI est la spécification d'un framework pour décrire et découvrir des services Web. En particulier, UDDI définit une structure de donnée et une API d'interrogation.
- LARKS : Une simple recherche par mot-clé ou via une classification ne répond pas tous les besoins. Des recherches ont donc été réalisées afin d'utiliser des mécanismes plus intelligents, incluant la connaissance du monde (le but est de détecter que l'annonce : journal financier correspond à la requête : cours récent de bourse).
- DAML-S : DAML-S est une ontologie de service dont les instances représentent des services Web , l'idée est également de pouvoir faire des recherches sur le contenu du service plutôt que simplement sur la base de mot-clé.

1.4.5 Les travaux relatifs

Parmi les travaux relatifs à la découverte sémantiques des services Web géographiques nous citons :

- l'approche proposée par E. R. Sacramento dans [SAC09], qui utilise les annotations sémantiques pour la découverte automatique et la composition de SWG. La thèse traite d'abord de la façon de combiner la sémantique fournies par les catalogues de métadonnées avec la sémantique fourni par des ontologies pour produire des services catalogue capable de faciliter la recherche des services appropriés du web. Puis, il aborde la façon de composer des services Web à partir des services Web de base OGC de manière automatique.

- SWING (Semantic Web Services Interoperability géospatiales décision Making) dans [DUM07] vise à déployer la technologie SWS dans les applications géospatiales. En particulier, il répond à deux grands obstacles qui doivent être surmontés : réduire la complexité de la création de descriptions sémantiques et d'augmenter le nombre de services décrits sémantiquement. Le projet permettra de développer des méthodes et des outils qui peuvent masquer la complexité d'automatiser la création les descriptions sémantiques. SWING a pour but d'établir un dialogue ouvert, facile à utiliser SWS cadre approprié d'ontologies et de déduction pour les outils d'annotation, de la découverte, la composition, et l'invocation de services Web géographiques.

- L'approche proposée par [YUE06] permet d'automatiser la composition des services Web géographiques en utilisant la sémantique géographique dans l'architecture orientée services (SOA). Il montre comment les ontologies fondées sur la sémantique géographique sont utilisés dans un prototype de système pour permettre la découverte automatique, l'accès et le chaînage des services Web géographique. Une étude de cas du processus de chaînage pour calculer un indice de susceptibilité de glissement de terrain illustre l'applicabilité de l'ontologie axée sur la composition automatique de services Web pour des applications géographiques.

1.5 Conclusion

Dans ce chapitre, nous avons mis en relief les différents concepts de bases exploités dans notre travail à l'image des systèmes d'information géographiques, les services Web, les systèmes de gestion des bases de données géographiques et les services Web géographiques. Les langages proposés par le consortium W3C permettant aux fournisseurs de décrire leurs services Web, en assurant la description du service à travers le WSDL et sa publication dans le registre UDDI ainsi que sa découverte en consultant ce registre. L'invocation du service Web est assurée à travers l'échange de messages entre le client et le fournisseur en utilisant le SOAP.

L'intégration du Web sémantique dans la communauté des services Web a apporté, entre autres, un niveau sémantique dans la description des services Web (on parle alors de services Web sémantiques). Les acteurs de ces domaines ont proposé plusieurs travaux sur la base de l'enrichissement de la description des services Web par l'aspect sémantique. Pour la description des

SWG, nous avons présenté le WSDL, les métadonnées et la norme ISO19119, pour la publication le registre UDDI et Xmethode, pour la découverte les algorithmes de « matching » à l'image de LARKS, DAML-S et l'algorithme de « matching » de l'UDDI.

Pour renforcer l'aspect sémantique, nous enrichissons la description des SWG par des informations concernant la qualité des SWG. Elle est étudiée dans le contexte des SWG dans le chapitre suivant qui décrit la qualité des services Web géographiques.

Chapitre 2: La qualité des services Web géographiques

Chapitre 2 : La qualité des services Web géographiques

Avec la croissance du World Wide Web est venu l'idée que les méthodes actuellement disponibles pour la découverte des services Web géographiques souvent insuffisantes. Aujourd'hui, les méthodes d'extraction sont généralement limitées à des recherches par mots clés. Il s'avère nécessaire d'exploiter des connaissances sémantiques pour satisfaire l'utilisateur. Cette satisfaction peut être assurée à travers :

- la qualité joue un rôle primordial pour satisfaire les besoins explicites ou implicites de l'utilisateur [TRI07, GUT07, BET05];
- la personnalisation qui facilite l'expression du besoin utilisateur et rend l'information sélectionnée intelligible à l'usager et exploitable [KOS7, BOU04].

Ce chapitre consiste à présenter le premier élément assurant la satisfaction de l'utilisateur. Il est réservé à présenter en premier lieu la notion de qualité d'ordre générale en spécifiant la notion de qualité des services et les indicateurs de qualité. En deuxième lieu, il consiste à décrire la qualité des SWG à travers la qualité des données géographiques, la qualité des services Web et la qualité d'usage ainsi que quelques travaux relatifs.

2.1 Qualité de service

La qualité en règle générale se définit comme : « l'ensemble des propriétés et caractéristiques d'un produit ou d'un service qui lui confère l'aptitude à satisfaire des besoins exprimés ou implicites. » (Norme ISO 8402 :1994) [TRI07]. La qualité de service représente un concept abstrait et diffus, à multiples facettes [BRE03].

Le concept de qualité de service représente un domaine en pleine évolution. Il a une importance croissante dans l'économie mondiale, la question de la mesure et de la gestion de la qualité de service a été sujette à de nombreux débats. Dans un premier temps, cette section aborde la qualité des services traditionnels, par la suite, la qualité des services électroniques, enfin, la qualité des services Web.

2.1.1 La qualité du service traditionnel

Contrairement à la qualité des biens, que l'on peut mesurer objectivement par des indicateurs comme la durabilité ou le nombre de défauts de fabrication, la qualité de service représente un construit abstrait et diffus en raison des trois caractéristiques associées aux services : intangibilité, hétérogénéité et inséparabilité ([EIG87] cité dans [BRE03]). Les fondations de la théorie de la qualité de service reposent sur la littérature concernant la qualité des produits et la satisfaction des consommateurs. Les premières conceptualisations se sont basées sur le paradigme

de la disconfirmation utilisé dans la littérature sur les produits physiques; ce paradigme stipule que la qualité résulte d'une comparaison entre ce qui est perçu et la performance attendue. Dans le but d'adapter le paradigme de la disconfirmation à la mesure de la qualité de service, Grönroos dans [GRO82] identifie deux dimensions de la qualité perçue : la qualité fonctionnelle et la qualité technique. La qualité fonctionnelle représente le processus de livraison du service, c'est à dire les perceptions des consommateurs concernant les interactions qui ont lieu durant le processus de service. La qualité technique reflète le résultat de l'acte de service, c'est à dire ce que le consommateur reçoit durant la rencontre de service. La qualité de service est le résultat de la différence entre les attentes du consommateur (i.e. ce qu'il considère comme devoir être le service offert par la firme) et l'évaluation de la performance réelle du service. La qualité perçue de service est conceptualisée et opérationnalisée comme la différence ou l'écart entre attentes et perceptions des consommateurs, différence vue sous l'angle de son amplitude et de sa direction [BRE03]. Le modèle SERVQUAL détaillé dans [PAR88] assure la mesure de la qualité de service par l'évaluation des écarts.

2.1.2 Qualité de service électronique

Le service électronique a vu le jour avec l'apparition de d'Internet. Les services électroniques sont classés en plusieurs catégories, le plus utilisé est celui du commerce électronique. Ce service est un site marchand représenté par un site de vente au détail sur lequel le consommateur peut naviguer, évaluer, commander et acheter un produit ou un service. Il s'agit d'une version électronique d'un magasin physique à la différence près que toutes les transactions et les activités reliées se déroulent dans un environnement virtuel [BRE03].

La première définition formelle de la qualité de service électronique a été donnée par Zeithaml, Parasuraman et Malhotra dans [ZET00]. Ils la définissent comme le degré selon lequel un site Web facilite un magasinage, un achat et une livraison efficace et efficiente des produits ou services. Il s'agit d'une qualité transactionnelle, elle inclut des éléments d'évaluation pré et post expérience de service. Les chercheurs ont identifié un certain nombre de critères que les consommateurs prennent en compte lorsqu'ils évaluent les sites Web en général et la qualité de service délivrée par ces sites plus particulièrement. Ces éléments sont [BRE03]

- *La disponibilité et le contenu de l'information*: La disponibilité et la profondeur de l'information sont fréquemment mentionnées comme des raisons importantes de l'achat en ligne. Utilisé comme critère d'évaluation d'un site Web en particulier, fournir suffisamment d'information pour pouvoir comparer les produits et faire une sélection semble être un facteur important. En terme de contenu de l'information, la capacité de comparer les prix et la qualité de l'information fournie (précision, actualisation, crédibilité, ...) augmentent la satisfaction relative, à la fois, à l'expérience de consommation et à l'achat du produit, ce qui

accroît la probabilité de revisite du site et d'achats répétés. De plus, lorsque l'utilisateur peut contrôler le contenu, l'ordre et la durée de l'information (temps où l'information est présente) relative à un produit, sa capacité à intégrer, se rappeler et par conséquent à utiliser l'information est améliorée. La qualité et la quantité des informations disponibles sur un site Web commercial est une dimension qui n'a pas été directement explorée dans les évaluations de qualité de services traditionnelles. Cette dimension souligne le fait qu'un montant adéquat et exact d'informations est considéré comme un élément clé du service fourni par le vendeur en ligne. L'importance de cette dimension peut s'expliquer par le manque d'interaction avec des « vrais » vendeurs, obligeant le cyber-consommateur à se reposer sur ses propres capacités à localiser et acheter les produits ou services recherchés.

- *La facilité d'utilisation:* Les transactions utilisant Internet semblent complexes et peuvent intimider de nombreux consommateurs, la facilité d'utilisation d'un site Web s'impose comme un élément important de la qualité de service électronique. La facilité d'utilisation a longtemps été appelée usability (utilisabilité) dans le contexte d'Internet. Le moteur de recherche du site, la vitesse de téléchargement des pages et l'organisation du site constituent les éléments clés qui affectent « l'utilisabilité » du site.
- *Le design :* Les chercheurs se sont intéressés à l'impact du style graphique du site sur les perceptions des consommateurs relatives au shopping en ligne. Le style graphique comprend des éléments comme la couleur, la taille, le type de la police, les photographies, les graphiques et les animations, etc.
- *La fiabilité / le respect des engagements:* La dimension dominante dans les évaluations de la qualité de service traditionnelle, à savoir la fiabilité, est aussi un élément qui semble important dans les évaluations de qualité de service électronique. Les évaluations de la fiabilité prédisent le plus fortement la satisfaction des consommateurs et la qualité perçue, et arrivent au deuxième plan pour prédire la fidélité et les intentions de réachat sur le site. Dans un contexte traditionnel, la fiabilité est définie comme « la capacité pour le prestataire de service d'offrir le service promis de manière exacte, précise et digne de confiance ». Cela se traduit sur le Web par le respect des délais de livraison, l'exactitude de la commande, la précision dans la présentation de produits et les autres aspects du respect des engagements.
- *La sécurité et le respect de la vie privée :* La sécurité et le respect de la vie privée sont des critères clés dans l'évaluation des services en ligne. La sécurité implique de protéger l'utilisateur des risques de fraude et de pertes financières par l'utilisation de leur carte de crédit sur le site. Le respect de la vie privée implique la protection des données personnelles et de ne pas partager ou revendre avec d'autres sites, ces informations collectées sur les

consommateurs au cours de l'expérience de service. Cela implique de préserver l'anonymat des consommateurs et de demander leur consentement avant toute transmission de données personnelles (à titre gratuit ou payant). Cette dimension apparaît être une dimension unique dans un contexte de commerce en ligne. A cause de l'absence de contact interpersonnel sur Internet, le consommateur accorde une grande importance à la sécurité et à la confidentialité des transactions en ligne.

2.1.3 Qualité de services Web

La qualité du service a été un domaine de recherche actif pour plusieurs domaines. Le terme « qualité du service » a été employé pour exprimer des conditions non fonctionnelles pour différents secteurs tels que la communauté de la recherche de réseau et dans les issues en temps réel [SHU03]. Dans l'objectif d'assurer une qualité de service (QoS) adaptée à l'utilisateur de nombreuses propriétés non fonctionnelles doivent être exploitées durant le déploiement ou l'exécution des services. Ces propriétés de QoS révèlent des caractéristiques diverses et s'expriment sous différentes formes, car elles peuvent être relatives aux services mêmes, notamment en terme (i) de performance (fiabilité, temps d'exécution attendu, etc.), (ii) au contexte d'exécution (temps de latence, débit, etc.) ou (iii) aux besoins de l'utilisateur (qualité du résultat, coût économique du calcul, etc.)[ACH08].

Puisque les services Web sont fournis par des tiers et être appelés dynamiquement au-dessus de l'internet, leur QoS peut varier considérablement. Par conséquent il est important d'avoir un cadre pour capturer le QoS donné par le fournisseur et le QoS exigé par le client, et finalement la correspondance entre les deux en découvrant le meilleur service Web selon la QoS exigée [SHU03]. L'ISO 8402 (une partie de standard de qualité de l'ISO 9000 (ISO9000 2002)) décrit la qualité comme « totalité de dispositifs et de caractéristiques d'un produit ou d'un service qui concernent sa capacité de satisfaire aux besoins indiqués ou implicites». Nous rejoignons la définition de la qualité de service présenté par R Shunig dans [SHU03], elle est définie comme ensemble d'attributs non fonctionnels qui peuvent effectuer la qualité du service offert par un service Web. Dans la littérature nous trouvons plusieurs classifications de la QoS entre autres :

2.1.3.1 Classification 1 : Selon A Mani et al [MAN02]

Les principaux critères de qualité des services Web peuvent être comme suit :

- *Disponibilité*: la disponibilité est l'aspect de qualité de si le service Web est présent ou prêt pour l'usage immédiat. La disponibilité représente la probabilité qu'un service est disponible.
- *Accessibilité*: l'accessibilité est l'aspect de qualité d'un service qui représente le degré qu'il est capable de servir une demande de service Web. Elle peut être exprimée comme mesure de probabilité dénotant l'indice de réussite ou la possibilité d'une instanciation réussie de service à

un moment. Il pourrait y avoir des situations quand un service Web est disponible mais non accessible.

- *Intégrité*: l'intégrité est l'aspect de qualité de la façon dont le service Web maintient l'exactitude de l'interaction en ce qui concerne la source. L'exécution appropriée des transactions de service Web fournira l'exactitude de l'interaction. Une transaction se rapporte à un ordre des activités à traiter comme unité simple de travail. Toutes les activités doivent être accomplies pour effectuer une transaction réussie. Quand une transaction n'accomplit pas, toutes les modifications apportées sont compensées.
- *Exécution*: l'exécution est l'aspect de qualité du service Web, qui est mesuré en termes de sortie et latence. Une sortie plus élevée et des valeurs plus basses de latence représentent la bonne exécution d'un service Web.
- *Fiabilité*: la fiabilité est l'aspect de qualité d'un service Web qui représente le degré d'être capable de maintenir le service et la qualité de service. Le nombre d'échecs par mois ou année représente une mesure de fiabilité d'un service Web. Dans un autre sens, la fiabilité se rapporte à la livraison assurément et commandée pour des messages envoyés et reçus par des demandeurs et des prestataires de service Web.
- *Normalisation*: la normalisation est l'aspect de qualité du service Web dans la conformité avec les normes.
- *Sécurité*: la sécurité est l'aspect de qualité du service Web de fournir la confidentialité et la non-répudiation en authentifiant les parties concernées, les messages de chiffrement, et fournir le contrôle d'accès. La sécurité a une grande importance parce que l'invocation de service Web se produit au-dessus de l'Internet public. Le fournisseur de services peut avoir différentes approches et niveaux de sécurité selon le demandeur de service.

2.1.3.2 Classification 2: Selon R Shuping [SHU03]

Il y a beaucoup d'aspects de QoS importants pour des services Web. Nous commençons à les organiser en catégories de QoS. Chaque catégorie doit avoir un ensemble de paramètres ou de mesures quantifiables. Pour faciliter la description, les catégories sont groupées dans différents types.

1. Temps d'exécution relatif à la QoS

- *Évolutivité*: la capacité d'augmenter la capacité informatique du système informatique du fournisseur de services et de la capacité du système de traiter plus d'opérations ou de transactions en période donnée.
- *Capacité* - limite des demandes concourantes d'exécution garantie.
- *Performance* : une mesure de la vitesse en remplissant une demande de service. Elle est

mesurée par ([GUN98] cité dans [SHU03]):

- Temps de réponse : le maximum (moyen ou minimum) de temps nécessaire pour remplir une demande de service (liées à la capacité).
- Latence: le temps pris entre l'arrivée de la demande de service et la satisfaction de la demande.
- Sortie - le nombre de demandes de service réalisées sur une période de temps. La sortie est liée à la latence et à la capacité.
- Exécution - une mesure de la vitesse en accomplissant une demande de service.
- Fiabilité - la capacité d'un service de remplir ses fonctions requises dans des conditions indiquées pendant une période spécifique.
- Flexibilité / robustesse: c'est la mesure dans laquelle un service peut fonctionner correctement dans la présence d'entrées invalides, incomplets ou contradictoires.
- Manipulation d'exception: puisqu'il n'est pas possible que le concepteur de service spécifie tous les résultats et solutions de rechange possibles (particulièrement avec de divers cas spéciaux et possibilités imprévues), des exceptions peuvent être prévues. La manipulation d'exception est comment le service manipule ces exceptions.
- Exactitude: définit le taux d'erreur produit par le service.

2. Support de transaction relatif à la QoS

- Intégrité: des transactions peuvent être groupées dans une unité afin de garantir l'intégrité des données opérées par ces transactions. L'unité peut être réussie où toutes les transactions en unité « commettent » ou tous « roulez en arrière » à leur état original en cas d'échec de transaction. Ceci est décrit par les propriétés ACIDES : Atomicité (s'exécute entièrement ou pas du tout), uniformité (maintient l'intégrité des données), isolement (individuellement les transactions fonctionnent comme si aucune autre transaction n'est présente) et longévité (les résultats sont persistants)).

3. Gestion de la configuration relative aux coûts de la QoS :

- Réglementation: il s'agit d'une mesure de la façon dont le service est conforme à la réglementation.
- Norme (supported standards) - une mesure de si le service est conforme aux normes (par exemple normes propres à l'industrie).
- Cycle de stabilité/changement: une mesure de la fréquence du changement s'est rapportée au service en termes de son interface et/ou exécution.
- Coût: c'est une mesure du coût entraîné en demandant le service.

- Perfection: une mesure de la différence entre l'ensemble de dispositifs spécifiques et l'ensemble de dispositifs mis en application.

4. Sécurité liée à la QoS

- Authentification: comment le service authentifie-t-il les directeurs (des utilisateurs ou d'autres services) qui peuvent accéder au service et aux données ?
- Autorisation: comment le service autorise-t-il les directeurs de sorte que seuls eux peuvent accéder aux services protégés?
- Confidentialité: comment le service traite-t-il les données, de sorte que seulement les principaux autorisés puissent accéder ou modifier aux données ?
- Responsabilité - le fournisseur peut-il être prise responsable de leurs services ?
- Traçabilité et contrôlabilité: est il possible de tracer l'histoire d'un service quand une demande a été entretenue.
- Chiffrement de données: comment le service chiffre-t-il des données ?
- Non-répudiation: le principal ne peut pas refuser la demande d'un service ou des données après le fait.

2.1.3.3 Classification 3 : Selon L. Médini [MED05]

Les critères de qualité des services sont indiqués au niveau de contrat de service. Le contrat représente un engagement qui formalise la relation de service, il doit être rédigé dans un langage neutre. Le contrat est composé de six parties à savoir :

- Identification des parties ;
- Fonctions des services ;
- Interface du service ;
- Qualité du service ;
- Cycle de vie du service et du contrat ;
- Description des termes de l'échange.

Dans son document L. Medini définit la qualité de service Web comme un ensemble de propriétés opérationnelles liées à la réalisation de la prestation, cet ensemble est organisé comme suit :

- Périmètre de la prestation
 - caractère optionnel de la prestation ;
 - droits et obligations du client ;
 - conformité aux normes et standards.

- Qualité de fonctionnement
 - dimensionnement des objets manipulés ;
 - exactitude ;
 - précision ;
 - performance,
 - accessibilité de l'application
- Sécurité
 - authentification ;
 - contrôle d'accès ;
 - confidentialité ;
 - intégrité ;
 - non-répudiation.
- Robustesse
 - fiabilité ;
 - disponibilité ;
 - gestion d'arrêt de l'application prestataire ;
 - gestion des transactions.

L'étude la qualité de service Web est un domaine de recherche en pleine évolution. Parmi les plusieurs travaux de ce domaines nous citons le travail de R.Shuping présente dans [SHU03] qui présente un modèle de découverte des services Web basé sur la QoS dans l'objectif est de donner plus de confiance au client. E. Al masri et all dans [MAS07] propose un service (the Web Service Relevancy Function (WsRF)) pour l'évaluation d'un classement pertinente d'un service Web particulier basé sur la QoS, Okkyung Choi et all dans [OKK07] étudie le modèle standard de la découverte des services Web en indiquant ces inconvénients et proposant une approche de la découverte sémantique des services Web basé sur la QoS en appliquant les principes du Web sémantiques et Mourad Ouzzani et all proposent dans [OUZ04] une infrastructure de requête traitant les services Web comme des objets de première classe. Les requêtes sont évaluées à travers l'invocation des opérations des services Web. Cette solution est présentée à travers un modèle d'optimisation de requête basé l'agrégation des paramètres de la QoS.

2.1.4 Qualité des services de réseaux

La qualité de service QoS (Quality of Service) peut être définie comme le degré de satisfaction d'un utilisateur des services fournis par un système de communication. La QoS est définie comme la capacité d'un élément du réseau (ex : routeur, noeud ou une application) de

fournir un niveau de garantie pour un acheminement des données [MER05].

Le RFC 2386 ([CRA98] cité dans [MER05] se caractérise la QoS comme un ensemble de besoins à assurer par le réseau pour le transport d'un trafic d'une source à une destination. Ces besoins peuvent être traduits en un ensemble d'attributs pré-spécifiés et mesurables en terme de :

- Délai de bout en bout,
- Variance de délai (gigue),
- Bande passante,
- Pertes de paquets.

Suivant le type de l'application, les besoins de QoS sont différents. Par exemple, pour les applications temps réel, comme la voix et la vidéo, le délai de bout en bout d'un paquet doit être limité, autrement le paquet est inutile. Les applications non temps réel, comme le transfert de fichier ou la messagerie, quant à elles se focalisent sur la fiabilité des communications [MER05].

2.2 Les indicateurs de qualité

D'après l'encyclopédie Wikipedia (<http://fr.wikipedia.org/wiki/Indicateur>), un **indicateur** est un outil d'évaluation et d'aide à la **décision** (pilotage, ajustements et rétro-correction) grâce auquel on va pouvoir mesurer une situation ou une tendance, de façon relativement objective, à un instant donné, ou dans le temps et/ou l'espace.

Un indicateur se veut être une sorte de *résumé* d'informations complexes offrant la possibilité à des acteurs différents (scientifiques, gestionnaires, politiques et citoyens) de dialoguer entre eux. L'indicateur (qualitatif ou quantitatif) décrit généralement un état, une pression et/ou une réponse ne pouvant être appréhendés directement. Un indicateur peut en agréger d'autres. Pour un indicateur agrégé, on parle plus souvent d'indice. Il doit exister une relation causale entre le fait mesuré (*indiqué*) et l'indicateur. L'utilité d'un indicateur dépend d'abord de sa capacité à refléter la réalité, mais aussi de sa simplicité d'acquisition et de compréhension.

Pour une mise sous contrôle de la politique qualité, une instrumentation est nécessaire représenté par les indicateurs de qualité. L'orientation confirmée des normes ISO vers la satisfaction du client rend inévitable la mesure de la dite satisfaction (dite qualité perçue), qui ne peut se confondre avec la mesure de la qualité livrée. Cette mesure passe souvent par des enquêtes auprès des clients [BOU00].

Mettre en place un système qualité, ce n'est ni plus ni moins que s'efforcer de mettre l'activité de l'entreprise sous contrôle au sens des sciences de l'ingénieur transposé au domaine de la gestion. Une activité est sous contrôle lorsque son déroulement est orienté et constamment ajusté de façon à atteindre l'objectif prévu avec une marge d'erreur prédéterminée. La démarche de mise sous contrôle est habituellement représentée dans la qualité par la "roue de DEMING" [BOU00]

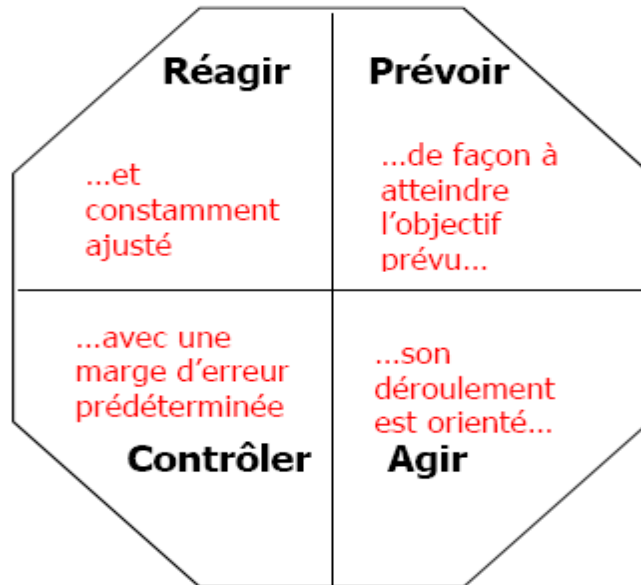


Figure 09 : Roue de DEMING [BOU00]

Au verbe "**PREVOIR**" correspond la notion d'objectif prévu et plus précisément l'exigence d'une définition (préalable) de la politique et des objectifs qualités.

Au verbe "**AGIR**" correspond un déroulement orienté de l'action à travers des "plans d'action".

Au verbe "**CONTROLER**" correspond la possibilité d'apprécier la marge d'erreur, pourvu qu'un système d'information et de mesure ait été mis en place.

Au verbe "**REAGIR**" correspond enfin l'ajustement de l'action à travers des "actions correctives".

Pour que les indicateurs réagissent afin de réguler l'action dans le sens des orientations fournies par la politique et les objectives qualités. Il faut donc trouver la meilleure cohérence possible entre les objectifs et les indicateurs pour que cette réaction soit opportune. Le choix des indicateurs constitue une "**instrumentation**", une quantification, des objectifs pour rendre la mise sous contrôle (au sens de la gestion) efficace [BOU00].

Le choix des indicateurs va donc comprendre deux phases successives :

1. une clarification du système d'objectifs
2. l'instrumentation à proprement parler de ce système d'objectifs à travers des données quantifiables.

Un indicateur doit être techniquement et conceptuellement apte à mesurer, avec une précision acceptable, le phénomène qu'il est censé mesurer et demeurer pertinent dans le cadre du déploiement actuel de la politique qualité.

Les qualités nécessaires pour y parvenir sont de 3 ordres [BOU00]:

1. **La qualité d'usage** : nous pourrions dire, un ensemble de qualités de bon sens qu'il est malheureusement rare de réunir toutes ensemble, dès lors que le champ à instrumenter intègre des aspects humains, ce qui est le cas très général de la qualité ;

2. **La qualité métrologiques** : un indicateur est un instrument de mesure, il doit satisfaire aux qualités requises de tout instrument de mesure ;
3. **La qualité systémique** : l'indicateur s'insère dans un ensemble visant à assurer une conduite du système qualité dans la bonne direction (celle de la politique définie).

2.3 La qualité des services Web géographiques

Dans le domaine de recherche d'information les données de la dimension «Qualité» décrivent la qualité attendue ou espérée par l'utilisateur qui sera confrontée à la qualité effective produite par le système de recherche d'informations afin de restreindre l'espace de recherche [KOS07]. Selon les travaux de M. Bouzeghoub et D Kostadinov[KOS07,BOU04], les informations dans cette dimension sont représentées par des facteurs de qualité en trois catégories de facteurs de qualité selon le type d'objets auxquels ils font référence sont distinguées :

- Facteurs sur le contenu des données : cette catégorie concerne la qualité désirée des objets de contenu. Des exemples de facteurs de cette catégorie sont la fraîcheur et l'exactitude des données.
- Facteurs sur le contenant des données : catégorie regroupe les facteurs de qualité des conteneurs des données (sources). La plupart des facteurs de cette catégorie résultent de l'expérience des autres utilisateurs (ex. le niveau de confiance dans la source, la fiabilité etc.). Ces facteurs sont utilisés pour filtrer les sources de données lorsque leur nombre est trop important.
- Facteurs sur les processus : délivrent ou notifient l'arrivée des résultats. Les facteurs de cette catégorie mesurent aussi bien les performances des processus (ex. temps de réponse) que leur capacité de produire tous les résultats pertinents (rappel) et uniquement les résultats pertinents (précision).

Avec la projection de cette classification dans notre cas nous pouvons conclure que : les facteurs sur le contenu des données peuvent être représenté par les critères de qualité des données géographiques, les facteurs sur le processus par les critères de qualités des services Web (détaillé dans 2.1.3) et les facteurs sur le contenant des données par les critères de qualité d'usage (ce que veut l'utilisateur).

2.3.1 Qualité des informations géographiques

La qualité des données reste un domaine de recherche sur l'information géographique et spatiale. En effet, si ce domaine de recherche a été très actif dans le passé notamment en raison de la croissance des échanges de données et de l'utilisation des SIG en tant qu'outil d'aide à la décision

[VER99], il s'avère aujourd'hui que des problèmes persistent notamment en termes de gestion et d'exploitation [GUT07]. La qualité des données localisées doit permettre de renforcer les démarches d'analyse spatiale, d'aide à la décision, d'archivage, et de favoriser les échanges entre partenaires et la consultation des données par le public. De nombreux enjeux accompagnent la notion de qualité des données géographiques notamment [TRI07] :

- Un enjeu d'échanges : participer au développement des partenariats et des usages. La politique partenariale prendra tout son sens si les échanges de données localisées sont effectivement réalisés.
- Un enjeu organisationnel : participer à une administration rigoureuse des données localisées et développer l'usage des métadonnées. Face à la quantité et à la diversité des utilisateurs et des données localisées produites, il devient nécessaire de connaître la provenance des informations à traiter. C'est pourquoi, des critères de qualité deviennent indispensables aux données localisées, dès leur saisie. Ces informations supplémentaires permettent alors d'exploiter des données de sources différentes en toute sécurité pour les utilisateurs pour mieux en maîtriser les limites. Ainsi, la production ou l'utilisation de données localisées de qualité participe pleinement à une administration rigoureuse des données localisées au sein d'un service. Si les métadonnées n'améliorent pas à proprement parler la qualité des données, elles permettent en revanche de décrire cette qualité et aident l'utilisateur à choisir un jeu de données correspondant à ses propres besoins.
- Un enjeu économique : maîtriser ses besoins et chercher à rentabiliser les données.
- Un enjeu de professionnalisation : développer la rigueur dans les métiers.

La normalisation représente le remède des problèmes posés à cause le très grand nombre d'acteurs qui utilisent, valorisent, enrichissent et font circuler l'information géographique. Elle est donc un enjeu essentiel pour la valorisation de l'information parce qu'elle permet la transparence de ces flux, ces réemplois et ces enrichissements [TRI07]. La norme ISO 9000 définit la qualité d'un système comme "la totalité des propriétés et caractéristiques qui lui confère l'aptitude à satisfaire des besoins explicites ou implicites". Aujourd'hui, la qualité des données est exprimée en fonction de "fitness for use / adéquation à l'usage » [VER99]. Ceci signifie que pour définir la qualité des données, il est nécessaire de disposer d'informations sur les données utilisées et sur les besoins des utilisateurs [DEV05]. Dans ce contexte, nous définissons la qualité comme la proximité entre les caractéristiques des données et les besoins d'un utilisateur pour une application donnée à un instant donné [GUT07].

2.3.1.1 Normalisation de la qualité des informations géographiques

La qualité occupe une importante place dans le standards ISO/TC 211, à travers 03 normes et un projet de spécification techniques à savoir [SER06b] :

- la norme ISO 19113 : Cette norme assure la définition des éléments principales de la qualité des données;
- la norme ISO 19114: est consacrée aux procédures d'évaluation de la qualité. Elle définit les moyens d'exprimer les mesures de la qualité, que ce soit dans les rapports d'évaluation ou dans les métadonnées;
- la norme ISO 19115: précise la structure conceptuelle de métadonnées. cette structure conceptuelle prend en compte les différentes composantes de qualité définies par la norme ISO 19113;
- La norme ISO 19138 (Projet préliminaire de spécifications techniques): décrit l'ensemble des indicateurs de qualité.

2.3.1.1.1 La norme ISO 19113 : Les critères de qualité

La qualité des données est essentielle. Elle doit assurer la fiabilité des processus s'appuyant sur les informations. Les éléments composant la qualité ne sont pas tous des métriques quantifiables et la propagation des éléments non quantifiables est encore mal connue [SER06b]. Dans les phénomènes géographiques, la qualité des données peut être différenciée suivant l'espace, le temps et la thématique d'application. Pour chacune de ces dimensions, diverses composantes de qualité sont identifiées. Ces critères sont définis selon les aspects spatiaux qui permettent d'évaluer la pertinence des données par rapport au contexte géographique [GUT07]. Ils sont décrits selon la norme ISO 19113 qui assure la description de la qualité des données géographiques à l'aide de 7 critères que l'on peut décomposer en deux grandes classes : les critères quantitatifs et les critères qualitatifs. Le septième critère qui définit la qualité spécifique n'entre, bien sûr, dans aucune de ces deux classes. Les critères définis dans la norme ISO 19113 sont donc les suivants [TRIO7,SER06b] :

- **Critères quantitatifs**
 - Précision géométrique : Elle définit les écarts de valeurs de position respectives entre les données étudiées et le terrain nominal (abstraction du monde réel servant de jeu de données référence).
 - Exhaustivité et complétude : Elle décrit les relations entre les objets représentés dans l'ensemble des données et une abstraction de l'ensemble des objets du monde réel.

- Précision sémantique : Elle est définie comme la différence entre une mesure et une autre mesure comparable et connue pour être plus exacte.
- Cohérence logique : Elle se rapporte à toutes les règles logiques sur les structures et les attributs des données spatiales et elle décrit la compatibilité d'un élément du jeu de données avec les autres.
- **Critères qualitatifs**
 - Actualité : Elle renseigne sur la gestion des dates d'observation des données, des types de mise à jour et des périodes de validité.
 - Généalogie : Elle offre des informations sur la source de données, l'acquisition et les transformations ou analyses exécutées. Elle identifie les informations primordiales afin de reconstituer l'histoire d'un jeu de données.
- **Critère spécifique**
 - Qualité spécifique : La qualité spécifique permet à l'utilisateur de définir ses propres critères de qualité si les critères officiels définis par la norme ne répondent pas à ses attentes. C'est en quelque sorte un "critère personnalisé".

2.3.1.1.2 La norme ISO 19114 : Mesurer les critères de qualité

Mesurer la qualité revient à effectuer des actions de mesure, d'examen, d'essai, de calibrage d'une ou plusieurs caractéristiques du jeu de données permettant d'évaluer l'écart entre le jeu de données et le terrain nominal. Cette étape importante est prise en charge par la norme ISO 19114 qui présente un cadre d'évaluation de la qualité applicable aux jeux de données et précise le type d'information devant être consigné dans la procédure qualité [TRI07].

Selon la norme ISO 19114 [TRI07], la mesure des écarts entre le lot de données et le monde nominal qu'il représente peut être réalisée par :

- Comparaison ;
- Mesure sur les objets ;
- Rassemblement statistique ;
- Validation ou rejet par rapport à un domaine de validité ;
- En rendant compte.

Pour chaque mesure, on précisera les éléments suivants :

- Le domaine d'application (le niveau, le type d'objet, l'emprise géographique, l'emprise temporelle) ;

- Le type de mesure ;
- La méthode de comparaison utilisée ;
- Le résultat ;
- Le type de valeur ;
- L'unité de valeur ;
- La date

Des informations complémentaires pourront être fournies selon les cas :

- La population de l'échantillon ;
- Le stockage du résultat ;
- La valeur du seuil d'acceptation du contrôle

2.3.1.1.4 La norme ISO 19115 : Les métadonnées des informations géographiques

La norme ISO 19115 est la référence internationale concernant l'échange pour l'information géographique [BAD03]. Cette norme permet de créer différents modèles, selon un point de vue différent à partir d'éléments standards ou obligatoires de la norme et d'autres éléments propres à l'application [GUT07a]. Pour faciliter l'accès à l'information, les éléments de la norme ont été regroupés en 4 groupes [BAD03]:

- Thématique (Quoi);
- Spatial (Où);
- Temporel (Quand);
- Contact (Qui).

La norme ISO 19115 fournit une structure pour la description des données géographiques numériques sous forme d'un modèle conceptuel de métadonnées. Ce modèle doit permettre une description si possible exhaustive de toutes les informations concernant les données géographiques. Les éléments des métadonnées y sont définis avec leur type, les relations qui les unissent et les conditions associées. Dans cette norme internationale, les métadonnées pour les données géographiques sont réparties en packages (Figure 10) [GUT07a].

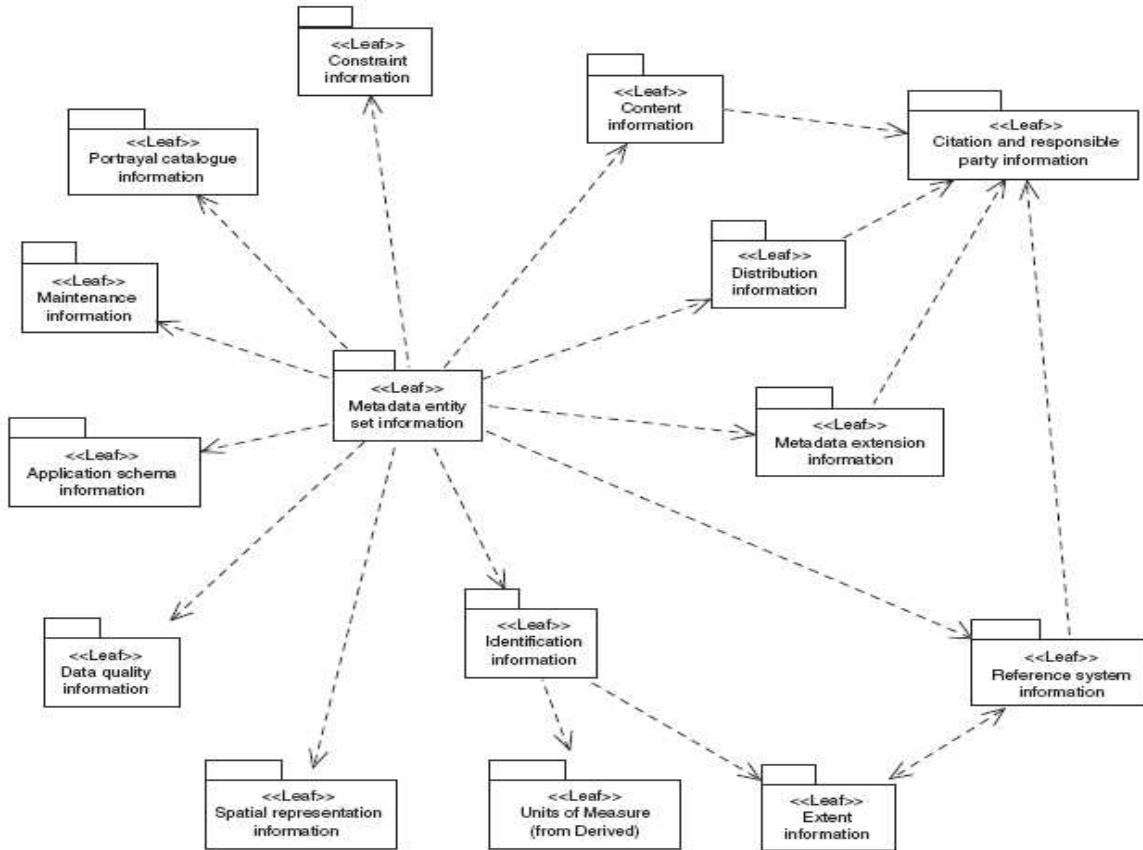


Figure 10 : Packages des métadonnées géographiques. Norme ISO 19115 [ISO03]

Chaque package contient une ou plusieurs entités (Classes UML) (Figure 11) qui peuvent être spécifiées en sous-classes ou généralisées (super classes) [GUT07a].

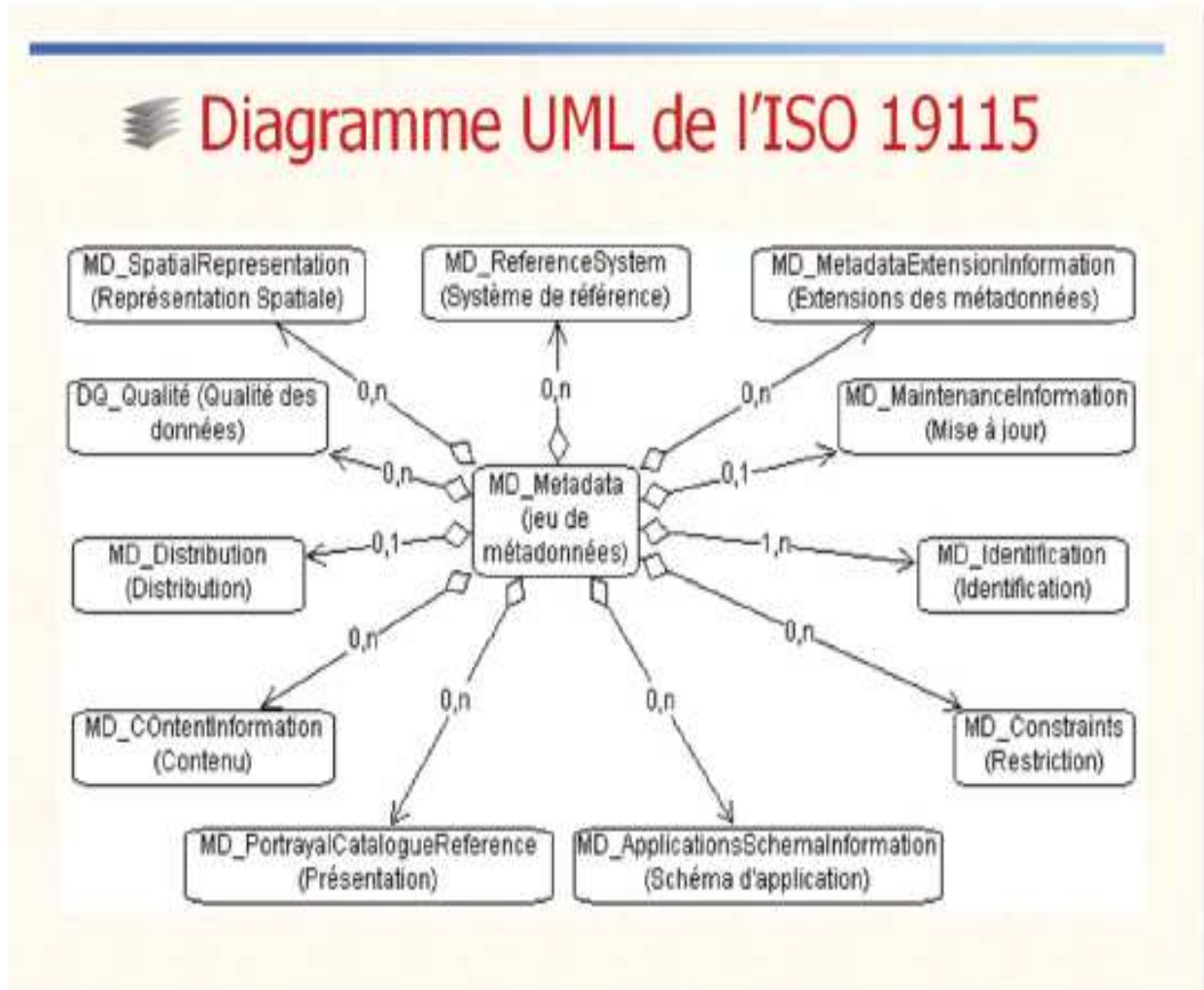


Figure 11: Diagramme décrivant la sémantique de la norme ISO 19115 [FRA08,MAR01,ISO03].

Le diagramme présenté dans la figure 11 est composé de [GUT07a] :

- *Métadonnées (MD_Metada)*: représente le jeu de métadonnées comprenant les informations d'identification importantes et étant en relation (agrégation) avec les autres classes principales.
- *Identification (MD_identification)*: informations permettant d'identifier les données décrites par les métadonnées.
- *Contraintes d'information (MD_Contrains)*: informations concernant les restrictions sur les données.
- *Qualités des données (DQ_DataQuality)*: informations concernant la qualité et la provenance des données telles que l'exactitude, la cohérence et la précision des données,...
- *Mise à jour (MD_MaintenanceInformation)*: informations sur la fréquence et les possibilités de mise à jour des données.
- *Système de référence (MD_ReferenceSystem)*: description des systèmes de référence spatiale

et temporelle utilisés dans le jeu de données.

- *Représentation spatiale (MD_SpatialRepresentation)*: informations contenant les mécanismes utilisés pour représenter l'information spatiale.
- *Contenu d'information (MD_ContentInformation)*: description du contenu du jeu de données. Renvoi au catalogue des objets, au modèle de données ou à la description des données. Le contenu de ces catalogues ne fait toutefois pas partie des métadonnées.
- *Interprétation du catalogue d'information* : informations permettant l'identification du catalogue utilisé (MD_PortrayalCatalogueReference).
- *Distribution d'information (MD_Distribution)*: informations concernant le fournisseur, le format de données, le support de transfert numérique, ...
- *Extension des métadonnées (MD_ExtensionInformation)*: informations concernant les extensions spécifiées par les utilisateurs.
- *Schéma d'application (MD_ApplicationSchemaInformation)*: description du logiciel d'application utilisé pour la génération des données.

Les composantes métadonnées présentées ci-dessus sont fournies par les producteurs de données lors de la livraison de lots de données, et sont censés répondre également aux besoins généraux de l'utilisateur [GUT07a].

L'étude de la qualité des données géographiques fait l'objet de plusieurs travaux entre autres :

- Le projet MDweb (<http://www.mdweb-project.org/>) qui représente un outil générique, multi-langues, multi-normes, destiné au catalogage et à la localisation des données et documents (spatiales et non spatiales). Cette application est conçue pour constituer, gérer, administrer et consulter des catalogues de données via le web et d'accéder si besoin aux données et documents référencés. MDweb s'appuie sur les standards internationaux de métadonnées (ISO 19115 et 19139) et de communication (OGC – Catalogue service Specification – Z39.50) qui permettent d'assurer des interrogations entre catalogues. MDweb répond aujourd'hui aux besoins de fournitures de métadonnées dans le contexte des infrastructures de données spatiales telles que spécifiées au niveau français (initiatives géocatalogue et géoportail) et européen (Directives INSPIRE).
- S. Bard propose dans [BAR04] une méthode d'évaluation de la qualité de données géographiques généralisées qui porte sur la qualification et la quantification de l'information géographique à l'issue d'une dérivation. Il s'agit d'identifier une ou plusieurs mesures de qualité permettant de déterminer l'adéquation du produit dérivé avec les besoins qui ont nécessité la dérivation. Il place dans une approche de dérivation automatique de données

(cartes ou base de données) en fonction d'un besoin utilisateur (spécifications).

- Le travail présenté par C.Gutiérrez et al dans [GUT07] entre dans le cadre de la gestion temps réel des données spatio-temporelles issues des capteurs, il s'agit de la structuration des données en temps réel, de leur interrogation ou encore de leur communication intelligente à l'utilisateur. Plus particulièrement l'évaluation de la qualité des données spatio-temporelles temps réel se pose également. Elles proposent une analyse des métadonnées en temps réel notamment pour l'évaluation de la qualité dans un contexte d'aide à la décision en situation critique, sachant que les métadonnées géographiques, notamment concernant la qualité des données, sont normalisées, les critères définis ne prennent pas en compte la qualification de données dynamiques issues d'objets mobiles, agiles ou de mesures temps réel exploitées dans des applications "en ligne" comme la surveillance de phénomènes naturels ou industriels.

2.3.1.1.5 La norme ISO 19138 (Projet préliminaire de spécifications techniques)[SER06b]:

Cette norme sert à définir les informations nécessaires pour la description des indicateurs de qualité et fournir la description de la liste des indicateurs de qualité. Ces indicateurs de qualité sont, sans doute, un facteur d'interopérabilité, mais les exigences des utilisateurs sont telles que chaque communauté doit pouvoir décrire ses propres indicateurs. La norme l'ISO 19138 permettra l'émergence des registres communautaires des indicateurs, simplifiant ainsi l'approche de la qualité par les organisations pour qui la production de données géographiques est une activité secondaire. L'implémentation de ces registres a besoin d'un standard actuellement en cours de développement.

2.3.2 La qualité d'usage

La qualité d'usage dans notre contexte représente les éléments demandés par l'utilisateur en termes de qualité à travers sa requête. La dimension « Qualité » joue un rôle très important dans le domaine de la personnalisation. Les données de cette dimension décrivent la qualité attendue ou espérée par l'utilisateur. Ces données seront confrontées à la qualité effective produite par le système de recherche d'informations afin de restreindre l'espace de recherche [KOS07]. (La qualité d'usage sera détaillée dans le chapitre suivant).

2.3.4 Les travaux relatifs

Une tendance actuelle consiste à exploiter l'aspect qualité dans la découverte des SWG. Dans ce cadre plusieurs travaux sont effectués, entre autres nous citons :

- l'approche proposée par E. R. Sacramento dans [SAC09], qui utilise les annotations sémantiques pour la découverte automatique et la composition de SWG. La thèse traite d'abord de la façon de combiner la sémantique fournies par les catalogues de métadonnées avec la sémantique fourni par des ontologies pour produire des services catalogue capable de faciliter la recherche des services appro

- priés du web. Puis, il aborde la façon de composer des services Web à partir des services Web de base OGC de manière automatique.
- L'approche proposée par [AKT05] qui prend en compte les métadonnées descriptives, à savoir la qualité des attributs de service, dans le processus de découverte en faisant correspondre entre la qualité de service et celle présentée dans le profil en distinguant les services géographiques qui correspondent aux besoins des clients;
- Le travail proposé par [DON08] concerne les SW avec la possibilité de générer dynamiquement des informations de qualité pour les résultats des analyses spatiales en tenant compte de la qualité des données d'entrée;
- Dans le système proposé par [ALA06], les calculs d'affectation spéciale sont établis par les capacités de base correspondant à des paramètres de QoS. Le projet Geospatial Services Ordering Metric (GSOM) est utilisé pour l'évaluation de qualité de service et de mise en confiance. Il représente une architecture complète pour une sémantique consciente pour la sécurité, la découverte et l'orchestration des SWG.

2.4 Conclusion

Ce chapitre nous a permis de mettre en évidence l'aspect qualité de service. Tout d'abord, nous avons décrit les différentes notions de qualité de service citées dans la littérature. Ensuite, nous avons présenté la qualité des services Web géographiques en indiquant un ensemble de types et techniques et de normes spécifiques à la qualité des données géographiques et des services Web.

Les travaux relatifs à la qualité des données géographiques, la qualité des services Web et celle des services Web géographiques montrent de l'intérêt pour la prise en compte de la qualité dans la découverte des services Web géographiques. Il existe trois types de qualité (qualité de données, qualité de services et qualité d'usage) nécessaires pour prendre en charge la qualité des SWG.

Pour compléter notre étude concernant les connaissances sémantiques nécessaires pour la satisfaction de l'utilisateur, nous réservons le prochain chapitre à l'étude du profil utilisateur qui spécifie la qualité d'usage.

Chapitre 3: Le profil utilisateur

Chapitre 3 : Profil utilisateur

Un nombre croissant de Services Web Géographiques (SWG) sont conçus pour interopérer des informations spatiales sur le Web. Les SWG sont en train de changer la manière dont les informations spatiales, les systèmes et les applications sont conçus, développés et déployés [ZHA07]. Au fur et à mesure que les SWG se multiplient, la difficulté de la découverte de service s'accroît. La maîtrise de cette phase nécessite la personnalisation de la sélection des services Web géographiques désirés par l'utilisateur. La personnalisation est une réponse efficace au problème de la surcharge d'informations, d'une part, en injectant dans les requêtes des critères de filtrage plus restrictifs, et, d'autre part, en reformulant éventuellement les requêtes pour mieux tenir compte des centres d'intérêt et des préférences de l'utilisateur. Une sélection plus restrictive des données permet de limiter la surcharge d'informations et une reformulation de certaines parties de la requête permet de mieux cibler les résultats. Par exemple, la requête d'un utilisateur qui cherche des livres sur « Java » peut être personnalisée en sachant qu'il s'intéresse à la programmation et non à la géographie (centre d'intérêt), qu'il préfère lire en anglais (préférence de langue) et qu'il cherche les éditions les plus récentes avec une bonne critique (préférences de qualité). L'ensemble des données décrivant l'utilisateur et ses préférences sont souvent regroupés et stockés dans un profil utilisateur. Dans un tel contexte, la requête utilisateur exprime une demande d'information particulière tandis que le profil représente la partie relativement stable des besoins de l'utilisateur qui doit être prise en compte lors de l'évaluation de la requête [BOU07, KOS07].

Dans ce chapitre nous faisons un survol sur les systèmes de recherche d'informations, le profil utilisateur (profil, contexte et préférences). Une étude de la dimension qualité est présentée en indiquant les éléments de chaque facteur de cette dimension ainsi qu'une présentation de la requête géographique.

3.1 Système de recherche d'information (SRI)

Les progrès des technologies de l'information, le large développement de nouveaux supports de média informatique et l'amélioration des capacités de stockage sont les rouages essentiels de l'essor d'Internet. Au delà des progrès et libertés apportés aux procédés d'éditions de l'information, cette innovation porte essentiellement sur l'ensemble des services de communication et d'accès à l'information. Elle a marqué le début d'une nouvelle ère de communication, d'une nouvelle société axée principalement autour de l'information. Dès lors, l'accès à l'information est devenu un enjeu capital et stratégique : acquérir l'information pertinente, au bon moment, dès qu'elle est disponible est une nécessité pour tous, dans tous les domaines de la vie active. Le Web par exemple contient une quantité énorme d'information et n'a pas cessé d'augmenter; l'élaboration de systèmes automatisés pour gérer ces masses de données est devenue très nécessaire [ZEM08].

La Recherche d'Information (RI), domaine déjà ancien, est une branche en informatique qui s'intéresse à l'acquisition, l'organisation, le stockage et la recherche des informations. Elle propose des outils, appelés systèmes de recherche d'information (SRI), dont l'objectif est de capitaliser un volume important d'information et d'offrir des moyens permettant de localiser les informations pertinentes relatives à un besoin en information d'un utilisateur exprimé à travers une requête [ZEM08].

La personnalisation est un processus qui change la fonctionnalité, l'interface, la teneur en information, ou l'aspect d'un système pour augmenter sa pertinence personnelle en fonction des caractéristiques socio-démographiques déclarées de l'utilisateur (sexe, âge, lieu de résidence, etc.) et/ou de son comportement observé contenu dans ce que l'on nomme le *modèle* utilisateur. Ce modèle décrit toute information sur l'utilisateur, comme ses préférences, ses centres d'intérêts, ses besoins en information et son environnement de recherche. Indépendamment de l'objectif applicatif visé, nous identifions trois principaux aspects à promouvoir dans les systèmes d'accès personnalisés

1. La capacité à identifier l'intention conceptuelle de l'utilisateur.
2. L'exigibilité du processus de sélection de l'information en vue de s'adapter au contexte d'utilisation courant.
3. et l'intelligibilité des interactions utilisateur - système.

3.1.1 Les fondements de la recherche d'information

La définition de la RI donnée par [VAN86] (cité dans [ZEM08]) dans sa forme originelle :

"The user expresses his information need in the form of a request for information. Information retrieval is concerned with retrieving those documents that are likely to be relevant to his information need as expressed by his request".

Cette définition fait apparaître deux notions clés que nous introduisons dans ce qui suit :

- Document : Un document peut être un texte, un morceau de texte, une page web, une image, une vidéo, etc. On peut appeler document toute unité qui peut constituer une réponse à un besoin informationnel de l'utilisateur.
- Requête : Une requête constitue l'expression du besoin en information de l'utilisateur. Elle représente l'interface entre le SRI et l'utilisateur.

L'objectif fondamental d'un processus de RI est de sélectionner les documents "*les plus proches*" du besoin en information de l'utilisateur décrit par une requête. Pour cela, le système de recherche regroupe un ensemble de méthodes et procédures permettant la gestion des collections de

documents stockés sous forme d'une représentation intermédiaire permettant de refléter aussi fidèlement que possible leurs contenus sémantiques. L'interrogation de la collection de documents à l'aide d'une requête nécessite la représentation de cette dernière sous une forme unifiée compatible avec celles des documents. Ces fonctionnalités sont représentées à travers le processus global de la RI, communément nommé processus en U [KEL92] et schématiquement illustré par la figure 12.

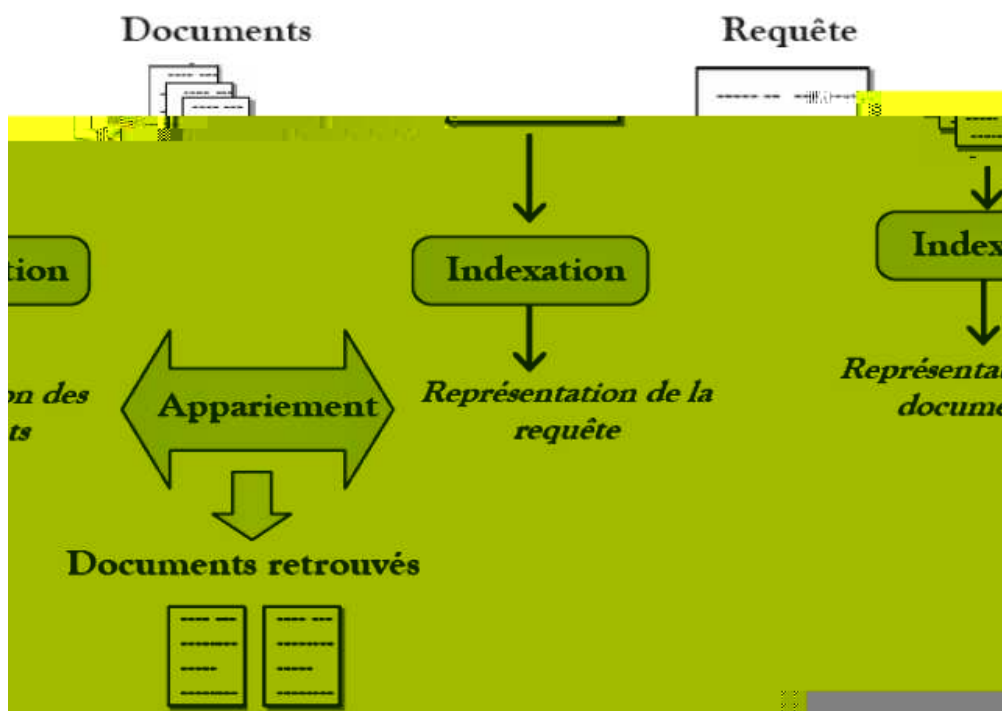


Figure 12 : Processus en U de la recherche d'informations

Le déroulement de ce processus induit deux principales phases : indexation et appariement requête/document.

- **L'indexation** est une étape très importante dans le processus de RI. Elle consiste à déterminer et à extraire les termes représentatifs du contenu d'un document ou d'une requête. La qualité de la recherche dépend en grande partie de la qualité de l'indexation. Le résultat de l'indexation constitue, ce que l'on nomme le **descripteur** du document ou de la requête. Ce dernier est souvent une liste de termes ou groupe de termes significatifs pour l'unité textuelle correspondante, généralement assortis de poids représentant leur degré de représentativité du contenu sémantique de l'unité qu'ils décrivent. Les descripteurs des documents (mots, groupe de mots) sont rangés dans un catalogue appelée dictionnaire constituant le **langage d'indexation**. Techniquement, l'indexation peut être manuelle, automatique ou semi-automatique. Les termes extraits des documents ne jouent pas le même rôle dans la représentation de ces derniers, en ce sens où ils n'ont pas le même degré

d'importance. Pour caractériser ce degré de discrimination, il est courant en RI, d'affecter à chaque terme un poids. Cette étape est primordiale dans le processus d'indexation correspond au **processus pondération** [ZEM08].

- **L'appariement document-requête** : Le processus d'appariement document-requête permet de mesurer la pertinence d'un document vis-à-vis d'une requête. De manière générale, à chaque réception d'une requête, le SRI calcule un score de pertinence (similarité vectorielle, probabiliste, etc.). Le processus d'appariement est étroitement lié au processus d'indexation et de pondération des termes. Il existe deux méthodes d'appariement :
 - *Appariement exact* («*exact match retrieval*») Le résultat est une liste de documents respectant exactement la requête spécifiée avec des critères précis. Les documents retournés ne sont pas triés [SAL71] (cité dans [ZEM08]).
 - *Appariement approché* («*best match retrieval*») Le résultat est une liste de documents sensés être pertinents pour la requête. Les documents retournés sont triés selon leur score de pertinence vis-à-vis de la requête [ROB76] (cité dans [ZEM08]).

3.1.2 Les types de systèmes de recherche d'information (SRI)

Les efforts continus des chercheurs en RI ont permis jusqu'à présent d'améliorer sans cesse les performances et la qualité des services d'accès à l'information. Dans ce paragraphe nous faisons un survol des principales évolutions dans le domaine de la RI : de la première génération de systèmes de recherche d'information (SRI) dits *classiques* à la RI *adaptive*, puis récemment à la RI personnalisée [ZEM08].

3.1.2.1 Le système de recherche d'information classique

Le SRI classique apparue dans les années soixante, a une vision orientée système, en ce sens où la recherche des informations pertinentes se base uniquement sur l'appariement des documents avec la requête soumise par l'utilisateur. Toutefois, cette vision de l'accès à l'information suppose que l'utilisateur est extérieur au système de recherche. La différence de vocabulaire entre les termes choisis par l'utilisateur pour formuler sa requête et les termes utilisés pour représenter les documents engendrent un défaut d'appariement. Ce dernier est à l'origine d'une dégradation des performances de recherche et particulièrement dans le cas d'accroissement continu des sources d'information hétérogènes et la diversité des utilisateurs.

3.1.2.2 Le système de recherche d'information adaptatif

Les approches adaptatives des SRI exploitent les diverses sources d'évidence (documents jugés, termes pertinents, etc.) pour aider et assister l'utilisateur à retrouver les informations

pertinentes à son besoin.

3.1.2.3 Le Système de Recherche d'Information personnalisé

Les approches dites de personnalisation exploitent les caractéristiques informationnelles spécifiques de l'utilisateur dans les processus d'accès à l'information.

3.2 Profil Utilisateur

La personnalisation de l'information se définit, entre autres, par un ensemble de préférences individuelles représentées par des couples (attribut, valeur), par des ordonnancements de critères ou par des règles sémantiques spécifiques à chaque utilisateur ou communauté d'utilisateurs. Ces préférences servent à décrire le centre d'intérêt de l'utilisateur, le niveau de qualité de données qu'il désire ou les modalités de présentation de ces données. L'ensemble de ces informations est représenté dans un modèle d'utilisateur appelé souvent profil. Un profil regroupe l'ensemble de connaissances nécessaires à une évaluation efficace des requêtes et à une production d'une information pertinente adaptée à chaque utilisateur. Le contenu du profil utilisateur varie selon les approches et les spécifications. Dans le domaine de l'IHM, le profil contient, par exemple, des informations permettant l'affichage des résultats selon les préférences de l'utilisateur. La notion de profil est souvent liée à celle de préférences et de contexte. En effet, les préférences de l'utilisateur font partie intégrale de son profil. En plus la description de l'utilisateur et ses préférences peut changer en fonction du contexte dans lequel il évolue [BOU05].

Un profil utilisateur représente l'ensemble des informations décrivant l'utilisateur. Un *contexte* représente les données décrivant l'environnement d'interaction entre un utilisateur et un système. Une *préférence* est une expression permettant de hiérarchiser l'importance des informations dans un profil ou un contexte. Pour décrire l'utilisateur, le contexte et les préférences, D. Kostadinov propose dans [KOS07] des méta modèles génériques. Le terme « générique » ne veut pas dire un ensemble exhaustif d'attributs, mais plutôt une liste de concepts de haut niveau pouvant être spécialisés, affinés et instanciés dans chaque environnement de personnalisation. Il y a toutefois une différence importante entre les méta modèles de profil et de contexte, d'une part, qui caractérisent les individus et les situations dans lesquelles ils interagissent avec un système d'information, et le méta modèle de préférences, d'autre part, qui décrivent une typologie de préférences utilisées dans les deux premiers.

Les trois méta modèles doivent satisfaire les exigences suivantes :

- ils doivent être capables d'acquérir les principales catégories des connaissances utilisées dans les systèmes de personnalisation actuels,

- ils doivent être indépendants de tout système de gestion des données et de toute technologie,
- la spécialisation, la généralisation et l'instanciation de ces modèles doit être facile,
- ils doivent être ouverts, facilement extensibles à d'autres types de connaissances et d'autres types de préférences.

3.2.1- Approches de représentation du profil utilisateur

La représentation de l'utilisateur à travers la notion de profil permet de mieux comprendre certains mécanismes cognitifs, notamment ceux permettant de percevoir le concept subjectif de la pertinence et au-delà, cibler ses besoins spécifiques dans le but d'améliorer les performances de recherche. Un modèle de représentation permet d'organiser ces éléments afin de faciliter leur exploitation dans le processus d'accès à l'information. On distingue quatre principales approches de représentation : ensembliste, connexionniste, conceptuelle et multidimensionnelle [ZEM08].

3.2.2 Méta modèle de profil utilisateur

La classification, l'organisation et la structuration des données des profils est un élément clé si on veut avoir une vision globale de la personnalisation. Différents travaux ont abordé cet aspect sans le couvrir dans son ensemble. Par exemple dans [AMA99], les auteurs proposent un modèle de profil pour les utilisateurs d'une bibliothèque digitale composé de cinq catégories : données personnelles, données collectées, données de livraison, données de comportement et données de sécurités. La figure 13 donne un aperçu de la structure générique d'un profil. Chaque dimension est constituée d'un ensemble d'attributs dont les valeurs peuvent être simples ou complexes. Certaines dimensions sont organisées en sous-dimensions selon la nature de leurs attributs. Un attribut est défini par un nom, un type, une expression de préférence et une sémantique.

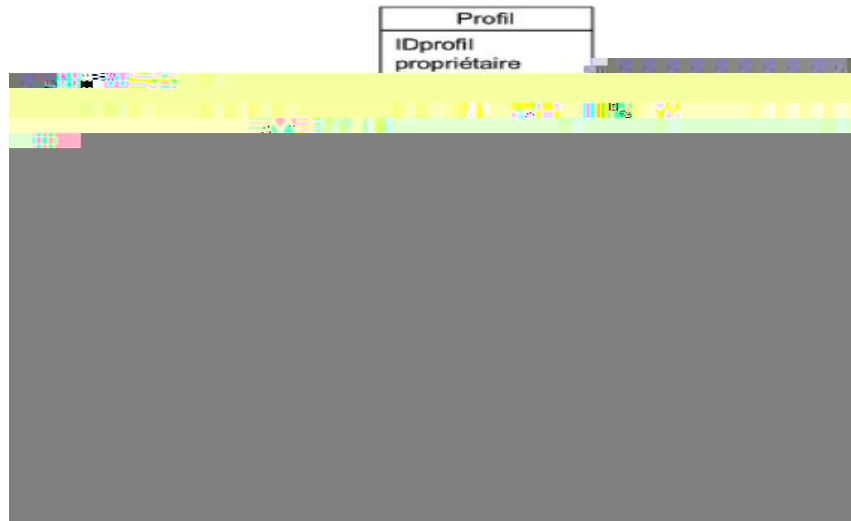


Figure 13 : Méta-modèle du profil utilisateur [KOS07]

Le profil utilisateur présenté entre le cadre de **l'approche multidimensionnelle**, est identifié à travers 5 principales dimensions : domaine d'intérêt, données personnelles, données de qualité, données de livraison et données de sécurité.

3.2.2.1 Domaine d'intérêt

Le domaine d'intérêt est la dimension centrale du profil utilisateur. Cette dimension regroupe tous les attributs qui concernent les *objets de contenu* (informations ciblées). Elle peut définir aussi bien le domaine d'expertise et le niveau de qualification de l'utilisateur dans un domaine particulier que le contenu auquel s'intéresse l'utilisateur. Le domaine d'intérêt peut être représenté de différentes manières (Figure 14) :

- par un vecteur de mots clés,
- par un ensemble de formules logiques,
- par un historique des interactions entre l'utilisateur et le système,
- par ontologies.

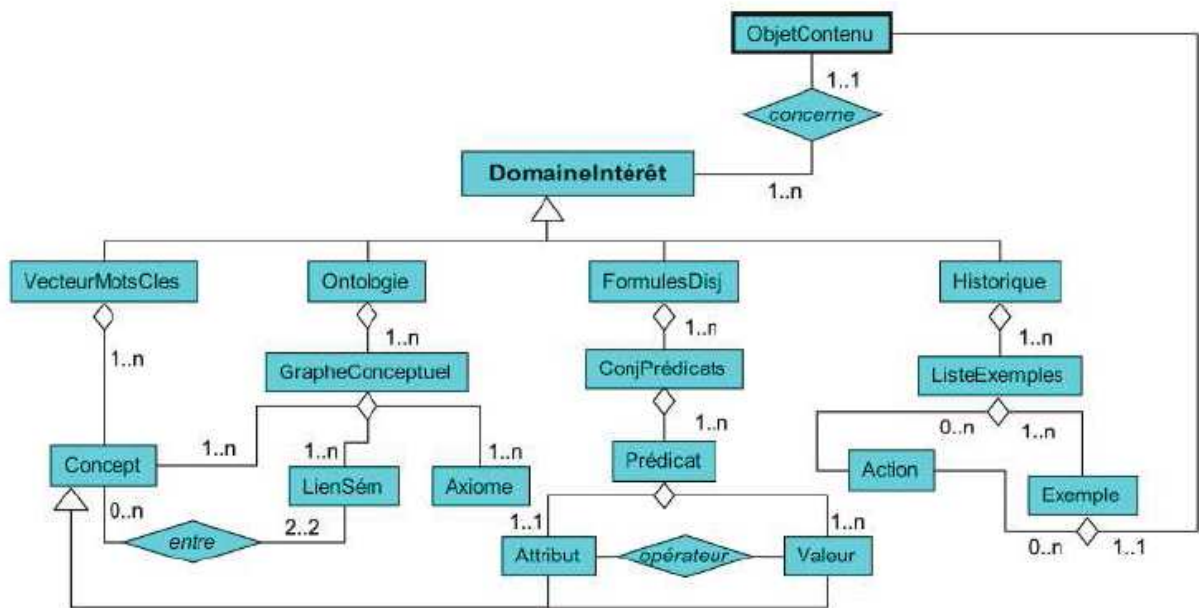


Figure 14 : Méta modèle de la dimension domaine d'intérêt [KOS07]

3.2.2.2 Données personnelles

Les données personnelles sont la partie statique du profil. Elles contiennent des informations qui décrivent l'utilisateur et ne dépendent pas du système à interroger. Ces informations sont classées selon trois catégories (Figure 15):

1. Données personnelles : catégorie (identité) est composée d'un ensemble d'attributs d'identification de l'utilisateur. Des exemples de tels attributs sont le nom et le prénom de l'utilisateur, son adresse, son numéro de téléphone ou de fax, son adresse email etc.,
2. Données démographiques : attributs démographiques comme par exemple la date de naissance de l'utilisateur, son genre, son revenu, son état civil, le nombre d'enfants etc.
3. Données des contacts : les contacts de l'utilisateur représentent son carnet d'adresse. Un élément de ces contacts est représenté par les données personnelles d'un autre utilisateur.

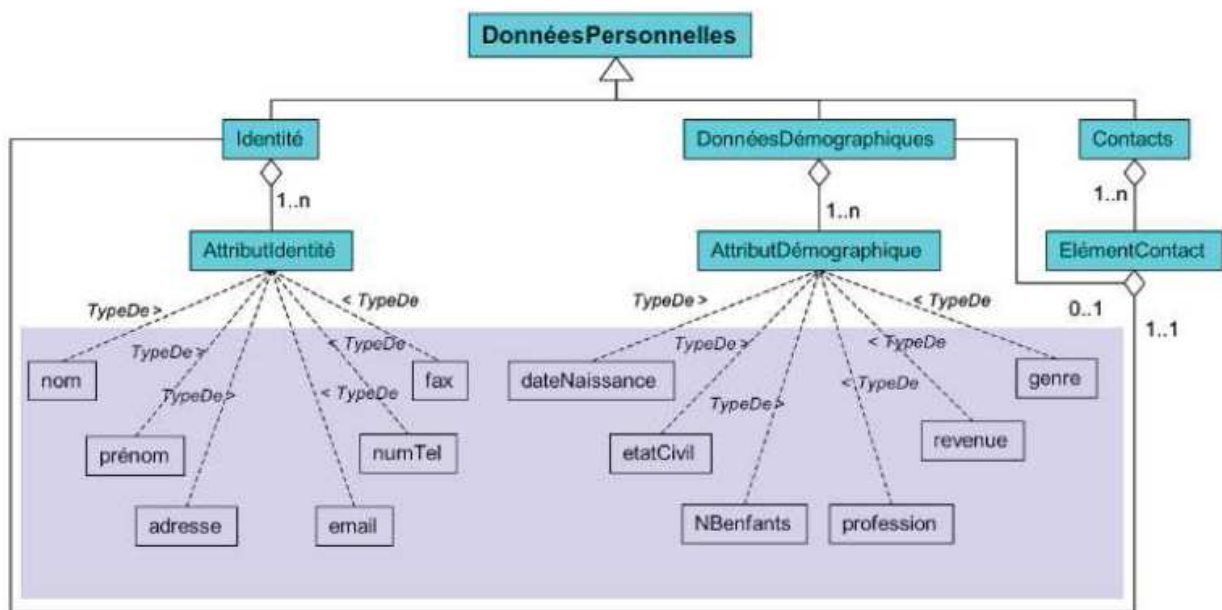


Figure 15 : Méta modèle de la dimension données personnelles [KOS07]

3.2.2.3 Qualité

La dimension « Qualité » joue un rôle très important dans le domaine de la personnalisation. **Les données de cette dimension décrivent la qualité attendue ou espérée par l'utilisateur qui sera confrontée à la qualité effective produite par le système de recherche d'informations afin de restreindre l'espace de recherche.** Les informations dans cette dimension sont représentées par des facteurs de qualité (Figure 16). Trois catégories de facteurs de qualité selon le type d'objets auxquels ils font référence sont distinguées :

1. Facteurs sur le contenu des données : cette catégorie concerne la qualité désirée des objets de contenu. Des exemples de facteurs de cette catégorie sont la fraîcheur et l'exactitude des données;
2. Facteurs sur le contenant des données : catégorie regroupe les facteurs de qualité des conteneurs des données (sources). La plupart des facteurs de cette catégorie résultent de l'expérience des autres utilisateurs (ex. le niveau de confiance dans la source, la fiabilité etc.). Ces facteurs sont utilisés pour filtrer les sources de données lorsque leur nombre est trop important;
3. Facteurs sur les processus : délivrent ou notifient l'arrivée des résultats. Les facteurs de cette catégorie mesurent aussi bien les performances des processus (ex. temps de réponse) que leur capacité de produire tous les résultats pertinents (rappel) et uniquement les résultats pertinents (précision).

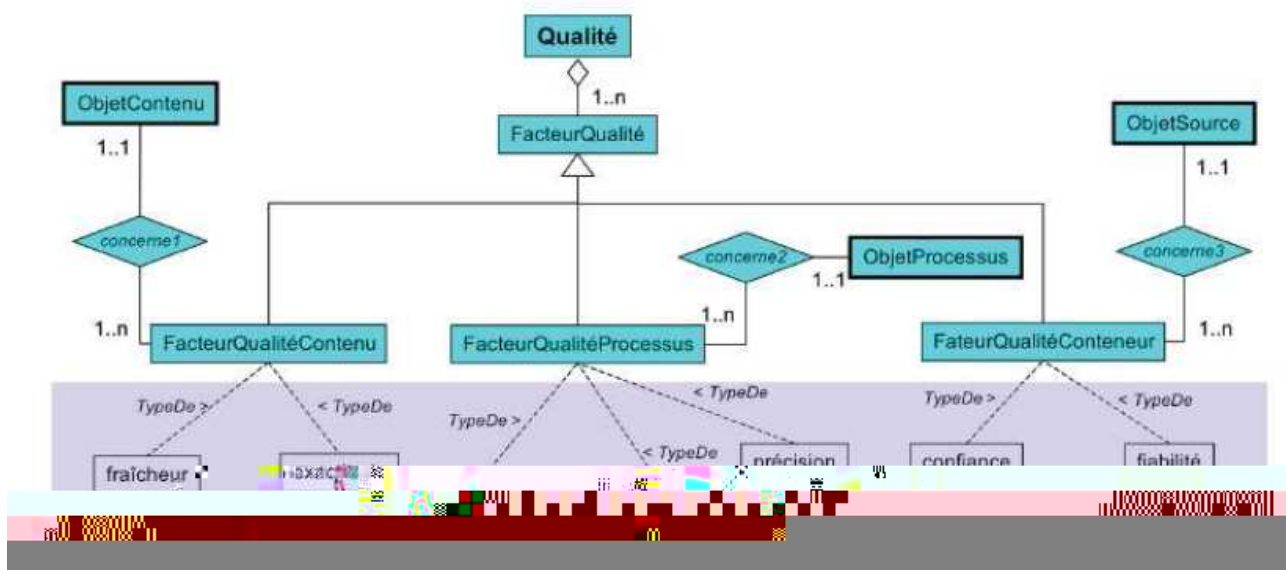


Figure 16 : Méta modèle de la dimension Qualité [KOS07]

3.2.2.4 Données de livraison

Les données de livraison concernent d'abord tout ce qui est lié aux modalités de présentation des résultats en fonction de la plateforme de l'utilisateur, de la nature et du volume des informations délivrées, des préférences esthétiques ou visuelles de l'utilisateur. À ces modalités de présentation, on peut ajouter les modalités de livraison et de notification, décrivant le moment et le moyen de livraison des résultats et la manière de notifier cette livraison (différé, immédiat par exemple).

Les données de livraison sont divisées en deux classes (Figure 17) :

1. des caractéristiques des objets de contenu délivrés : le format des documents, leur nombre et leur taille ou encore la mise en page des données;
2. des informations sur les interactions entre le système et l'utilisateur : des informations sur le moment et le moyen des interactions entre l'utilisateur et le système.

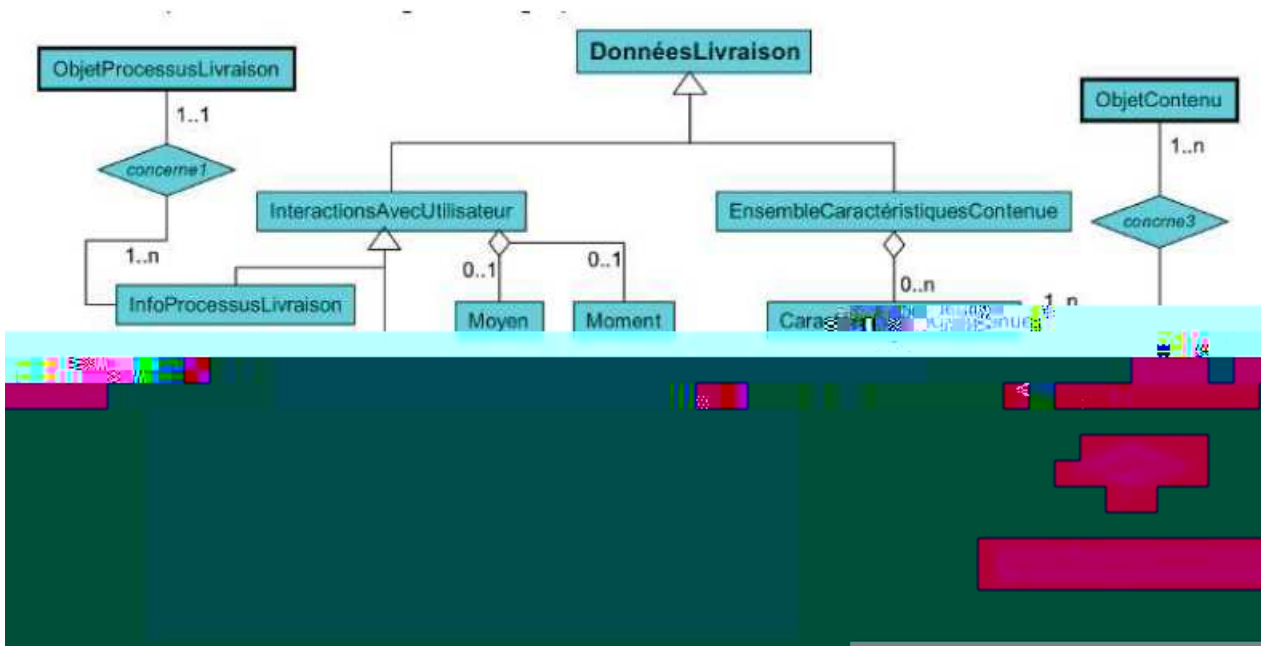


Figure 17 : Méta modèle de la dimension données de livraison [KOS07]

3.2.2.5 Données de sécurité

La sécurité est une dimension fondamentale du profil. Elle peut concerner les données que l'on interroge ou modifie, les informations que l'on calcule, les requêtes utilisateurs ou les autres dimensions du profil. Trois types de sécurité en fonction des objets concernés sont distingués (Figure 18) :

1. la sécurité du profil utilisateur : se fait à travers des autorisations d'accès aux informations du profil. Les utilisateurs d'un profil, qui peuvent être le système d'information ou les autres utilisateurs particuliers, se voient attribuer des droits d'accès ou des paquets de droits d'accès (rôle) afin de pouvoir interroger les données de ce profil.
2. la sécurité des résultats : concerne les objets de contenu qui sont délivrés à l'utilisateur. Elle est exprimée en indiquant le niveau de sécurisation des résultats (sur une échelle prédéfinie) ou à travers des expressions dans un formalisme existant.
3. la sécurité du processus de production des résultats : exprime la volonté de l'utilisateur de cacher un traitement sensible qu'il effectue.

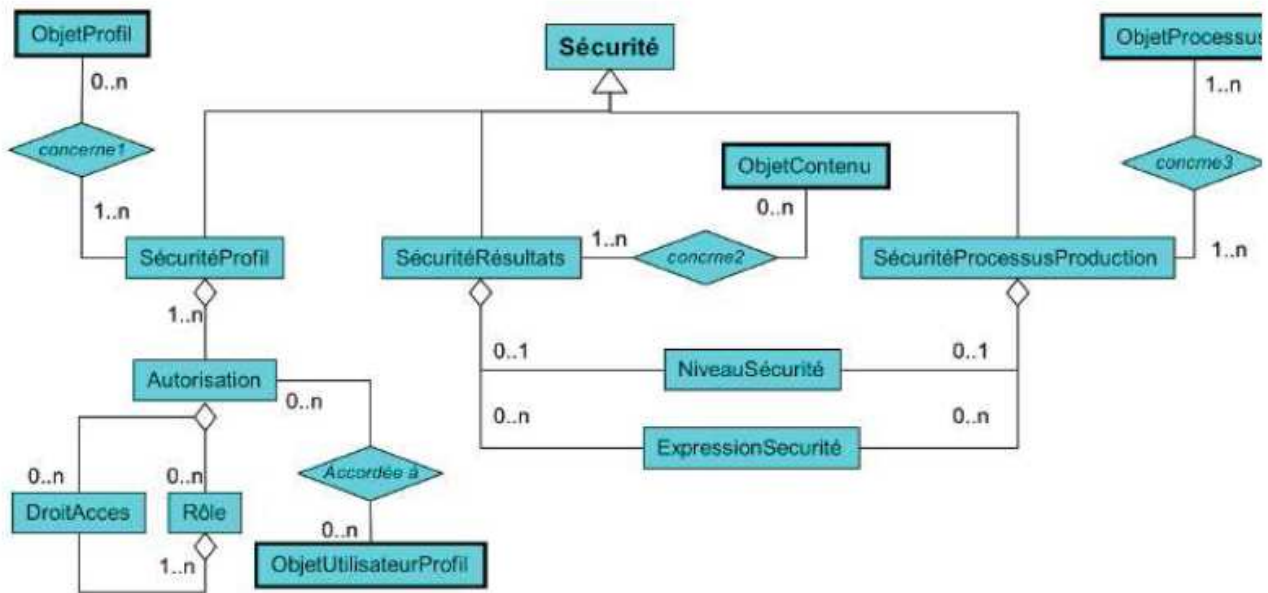


Figure 18 : Méta modèle de la dimension sécurité [KOS07]

3.2.2.6 Dimensions et sous-dimensions du modèle de profil utilisateur

M. Bouzeghoub présente dans [BOU04] un ensemble de dimensions ouvertes, capables d'accueillir la plupart des informations caractérisant un profil. Il distingue principalement sept dimensions auxquelles il a rajouté un huitième constituant le pot-pourri des informations inclassables:

1. les données personnelles,
2. le centre d'intérêt,
3. l'ontologie du domaine,
4. la qualité attendue des résultats délivrés,
5. la customisation (données de livraison),
6. la sécurité et la confidentialité,
7. le retour de préférences (feedback),
8. les informations diverses.

L'ensemble de ces dimensions est représenté par le schéma de classes de la figure 19. Nous remarquerons la subdivision des dimensions en sous-dimensions pour mieux faire apparaître les différentes catégories d'information définissant les profils. Ces sous-dimensions peuvent elles mêmes être décomposées par un affinage ultérieur. Chaque dimension ou sous-dimension est décrite par un ensemble d'attributs ou d'expressions types.

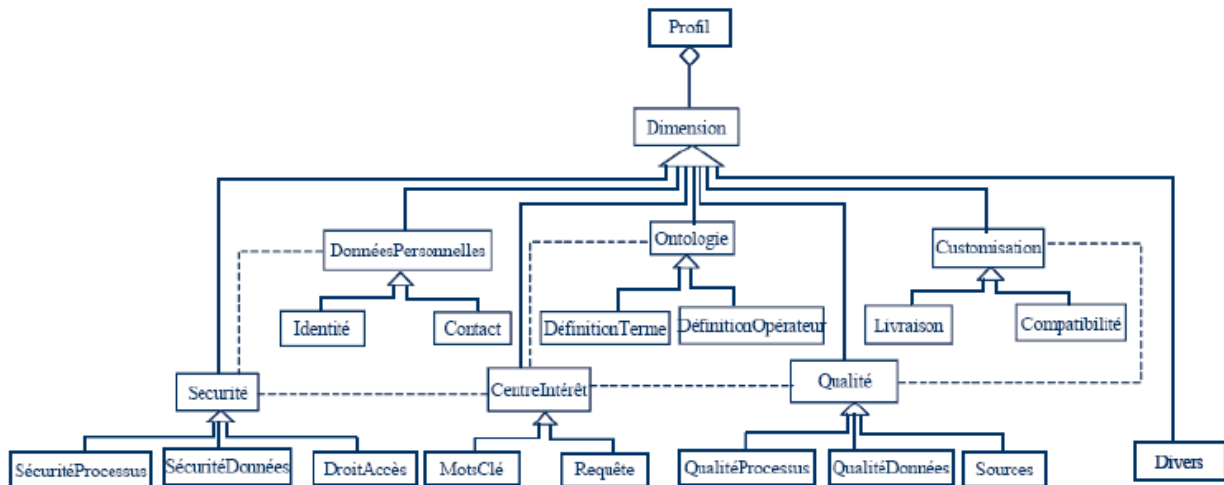


Figure 19 : Dimensions et sous-dimensions du modèle de profil [BOU04]

3.2.3 Méta modèle de contexte

Le contexte regroupe toutes les informations relatives à l'environnement dans lequel se fait l'interaction entre l'utilisateur et le système d'information. Le contexte est composé d'un ensemble de dimensions où chaque dimension contient des attributs simples ou composés (sous dimensions). Les principales dimensions du contexte sont : la dimension spatiale, la dimension temporelle et l'équipement utilisé par l'interaction (Figure 20). Le contexte permet de situer l'utilisateur dans le temps et dans l'espace. Ceci est fait à travers les dimensions temporelle et spatiale. La dimension temporelle indique le moment de l'interaction. Ce moment peut être représenté dans un format de date standard ou en utilisant un libellé qui indique la période du temps (matin, soir, ...). La dimension spatiale désigne l'endroit dans lequel se trouve l'utilisateur. Cet endroit peut être statique (à la maison) ou dynamique (en voiture). La dimension spatiale peut contenir les coordonnées précises de l'utilisateur ou le libellé général de la localité dans laquelle il se trouve (travail, maison, voiture, ...). Comme pour la dimension temporelle, les libellés de la dimension spatiale peuvent changer d'une application à une autre. En plus de l'aspect spatio-temporel, le contexte contient des informations sur l'équipement de l'utilisateur. Cet équipement peut être représenté par des caractéristiques du matériel et/ou du logiciel dont se sert l'utilisateur. Des exemples de caractéristiques du matériel sont la taille de l'écran, la capacité de la mémoire, la rapidité du processeur etc. En ce qui concerne le logiciel, il peut être décrit par le système d'exploitation de la machine de l'utilisateur ou encore les logiciels de bureautique qu'il utilise. Les caractéristiques de l'équipement servent à rendre les résultats exploitables sur la plateforme de l'utilisateur [KOS07, BOU04].

Il est intéressant à remarquer que le contenu du contexte est très proche de celui de la dimension « Données de livraison » du profil utilisateur. Le contexte détermine les valeurs du

profil utilisateur qui peuvent être prises en compte, mais si plusieurs options sont possibles, le choix de celle qui sera prise est fait en fonction des préférences qui se trouvent dans le profil. Par exemple si plusieurs formats de fichiers sont compatibles avec le système de l'utilisateur, alors le profil décidera lequel choisir en fonction des préférences de l'utilisateur. Il est évident que le choix dicté par le profil peut être contextualisé i.e. changé d'un contexte à un autre [KOS07].

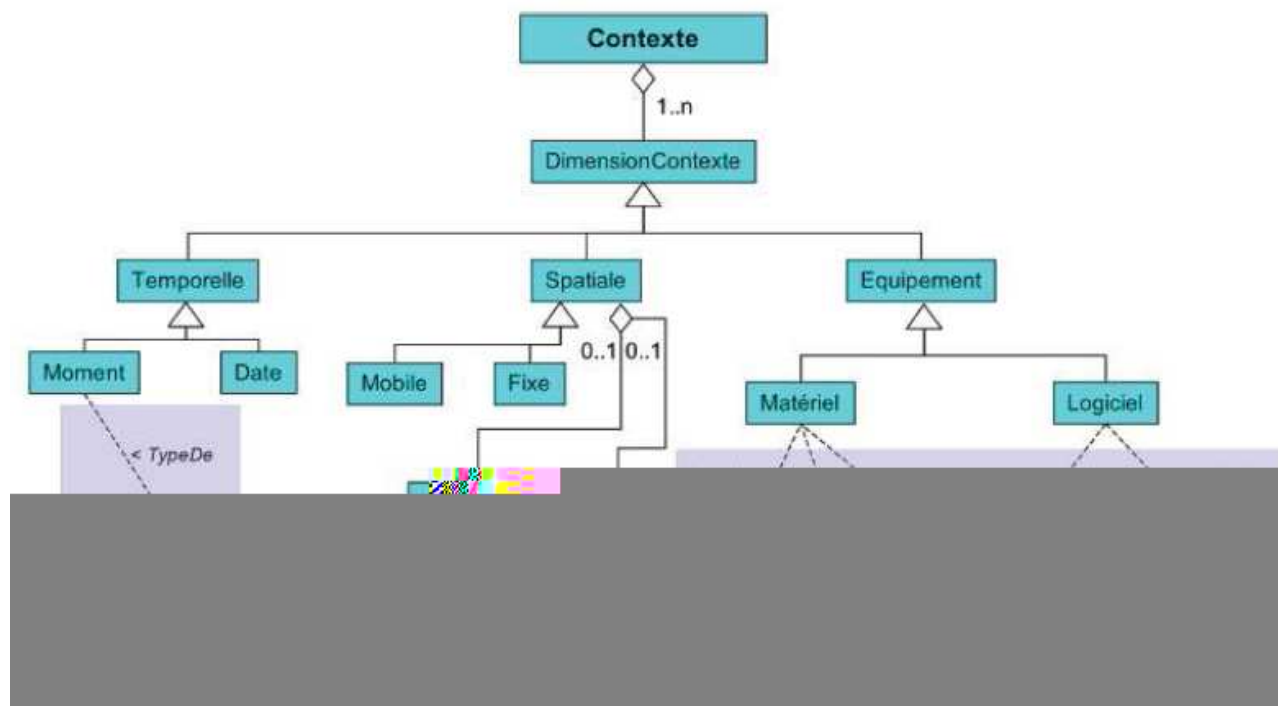


Figure 20 : Méta modèle du Contexte [KOS07]

3.2.4- Méta modèle de préférences

Le méta modèle de préférence décrit les types de préférence qu'on peut trouver dans les profils et dans les contextes. Ce modèle n'est pas instanciable séparément des deux autres. Il fournit une palette de préférences qui peuvent être utilisées pour hiérarchiser l'importance des connaissances définies dans les profils et les contextes. Il y a deux types de préférences : des préférences simples et des préférences composées (Figure 21).

Une préférence simple est définie par un seul concept caractérisant un ou plusieurs objets. Ce concept peut être un poids, un score, une fonction d'utilité, un prédicat ou tout mot clé introduisant une appréciation sur un objet ou un ordre d'intérêt entre deux ou plusieurs objets. Selon le nombre d'objets caractérisés, les préférences simples peuvent être unitaires, binaires ou ensemblistes.

Une préférence composée est une expression de deux ou plusieurs *préférences*

intermédiaires. Une préférence intermédiaire peut être une préférence simple ou une préférence composée. La combinaison de préférences peut se faire de façon indépendante ou prioritaire.

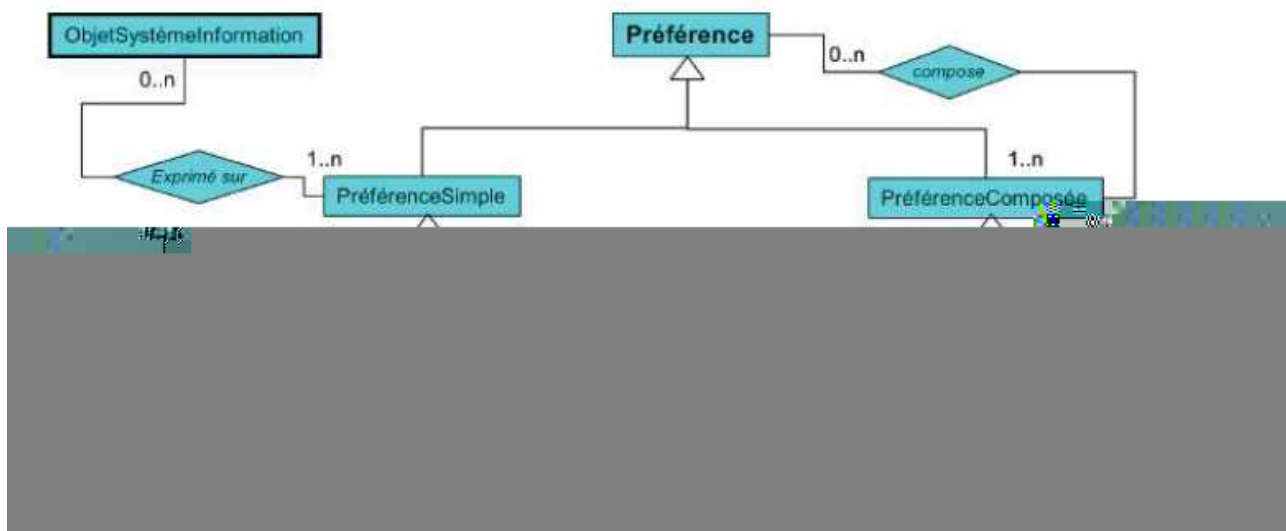


Figure 21 : Méta modèle de préférences [KOS07]

3.2.5 Relations entre le profil, le contexte et les préférences

L'obtention d'un modèle complet décrivant l'utilisateur et/ou le contexte passe obligatoirement par l'identification des relations qui existent entre les trois composants profil, contexte et préférences. La figure 22 illustre les principaux liens qui existent entre le profil, le contexte et les préférences. En plus de l'agrégation, trois catégories d'associations sont utilisées pour construire le modèle de cette figure : « *type de* », « *défini sur* » et « *contextualisé selon* ». Les liens de la première catégorie permettent de différencier les préférences du profil de celles du contexte. La seconde catégorie de liens sert à relier les préférences aux éléments sur lesquels elles sont exprimées. Finalement, les associations « *contextualisé selon* » assurent la relation entre le contexte et le contenu du profil.

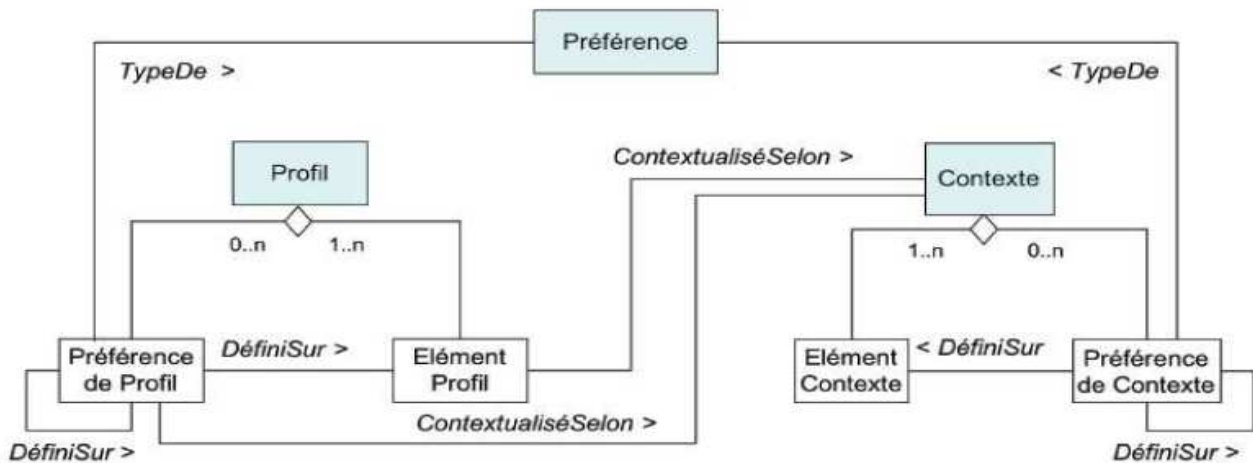


Figure 22 : Associations entre profil, contexte et préférences [KOS07]

3.2.6 Gestion de profils

La gestion de profils implique la mise en place d'un ensemble d'outils qui facilitent la définition et l'évolution des profils. Les principales opérations de gestion de profils sont [KOS07] :

- **Instanciation** : cette opération permet de créer des profils particuliers.
- **Importation** : cette opération permet de rendre un profil utilisable sur un système qui n'est pas celui qui a servi à le créer. Elle implique un test de validité du profil par rapport au modèle de profil reconnu par le système (validité sémantique).
- **Validation** : la validation d'un profil permet de déterminer quelle partie de ce profil peut être satisfaite sur un système particulier. Elle implique un test par rapport à l'environnement technique dans lequel le profil sera utilisé (validité opérationnelle).
- **Intégration** : l'intégration permet de combiner le contenu de deux ou plusieurs profils dans un seul. Cette opération peut être appliquée sur les différents profils du même utilisateur ou sur les profils de différents utilisateurs. Dans le premier cas, l'objectif est d'uniformiser la représentation de l'utilisateur, alors que dans le second cas il s'agit de créer le profil d'une communauté d'utilisateurs.
- **Adaptation** : l'adaptation d'un profil se fait toujours par rapport à un contexte. Elle permet d'obtenir le contenu du profil par rapport au contexte dans lequel se trouve l'utilisateur.
- **Appariement** : l'appariement de profils comprend un ensemble d'opérations de base qui sont le matching, la différence et l'équivalence de profils. Ce sont des opérations qui peuvent être utilisées comme briques de base pour obtenir les autres opérations.

3.2.6.1- Instanciation des modèles

L'instanciation est un processus qui permet de créer des instances des différents modèles et d'adapter leur contenu à des besoins particuliers. Nous avons deux niveaux d'instanciations [KOS07]:

- Instanciation des méta-modèles pour avoir les modèles.
- Instanciation des modèles pour avoir les instances;

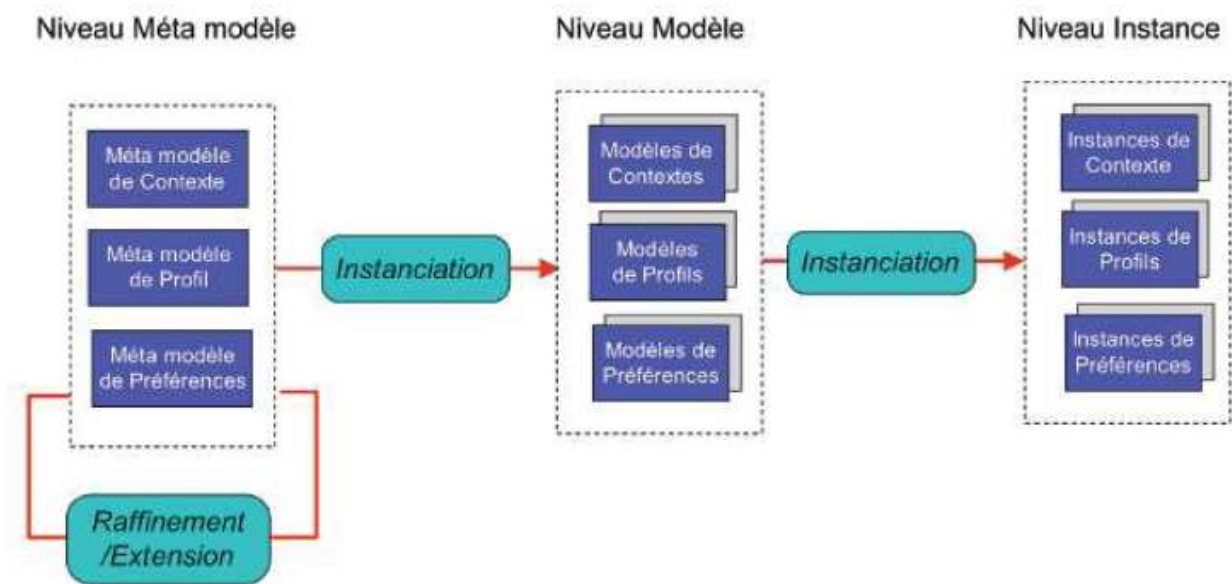


Figure 23 : Niveau d'instanciation [KOS07]

3.3 Détail de la dimension qualité

Les facteurs de qualité sont utilisés comme des attributs dans la description des données et les préférences des utilisateurs sur ces attributs de qualité s'expriment sous forme de poids ou de seuils. La qualité peut d'intégrer dans certaines activités de personnalisation entre autres nous présentons [BOU07]

- sélection de sources de données: Quand plusieurs sources de données fournissent les mêmes types de données, la sélection de sources consiste à choisir les sources qui fournissent les « meilleures » données afin de répondre à une requête utilisateur.
- sélection de plans d'exécution : Dans le contexte de BD relationnelles, les facteurs de qualité peuvent être exploités durant la compilation d'une requête pour choisir le meilleur plan d'exécution. Plus précisément, différents plans accédant à différentes sources sont comparés et ordonnés par rapport à la qualité des données. (Sélection du traitement)

Les préférences de qualité de l'utilisateur peuvent être exprimées de différentes manières

selon les besoins des applications qui les utilisent. Elles peuvent concerner **un facteur particulier** ou un **ensemble de facteurs** [BOU07]. Afin de concrétiser l'importance de la dimension «qualité» citée dans 2.2.3, l'utilisateur doit être capable de comprendre la signification de chaque facteur et des métriques utilisées pour l'évaluer.

Les trois catégories de facteurs de qualité seront détaillées comme suit :

3.3.1 Facteurs sur le contenu des données

Cette catégorie concerne la qualité désirée des objets de contenu. Ces facteurs peuvent pris en charge par la norme ISO 19 113 dans le des données géographiques indiquée dans [TRI07, SER06]. Dans la suite, une synthèse des familles de facteurs de qualité présentée par Bouzeghoub dans [BOU07] : la fraîcheur et l'exactitude.

3.3.1.1 Fraîcheur

Intuitivement, le concept de fraîcheur introduit l'idée d'âge des données :

- Les données sont-elles suffisamment fraîches pour les utilisateurs ?
- Les données sont-elles obsolètes ?
- Une certaine source de données a-t-elle les données les plus récentes ?
- Quand les données ont-elles été produites ?

Le Tableau 05 montre les métriques de fraîcheur proposées dans la littérature ; une étude détaillée est présentée dans [BOU07].

Facteur	Métrique	Description
Actualité	Actualité	Le temps passé depuis l'extraction des données
	Obsolescence	Le nombre de transactions de mise à jour d'une source depuis l'extraction des données
	Ratio de fraîcheur	Le taux de données qui sont à jour
Age	Age	Le temps passé depuis la création ou la mise à jour des données

Tableau 05: Métriques de fraîcheur [BOU07]

3.3.1.2 Exactitude

L'exactitude est liée à la correction et la précision avec laquelle les données du monde réel sont représentées dans un système d'information :

- Les données, correspondent-elles au monde réel ?
- Comportent-elles des erreurs ?

- Le niveau de détail des données est-il acceptable ?

Le Tableau 06 liste les métriques d'exactitude proposées dans la littérature ; une étude détaillée est présentée dans [PER06].

Facteur	Métrique	Description
Correction sémantique	Taux de correction sémantique	Le taux de données qui ne correspondent pas au monde réel
	Taux de valeurs inexistantes	Le taux de données qui n'existent pas dans le monde réel
	Taux de valeurs inexacts	Le taux de données qui contiennent des erreurs de représentations pour certains attributs
	Degré de correction sémantique	Le degré de confiance dans la correction sémantique des données
	Déviations de correction sémantique	La distance sémantique entre une données et son correspondant dans le monde réel
Correction syntaxique	Taux de correction syntaxique	Le taux de données qui satisfont les règles syntaxiques
	Déviations de correction syntaxique	La distance syntaxique entre une donnée et une valeur de référence considérée comme correcte
Précision	Échelle	La précision associé à l'échelle de mesure
	Erreur standard	La déviation standard d'un ensemble de mesure
	Granularité	Le nombre d'attributs utilisés pour représenter un concept simple

Tableau 06. Métriques d'exactitude [PER06]

3.3.2 Facteurs sur les processus

Les facteurs de cette catégorie mesurent aussi bien les performances des processus (ex. temps de réponse) que leur capacité de produire tous les résultats pertinents (rappel) et uniquement les résultats pertinents (précision). Le processus dans notre travail est considéré comme un service Web puisque les services Web sont fournis par des tiers et être appelés dynamiquement au-dessus de l'Internet, leur QoS peut varier considérablement. Par conséquent il est important d'avoir un cadre pour capturer le QoS donné par le fournisseur et le QoS exigé par le client, et finalement la correspondance entre les deux en découvrant le meilleur service Web selon la QoS exigée [SHU03]. Dans la littérature nous trouvons plusieurs classifications de la QoS [SHU03, MAN02, MED05]. D'après l'étude effectuée des différentes classifications des critères de qualités de services nous pouvons regrouper les qualités de services Web dans les catégories suivantes :

3.3.2.1 Qualité de performance

1. **accessibilité** : l'accessibilité est l'aspect de qualité d'un service qui représente le degré qu'il est capable de servir une demande de service de Web. Elle peut être exprimée comme mesure de probabilité dénotant l'indice de réussite ou la possibilité d'une instanciation réussie de service à un moment. Il pourrait y avoir des situations quand un service Web est disponible mais non accessible.
2. **Fiabilité** : la fiabilité est l'aspect de qualité d'un service Web qui représente le degré d'être capable de maintenir le service et la qualité de service. Le nombre d'échecs par mois ou année représente une mesure de fiabilité d'un service Web. Dans un autre sens, la fiabilité se rapporte à la livraison assurément et commandée pour des messages envoyés et reçus par des demandeurs et des prestataires de service.
3. **Intégrité** : l'intégrité est l'aspect de qualité indiquant la façon dont le service Web maintient l'exactitude de l'interaction en ce qui concerne la source. L'exécution appropriée des transactions de service Web fournira l'exactitude de l'interaction. Une transaction se rapporte à un ordre des activités à traiter comme unité simple de travail. Toutes les activités doivent être accomplies pour effectuer une transaction réussie. Quand une transaction n'accomplit pas, toutes les modifications apportées sont compensées.
4. **Disponibilité** : la disponibilité est l'aspect de qualité indiquant si le service Web est présent ou prêt pour l'usage immédiat. La disponibilité représente la probabilité qu'un service est disponible.
5. **Exactitude** : définit le taux d'erreur produit par le service. (Question – Combien d'erreurs le service produit sur une période de temps ?)
6. **Normalisation** : la normalisation est l'aspect de qualité du service Web dans la conformité avec les normes.
7. **Coût** : c'est une mesure du coût entraîné en demandant le service.

3.3.2.2 Qualité liée au contexte d'exécution

1. **Temps d'exécution** : Le temps maximum garanti demandé pour compléter une requête du service.
2. **latence** : Temps pris entre l'arrivée de la requête du service et la réponse émise par le service.

3. Sortie : le nombre de demandes réalisées de service sur une période de temps. La sortie est liée à la latence et à la capacité.

3.3.2.3 Qualité de sécurité

1. Authentification : Comment le service authentifie-t-il des principaux (des utilisateurs ou d'autres services) qui peut avoir accès au service et des données ?
2. Confidentialité : Comment le service traite-t-il les données, pour que seulement les principaux autorisés puissent avoir accès ou modifie les données ?
3. Traçabilité et vérifiabilité : Est-il possible de tracer l'histoire d'un service lorsqu'une demande a été entretenue?

Remarque: Nous pouvons trouver d'autres facteurs de qualité de services Web et d'autres classifications. Existe-t-il une norme pour la qualité des services ?

3.3.3 Facteurs sur le contenant des données

Cette catégorie regroupe les facteurs de qualité des conteneurs des données (sources). La plupart des facteurs de cette catégorie résultent de l'expérience des autres utilisateurs (ex. le niveau de confiance dans la source, la fiabilité etc.). Ces facteurs sont utilisés pour filtrer les sources de données lorsque leur nombre est trop important.

3.3.3.1 Popularité

La popularité fait référence au degré d'utilisation effective de l'information par les utilisateurs. Ce facteur présente un grand intérêt car il permet de capturer des items qui sortent du champ habituel de contenu auquel s'intéresse l'utilisateur a priori. Il peut donc contribuer à contrecarrer l'effet dit « entonnoir », en apportant des items inattendus sous l'angle de leur strict « contenu », et néanmoins appréciés pour leurs qualités reconnues par toute une communauté d'utilisateurs. Le Tableau 7 liste les métriques de la popularité proposées dans la littérature ; une étude détaillée est présentée dans [SCH01]

Facteur	Métrique	Description
Popularité	Nombre de consultations	Le nombre de consultations des données
	Rating	L'évaluation ou recommandation donnée par un groupe d'utilisateurs
	Nombre de citations	Le nombre de citations référençant les données
	Quantité d'achats	Le nombre d'achats

Tableau 07: Métriques de popularité [SCH01]

3.4 Requête utilisateur

D'après l'encyclopédie wikipedia , en informatique, une requête (en anglais request) est une demande de traitement. Le terme est notamment employé dans le contexte:

- des bases de données, une requête correspond à l'interrogation d'une base pour en récupérer une certaine partie des données ;
- des protocoles client-serveurs, une requête est le message initial envoyé par le client vers le serveur. Le message du serveur étant la réponse

Dans notre contexte la requête représente une demande pour trouver un service web géographique satisfaisant l'utilisateur. Dans cette section nous présentons la classification des requêtes proposée par A.Yann dans sa thèse [YAN06], la requête géographique et la forme canonique d'une requête d'exploitation des données géographiques.

3.4.1 Classification de requêtes selon A. YAN

Cette classification propose les catégories de requêtes suivantes [YAN06]:

3.4.1.1 Requetes de recherche de traitements

Un utilisateur cherche un ou plusieurs traitements, en vue de satisfaire un besoin précis et bien défini, ou au contraire vague. Voici des exemples de requêtes.

- Ou sont disponibles les programmes d'appariement ?
- Quels algorithmes de détection de contour ont été développés à l'IGN en 1995 ?
- Quels SIG possèdent la fonctionnalité "buffer" ?
- Quels sont les algorithmes de calcul de flux sur un réseau ?
- Quels sont les traitements utilisables sur les objets "bâtiments" ?
- Quels sont les programmes du COGIT qui utilisent l'API GeOxygene ?

Certaines requêtes sont plus complexes que d'autres. Plutôt qu'une simple sélection dans la base de métadonnées, ces requêtes requièrent la mise en œuvre des raisonnements.

3.4.1.2 Requetes pour connaître les traitements

L'analyse des requêtes typiques exprimées par les utilisateurs montre que pour un traitement donné les besoins d'information portent sur cinq thèmes principaux : les métadonnées qui l'identifient (nom, date, auteur, etc.), "ce qu'il fait", "comment il fonctionne", "comment l'utiliser" et "quelle évaluation en est faite".

Voici un exemple de requêtes pour chacun des cinq thèmes en question.

- Quel chercheur a développé le plus de traitements au laboratoire COGIT en 2005 ?
- Quels sont les requêtes topologiques permises par Arcview 8 ?

- Sur quelle théorie mathématique repose l'algorithme Accordéon ?
- Comment faire un changement de projection Lambert 2 sous Geoconcept ?
- Le programme de détramage « planche mère » est-il rapide ?

Dans notre travail nous faisons appel à une requête qui appartient en même temps au:

- premier type puisqu'elle consiste à découvrir un SWG (chercher un traitement) convenable aux besoins de l'utilisateur;
- deuxième type car elle est basée sur les métadonnées de description de SWG.

3.4.1.3 Requêtes de l'utilisation des traitements

Une des questions les plus couramment posées est probablement "comment utiliser ce traitement ?". C'est également une des questions auxquelles il est le plus difficile de répondre. D'abord parce qu'elle mobilise de nombreuses connaissances, souvent tacites et non liées directement au traitement, ensuite parce que la réponse dépend du contexte d'utilisation.

3.4.2 Requête géographique [SER05, POI06]

L'objectif des requêtes spatiales est de sélectionner, de localiser, ou de manipuler des objets répondant à un ou plusieurs critères. Les requêtes spatiales sont définies par:

- la nature et la structure de l'information d'entrée;
- l'objet de la requête;
- la formulation et la structuration de la requête

Les requêtes spatiales prennent place dans le contexte des systèmes d'information géographique. Ainsi elle hérite des concepts de structuration et de normalisation associés. La structuration et la normalisation des données garantissent la qualité et la faisabilité de l'interrogation, La rigueur de saisie de l'information géographique garantit également la bonne mise en œuvre des requêtes spatiales.

Les requêtes peuvent être classées en trois groupes:

- les requêtes thématiques portant sur les attributs des objets spatiaux;
- les requêtes géométriques portant sur la géométrie des objets;
- les requêtes topologiques portant sur les relations entre objets.

Ces requêtes peuvent classées en familles comme suit :

- Requêtes d'identification : de façon directe sur une carte ou un graphique (pointage avec la

souris);

- Requêtes de localisation, analyse spatiale, analyse temporelle : de façon indirecte par l'écriture de requêtes spatiales: en utilisant la langage SQL (langage de requête des BD), enrichie avec des opérateurs spatiaux ou le langage de requête visuel (icônes, métaphore).

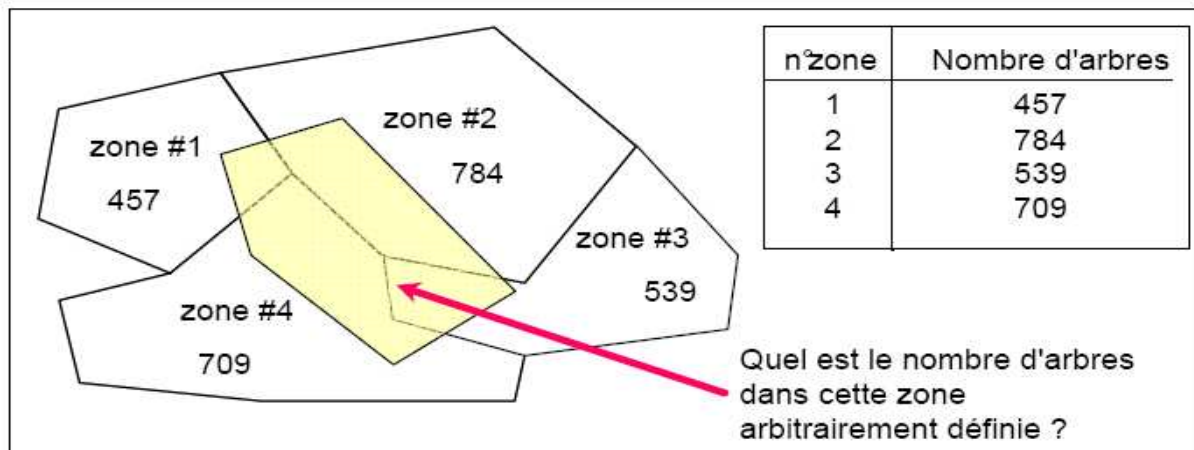


Figure 24 : Exemple de requête d'analyse spatiale [SER05]

3.4.3 Forme canonique des requêtes géographiques:

Pour compléter le traitement de recherche de données dans le cadre de la sémantique des services Web Géographiques (Geospatial Semantic Web), Deux points essentiels à prendre en charge : (1) une forme canonique plausible pour poser des questions sur les données géospatiales (2) des méthodes pour évaluer la sémantique des sources de données disponibles pour déterminer si leur structure sémantique peut être exploitée dans toute requête géographique. La forme canonique proposée par Max J. Egenhofer dans [EGE02] est la suivante :

<geospatial request> ::= <geospatial constraint> [<logical connective> <geospatial request>]

où de multiples contraintes géographiques peuvent être une partie de conjonctions ou disjonctions en utilisant les connecteurs logiques "et" ou "ou" (<logical connective>:: = "et" | "ou"). La requête géographique ressemble à la partie WHERE de la requête spatial SQL [EGE94]. Chaque contrainte géographique est composée de trois parties, deux termes géographiques qui sont liés par un comparateur de données géographiques. Les requêtes sont présentées comme suit :

<geospatial constraint> ::= <geospatial term><geospatial comparator><geospatial term>

<geospatial comparator>:: =! sur la base du rapport d'ontologie géospatiale utilisés

<geospatial term>:: = <geospatial Classe> | <geospatial label>

<geospatial Classe>:: =! sur la base d'une fonction de l'ontologie géospatiales

<geospatial label>:: =! sur la base d'une nomenclature géospatiales

La sémantique des données géographiques de comparaison est capturée à partir d'une ontologie de relations géospatiales; les termes géographiques faisant référence à d'autres classes géographiques.

3.5 Conclusion

Dans ce chapitre nous avons présenté un survol de la recherche d'information en spécifiant d'une part le profil utilisateur. Ce dernier est présenté selon D. Kostadinov dans sa thèse [KOS07] selon 05 dimensions : domaine d'intérêt, données personnelles, données de qualité, données de livraison et données de sécurité. La plus importante dans notre contexte est celle de la qualité dont les données décrivent la qualité attendue ou espérée par l'utilisateur qui sera confronté à la qualité des services web géographiques. Et d'autre part la requête utilisateur en indiquant ses différents types selon A. YAN dans [YAN06] ainsi que ses aspects géographiques.

Dans l'objectif d'optimiser la découverte des services web géographiques en réduisant l'espace de recherche et augmentant le nombre de services pertinents. Nous avons opté à proposer comme première contribution une approche de découverte sémantique des services Web géographiques à base de métadonnées en utilisant les systèmes multi-agents. La description de cette approche fait l'objet du prochain chapitre.

***Chapitre 4: Une approche de la
découverte sémantique des services Web
géographiques basée agent***

Chapitre 4 : Une approche de la découverte sémantique des SWG basée agent

Notre approche de découverte sémantique des services Web géographiques fait appel à quatre domaines de recherche importants à savoir les SIG, le Web sémantique, les services Web et les systèmes multi-agents. Les SIG représente une richesse comme outils de traitement de données géographiques. Le Web sémantique s'intéresse principalement aux informations statiques disponibles sur le Web et les moyens de les décrire de manière intelligible pour les machines. Les services Web, quant à eux, ont pour première préoccupation l'interopérabilité entre applications via le Web en vue de les rendre plus dynamique. Les systèmes multi-agents sont devenus un paradigme incontournable pour la modélisation des systèmes complexes nécessitant un haut degré d'interopérabilité.

L'approche proposée repose sur la coopération d'un ensemble d'agents utilisant les capacités sémantiques des services Web géographiques à travers les catalogues basés sur les métadonnées selon la norme ISO19119. Les agents de notre architecture sont conformes aux propriétés générales des agents (Portabilité, Autonomie, Stabilité, Persistance, Flexibilité, Sociabilité).

Ce chapitre est composé de deux parties pour décrire l'approche proposée. La première partie est consacrée à la présentation des différents concepts utilisés dans notre travail à l'image des services Web sémantiques et particulièrement les ontologies, les catalogues de métadonnées, les systèmes multi-agents ainsi de quelques travaux basés sur les SMA dans le domaine de SIG et le domaine de SW. La deuxième partie fait l'objet de la description de notre approche de la découverte sémantique des SWG basée agent.

4.1 Concepts exploités

Dans notre travail nous avons fait appel à plusieurs concepts à savoir :

4.1.1 Les services Web sémantiques

Les services Web sémantiques [DAM02, FEN02, MCI01] se situent à la convergence de deux domaines de recherche importants qui concernent les technologies de l'Internet : le Web sémantique et les services Web. Les différentes étapes liées à la fourniture d'un service Web (par exemples, découverte et composition) se limitent l'usage des services Web aux utilisateurs humains plutôt qu'aux machines. En effet, de nombreuses connaissances, indispensables pour l'automatisation des services, sont soit absentes, soit décrites pour être interprétées et exploitées par des humains. Il en résulte un rôle prédominant pour le programmeur humain. Il semble donc nécessaire de tendre vers des services intelligibles pour des machines : c'est le concept de service Web sémantique. De manière générale, l'objectif visé par la notion de services Web sémantiques est de créer un Web sémantique de services dont les propriétés, les capacités, les interfaces et les effets

sont décrits de manière non ambiguë et exploitable par des machines et ce en utilisant les couches techniques sans pour autant en être conceptuellement dépendants [KEL03]. L'approche du service Web sémantique est décomposée en trois parties qui sont [BOG07]:

- développement des ontologies pour décrire les données manipulées par des services Web,
- définition de la sémantique des services Web,
- développement d'ontologie des services Web.

4.1.1.1 Les ontologies

Les ontologies sont utilisées, entre autres, pour représenter les données disponibles sur le Web. Les motivations sont la réutilisation, le partage et la communication des connaissances entre systèmes, afin d'intégrer de multiples sources d'information. Un de leurs buts est de faciliter la recherche et l'extraction de connaissances en permettant d'inférer des connaissances. La construction d'une ontologie permet de clarifier les concepts d'un domaine, son but est d'éviter autant que possible les ambiguïtés sémantiques [BOG07]. F. Gandon, dans [GAN02], décrit ainsi ce domaine de recherche : « l'ontologie est un nouvel objet de l'intelligence artificielle, qui a récemment atteint sa maturité en devenant un puissant outil conceptuel pour la modélisation de connaissances ». Une ontologie est suivant le cas soit :

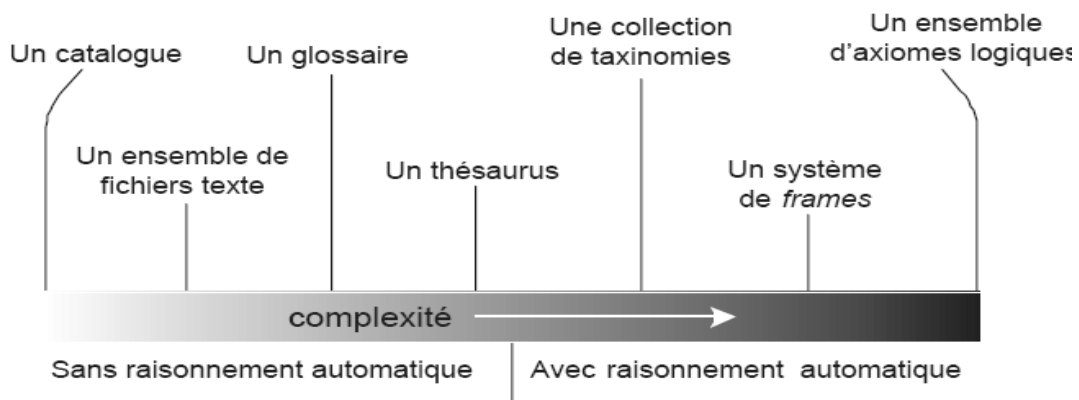


Figure 25: Différentes acceptations du terme "ontologie" (extrait de [TRO04], d'après [SMI01])

4.1.1.1.1 Les composants d'une ontologie

Une ontologie peut être vue comme un treillis de concepts et de relations entre ces concepts destinés à représenter les objets du monde sous une forme compréhensible aussi bien par les hommes que par les machines. Une ontologie est constituée de concepts et des relations ainsi que des propriétés et des axiomes [BAH06].

- Les concepts sont des notions (ou objets) permettant la description d'une tâche, d'une fonction, d'une action, d'une stratégie ou d'un processus de raisonnement, etc. Ils peuvent

être abstraits ou concrets, élémentaires ou composés, réels ou fictifs. Habituellement, les concepts sont organisés en taxinomie. Un taxonomie est une hiérarchie de concepts (ou d'objets) reliés entre eux en fonction de critères sémantiques particuliers.

- Les relations sont les liens organisant les concepts de façon à représenter un type d'interaction entre les concepts d'un domaine. Elles sont formellement définies comme un sous ensemble d'un produit de n ensembles. C'est à dire $R: C_1 \times C_2 \times \dots \times C_n$. Des exemples de relations binaires sont : sous-concept-de, connect-à, sorte de, etc.
- Les propriétés sont des restrictions des concepts ou des relations.
- Les fonctions sont des cas particuliers de relations dans lesquelles le N ième éléments de la relation est unique pour les $n-1$ précédents. Formellement, les fonctions sont définies ainsi $F: C_1 \times C_2 \times \dots \times C_{n-1}, C_n$.
- Les axiomes de l'ontologie permettent de définir la sémantique des termes (classes, relation), leurs propriétés et toutes contraintes quant à leur interprétation. Ils sont définis à l'aide de formules bien formées de la logique du premier ordre en utilisant les prédicats d'ontologie.
- Les instances sont utilisées pour représenter des éléments.

4.1.1.1.2 Méthodologie de l'ingénierie ontologique

Les méthodologies de l'ingénierie ontologique présentées par Fernandez et Corcho [FER99, COR03], aident l'utilisateur à construire des ontologies mais elles restent insuffisantes et aucune d'elles n'est standardisée, Mais, elles se rejoignent sur un processus général de construction d'ontologie. Ce dernier repose sur un enchaînement de quatre étapes (figure 26).

- **La conceptualisation**, sert à dégager à partir des données brutes les concepts et les relations entre ces concepts permettant de décrire de manière informelle les entités cognitives du domaine.
- **L'ontologisation**, est la formalisation partielle, indépendante des contraintes d'opérationnalisation, du modèle conceptuel en précisant la sémantique à l'aide d'axiomes et de contraintes sur les concepts et les relations. elle est semi-formelle, certaines connaissances ne pouvant pas forcément être formalisées à ce niveau.
- **L'opérationnalisation**, est une formalisation de l'ontologie conceptuelle à l'aide d'un langage de représentation de connaissances pour obtenir une représentation formelle des connaissances du domaine, basé sur l'ontologie.

- **L'implémentation**, est le passage du modèle formel de connaissances à son implémentation en machine. Cette étape peut imposer des contraintes lors de l'étape précédente.

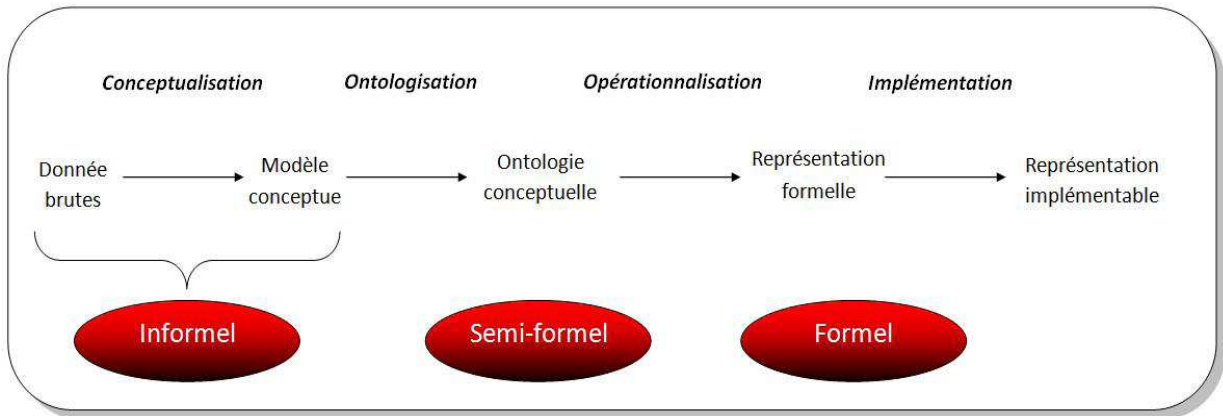


Figure 26 : Processus de construction d'ontologie

4.1.1.1.3 Les types d'ontologies

Les types d'ontologies les plus utilisés sont [BAH06]:

- Les ontologies générales: elles portent sur les concepts généraux qui se veulent indépendants d'un domaine ou d'un problème particulier, tels que les concepts de temps, d'espace, de notions mathématiques.
- Les ontologies du domaine: ce type exprime des conceptualisations spécifiques à des ontologies de domaines particuliers tout en étant générique pour ce domaine. Ces conceptualisations mettent des contraintes sur la structure et les contenus des connaissances du domaine.
- Les ontologies d'application: ces ontologies sont les plus spécifiques. Les concepts correspondent souvent aux rôles joués par les entités du domaine tout en exécutant une certaines activités. Elles peuvent contenir des extensions spécifiques telles les méthodes et les tâches. Elles contiennent toutes les définitions nécessaires pour décrire la connaissance requise pour une application particulière.

4.1.1.1.3.1 Les ontologies Géographiques

Une ontologie géographique représente des concepts géographiques, qui décrivent l'espace géographique ou des entités ou phénomènes de cet espace géographique. Ce qui fait la particularité des concepts géographiques, c'est justement leur relation à l'espace, auxquels ils sont intrinsèquement liés, de telle sorte qu'ils ne peuvent être dissociés du territoire physique qu'ils occupent. En conséquence, décrire des concepts géographiques implique de s'appuyer sur des éléments de représentation spatiale comme les relations métriques (relation de partie à tout), de

topologie (notions de contact et de connexité), de localisation, de distance, d'orientation (alignement, positionnement relatif d'objets), de forme, etc [BAD05].

Suivant le type de concept qu'on souhaite représenter, on peut, suivant F. Fonseca dans [FDC03 cité dans BAD05], distinguer plusieurs types d'ontologies géographiques :

- Les ontologies de « haut niveau » (top-level), qui représentent les concepts les plus généraux liés à l'espace, comme les relations de partie à tout, de topologie, de géométrie. Elles sont également accessibles de nombreuses ontologies basées sur les différentes normes relatives à l'information géographique, comme par exemple la norme ISO 19111 :2003 qui définit le référencement par coordonnées spatiales.
- Les ontologies spécifiques à certains domaines, (qui peuvent être comme dans notre cas des ontologies de domaine de tourisme).
- Les ontologies spécifiques à certaines tâches, comme l'interprétation d'images ou l'évaluation de la pollution sonore.

4.1.1.1.3.2 Les ontologies du domaine de tourisme

Le tourisme est un domaine qui se caractérise par une multitude de sources de données réparties (notamment sur le Web). Ces ressources de données souvent hétérogènes, distribuées et autonomes [JEL08]. Parmi les projets utilisant les ontologies du domaine de tourisme nous présentons **le projet EIFFEL-eTourisme et Web Sémantique (<http://www.projet-eiffel.org/>)**

L'objectif de ce projet est la mise en œuvre d'une plateforme logicielle permettant, autour d'une ontologie tourisme dédiée à un territoire :

- de sélectionner, classer, qualifier des contenus distribués sur le web ;
- d'enrichir automatiquement l'ontologie de nouvelles instances et relations (base de connaissance des ressources touristiques du territoire) en s'assurant de la cohérence de l'ontologie ;
- de traiter des requêtes mettant en œuvre le contenu de l'ontologie du territoire et les contenus Web indexés et d'offrir des fonctions de navigation dans la base de connaissance ;
- d'assister l'utilisateur dans la construction d'un voyage, séjour à partir des ressources sélectionnées.

Le référentiel sémantique, est constitué de deux niveaux distincts et complémentaires :

- une terminologie multilingue du tourisme ;
- une base de connaissance tourisme/territoire.

La terminologie multilingue du tourisme est constituée du thésaurus touristique multilingue (français, anglais, espagnol, italien) de l'Organisation Mondiale du tourisme.

La partie terminologie de l'ontologie est utilisée pour :

- l'indexation des contenus (métadonnées) ;
- la qualification fine des ressources touristiques décrites dans le référentiel ;
- la constitution du vocabulaire de base pour les outils linguistiques de catégorisation et de text mining ;
- l'intermédiation entre l'utilisateur et le moteur de recherche. L'utilisateur peut exprimer ses requêtes dans un langage étendu et dans sa langue. Les fonctions d'extension sémantique basées sur la terminologie permettant de répondre au mieux aux requêtes en identifiant les ressources touristiques et contenus documentaires répondant à la requête.

La terminologie est également utilisée comme outil de suggestion pour des thématiques proches des demandes faites par l'utilisateur.

La partie de l'ontologie décrivant les ressources touristiques, au sens large (activité, hébergement, logistique, patrimoine, territoires, voyages et itinéraires...) constitue la base de connaissances du territoire. La qualification fine des ressources touristiques et les relations décrites entre les ressources (proximité, suggestion, historique, étapes d'un voyage...) sont au cœur de la solution. La base de connaissances permet en effet de présenter les ressources touristiques dans leur contexte, de suggérer des voyages, itinéraires, séjours, activités aux utilisateurs. Elle permet au territoire de valoriser des ensembles cohérents de ressources, de créer des offres nouvelles et composites (route de l'huile d'olive, circuit des peintres, itinéraires de l'aventure...) et de piloter le marketing du territoire par d'autres biais que le prix.

Les descriptions d'objets touristiques issues du mécanisme d'extraction automatique sont contrôlées et enrichies par un moteur de raisonnement. Cela permet d'une part d'éliminer les descriptions incohérentes par rapport à l'ontologie (un rapport d'erreur est alors généré permettant une éventuelle réparation), et d'autre part de compléter les descriptions par inférence à partir d'un jeu de règles modélisant les connaissances implicites du domaine.

Dans notre travail nous exploitons l'ontologie qui comporte uniquement la partie terminologique. Celle-ci se présente sous forme de trois arborescences: les destinations, les types d'hébergement et les activités (Figure 27).

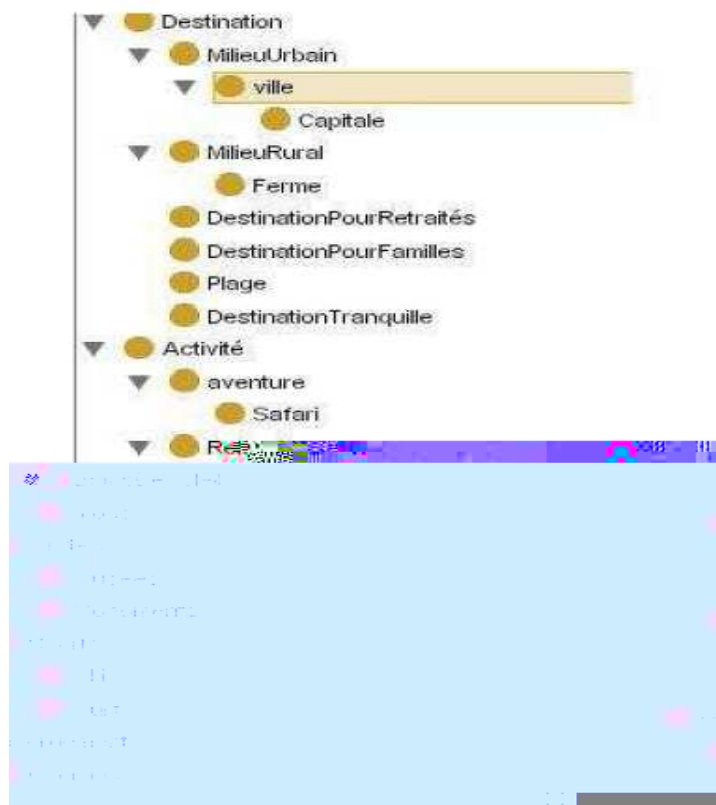


Figure 27 : Aperçu de l'ontologie « tourisme » comme elle apparaît sur l'éditeur protégé [JEL08]

L'ontologie sert à affiner les requêtes utilisateur. Il se peut, en effet, que la requête retourne peu ou pas de résultats. L'affinement se traduit souvent par l'enrichissement sémantique de la requête; il s'agit de transformer la requête en remplaçant ses termes par des termes plus généraux ou plus spécifiques. La partie taxonomique de l'ontologie est organisée sous forme de hiérarchie et l'enrichissement va consister en l'ajout des termes liés aux termes de la requête par des liens de généralisation/spécialisation. L'idée est d'ajouter les termes subsumants i.e., les ancêtres dans l'arborescence dans le cas de la généralisation et les termes subsumés i.e., les descendants dans l'arborescence dans le cas de la spécialisation [JEL08].

4.1.2 Les géocatalogues

Un géocatalogue est défini comme « organisation d'un ensemble de métadonnées sous la forme d'un recueil (base de données ou autre) facilitant leur consultation et leur diffusion ». Il a les objectifs suivants [CNI07]:

- décrire une donnée ou une série de données pour la connaître et la faire connaître (Niveau métadonnées),
- permettre d'organiser des métadonnées structurées de façon identique afin de les valoriser et d'en favoriser la diffusion et la consultation (Niveau géocatalogue);

pour finalement faciliter le partage et la mutualisation des données géographiques proprement dites.

Les métadonnées et les géocatalogues sont des outils complémentaires qui contribuent à améliorer la gestion et l'administration des données au sein d'un organisme.

Il existe aujourd'hui plusieurs outils disponibles en matière de catalogage, certains propriétaires (ArcCatalog* de ESRI, etc.), d'autres libres (GéoSource*, MDWeb*, etc.). Lors de la mise en place d'un géocatalogue, il ne s'agit pas de réinventer ce qui existe, mais d'engager une réflexion plus large qui implique à la fois une approche technique et organisationnelle où la prise en compte de la dimension humaine est capitale [CNI07].

Ce travail s'organise autour de trois éléments principaux [CNI07] :

- 1 la modélisation et la structuration des métadonnées,
- 2 l'utilisation d'un système de stockage adapté et performant,
- 3 l'élaboration d'un cycle de vie optimal des métadonnées.

Les principales fonctions des services de catalogue sont la publication des métadonnées sur les ressources géographiques et l'offre des méthodes pour rechercher et interroger ces ressources. Le concept de catalogues géographiques est basé sur le partage de l'information entre les réseaux distribués. L'Open Geospatial Consortium (OGC, <http://www.opengeospatial.org>) travaille en collaboration avec l'organisation internationale de normalisation (ISO, <http://www.iso.org>) pour élaborer des normes pour assurer l'interopérabilité de l'information géographique et des services.

Parmi les normes nous citons la norme ISO 19115 pour les données géographiques et la norme ISO 19119 pour les services Web géographiques[LAR06].

Parmi les catalogues proposés par l'OCG nous citons [LAR06] :

- **INSPIRE Geo-Portal** (<http://eu-geoportal.jrc.it/>): INSPIRE (Infrastructure for Spatial Information in Europe): est une initiative lancée par la commission européenne, il représente le point d'accès aux données et services spatiale en europe.
- **IDEC Catalog** (<http://www.geoportal-idec.net/>): IDEC (Infraestructura de Dades Espacials de Catalunya) : est un portail de la région du catalogue en ESPAGNE.
- **Gigateway** (<http://www.gigateway.org.uk/>): est un service Web gratuit destiné à croître la sensibilisation et l'accès à l'information géographique dans le ROYAUME UNI.

Trois grands types d'utilisateurs des géocatalogues sont généralement admis [CNI07]:

- administrateur : personne qui coordonne et anime l'utilisation du géocatalogue. Validant la saisie effectuée par les catalogueurs, elle est garante de sa bonne utilisation et de la cohérence des informations,
- catalogueur : personne qui saisit et met à jour les métadonnées relatives aux données qu'il

produit et qu'il gère,

- lecteur : personne accédant au géocatalogue pour la consultation. Elle intervient comme utilisateur final de l'outil. Le travail de valorisation des métadonnées doit être adapté à ses besoins et à ses attentes.

4.1.3 Les systèmes multi-agents (SMA)

Dans cette section, nous faisons un survol sur les SMA en spécifiant la définition et les caractéristiques d'un agent, la définition des SMA et l'utilité des SMA, ainsi que la présentation les couplages SMA/SIG, SMA/SW et le SMA/SWG.

4.1.3.1 Qu'est ce qu'un agent?

Le concept d'agent a été l'objet d'études pour plusieurs décennies dans différentes disciplines. Il a été non seulement utilisé dans les systèmes à base de connaissances, la robotique, le langage naturel et d'autres domaines de l'intelligence artificielle, mais aussi dans des disciplines comme la philosophie et la psychologie. Aujourd'hui, avec l'avènement de nouvelles technologies et l'expansion de l'Internet, ce concept est encore associé à plusieurs nouvelles applications comme *agent ressource*, *agent courtier*, *assistant personnel*, *agent interface*, *agent ontologique*, etc. Dans la littérature, on trouve une multitude de définitions d'agents. Elles se ressemblent toutes, mais elles sont différentes selon le type d'application <pour laquelle est conçu l'agent. L'une des premières définitions de l'agent due à Ferber [FER95] :

Un agent est une entité autonome, réelle ou abstraite, qui est capable d'agir sur elle-même et sur son environnement, qui, dans un univers multiagents, peut communiquer avec d'autres agents, et dont le comportement est la conséquence de ses observations, de ses connaissances et des interactions avec les autres agents.

Il ressort de cette définition des propriétés clés comme l'*autonomie*, l'*action*, la *perception* et la *communication*. D'autres propriétés peuvent être attribuées aux agents. Nous citons en particulier la *réactivité*, la *rationalité*, l'*engagement* et l'*intention*. Récemment, Jennings, Sycara et Wooldridge [JEN98] ont proposé la définition suivante pour un agent :

Un agent est un système informatique, *situé* dans un environnement, et qui agit d'une façon *autonome* et *flexible* pour atteindre les objectifs pour lesquels il a été conçu.

Les notions "situé", "autonomie" et "flexible" sont définies comme suit :

- *situé* : l'agent est capable d'agir sur son environnement à partir des entrées sensorielles qu'il reçoit de ce même environnement. Exemples : systèmes de contrôle de processus, systèmes embarqués, etc.
- *autonome* : l'agent est capable d'agir sans l'intervention d'un tiers (humain ou agent) et contrôle ses propres actions ainsi que son état interne;

- *flexible* : l'agent dans ce cas est :
 - *capable de répondre à temps* : l'agent doit être capable de percevoir son environnement et élaborer une réponse dans les temps requis;
 - *proactif* : l'agent doit exhiber un comportement proactif et opportuniste, tout en étant capable de prendre l'initiative au "bon" moment;
 - *social* : l'agent doit être capable d'interagir avec les autres agents (logiciels et humains) quand la situation l'exige afin de compléter ses tâches ou aider ces agents à accomplir les leurs.

Bien entendu, certains agents auront des caractéristiques additionnelles et, pour certains types d'applications, certains attributs seront plus importants que d'autres.

4.1.3.2 Les systèmes multi-agents (SMA)

Pour Jennings, cité dans [WOO02, JAR05, JEN98], un système multi-agents est un système formé de plusieurs composants autonomes qui sont caractérisés par :

- Chaque agent a des informations ou des capacités de résolution de problèmes limitées, ainsi chaque agent a un point de vue partiel,
- Il n'y a aucun contrôle global du système multi-agents,
- Les données sont décentralisées,
- Le calcul est asynchrone.

4.1.3.2.1 Utilité des systèmes multi-agents

Certains domaines requièrent l'utilisation de plusieurs entités, dont, les systèmes multi-agents qui sont très bien adaptés. Parmi ces domaines nous citons par exemple [CHA08]:

- les systèmes géographiquement distribués comme le contrôle aérien, les bases de données coopératives distribuées, etc.
- les systèmes et les données qui s'y rattachent appartiennent à des organisations indépendantes qui veulent garder leurs informations privées et sécurisées pour des raisons concurrentielles. Par exemple, la majorité des missions maritimes se font maintenant en collaboration avec plusieurs pays.
- les systèmes de résolution des problèmes possédant de multiples méthodes de résolution, de multiples perspectives et/ou de multiples solveurs.

Cette adaptation est due à plusieurs raisons. Entre autres nous citons les suivantes :

- La nature distribuée du cycle de vie des services Web.
- Les agents permettent de partager des connaissances (description sémantique).
- Les agents dialoguent entre eux et négocient la fourniture des services (publier, rechercher, exécuter et offrir l'information).

- Les agents peuvent travailler en parallèle lorsque l'environnement matériel et logicielle permet (La modularité). C'est-à-dire, la publication et la découverte ainsi que l'exécution d'un service peuvent être réalisées simultanément d'une façon modulaire.

La nature de notre approche qui représente un système géographiquement distribué et rattaché à des organisations différentes et les raisons d'adaptabilité des SMA, nous optons à modéliser notre approche en se basant sur les SMA.

4.1.3.2.2 Couplage SMA/SIG

L'utilisation simultanée du système multi-agents et du système d'informations géographiques est une nouvelle tendance. La littérature offre peu d'exemples de couplage de ces deux types de représentations de la réalité. D'une part l'utilisation d'une plateforme multi-agents permet de disposer de ce types de primitives de modélisation spécifiques qui sont aujourd'hui absentes des logiciels d'information géographique et d'autre part l'utilisation d'un SIG permet la validation des modèles en comparant données obtenues par le calcul et la simulation avec les relevés effectués sur le terrain ce qui n'est pas permet dans les systèmes multi-agents [URB06,GON04]. Ce couplage SMA/SIG représente une boucle constructive permettant l'élaboration de la connaissance de l'espace : la compréhension du système spatial permet de construire le modèle; la simulation du fonctionnement permet de préciser les processus spatiaux [BON00].

4.1.3.2.3 Couplage SMA/SW

Ces dernières années nous assistons à un rapprochement entre les SMA et les services Web. Ce rapprochement se manifeste selon différentes directions dont les principales sont les suivantes:

- L'utilisation des SMA en tant qu'entité médiatrice dans le modèle fonctionnel des services Web. La médiation intervient à plusieurs niveaux [ELF04]. Par exemple, Dans les travaux [MOS03], les auteurs proposent des SMA «proxy» qui facilitent la localisation de services Web tandis que dans [SHE03] des SMA de planification et de composition de services Web sont présentés.
- L'utilisation des services Web comme cadre architectural et technologique pour mettre en place des SMA accessibles à travers le Web. Dans ce type d'application on retrouve des applications agents qui offrent leurs capacités à travers des services Web [ELF04].

4.1.3.2.3 Couplage SMA/SWG

L'informatique géographique d'aujourd'hui nécessite une grande capacité humaine. Les anciens systèmes (systèmes SIG) sont inappropriés pour les environnements repartis et hétérogènes

à l'image d'Internet. Pour remédier ce problème, une tendance actuelle consiste à exploiter les systèmes multi-agents qui représentent une architecture intelligente qui peut fonctionner dans des situations compliquées. Plusieurs travaux entre dans le cadre de couplage SMA/SWG parmi lesquels nous citons :

1. Le travail présente par Li-Li He dans [LI-05] qui propose une approche pour mettre en œuvre des systèmes à grande échelle, compliqués, distribués et auto adaptables. Spécifiquement il exploite les systèmes multi-agent pour faciliter la construction de services géographique intelligents distribués. Cette approche avait été partiellement appliquée dans un projet réel succès.
2. LIN Hui propose dans [LIN06], une nouvelle architecture de l'application de location de véhicule complexe basée service (complex vehicule Location Based Services (LBS)), de navigation et de système Intelligent de transport basé sur l'architecture SMA. L'implémentation et la performance de la méthodologie mash-up sur la gestion urbaine des camions ont été introduites. L'évaluation et le test du système final de cette approche donnent des horizons à exploiter les SMA basée mash-up.

4.2 Modèle proposé

L'approche proposée repose sur la coopération d'un ensemble d'agents utilisant les capacités sémantiques des services web géographiques à travers les catalogues basés sur les métadonnées selon la norme ISO19119. Les agents de notre architecture sont conformes aux propriétés générales des agents (Portabilité, Autonomie, Stabilité, Persistance, Flexibilité, Sociabilité). Dans cette section nous allons présenter notre approche en indiquant l'architecture du modèle proposé suivie de l'aspect architectural et comportemental de chaque composant. Les protocoles de communication représentent la dernière partie de cette section.

4.2.1 Architecture générale

L'architecture du modèle proposé est représentée dans la figure suivante :

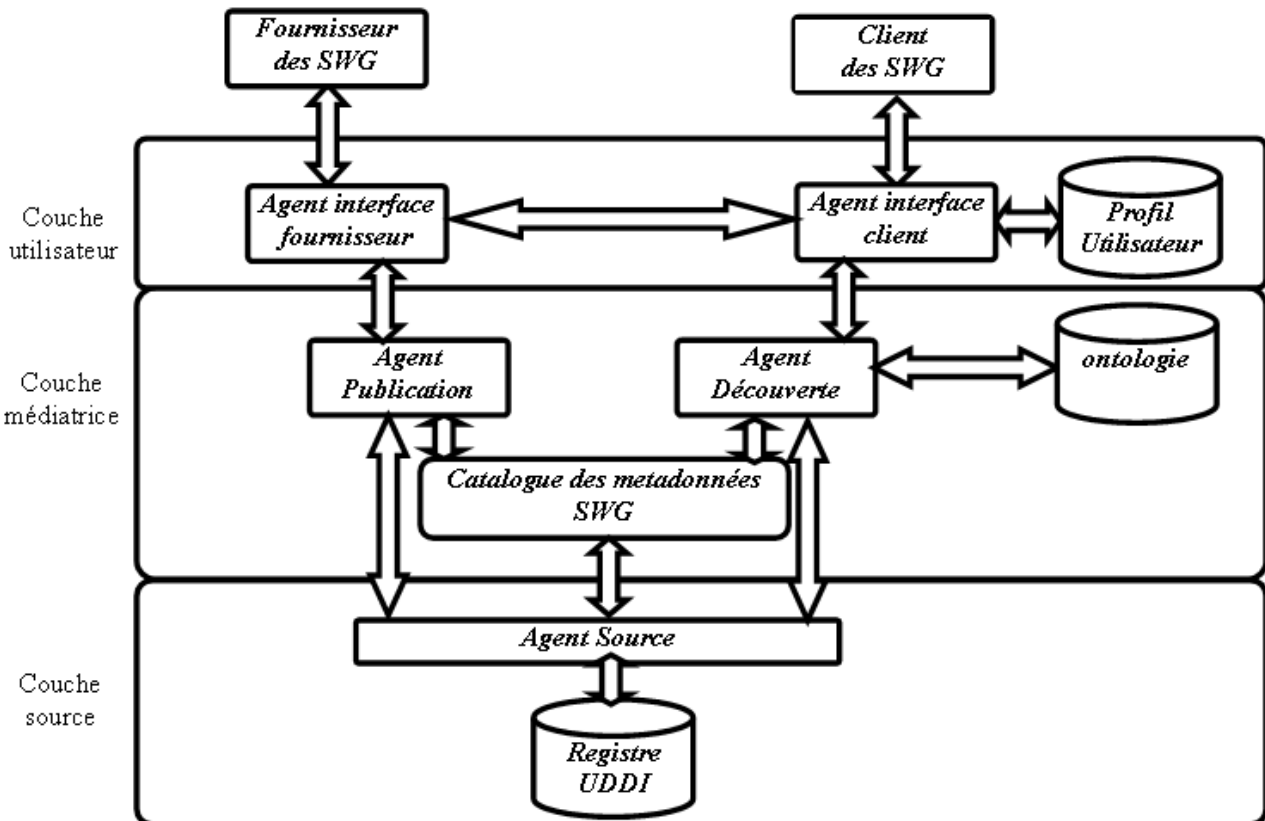


Figure 28: Modèle sémantique basé agent pour la découverte des SWG

4.2.2 Architecture et fonctionnement des composants

Les composants essentiels de notre architecture sont :

4.2.2.1 Profil utilisateur

Dans notre système nous avons utilisé les profils réflexifs basés sur les ontologies de domaine, pour orienter le client dans sa requête ce qui mène à l'optimisation de la découverte des services web géographique. Dans ce travail nous utilisons le principe proposé dans [ZAY07] qui consiste à assurer l'adaptation dynamique à l'utilisateur. Cette adaptation est effectuée à travers un algorithme dont le but est de prendre en compte les caractéristiques de l'utilisateur dans la requête avant son évaluation afin de proposer des résultats plus pertinents du point de vue de l'utilisateur. En premier lieu nous utilisons un sous ensemble des principales dimensions du profil utilisateur présentées dans [KOS07] à l'image de domaine d'intérêt, données personnelles.

Le mécanisme d'adaptation de la requête utilisateur peut être schématisé comme suit :

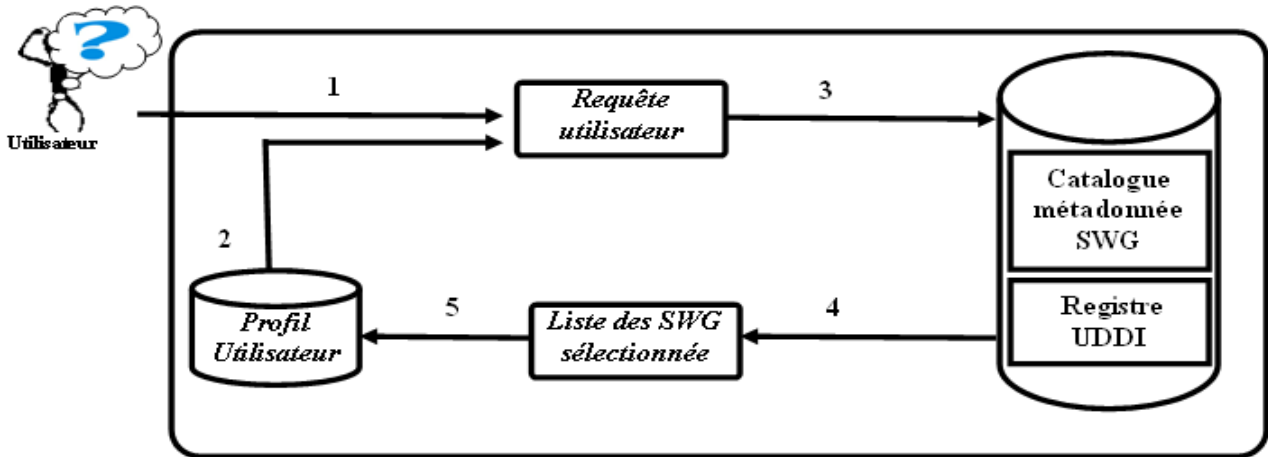


Figure 29 : Mécanisme d'adaptation de requête utilisateur

Ce mécanisme est composé des étapes suivantes :

1. Construction de requête utilisateur : Durant laquelle l'utilisateur fait introduire sa requête d'une manière libre en utilisant un langage naturel ou bien orienté en utilisant un langage fonctionnel. (Agent interface client)
2. Enrichissement de la requête utilisateur : Durant laquelle la requête sera enrichie en fonction des caractéristiques de l'utilisateur en utilisant les connaissances de son profil (agent interface client)
3. Découverte de service : la phase de découverte de SWG consiste à:
 - trouver les mises en correspondance entre les catégories des SWG et la requête: la correspondance entre les entrées du catalogue de métadonnées des catégories de SWG et le traitement de la requête est assuré à travers l'utilisation d'un algorithme de «matching»;
 - trouver les mises en correspondance entre les fonctionnalités des SWG et celle de la requête: le résultat de la phase précédente sera exploité dans la phase de découverte au niveau du registre UDDI pour restreindre l'espace de recherche à la classe concerné en appliquant l'algorithme de «matching». (Agent découverte et agent source)
4. Affichage de la liste des SWG pertinents (agent interface client).
5. La mise à jour du profil utilisateur selon le SWG choisi par l'utilisateur (agent interface client).

4.2.2.2 Base de connaissances

Dans notre approche nous utilisons une base de connaissances conservant d'une part la description des SWG à travers deux niveaux de description. Le premier niveau concerne le catalogue des classes des SWG en se basant sur la norme ISO19119, le deuxième niveau concerne le registre UDDI à travers le WSDL. Et d'autre part facilite la découverte de SWG en exploitant l'algorithme de « matching » au niveau catalogue de métadonnées et du registre UDDI. Cette base peut être schématisée comme suit :

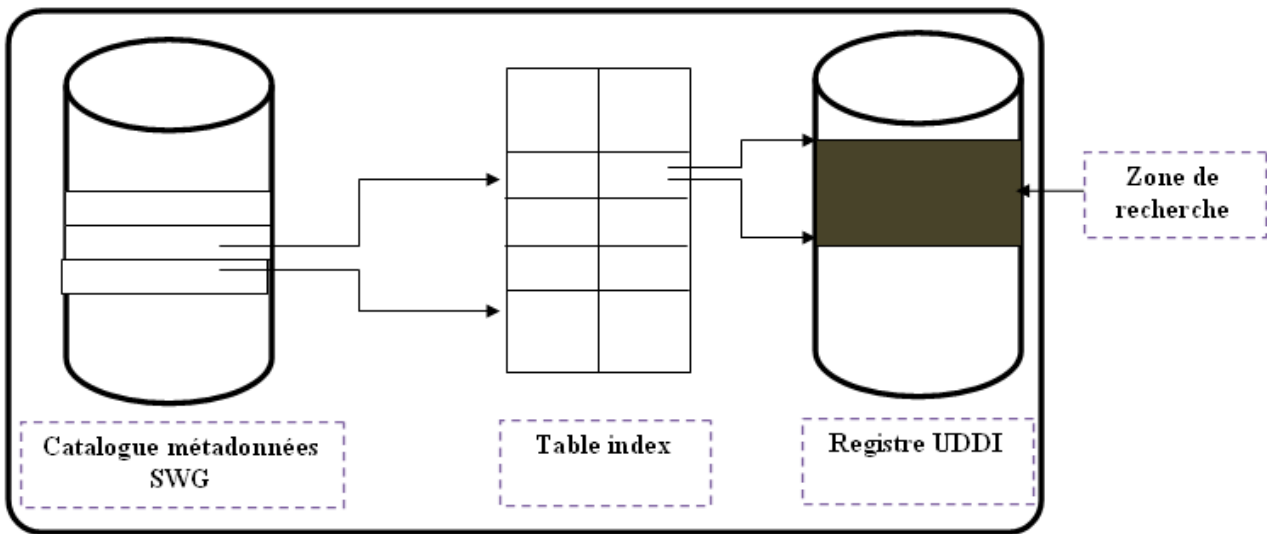


Figure 30 : Architecture de la base de connaissances

4.2.2.2.1 Catalogue des métadonnées des SWG

D.Gorga [GOR03] indique que la capacité des métadonnées à faciliter l'accès aux ressources en ligne dépend grandement de l'existence d'un standard, doté de propriétés. Dans la section 1.1.3 du chapitre 1, nous avons indiqué que les fonctionnalités d'un SIG peuvent être classées en 5 classes (les 5 A), et nous avons précisé que les services Web géographiques ont pour objectif de déplacer des grosses applications SIG des machines clientes vers des serveurs qui ne fourniraient alors que les fonctionnalités demandées à un instant donné ainsi que la normalisation proposée par l'ISO pour les services Web Géographique. A partir de là et dans l'objectif de restreindre l'espace de recherche un SWG, nous jugeons qu'il est nécessaire d'appliquer la classification des SWG au niveau des catégories des SWG en se basant sur la norme ISO19119 avec une interprétation adaptée aux classes (sous classes) des services Web géographiques.

4.2.2.2.1 Modèle de métadonnées des catégories des SWG

Notre modèle de métadonnées des catégories des SWG peut être présenté par le diagramme de classes suivant :

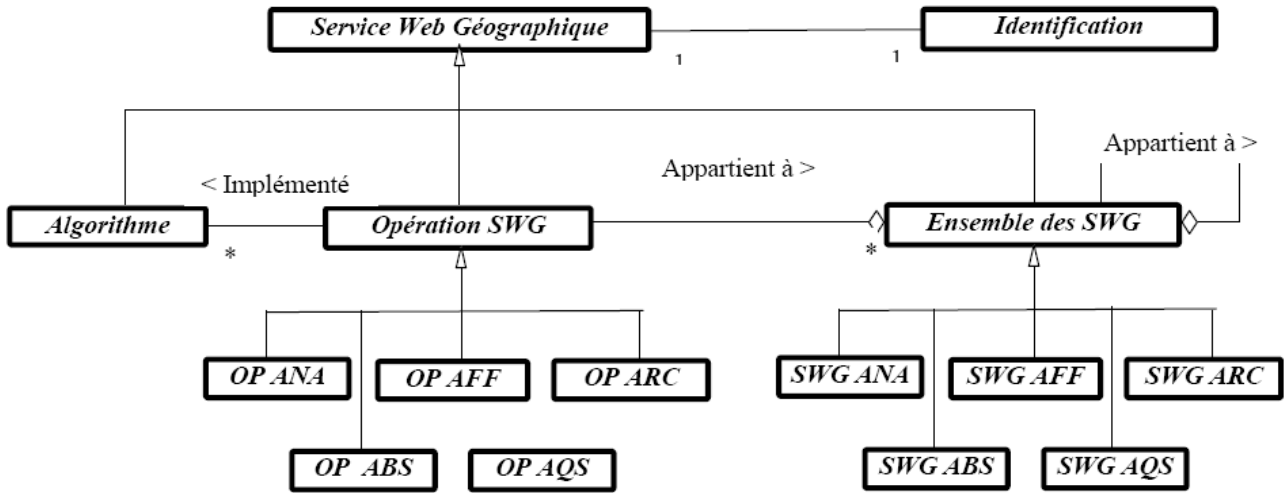


Figure 31: Diagramme de classes du service Web géographique

Le diagramme de classe proposé pour représenter les différentes classes du service Web géographiques est inspiré du diagramme de classes principales du modèle de métadonnées [YAN06], Il représente un cas particulier de ce dernier. L'interprétation de ce diagramme est la suivante :

1. **Service Web Géographique**: est une classe abstraite désignant tout traitement ou ensemble de traitements assurés par le service Web Géographique.
2. **Algorithme** : “Suite finie d’opérations élémentaires constituant un schéma de calcul ou de résolution d’un problème”, dans notre contexte, un algorithme est “une description de résolution de problème destinée à être implémentée sous forme d’opération service Web Géographique (Opération SWG).
3. **Opération SWG** : est une classe abstraite désignant les opérations de services Web Géographiques (Op SWG ANN, Op SWG AFF, Op SWG ABS, Op SWG ACQ, Op SWG ARC), Elle présente la particularité d’être invocable via le Web. Une Opération SWG est fournie par un Service Web Géographique , dont la description – typiquement en WSDL – spécifie les modalités d’invocation (protocole Web employé, port, format des messages, etc.).
4. **OP SWG AFF** : Opération SWG qui est utilisé pour assurer l’opération d’affichage (idem pour les autres opérations : Op SWG ANN, Op SWG ABS, Op SWG ACQ, Op SWG ARC).

L’identification (la découverte) des SWG est basée sur les éléments des métadonnées du catalogue. Ces éléments sont indiqués dans le tableau 08.

N°	Désignation du champ	Utilisation
01	Identifiant de la métadonnée	Numéro unique permettant de différencier chaque métadonnée par adhérent au sein du Géocatalogue
02	Titre du service	Libellé de la classe (sous classe) du service. S'affiche dans les tableaux de résultats,...
03	Langue de la métadonnée	Par défaut, la langue est « FRA ». La valeur « fre » demandée par INSPIRE (ISO 629-2/B) est également gérée par le Géocatalogue.
04	Date de création / mise à jour de la métadonnée	Date à laquelle la fiche a été créée ou actualisée.
05	Jeu de caractères des métadonnées	Jeu de caractères du fichier contenant les métadonnées
06	Résumé sur le service	Texte décrivant en quelques lignes le contenu de la classe (sous classe) du service. Cette information s'affiche dans le tableau de résultats.
07	Catégorie des données / services	Principale(s) thématique(s) de la donnée selon une typologie définie par la norme 19115
08	Niveau de sécurité des services	Information permettant d'indiquer si la classe (sous classe) de service est accessible avec ou sans contrainte
09	Date de création ou de publication de la donnée	Information afin de dater les données disponibles dans le Géocatalogue
10	Point de contact sur la métadonnée	Nom d'un organisme (ou d'une personne) assurant la fonction de contact pour la métadonnée.
11	Niveau hiérarchique du service	Fixé à une classe (sous classe) de service
12	Type de service	Décrit le type de la classe (sous classe) du service et permet de gérer les regroupements par service.
13	Paramètres des services	Contient au minimum la position du premier service d'une classe (sous classe) dans le registre UDDI.

Tableau 08: Les éléments des métadonnées représentant les classes des SWG

4.2.2.2.2 La table d'index

La table d'index a pour objectif de restreindre l'espace de recherche et en limitant le nombre des SWG concernés par l'application de l'algorithme de « matching ». Ce permet

d'optimiser le traitement de la requête et de restreindre l'interrogation du registre UDDI uniquement aux SWG appartenant à la catégorie indiquée. Initialement l'idée de base est inspirée de l'idée d'indexation des tables de base de données relationnelles, elle peut être enrichie par d'autres aspects sémantiques dans la même classe ce qui donne une base de lien sémantique.

4.2.2.2.3 Le registre UDDI

Dans notre approche nous conservons la structure de l'UDDI (Universal Description, Discovery and Integration) en assurant l'accès direct à ce registre à travers une table d'index qui assure la liaison entre le catalogue et l'UDDI.

4.2.3 Agents de l'architecture

Les agents identifiés dans notre architecture illustrée dans la figure 28 sont détaillés comme suit :

4.2.3.1 Agent interface client

Il est chargé de gérer le dialogue entre le demandeur de service et notre architecture. La structure de cet agent contient les composants suivants :

1. Le module de communication: ce module reçoit les requêtes du client et les transfère au module de traitement et reçoit les résultats.
2. Le module de traitement: il reçoit des données du module de communication et fait extraire les informations nécessaires à la découverte du service (type de classe (sous classe)) du service demandé ainsi que ses paramètres) et mis à jour le profil de l'utilisateur en fonction des requêtes présentés par le client. Il transforme les résultats reçus pour être exploité par le client.
3. Le module de communication inter-agents: il reçoit du module de traitement le type de service demandé ainsi que ses paramètres. Il les transfère également aux agents communicant avec : l'agent découverte et l'agent interface fournisseur.

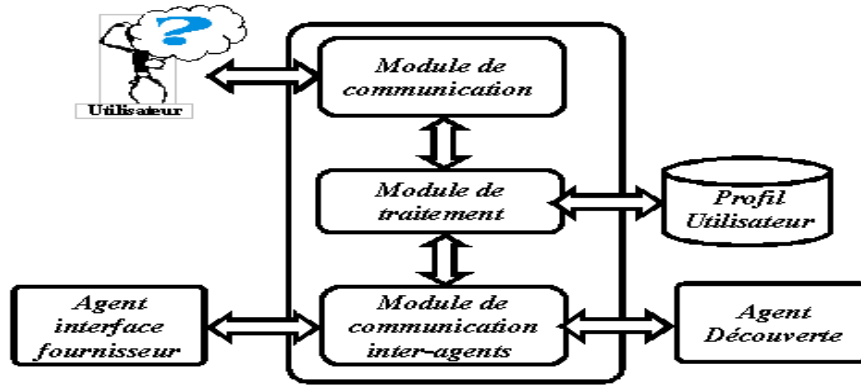


Figure 32: Architecture d'agent interface client

4.2.3.2 Agent interface fournisseur

Il assure le dialogue entre le fournisseur du service et notre architecture. La structure de cet agent contient les composants suivants :

1. Le module de communication: ce module reçoit la description des SWG (niveau catalogue et niveau registre UDDI) et les transfère au module de traitement et reçoit les résultats.
2. Le module de traitement: il reçoit des données du module de communication et fait les contrôles de cohérence de la description des SWG.
3. Le module de communication inter-agents: il reçoit du module de traitement le type de service demandé ainsi que ses paramètres. Il les transfère également aux agents communicants avec: l'agent publication et l'agent interface client.

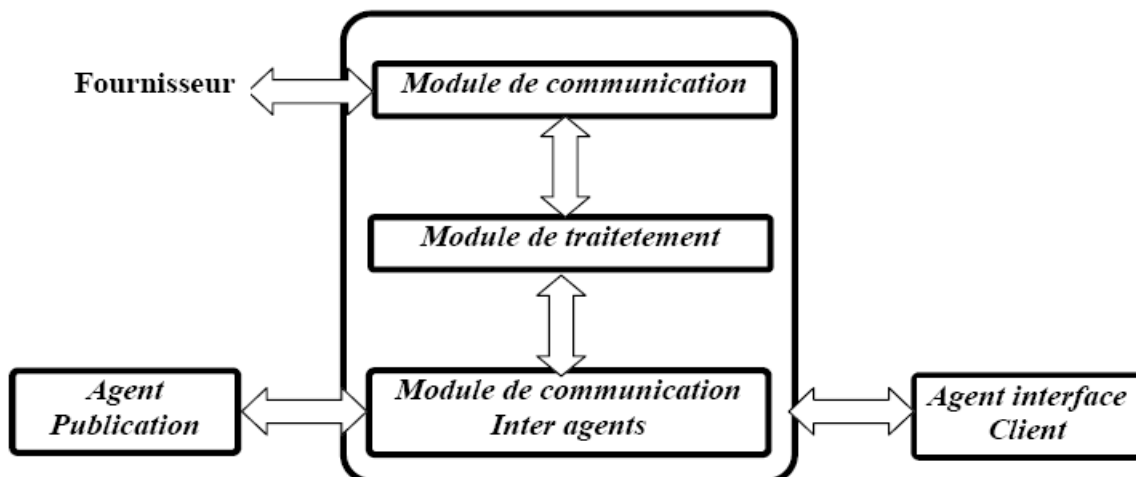


Figure 33: Architecture d'agent interface fournisseur

4.2.3.3 Agent découverte

Cet agent assure la découverte des services web géographiques, sa structure contient les composants suivants :

1. Le module de traitement : il reçoit les informations nécessaires à la découverte du service, le type de service demandé ainsi que ses paramètres. Pour trouver la position appropriée de la classe (sous classe) dans le catalogue nous faisons appel à l'algorithme de « matching » [PAO02]. C'est une application de cet algorithme au niveau métadonnées. Elle sert à trouver les mises en correspondance entre les critères de la catégorie du SWG et celles de la requête utilisateur. Cette mise en correspondance est assurée par la comparaison entre les caractéristiques de la catégorie du SWG recherché présentés dans la requête utilisateur et celles stockées dans le catalogue des métadonnées. La comparaison est basée sur l'ontologie du domaine, elle peut avoir quatre modes de résultat à savoir
 - Mode exact : le service correspond parfaitement à la requête.
 - Mode plug in : le service annoncé peut correspondre (être branché) à la requête. Par exemple, l'utilisateur désire un service triant une liste d'entiers. Le médiateur propose un service de tri de liste d'entiers et de strings.
 - Mode Subsume : retourne une offre si elle est incluse dans une demande (demande > offre).
 - Mode Fail aucune correspondance entre l'offre et la demande (échec de matching).

L'algorithme de comparaison utilisé à la fois en mode Plug-In et en mode Subsume utilise la fonction Englobe [BOU 06] donnée ci-dessous : Englobe (O, D) retourne vrai si l'offre O subsume (englobe) la demande D et faux sinon.

```
Fonction Englobe (E1 : chaîne, E2 : chaîne) : booléen
% Cette fonction retourne vrai si E1 englobe E2 faux sinon
% E1 est un élément du catalogue de métadonnées des SWG (offre)
% E2 est un élément de la requête (Demande)
% A représente l'ontologie (sous forme arborescente)
% On utilise les fonctions de haut niveau suivantes:
% Père(E) : retourne le père de E dans A
% Racine(A) : retourne la racine de A
Variables
SommetCourant : UnSommet % Sommet de A en cours d'examen
LesAncêtres : EnsembledeSommets % Les ancêtres de E2
Début
```



```

LesAncêtres ← ∅

Si E2 = racine(A) Alors
    % E2 n'a pas d'ancêtre et ne peut pas être englobé
    LesAncêtres ← ∅

Sinon
    SommetCourant ← Père(E2)
    LesAncêtres ← Père(E2)
    Tant Que (SommetCourant <> Racine(A)) Faire
        SommetCourant ← Père(SommetCourant)
        % «+» désigne l'ajout d'un nouvel élément
        % dans l'ensemble LesAncêtres
        LesAncêtres ← LesAncêtres + SommetCourant
    Fin Tant Que
Fin Si

Englobe ← (E1 ∈ lesAncêtres)
Fin
    
```

Figure 34 : Fonction Englobe [BOU06]

L'agent applique un test de subsomption sur les éléments de l'offre et de la requête, ensuite nous attribuons un score pour chaque mode de matching : Exact (score=3), PlugIn (score=2), Subsume (score=1), Fail (score=0), en appliquant la procédure degreeOfMach et la fonction GetScore.

```

Procédure degreeOfMatch(E1,E2 : chaîne )
% Cette Procédure retourne résultat de comparaison
% E1, E2 sont les éléments de la demande et de l'offre respectivement
Début
    Si E1 = E2 Alors Return Exact
    Sinon Si Englobe(E2, E1) Alors Return PlugIn
        Sinon Si Englobe(E1, E2) Alors Return Subsume
            Sinon Return Fail
        Fin Si
    Fin SI
Fin Si
Fin
    
```

Figure 35 : Procédure de degré de « matching »

```

Fonction GetScore (rel : chaîne) : Entier
% Cette Fonction retourne le score de matching
Val ←0
Début
    Si rel = "Exact" Alors val ←3
    Sion Si rel = "PlugIn" Alors val←2
        Sinon Si rel = "Subsume" Alors val←1
            Sinon Si rel = "Fail" Alors val←0
        Fin Si
    Fin SI
Fin Si
GetScore ← val
Fin
    
```

Figure 36: Fonction d'affectation de valeur de « Matching »

La valeur globale du « matching » entre l'offre O et la demande D égale à la somme des valeurs de « matching » des paires des concepts de l'offre et la demande.

2. Le module de communication inter-agents : il communique les résultats obtenus aux agents communicants avec l'agent source et l'agent interface client.

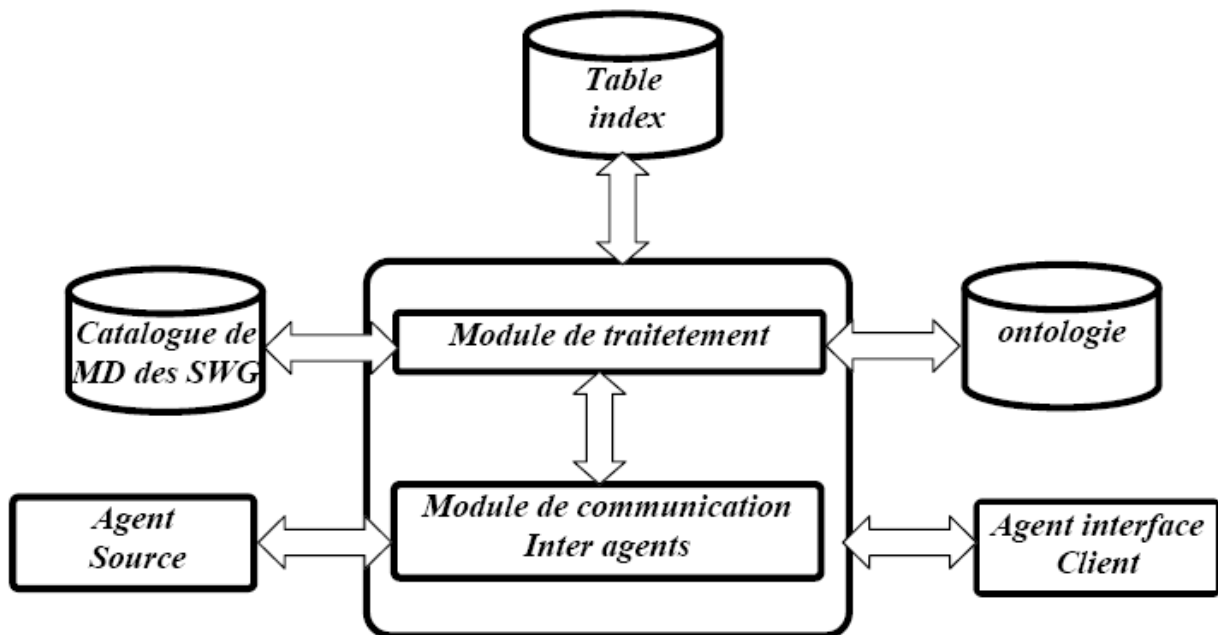


Figure 37: Architecture d'agent découverte

4.2.3.3 Agent publication

L'agent de publication assure l'insertion des métadonnées représentant les classes des services Web géographiques dans le catalogue. Il est composé essentiellement des modules suivants :

1. Le module de traitement : il reçoit la métadonnée du service à publier ainsi que les éléments nécessaires pour l'insérer dans l'UDDI) et mis à jour le catalogue des métadonnées des SWG.
2. Le module de communication inter-agents : il communique les résultats obtenus (confirmation de la publication du service et la position dans la table index) aux agents communicants avec l'agent interface fournisseur et l'agent source.

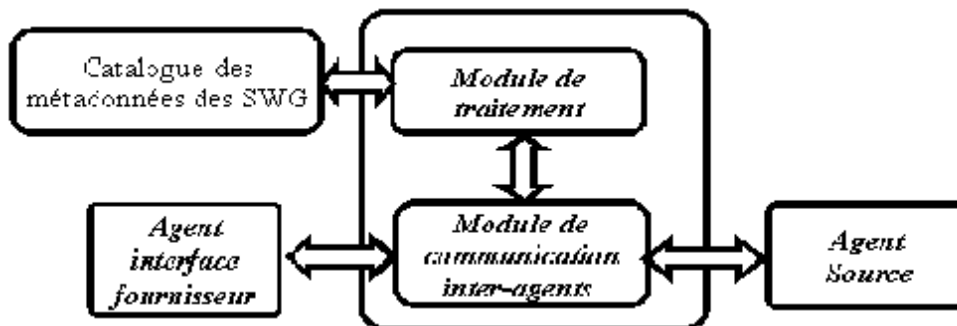


Figure 38: Architecture d'agent publication

4.2.3.4 Agent source

L'agent source assure l'insertion des services web géographiques dans l'UDDI, il est composé des modules suivants :

1. Le module de traitement : dans le cas de la découverte, Il reçoit la position dans la table d'index pour trouver l'accès convenable dans UDDI. Nous appliquons aussi l'algorithme de « matching » **basée sur le mécanisme d'interrogation du registre UDDI**. Le registre UDDI fournit une API d'interrogation de son registre. Elle permet une recherche par mot-clé, par tModel et donc, par voie de conséquence, une recherche par spécification (ex : classification, service utilisant WSDL,...). On peut comparer l'information disponible à celle fournie par un annuaire :
 - les pages blanches : liste alphabétique des différentes compagnies, d'un numéro de contact ainsi que des différents services proposés,

- les pages jaunes : classification des services et des compagnies suivant une taxonomie standardisée ou définie par l'utilisateur,
- les pages vertes : indications techniques pour faire appel au service, joindre le fournisseur.

Tous les services pertinents seront stockés dans un registre, ce dernier sera initialiser à chaque requête. Dans le cas de la publication ; ce module insère le service dans l'UDDI et met à jour la table d'index.

2. Le module de communication inter-agents : il communique les résultats obtenus (confirmation de la publication du service et les informations nécessaires à l'invocation du service) aux agents communicants avec l'agent publication et l'agent découverte.

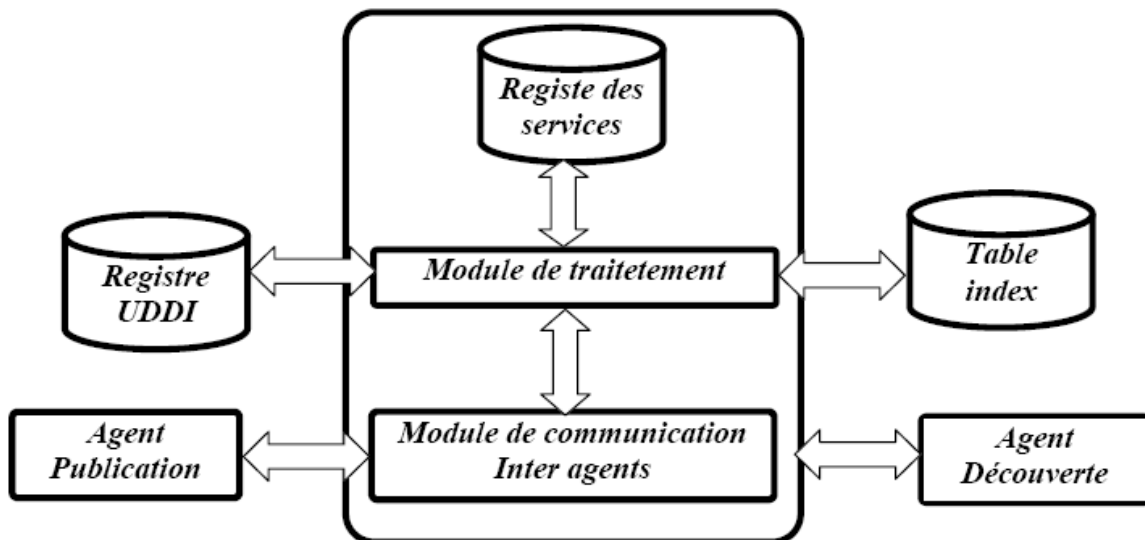


Figure 39: Architecture d'agent source

4.2.4 Protocoles de communication entre agents

Dans notre système, les agents communiquent en échangeant des messages basés sur le langage de communication FIPA ACL. Celui-ci étant fondé sur KQML excepté au niveau des noms des primitives de communication, il distingue le langage externe qui est associé à la signification prévue du message, du langage interne (de contenu) qui lui dénote l'expression à la quelle s'appliquent les croyances, les désirs et les intentions des interlocuteurs [MAZ01].

4.2.4.1 Protocole de publication de SWG

Dans l'architecture proposée, un fournisseur de service Web géographique assure la

publication de son service Web géographique au niveau de l'UDDI en utilisant le WSDL et les métadonnées au niveau du catalogue des services Web géographiques. La figure 40 représente le diagramme de séquences du protocole de publication du SWG. La phase de publication est effectuée à travers les opérations suivantes:

1. Demande de publication : cette demande est assurée par une interface graphique assurant la saisie de la description du SWG (Niveau catalogue des métadonnées et niveau registre UDDI) ;
2. Transfert de la description : l'agent interface fournisseur génère et envoie la description vers l'agent de publication;
3. Enregistrement des métadonnées dans le catalogue : L'agent de publication prend en charge de la mise à jour du catalogue des métadonnées en effectuant soit une nouvelle saisie, soit une modification.
4. Transfert de la description du SWG au registre UDDI : après le décryptage de description du catégorie SWG et la description du SWG. Cette dernière est transférée par l'agent publication à l'agent source.
5. Enregistrement de la description du SWG dans le registre UDDI : cette phase est assurée par l'agent source. Elle concerne l'enregistrement d'un nouveau service ou bien la modification d'un service existant, avec la mise à jour de la table d'index.
6. Confirmation de l'enregistrement dans le registre UDDI : Une fois l'action est exécuté un message de confirmation de l'action réalisée sera transmis à l'agent publication de l'agent source.
7. Confirmation de l'enregistrement dans le catalogue de métadonnées : L'agent publication reçoit la confirmation de l'agent source qui sera transmise avec la confirmation de l'action réalisée au niveau catalogue à l'agent interface fournisseur.
8. Confirmation de la publication : l'agent interface fournisseur envoie la confirmation de la publication au fournisseur.

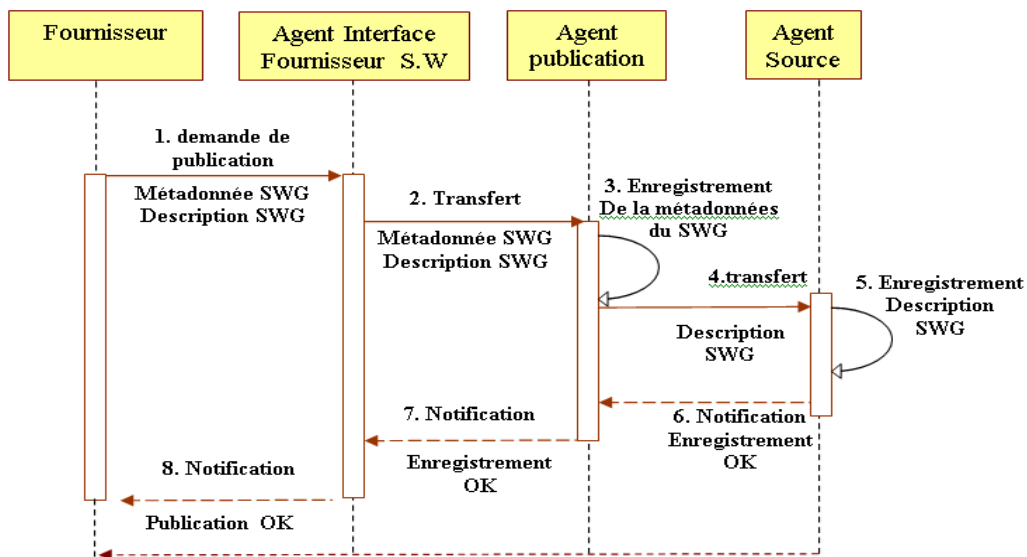


Figure 40 : Diagramme de séquences de publication de services Web géographique

4.2.4.2 Protocole de découverte de SWG

Le client du service présente sa requête suivant un profil d'utilisateur permettant d'optimiser sa demande. Le service demandé sera découvert suivant le diagramme de séquences du protocole de découverte du SWG présenté dans la figure 41. La phase de découverte est effectuée à travers les opérations suivantes :

1. Demande de service et adaptation de la requête : le client présente sa requête à l'agent interface client selon une interface graphique.
2. Adaptation de la requête et transfert des paramètres : l'agent interface client adapte la requête utilisateur selon son profil, puis il fait une reconnaissance de la requête en indiquant les paramètres concernant les catégories et ceux concernant le service et Il fait le transfert des paramètres à l'agent découverte.
3. Recherche de la métadonnée correspondante : L'agent découverte applique l'algorithme de « matching » au niveau catalogue.
4. Transfert des données pour l'accès au UDDI : l'agent découverte transmet les données nécessaires à l'accès au registre UDDI à l'agent source.
5. Recherche au niveau UDDI : l'agent source applique l'algorithme de « matching » au niveau registre et donne comme résultat une liste triée de SWG.
6. Transfert de la liste des services concernés de l'agent source à l'agent découverte.
7. Transfert de la liste des services concernés de l'agent découverte à l'agent interface client

8. Présentation des résultats au client.

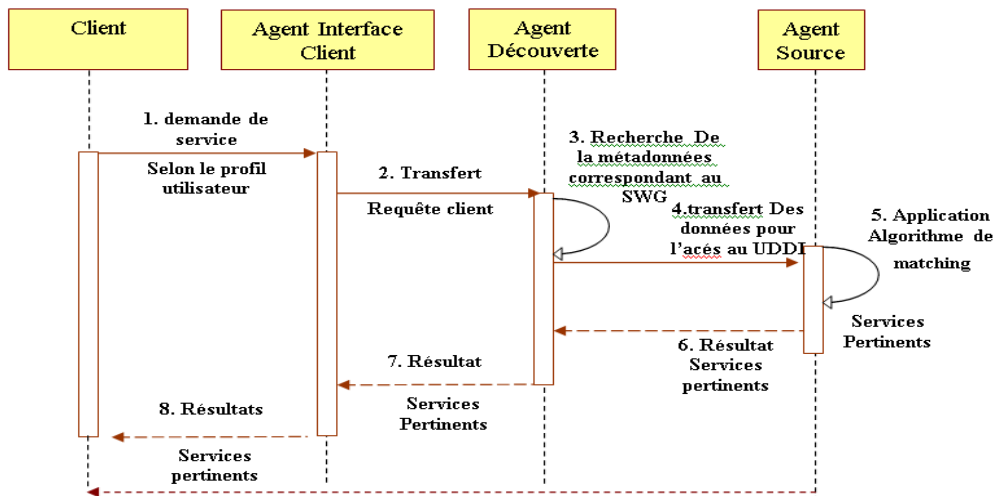


Figure 41 : Diagramme de séquences de découverte de services Web géographique

4.3 Conclusion

L'approche proposée dans ce chapitre représente une approche de découverte sémantique des services Web géographique en se basant sur les métadonnées à travers un système multi-agents. Ce travail concerne trois disciplines en pleine évolution à savoir : les systèmes d'information géographique, les services web géographique et le web sémantique en profitant des atouts des systèmes multi agents. Les métadonnées exploitées sont basées sur la norme ISO19119 avec une interprétation adaptée aux classes (sous classes) des SWG, initialement nous avons 5 classes (les 05 A). Le profil utilisateur joue un grand rôle dans la découverte sémantique des services à travers l'adaptation de la requête utilisateur. Cette dernière est basée essentiellement sur :

1. La description des SWG à deux niveaux : le premier concerne la description des classes (sous classes) des SWG suivant la classification des fonctionnalités des SIG (les 05 A). Cette description est basée sur la norme ISO 19119 à travers l'utilisation du catalogue de métadonnées. Le deuxième concerne le service Web lui même en exploitant le WSDL.
2. La découverte des SWG : dans la phase de découverte, cette solution permet de restreindre l'espace de recherche et d'augmenter le nombre de services pertinents à travers l'application de l'algorithme de « matching ». Ce dernier assure les correspondances sémantiques entre la description du SWG et les éléments de la requête utilisateur au niveau catalogue pour les éléments de description des classes (sous classes) et au niveau UDDI pour l'aspect fonctionnel en utilisant l'ontologie du domaine.

3. L'utilisation du paradigme des SMA : ce paradigme facilite la modélisation de notre modèle qui représente un système complexe nécessitant un haut degré d'interopérabilité.

Dans l'objectif d'optimiser notre approche nous proposons l'enrichissement des métadonnées par les critères de qualité. Cet enrichissement fait l'objet du prochain chapitre

*Chapitre 5: Optimisation de la découverte
sémantique des services Web
géographiques à base de qualité*

Chapitre 5 : Optimisation de la découverte sémantique des Services Web

Géographiques (SWG) à base de qualité

Comme nous avons mentionné précédemment, les systèmes d'information géographiques actuels donnent accès à un grand nombre de sources hétérogènes et distribuées. La multiplication des ressources et l'accroissement des données nécessitent l'utilisation des services Web qui apparaissent comme solution pour assurer l'interopérabilité entre les différentes ressources. Ils sont capables de collaborer et de partager le contrôle des processus de données et d'informations entre les applications sur différentes plateformes. L'utilisation des services Web dans le domaine des données géographiques a donné naissance aux SWG. Au fur et à mesure que ces derniers se multiplient, la difficulté de la découverte de service s'accroît. La maîtrise de cette phase nécessite l'enrichissement de la description des SWG et le profil utilisateur par des aspects sémantiques. Ces derniers sont assurés à travers la qualité et la personnalisation.

Dans ce cadre nous présentons une approche qui optimise l'approche précédente en faisant appel à la qualité et la personnalisation. La qualité joue un rôle primordial pour satisfaire les besoins explicites ou implicites de l'utilisateur [TRI07, GUT07, BET05] et la personnalisation facilite l'expression du besoin utilisateur et rend l'information sélectionnée intelligible à l'utilisateur et exploitable [KOS07, BOU04]. Dans l'objectif de profiter pleinement de la qualité et la personnalisation et des normes assurant l'interopérabilité telle que l'ISO 19119 qui propose une taxonomie des SWG [YUQ08], nous proposons notre approche qui consiste d'une part à renforcer les métadonnées exploités dans [REZ09] par les critères de qualité aux différents niveaux (données, service et usage), et d'autre part à exploiter le modèle du profil proposé par [KOS07].

Le reste du chapitre est organisé de la façon suivante : la section 2 représente un survol sur le langage de modélisation UML qui sera exploité dans la présentation des différents modèles de notre approche. la section 3 présente notre approche de découverte sémantique des SWG. Cette partie de chapitre consiste à détailler les différents modèles de représentations des SWG, le profil utilisateur ainsi que la requête utilisateur. Une représentation formelle de la qualité ainsi que l'algorithme de « matching » utilisé « LARKS » et les différentes exploitées dedans font l'objet de la 4 eme section. Une conclusion du chapitre sera présentée dans la 5 eme section.

5.1 Le langage de modélisation UML (Uniform Modeling Language)

UML [BOO98] est une notation et un langage standardisé en 1997 par l'OMG (*Object Modeling Group*) (OMG UML) qui facilite la conception de programmes, ainsi que leur description pour des non-informaticiens. UML a la particularité de s'intéresser essentiellement à la modélisation c'est-à-dire la représentation des différents concepts qui interviendront dans l'écriture d'un logiciel [CRA03]. Il est devenu le langage de modélisation orienté objet de référence dans le

monde professionnel car il comble une lacune importante des technologies objet : il permet d'exprimer et d'élaborer des modèles objet, indépendamment de tout langage de programmation [PIE06]. Il a été pensé pour servir de support à une analyse basée sur les concepts objet. C'est également un langage graphique semi-formel orienté objet issu des meilleurs outils et pratiques du génie logiciel du début des années 90.

La version (UML 2.1.) définit treize types de diagrammes, répartis en trois catégories [LUD09]:

- Les **diagrammes de structure** incluent le diagramme de classes, le diagramme d'objets, le diagramme de composants, le diagramme de structure composite, le diagramme de paquetages et le diagramme de déploiement.
- Les **diagrammes de comportement** incluent le diagramme de cas d'utilisation, le diagramme d'activités, et le diagramme d'états-transitions.
- Les **diagrammes d'interaction** (dérivés de la catégorie des diagrammes de comportement) incluent le diagramme de séquences, le diagramme de communication, le diagramme de synchronisation (ou de temps), et le diagramme de vue d'ensemble des interactions.

Parmi les diagrammes proposés par UML, nous exploitons particulièrement le diagramme de classes qui est le plus employé et le mieux connu des diagrammes orientés objet; il est considéré comme le plus important des diagrammes du fait qu'il est le cœur de la modélisation objet.

5.1.1 Diagrammes de classes

Un diagramme de classes est défini par Piechocki comme une collection d'éléments de modélisation statique qui fait abstraction des aspects dynamiques et temporels [PIE06].

Les différents éléments représentables dans un diagramme de classes UML sont décrits par le méta-modèle d'UML défini par l'OMG dans le document de spécification de la superstructure d'UML. La figure 42 résume sous forme hiérarchique (un diagramme de classes UML traduit en français). Le langage UML 2.x organise les éléments des diagrammes de classes en trois grandes catégories [LUD09]:

- Les **classificateurs** (*classifiers*) ont des caractéristiques communes (des propriétés et/ou des opérations) et sont qualifiés par un espace de nommage unique (un nom). Les classificateurs sont généralisables et peuvent être liés entre eux par des relations.
- Les **relations** (*relationships*) sont les connexions logiques entre les classificateurs apparaissant dans un diagramme. Elles peuvent être dirigées (héritage, dépendance) ou non (association). Les relations ont une structure et un sens bien définis en fonction de leur type.
- Les **caractéristiques** (*features*) sont encapsulées dans les classificateurs. Elles définissent leurs propriétés (attributs) et leurs comportements (les méthodes) communs.

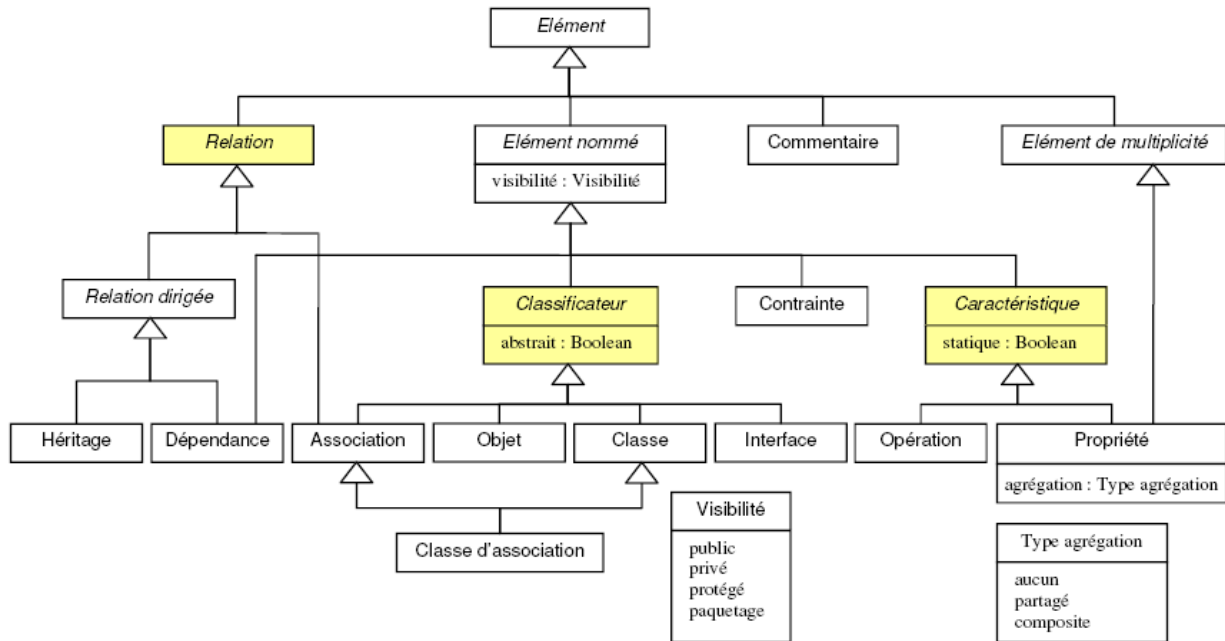


Figure 42 : Extrait simplifié du méta modèle pour les diagrammes de classes UML 2.X [LUD09]

5.1.1.1 Les classificateurs

Les classificateurs les plus employés dans les diagrammes de classes sont les **classes**. Une classe est représentée graphiquement sous forme d'un rectangle découpé en trois parties : le nom de la classe, les attributs et les méthodes encapsulées dans cette dernière. Les noms des classes portent par convention une majuscule en tête. Dans l'exemple suivant, la classe représentée est nommée « *Personne* » et possède trois attributs (« *nom* », « *date de naissance* » et « *age* ») et une méthode qui permet d'accéder à la valeur de l'attribut « *age* ». Pour ne pas surcharger les diagrammes de classes, les objets sont habituellement représentés à part dans des diagrammes d'objets permettant de mettre en valeur les objets instanciés et leurs liens effectifs.

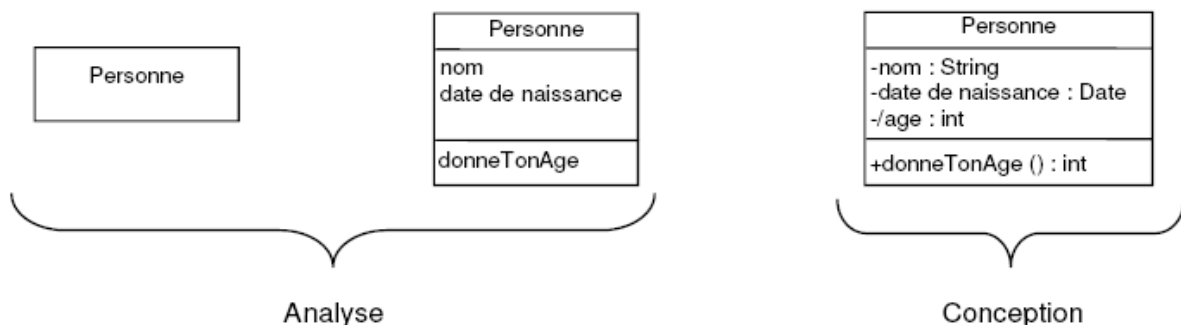


Figure 43 : Représentations graphiques plus ou moins détaillées d'une classe UML [LUD09]

Un type particulier de classe est la **classe abstraite**. Elle a une représentation similaire à celle d'une classe (concrète) mais ne peut être instanciée directement sous forme d'objets. Pour différencier graphiquement une classe abstraite d'une classe (qui ne l'est pas), son nom est écrit en italique ou précédé d'un stéréotype particulier. Les classes abstraites permettent de généraliser des

propriétés et un comportement commun à plusieurs autres classes plus spécialisées (reliées par des relations d'héritage). Les classes abstraites sont utilisées dès la phase d'analyse du problème et prennent tout leur sens au niveau de la phase de conception [LUD09].

5.1.1.2 Les relations

Les relations liant les classificateurs peuvent être de natures diverses. Une **association** est une relation sémantique entre des classes qui définissent un ensemble de liens [PIE06]. L'association est représentée sous forme d'une ligne entre les classes participantes. Le but d'une association peut être exprimé par un verbe ou un syntagme verbal décrivant comment les objets d'une classe sont en relation avec ceux d'une autre classe. Les associations entre plus de deux classes (**associations n-aires**) sont peu utilisées et sont rarement recommandées du fait de leur difficulté de déchiffrement et des erreurs qu'elles peuvent induire. Chaque extrémité d'une association (fin d'association) est une propriété dans le métamodèle d'UML (tout comme les attributs dans les classes). Une fin d'association définit le rôle de la classe liée (optionnel) ainsi qu'une multiplicité. Les multiplicités précisent combien d'objets des différentes classes interviennent dans la relation. Une multiplicité doit être affectée à chacune des classes participant à une association et doit suivre la syntaxe définie dans le tableau 09.

Symbole	Signification
1	une seule instance.
0..1	de zéro à 1 instance.
M..N	de M à N instances. $M < N$, M et N sont des entiers non négatifs.
*	de 0 à un nombre quelconque d'instances.
0..*	identique à *
1..*	de 1 à un nombre quelconque d'instances
M..*	de M à un nombre quelconque d'instances. M est un entier non négatif.

Tableau 09 : Multiplicités syntaxiquement correctes pour les propriétés en UML [LUD09]

Les **agrégations** et les **compositions** sont des associations binaires particulières et plus restrictives qui apportent le sens « est une partie de » à la relation. L'agrégation indique que les objets sont plus que des objets indépendants ayant des connaissances les uns des autres : ces objets (les composants) peuvent être **assemblés** ou **configurés** ensemble pour créer de nouveaux objets plus complexes (les agrégats). Un losange est reporté à l'extrémité de la relation du côté de l'agrégat pour différencier graphiquement une agrégation d'une simple association. La composition est une agrégation où la durée de vie des participants dépend de celle de la relation les reliant. Les objets (les composants) entrant dans la composition d'autres objets (les composites) ne peuvent exister seuls en dehors. Cette forme plus forte d'agrégation est indiquée graphiquement à l'aide d'un losange noir [LUD09].

5.2 Le modèle proposé

Dans cette section nous présentons l'aspect architectural et fonctionnel de notre approche. Nous spécifions le modèle de description des SWG, le modèle de description du profil utilisateur et celui de la requête utilisateur proposée.

5.2.1 Aspect architectural et fonctionnel

L'architecture générale du modèle proposé est la suivante :

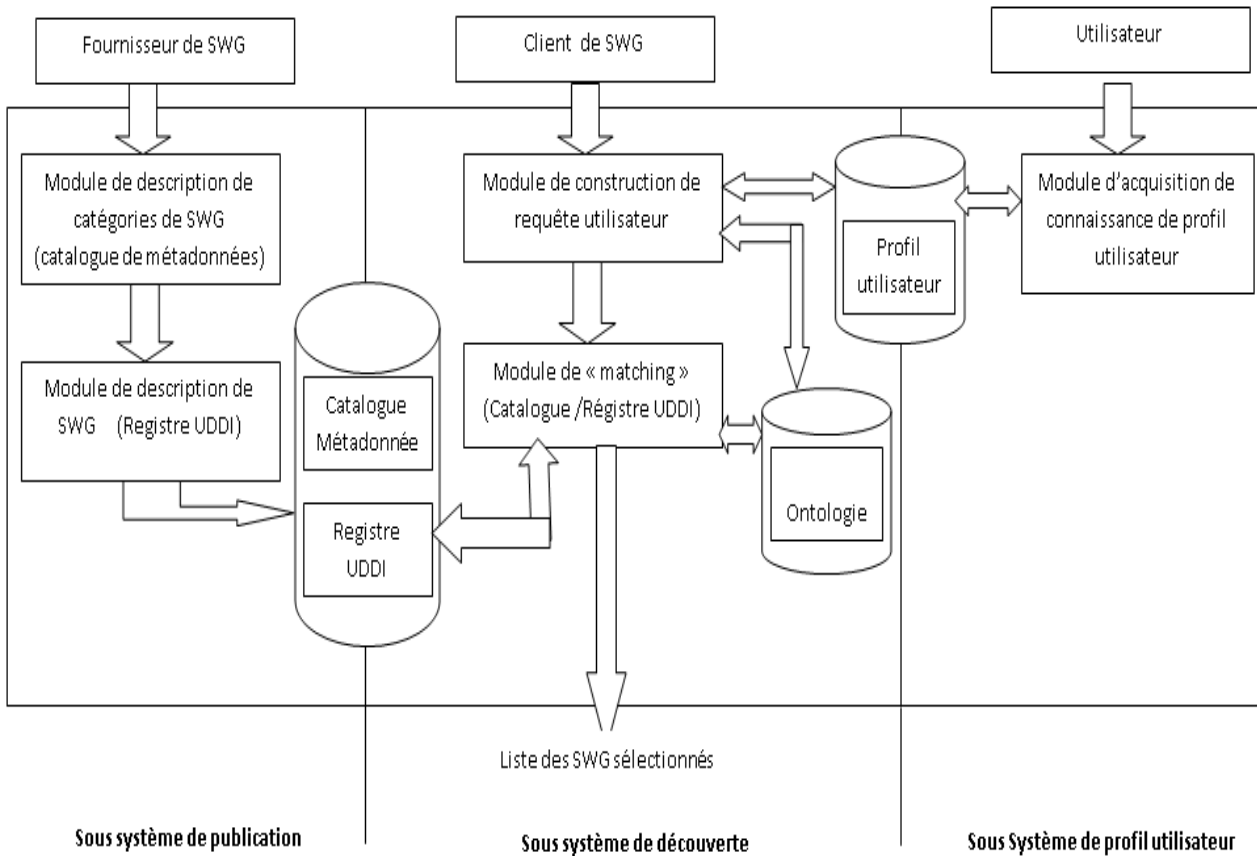


Figure 44 : Architecture générale du modèle proposé

Notre système contient les sous systèmes suivants :

5.2.1.1 Sous système de publication des SWG

La publication d'un SWG est assurée par les modules suivants:

- **Module de description de catégories des SWG** : ce module consiste à assurer l'annotation des SWG en spécifiant leur catégorie ainsi que leur qualification en matière de critères de qualité et leur enregistrement dans le catalogue de métadonnées.
- **Module de description de SWG** : ce module sert à compléter la description du SWG en utilisant le WSDL et leur enregistrement dans le registre UDDI. Une mise de la liaison entre le catalogue et le registre est effectuée par ce module.

5.2.1.2 Sous système de découverte des SWG

Ce sous système représente le noyau de notre travail. Il est composé des modules suivants :

- Module de construction de requête : dans la littérature de la personnalisation de la requête utilisateur, nous trouvons plusieurs approches [BOU06a] permettant d'exprimer les préférences comme l'opérateur Winnow (Best)[CHO02], l'opérateur Skyline [BOR01], la clause Prefer [LAC87] et la préférence SQL. Dans les approches citées précédemment, l'utilisateur est contraint d'écrire à chaque fois la requête complète qui définit son besoin d'information ce qui est un inconvénient non négligeable. L'objectif de ce module est de permettre la construction de la requête utilisateur et son enrichissement en se référant à son profil et aux deux ontologies (i) ontologie de qualité des SWG et (ii) ontologie du domaine. Une fois cette requête établie. Cette phase peut être semi automatique c'est-à-dire une requête initiale adaptée qui sera validée par l'utilisateur.
- Module de « matching » catalogue / registre UDDI : ce module sert à :
 - trouver les mises en correspondance entre les critères de qualité des SWG et celles de la requête: la correspondance entre les entrées du catalogue de métadonnées des catégories de SWG. Le traitement de la requête est assuré à travers l'utilisation d'un algorithme de «matching» en faisant appel à l'ontologie de qualité et l'ontologie du domaine ;
 - trouver les mises en correspondance entre les fonctionnalités des SWG et celle de la requête: le résultat de la phase précédente sera exploité dans la phase de découverte au niveau du registre UDDI pour restreindre l'espace de recherche à la classe concerné en appliquant l'algorithme classique de «matching» basé sur les pages blanches, jaunes et vertes du registre UDDI ;
 - présenter le résultat : ce résultat est une liste ordonnée de SWG selon leur pertinence. Les services sont tout d'abord classés selon leur catégorie (sous catégorie). Ensuite, dans les sous-ensembles (représentant les SWG d'une catégorie particulière) les SWG sont classés selon leur pertinence en termes de leurs éléments standards disponibles dans les différentes pages du registre UDDI. La pertinence est basée sur les préférences de l'utilisateur concernant les critères de qualités désirées et celles présentées dans la description des SWG. Le classement est basé sur la distance sémantique entre les concepts cités dans la catalogue de métadonnées des SWG et ceux de la requête utilisateur en se basant sur l'ontologie de la qualité des SWG et entre les concepts cité dans l'UDDI et ceux cités dans la requête utilisateur en se basant sur l'ontologie du domaine.

5.2.1.3 Sous système de profil utilisateur

Ce sous système consiste d'une part à acquérir les connaissances du profil utilisateur en se basant sur le modèle générique du profil utilisateur proposé par M. Bouzeghoub et D. Kostadinov dans leur travaux [BOU04, KOS07], en faisant une instanciation du modèle dans le cas de l'utilisateur des SWG dans un domaine bien spécifique. Selon notre approche, le profil utilisateur est décrit par 03 dimensions. La première représente l'identification de l'utilisateur à travers ses données personnelles. La seconde correspond au domaine d'intérêt en se basant sur une ontologie de domaine. La troisième concerne les critères de qualités des services préférés par l'utilisateur, indiqués par un ensemble de critères de qualité et leurs valeurs de préférence. Cette préférence est représentée par un poids dont la valeur entre 0 et 1 et un seuil. Il est composé d'un seul module: le module d'acquisition des connaissances de profil utilisateur.

5.2.1.4 Tableau de correspondance fonctionnelle entre les deux approches proposées

Notre approche de découverte de services Web géographiques à base de qualité représente une optimisation l'approche présentée dans le chapitre 04, sur le plan de découverte à travers l'enrichissement de la description des SWG par les critères de qualité ; mais elle est similaire à cette dernière point de vue fonctionnelle. Cette similarité est indiquée dans le tableau suivant :

N°	Sous système	Tâche	Module	Agent
1	Publication	Description des catégories des SWG	Module de description de catégories des SWG :	Agent Interface fournisseur
2	Publication	Enregistrement de la description de la catégorie SWG	Module de description de catégories des SWG :	Agent Publication
3	Publication	Description des SWG	Module de description de SWG	Agent Interface Fournisseur
4	Publication	Enregistrement de la description des SWG	Module de description de SWG	Agent Source
5	Découverte	Présentation et enrichissement de la requête utilisateur	Module de construction de requête	Agent Interface Client
6	Découverte	« Matching » au niveau du catalogue de métadonnées	Module de « matching » catalogue registre UDDI	Agent Découverte
7	Découverte	« Matching » au niveau du registre UDDI	Module de « matching » catalogue registre UDDI	Agent Source
8	Découverte	Présentation des résultats	Module de « matching » catalogue registre UDDI	Agent interface client
9	Profil utilisateur	Acquisition des connaissances du profil utilisateur	Module d'acquisition de connaissances de profil utilisateur	(prise comme une connaissance prête)

Tableau 10: tableau de correspondance fonctionnelle entre les deux approches proposées

5.2.2 Description du modèle proposé

L'objectif de notre modèle consiste à faciliter la découverte sémantique des SWG en se basant entre autre sur l'enrichissement de la description des SWG permettant d'une part aux fournisseurs de présenter leurs SWG d'une manière claire et visible, et d'autre part aux clients d'exploiter le service le plus convenable à travers des requêtes utilisateurs enrichies par les connaissances du profil utilisateur.

Il existe de nombreux registres sur le Web, ou des travaux de recherche, qui proposent un moyen de publier, de rechercher et de sélectionner les services Web. Cependant, les méthodes de recherche (par exemple par mots-clés) et la représentation des services ne suffisent pas à répondre à l'attente précise du client lors de la conception d'applications [LOP08]. Dans la littérature nous

constatons que la description des services Web (parmi eux les SWG) est effectuée en décrivant le service lui-même. En se basant sur la nature fonctionnelle des SWG tirés des SIG qui indique que les fonctions des SIG se résument dans les 05 A, nous proposons une description des SWG à deux niveaux:

1. la description de catégories des SWG basée sur la classification des SWG selon la norme ISO 19119 et les critères de qualité ;
2. la description des SWG en utilisant le WSDL.

5.2.2.1 Éléments nécessaires à notre modèle

Pour atteindre l'objectif cité précédemment, un ensemble de catégories d'informations est indispensable pour assurer une description riche des SWG et une découverte de service appropriée. Cet ensemble est composé du catalogue des catégories de SWG, de service, du domaine d'application et du profil utilisateur.

5.2.2.1.1 Le catalogue des métadonnées

Le catalogue de métadonnées consiste à conserver la description catégories des SWG. Cette description contient une partie basée sur la norme ISO 19119 pour permettre l'identification et la mise à jour des catégories des SWG, et une autre partie basée sur les critères de qualité qui permettent d'avoir des SWG pertinents comme résultat de la phase de découverte.

5.2.2.1.2 Le service

Pour sélectionner un service, il est important de connaître ses aspects fonctionnels. Cette catégorie d'information rassemble les méthodes proposées, les paramètres d'entrée et de sortie, le protocole à utiliser pour appeler le service, et la localisation du service (son URL). Ces connaissances sont assurées à travers l'utilisation du WSDL qui permet de décrire ce type d'information fonctionnelle.

5.2.2.1.3 Le domaine d'application

Une application couvre un ensemble de fonctionnalités qui peuvent être exécutées par des services Web qui ciblent des données et des opérations spécifiques à un domaine d'application, tel que la géomatique, le business-to-business ou le e-learning. Le fait d'associer la représentation du SWG (catégorie et service) à un domaine d'application dans lequel il peut intervenir et les fonctionnalités qu'il met en œuvre permet de faciliter la recherche et la sélection de services.

5.2.2.1.4 Le profil utilisateur

Cette catégorie d'information représente la source des connaissances qui contient des informations sur tous les aspects de l'utilisateur qui peuvent être utiles à l'efficacité des résultats de la découverte des SWG. Ces connaissances sont utilisées à l'enrichissement de la requête utilisateur

pour lui permettre d'avoir le service le plus approprié.

5.2.2.2 Vue d'ensemble du modèle proposé

Notre modèle **QGWSR-Model** (Quality Geospatial Web Service Representation Model) est composé de 4 packages UML comme indiqué dans la figure suivante :

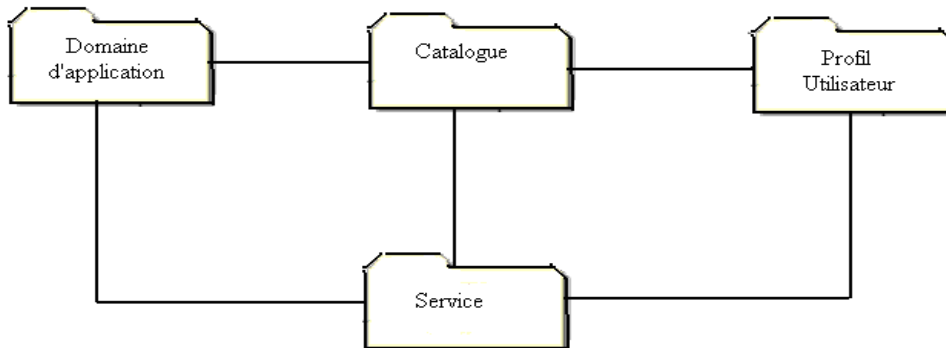


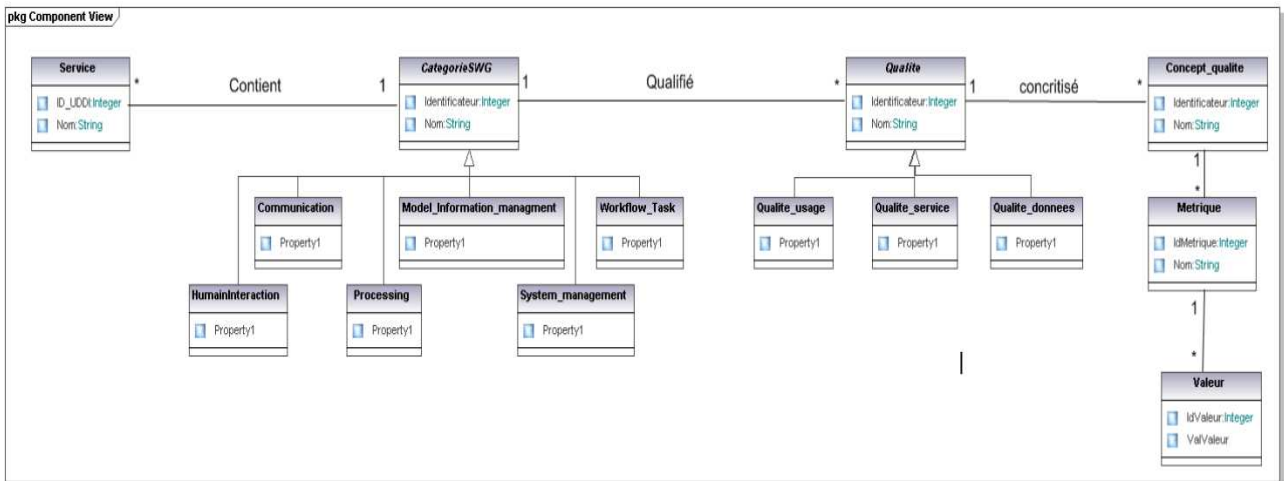
Figure 45 : Vue d'ensemble du QGWSR-Model

- **Catalogue:** ce package représente le package central du modèle, il représente la description des catégories des SWG en se basant sur les critères de qualité et la norme ISO 19119. Il permet aux fournisseurs de services Web de décrire leurs services Web en termes de catégorie de services, aux clients de découvrir les services les plus appropriés.
- **Service:** ce package représente les informations élémentaires d'un service issu de la technologie des services Web. Cette description est effectuée en utilisant le WSDL.
- **Domaine d'Application:** Le package du domaine d'application permet de décrire le domaine associé aux services représentés.
- **Profil utilisateur:** ce package intègre la description du profil utilisateur dans notre modèle afin de construire la requête utilisateur selon son profil.

Les sous-sections qui suivent décrivent chacun des packages composant le modèle de représentation de SWG (Catalogue, Service, Domaine d'application, Profil utilisateur). Les 03 premiers packages consistent à enrichir la description des SWG et le quatrième consiste à décrire le profil utilisateur et enrichir la requête utilisateur. Une dernière partie est consacrée aux liens existants entre les packages.

5.2.2.2.1 Le package catalogue

Pour la description du modèle du package catalogue, nous utilisons le diagramme de classes UML présenté dans la figure 46. Ce diagramme est composé de deux parties P1 et P2. La partie P1 est inspirée de la classification des SWG selon la norme ISO 19119 présenté dans [YUQ08] et la partie P2 représente notre proposition de qualifier les catégories de SWG par des critères de qualité.



Generated by UModel www.altova.com

Figure 46 : Diagramme de classes de modèle de description de catégorie des SWG.

Ce package est composé des classes suivantes :

- **Classe catégorie de SWG** : classe abstraite désignant les catégories des services Web Géographiques, elle peut être décrite selon deux facettes à savoir l'identification et la fonction.
- **Classe Human interaction**: cette classe assure la description des services destinés au contrôle et à la gestion d'interfaces utilisateur, de graphiques, de données multimédia, etc (Catalogue viewer, Geographic viewer, Geographic feature editor).
- **Classe Model/Information management** : cette classe décrit les services de gestion du développement, de traitement et de mémorisation de métadonnées, de concepts et de jeux de données (Feature access service, Map access service, Coverage access service, Catalogue service).
- **Classe Workflow/Task** : cette classe représente la description des services d'assistance aux utilisateurs dans le cadre de tâches et d'activités spécifiques (Chain definition service, Workflow enactment service).
- **Classe Processing** : cette classe décrit les services de traitement Web peuvent être configurés pour offrir tout type de fonctionnalité SIG à des clients reliés à un réseau, ce qui comprend l'accès à des calculs et/ou à des modèles de calcul préprogrammés qui s'appliquent à des données à référence spatiale.
- **Classe Communication** : cette classe représente la description des services de codage / décodage et de transfert de données au sein de réseaux de télécommunication.
- **Classe System management** : la description des services de contrôle des composants de systèmes d'applications et de réseaux (exemple : gestion de comptes d'utilisateurs) est assurée par cette classe.

- **Classe Service** : La classe service représente le SWG suivant la description WSDL. Elle comprend entre autres l'identificateur du registre UDDI auquel le service appartient.
- **Classe Qualité** : Classe abstraite désignant les catégories de qualité. Parmi ces importants attributs nous citons le poids de qualité qui représente l'importance de la qualité pour la catégorie de services.
- **Classe Qualité Données** : cette classe représente les critères de qualité des données géographiques (la norme ISO 19113).
- **Classe Qualité Service** : les critères de qualité des SWG lui-même sont représentés par cette classe.
- **Classe Qualité Usage** : cette classe représente les critères de qualité de profil utilisateur.
- **Classe Concept Qualité** cette classe représente un concept de qualité qui spécifié par un ensemble de métriques et un ensemble de valeurs.
- **Classe Métrique** : cette classe représente les indicateurs d'évaluations de qualité. Ces indicateurs représentent les méthodes de calcul des valeurs correspondantes aux critères.
- **Classe Valeur** : cette classe représente l'ensemble des valeurs des résultats des méthodes correspondantes aux métriques.

Dans notre travail nous proposons trois critères; un par classe de qualité étudiée à savoir :

- **Fraîcheur comme qualité de données**: Intuitivement, le concept de fraîcheur introduit l'idée d'âge des données: Les données sont-elles suffisamment fraîches pour les utilisateurs? Les données sont-elles obsolètes ? Une certaine source de données a-t-elle les données les plus récentes ? Quand les données ont-elles été produites ? La fraîcheur représente une famille de facteurs de qualité, chacun représentant un certain aspect de fraîcheur et ayant ses propres métriques. Nous distinguons deux facteurs de fraîcheur :
 - **Actualité** ([SEG90] cité dans [BOU07]): L'actualité mesure la distance ou le décalage entre l'extraction et la livraison de données. Par exemple, en regardant un solde bancaire, nous voulons savoir le moment où il a été extrait de la banque, peu importe quand il a été mis à jour.
 - **Age** [WAN96] : L'âge mesure la distance ou le décalage entre la création (ou mise à jour) et la livraison des données. Il est indépendant du moment de l'extraction des données. le concept de fraîcheur introduit l'idée d'âge des données, il peut avoir comme métriques : l'actualité, obsolescence, ratio de fraîcheur et l'âge comme indiqué dans le tableau suivant:

Facteur	Métrique	Description
Actualité	Actualité	Le temps passé depuis l'extraction des données
	Obsolescence	Le nombre de transactions de mise à jour d'une source depuis l'extraction des données
	Ratio de fraîcheur	Le taux de données qui sont à jour
Age	Age	Le temps passé depuis la création ou la mise à jour des données

Tableau 11 : Tableau des métriques de fraîcheur [BOU07]

- **L'accessibilité comme qualité de service:** l'accessibilité est l'aspect de qualité d'un service qui représente le degré qu'il est capable de servir une demande de service Web. Elle peut être exprimée comme mesure de probabilité dénotant l'indice de réussite ou la possibilité d'une instanciation réussie de service à un moment. Il pourrait y avoir des situations quand un service de Web est disponible mais non accessible. Selon [HAR05] l'accessibilité peut être représenté par
 - Le temps de réponse: le temps de réponse est une question préoccupante dans la performance de base de services Web. Le temps de réponse typique est un attribut de performance qui renvoie à la durée écoulée entre la fin de l'émission d'une demande à un service et le début de la réponse du service. Il représente le temps moyen que met le service Web à envoyer la réponse au client. La valeur de cet attribut peut être, soit renseignée par le fournisseur du service, soit donnée une fois que le service Web a été invoqué une première fois. Il est important de noter que cet attribut n'est donné qu'à titre indicatif mais peut aider le client lors de l'étape de sélection de services [BOC01, KAL03, RAN03]. L'évaluation du temps de réponse d'une requête R peut être représentée comme suit : Temps de réponse (R) = Temps d'exécution (R) + temps d'attente(R) [FLU03].
 - Disponibilité : la disponibilité est l'aspect de qualité de si le service Web est présent ou prêt pour l'usage immédiat. La disponibilité représente la probabilité qu'un service est disponible.
 - Prix : représente le montant à payer pour bénéficier du service Web.
- **La popularité comme qualité d'usage :** la popularité fait référence au degré d'utilisation effective de l'information par les utilisateurs. Ce facteur présente un grand intérêt car il permet de capturer des items qui sortent du champ habituel de contenu auquel s'intéresse

l'utilisateur a priori. Il peut donc contribuer à contrecarrer l'effet dit « entonnoir », en apportant des items inattendus sous l'angle de leur strict « contenu », et néanmoins appréciés pour leurs qualités reconnues par toute une communauté d'utilisateurs. La popularité peut être évaluée de différentes façons. Dans les contextes répandus d'accès en ligne aux items, la consultation et les téléchargements sont de bons indicateurs de popularité. Lorsque ces systèmes permettent aux utilisateurs d'évaluer les items de façon explicite (ratings), ces évaluations constituent une base intéressante pour évaluer la popularité, de même que les recommandations que les utilisateurs se font entre eux, de façon moins formelle dans les forums. De façon plus spécifique, dans le contexte de la publication académique, la notion de popularité se mesure au travers des citations croisées entre les publications [GOO01] ; en effet, cette activité de citation est une bonne indication de l'usage réel des documents par les membres de la communauté académique concernée. Les métriques de ce critère sont indiquées dans le tableau suivant :

Facteur	Métrique	Description
Popularité	Nombre de consultation	Le nombre d'exploitation de SWG
	Rating	L'évaluation ou recommandation donnée par un groupe d'utilisateur
	Nombre de citations	Le nombre de citations référençant le SWG

Tableau 12: Tableau des métriques de popularité

5.2.2.2.2 Le package Service

Le package nommé service rassemble les informations à publier afin que les concepteurs puissent utiliser les services. Nous nous sommes basés sur le modèle conceptuel de WSDL afin de construire ce package. La transcription de fichiers WSDL, exprimés en XML, en représentation orientée objet (telle que UML) a déjà fait l'objet de plusieurs travaux, tels que ceux de [MAR03]. Le package service est composé de cinq classes (Figure 47) : Methode, Binding, Parametre, Parametere_entrées et Parametre_sorties. Lorsque la publication concerne des services Web, les informations représentées dans ce package sont celles issues de la description WSDL, par conséquent peuvent être instanciées de manière automatique.

- **La classe Methode** : cette classe décrit les méthodes proposées par les services décrits.
- **La classe Binding**: lorsque le service est hébergé sur le Web, le fournisseur doit publier le moyen d'accès au service (i.e. le protocole à utiliser, tel que SOAP ou HTTP). La classe Binding représente ce type d'information. Une instance de cette classe peut être HTTP ou SOAP. Chaque méthode fournie par un même service peut être accessible par un protocole

différent. Le package service contient une relation entre la classe Methode et la classe Binding afin que les clients puissent connaître pour chacune des méthodes disponibles leur protocole d'accès.

- **Les classes Parametre, Parametere_entrees et Parametere_sortie:** les paramètres (classe Parametre) de chaque méthode sont représentés dans le modèle, par l'association reliant les classes Methode et Parametre. Un paramètre est décrit par son nom (attribut name) et son type (attribut type). Lors de l'appel de méthode il est nécessaire de différencier les paramètres d'entrée des paramètres de sortie (respectivement instances des classes Parametere_entrees et Parametere_sorties spécialisant la classe Parametre). Le package service est utile aux fournisseurs de services lors du processus de description des SWG. Ce package retranscrit, pour les services issus de la technologie des services Web, la description fonctionnelle réalisée en WSDL.

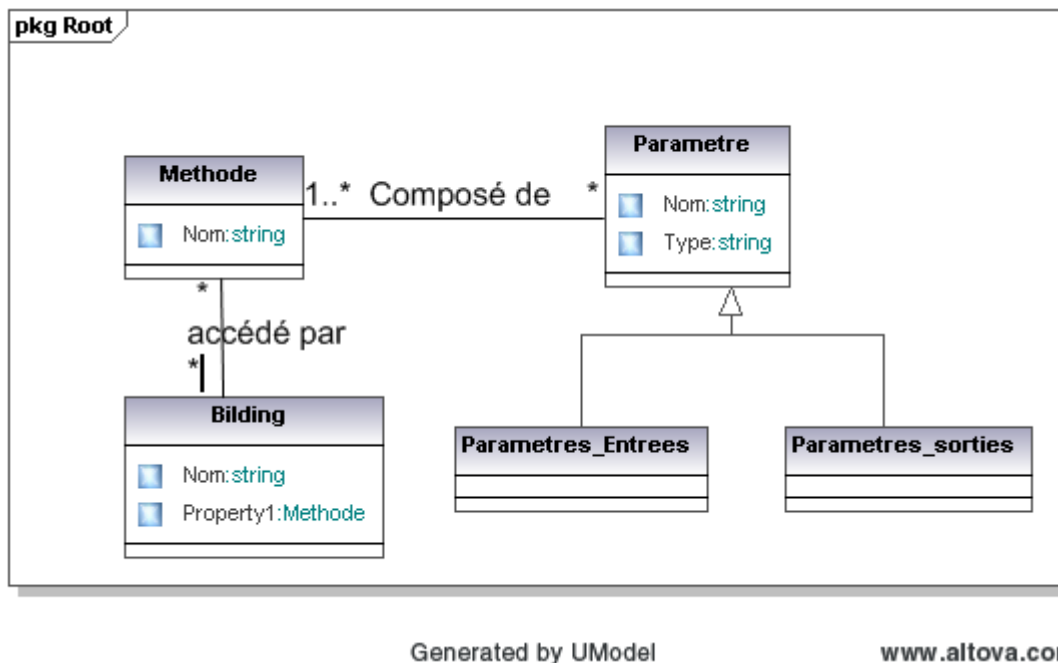


Figure 47 : Classes et associations du package Service

5.2.2.2.3 Le package Domaine d'application

Le package domaine d'application permet de caractériser le domaine d'application. L'utilisation de ce package par le fournisseur de services lors de leur publication consiste à ajouter un niveau de description supplémentaire. Ce package est composé de trois classes : Domaine, Application et Fonctionnalité.

- **La classe Domaine:** cette classe décrit le domaine d'application dans lequel le service intervient. L'instance de cette classe peut être un domaine tel que la géomatique, le travail collaboratif, le e-learning, etc.
- **La classe Application :** dans ce modèle nous associons à un domaine une ou plusieurs

applications (classe Application) faisant référence ici à des applications de type logiciel. L'association entre les classes Domaine et Application permet de représenter l'ensemble des applications pouvant intervenir dans un domaine donné.

- **La classe Fonctionnalité** : cette classe permet de représenter le fait qu'une application est développée à l'aide d'un ou plusieurs composants qui implémentent chacun une fonctionnalité spécifique. Cette classe décrit les fonctionnalités proposées par l'application.

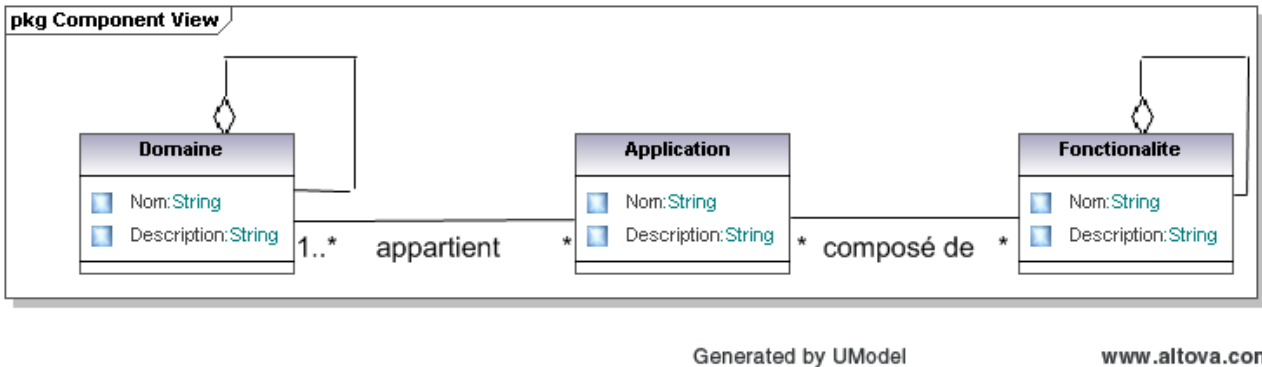


Figure 48 : Classes et associations du package Domaine d'Application

5.2.2.2.4 Le package Profil utilisateur

Le package Profil utilisateur regroupe les classes qui représentent l'information nécessaires pour décrire l'utilisateur afin d'adapter la requête utilisateur pour avoir le service le plus approprié ainsi que la classe requête qui représente la classe qui assure la liaison avec les classes de description de service. Il est représenté par le diagramme suivant :

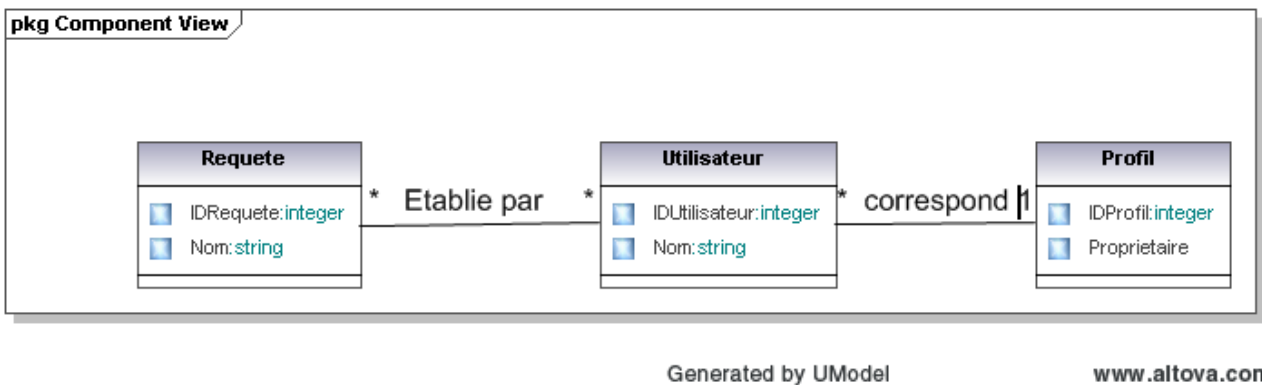


Figure 49 : Diagramme simplifié des classes de package profil utilisateur

- **La classe utilisateur**: cette classe représente les instances du profil utilisateur.
- **La classe Profil** : cette classe représente le profil utilisateur qui est inspiré des travaux de [KOS07]. Dans notre travail nous nous intéressons particulièrement à la dimension des données personnelles, la dimension du domaine d'intérêt représenté dans notre cas à travers une liste de mots clés et la dimension de qualité en ajoutant la préférence de l'utilisateur pour chaque concept de qualité. Ces classes sont l'issue de l'instanciation du méta-modèle du

profil utilisateur présenté dans le chapitre 3 (3.2.2). Nous obtenons le digramme suivant :

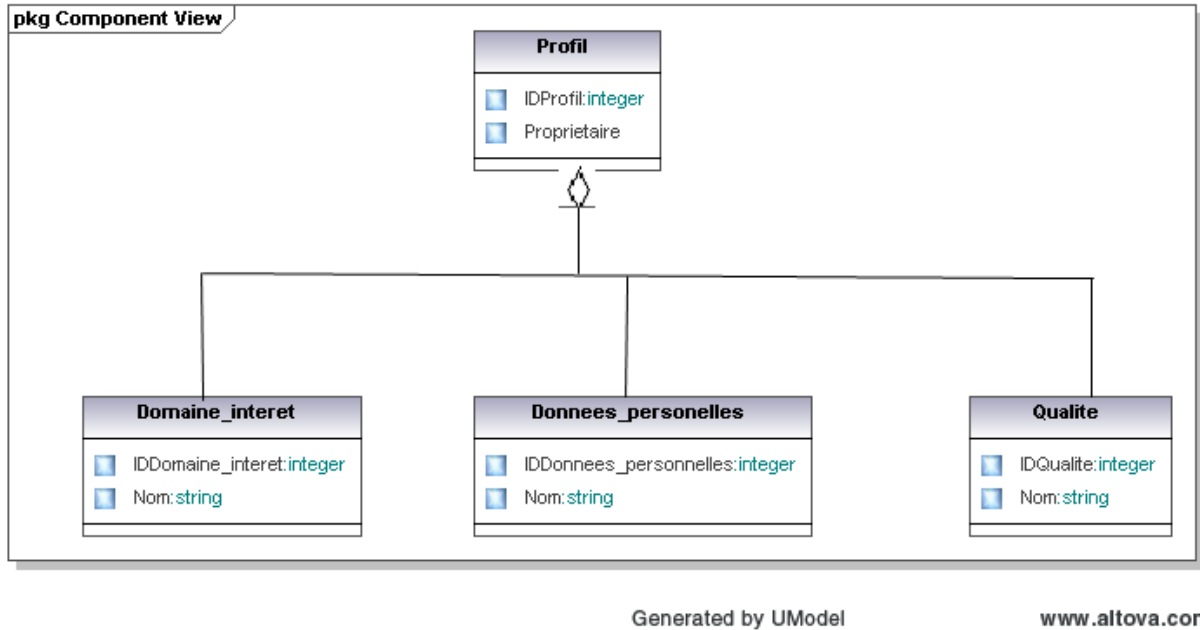


Figure 50 : Diagramme des classes de profil

- **Classe Profil :** cette classe représente les profils utilisateur à l'aide de l'identificateur et son propriétaire. Elle est composée de trois classes qui représentent les dimensions de profil utilisateur choisies dans notre travail.
- **Classe Domaine intérêt :** la classe du domaine d'intérêt est destinée à exprimer les centres d'intérêts de l'utilisateur. Ceci peut être fait par des **mots clés** ou par **des ontologies**. Les mots clés représentent des mots ou des phrases qui décrivent les informations recherchées. Ils définissent de façon simple le contenu ou le domaine des éléments ciblés. Dans cette catégorie nous mettons toute description textuelle des objets ciblés. Certaines approches associent aux mots clés des attributs exprimant l'importance du mot (dans le cas de mots clés leur fréquence) ou la pertinence de l'objet (dans le cas où le profil contient des éléments de l'espace de recherche attribut pertinent ou pas). L'un des problèmes les plus importants de la découverte des services Web est le grand nombre de services Web correspondant à la requête à cause de l'utilisation de termes trop généraux. Pour résoudre ce problème les termes employés dans les requêtes doivent être explicités de façon à donner aux mots un sens unique. Lorsqu'une requête devient trop spécifique, nous risquons d'avoir des résultats nuls et donc de décevoir l'utilisateur. Dans ce cadre on utilise des requêtes approximatives dont le but est d'élargir l'espace de recherche. Ces requêtes sont basées le plus souvent sur des calculs de similarité entre les termes et nécessitent une connaissance linguistique. Pour répondre aux problèmes cités plus haut, cette sous-catégorie comprend des synonymes des mots, leurs traductions dans les langues que l'utilisateur comprend et le sens qu'il donne à chacun de ces termes appelé plus souvent « contexte ». La notion de contexte est très

importante pour la réduction de l'espace de recherche. Le sens des termes doit être correcte sinon les résultats renvoyés à l'utilisateur en présence de son profil risquent d'être tous inintéressants. Ce contexte peut avoir également un impact sur l'évaluation des opérateurs de la requête. Le contexte est basé sur les ontologies pour définir les termes et les opérateurs employés. Dans la partie de définition des opérateurs, nous mettons la définition des outils nécessaires pour décrire le sens des expressions de préférence du client. Parmi les opérateurs on retrouve les opérateurs de base (égalité, supériorité, infériorité etc.), mais aussi des opérateurs plus complexes (Similaire, Environ, Entre, Meilleur que, Le plus grand, Le plus petit, Préféré devant, etc.). Ces opérateurs complexes peuvent être présentés par des formules de préférence où des fonctions l'expression desquelles dépend du contexte d'utilisation et de l'utilisateur [KOS04].

- **classe de données personnelles** : Elle représente la partie statique du profil. Elle contient des informations qui décrivent l'utilisateur et ne dépendent pas du système à interroger. Ces informations sont classées selon trois catégories:
 1. Données personnelles : catégorie (identité) est composée d'un ensemble d'attributs d'identification de l'utilisateur. Des exemples de tels attributs sont le nom et le prénom de l'utilisateur, son adresse, son numéro de téléphone ou de fax, son adresse email etc.,
 2. Données démographiques : attributs démographiques comme par exemple la date de naissance de l'utilisateur, son genre, son revenu, son état civil, le nombre d'enfants etc.
 3. Données des contacts : les contacts de l'utilisateur représentent son carnet d'adresse. Un élément de ces contacts est représenté par les données personnelles d'un autre utilisateur.
- **classe Qualité**: cette classe joue un rôle très important dans notre travail. Les données de cette classe décrivent la qualité attendue ou espérée par l'utilisateur qui sera confrontée à la qualité effective des SWG afin de restreindre l'espace de recherche. Elle est composée de trois catégories de qualité (détaillés dans le 2ème chapitre):
 1. **La qualité des données géographiques** : cette qualité est assurée à travers des critères définis selon les aspects spatiaux qui permettent d'évaluer la pertinence des données par rapport au contexte géographique [GUT07]. Ils sont décrits selon la norme ISO 19113 qui assure la description de la qualité des données géographiques à l'aide de 7 critères (03 critères quantitatifs, 03 qualitatifs, critère spécifique)[TRI07, SER06].
 2. **La qualité des services Web**: cette qualité exprime des conditions non fonctionnelles des services. Les propriétés de QoS révèlent des caractéristiques diverses et s'expriment sous différentes formes, car elles peuvent être relatives aux services mêmes, notamment

en terme (i) de performance (fiabilité, temps d'exécution attendu, etc.), (ii) au contexte d'exécution (temps de latence, débit, etc.) ou (iii) aux besoins de l'utilisateur (qualité du résultat, coût économique du calcul, etc.).

3. **La qualité d'usage:** la qualité d'usage dans notre contexte représente les éléments demandés par l'utilisateur en termes de qualité spécifiant les critères d'utilisation des services Web. Ces éléments spécifient l'environnement préféré par l'utilisateur. Parmi ces critères nous citons la popularité.
- **La classe requête utilisateur :** Les instances de cette classe permettent de décrire une requête utilisateur. Une requête est représentée par son nom. Elle est décrite par une agrégation de conditions ou chaque condition représente un aspect du service Web géographique. Pour faciliter la tâche de l'utilisateur durant l'établissement de sa requête, nous lui proposons un composant de construction de requête basé sur le profil utilisateur et l'ontologie de qualité et du domaine d'application à travers une interface Web conviviale. Ce composant propose à l'utilisateur une requête préalable en fonction de son profil et lui donne la possibilité de fixer le degré de la correspondance désiré entre sa requête et les offres de SWG possibles et ainsi de raffiner sa requête en utilisant un pseudo langage. Cette requête doit contenir des éléments représentant l'aspect sémantique et l'aspect syntaxique du service web géographique à découvrir. Elle doit avoir une structure permettant le «matching» au niveau du catalogue de métadonnées en fonction des critères de qualités, et le «matching » au niveau de registre UDDI en fonction des éléments de l'aspect fonctionnel du service à découvrir.

La requête utilisateur peut être construite en utilisant la grammaire suivante :

```
<Requête_utilisateur> ::= <Requête_fonctionnelle> <Requête_non_fonctionnelle>
<Requête_fonctionnelle> ::= <Fonctionnalité_requête> <P_fonctionnel_requête>
<Fonctionnalité_requête> := Archivage | Affichage | Acquisition | Analyse | Abstraction
<P_fonctionnel_requête> := <Paramètres_Entrées> < Paramètres_sorties>
<Requête_non_fonctionnelle> ::= (Where <Condition>) ?
<Condition> ::= <QCC> ('AND' <condition>)*
<QCC> ::= <Metrique_Q> <Operateur> <Valeur_q> [<Preference>]
                                                (<Op_Conectivité> <QCC>)*
<Operateur> ::= '=' | '<' | '<=' | '>' | '>=' | '<>' | <OP_ONT>
<Preference> := <Poids> <Seuil>
<Op_Conectivité> ::= 'AND' | 'OR'
```

Les composants de la requête sont décrits dans le tableau suivant :

Composants de la requête	Description
Requête_fonctionnelle	Indique la partie fonctionnelle de la requête c'est à dire l'aspect fonctionnelle en spécifiant les fonctionnalités et les paramètres.
Requête_non_fonctionnelle	Indique la partie non fonctionnelle de la requête c'est à dire l'aspect qualité de la requête.
Fonctionnalité_requête	Spécifique à la catégorie de service Web géographique, dans notre cas elle peut l'une des fonctionnalités suivantes : affichage, analyse, abstraction, archivage ou abstraction. Dans notre travail nous faisons appel à deux fonctionnalités : l'analyse et l'affichage.
P_fonctionnel_requête	Représente la liste des paramètres spécifiant l'aspect fonctionnel des services web Géographique.
Paramètres_entrées	Représente l'ensemble des paramètres d'entrées
Paramètres_sorties	Représente l'ensemble des paramètres de sorties
Condition	Représente les conditions simples ou composés de la requête.
QCC	Qualité de catégorie
Métrique_Q	Représente l'indicateur de qualité, dans notre cas les métriques sont des composants élémentaires.
Valeur_Q	Représente une valeur atomique de qualité,
Preference	Cette préférence est indiqué à travers le poids qui représente l'importance de QCC dans la requête et le seuil qui représente la bare au dela la QCC sera acceptée
OP_ONT	Opérateur liée à l'ontologie de qualité des SWG
?	L'élément est optionnel
	Ce symbole définit une alternative
*	L'élément est optionnel ou répétitif.
::=	L'élément à gauche du symbole est défini par les éléments de droites

Tableau 13 : Tableau de description des composants de la grammaire de la requête

Cette grammaire est un noyau qui pourra être complété en fonction du domaine d'application comme par exemples la liste des valeurs et les opérateurs (les opérateurs exploités sont les opérateurs de comparaisons, nous pouvons faire appel à d'autres opérateurs de comparaison géographiques capturés de l'ontologie du domaine d'application.

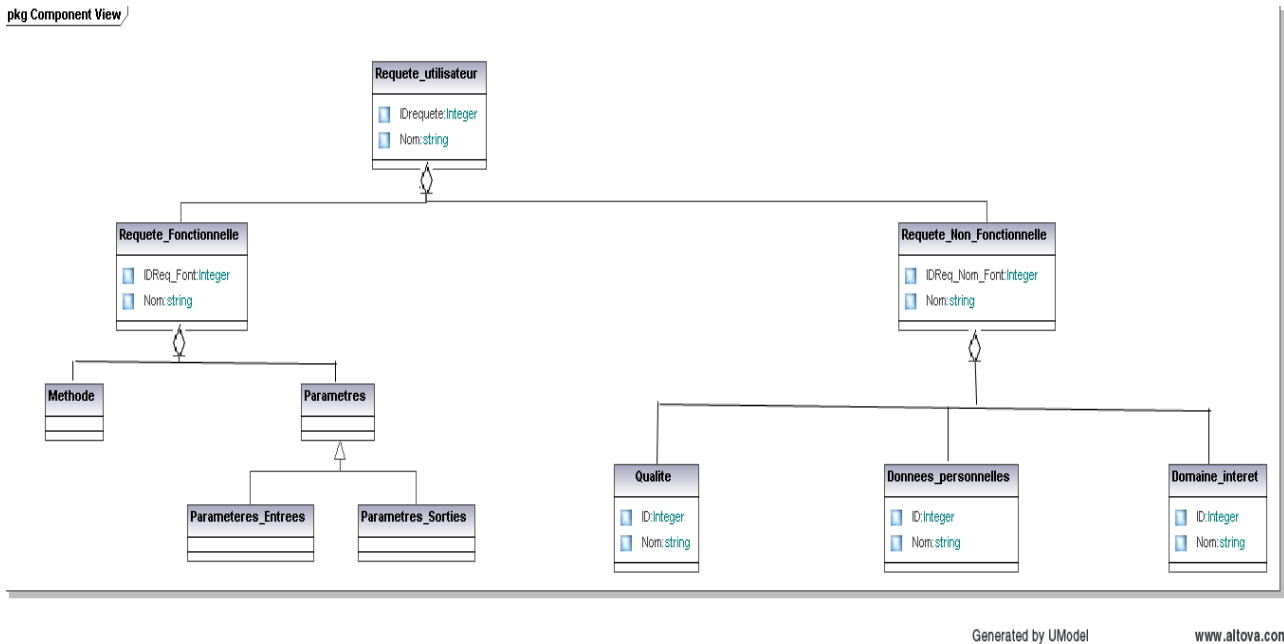


Figure 51 : Diagramme des classes de la requête utilisateur

5.2.2.2.5 Relations entre les packages

Dans notre modèle, il existe un ensemble de liens entre les classes des différents packages précédemment décrit. Ces packages sont classés en 02 catégories :

- Packages fournisseur : ces packages sont exploités pour la description des SWG, il contient le package catégorie, le package service et le package domaine d'application.
- Package client : ce package contient les classes de description des clients ainsi que celle de la requête.

La classe centrale de ces liens est la classe requête utilisateur qui assure le lien entre les packages fournisseurs et le package client. Ces relations de lien sont présentées comme suit :

1. Relation entre catégorie et domaine d'application : l'association reliant les classes de la catégorie SWG indique le domaine d'application des classes de catégories SWG. L'association entre les classes Catégorie concret et fonctionnalités sert à présenter les catégories concrètes dont les services réalisent des tâches spécifiques.
1. Relation entre catégorie et service : cette association présente l'idée de base de notre travail, elle permet de restreindre l'espace de recherche des services Web. Les informations concernant l'accès (classe Binding), les méthodes fournies (classe Methode), les paramètres d'entrée et de sortie (classes Parametres_entrees et Parametres_sorties) permettent d'instancier les classes correspondantes qui seront extraites de la description WSDL.
2. Relation entre catégorie et profil utilisateur : cette liaison représente la plus importante liaison dans notre travail, elle assure le lien entre les critères de qualité figurant dans la description des catégories SWG et ceux présentés dans la requête utilisateur.

3. Relation entre service et domaine d'application : cette liaison peut être directe entre la classe fonctionnalités du package domaine d'application et la classe methode du package service. Elle peut être remplacée par la liaison entre la classe catégorie concrète du package catégorie et la classe fonctionnalité du package domaine d'application.
4. Relation entre service et profil utilisateur : cette liaison concerne l'aspect fonctionnel du SWG. Elle est concrétisée à travers la liaison entre la classe fonctionnalité de la requête utilisateur et la classe methode du package service et la liaison entre les classes parametres de la requête utilisateur et le package service.

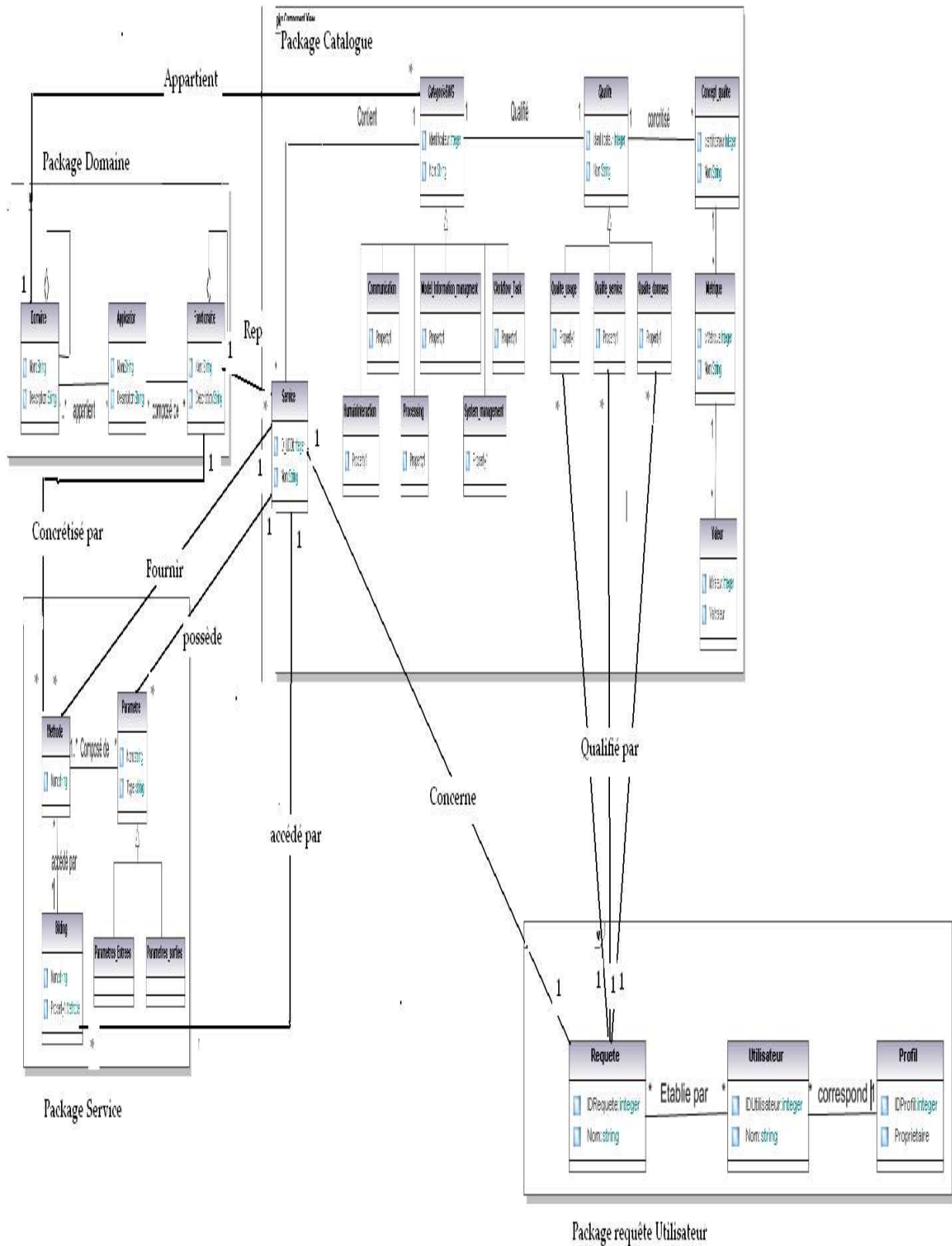


Figure 52 : Diagramme de classes représentant les relations entre les packages

5.3 Représentation formelle de la qualité

Pour assurer un « matching » efficace entre les critères de qualités des catégories des SWG et ceux de la requête utilisateur, nous proposons une représentation formelle de la notion de qualité comme suit :

5.3.1 Qualité des catégories des SWG

La qualité est représentée par un ensemble de critères. Cet ensemble contient trois type de critères à savoir : les critères de données, de services et d'usage. Chaque critère est représenté par une ou plusieurs dimensions ou chaque dimension a un poids. Ce dernier représente l'importance de la dimension pour la catégorie de service. La notion de qualité peut avoir la représentation formelle suivante:

- $Q = Q_d \cup Q_s \cup Q_u$: représente la qualité globale de la catégorie
 - Q_d : représente les critères de qualité des données tel que
 $Q_d = \{D_i (\text{Nom}, \text{Poids}) \quad i:1..N \}$
 - Q_s : représente les critères de qualité des services tel que
 $Q_s = \{D_i (\text{Nom}, \text{Poids}) \quad i:1..M \}$
 - Q_u : représente les critères de qualité des données tel que
 $Q_u = \{D_i (\text{Nom}, \text{Poids}) \quad i:1..P \}$
le poids attribué à D représente le degré d'importance de la dimension, il est entre 0 et 1.
- Q_d , Q_s et Q_u sont disjoints
- La relation sémantique entre les différentes dimensions d'une qualité est assurée à travers une ontologie de concepts de qualité.

5.3.2 Qualité de la requête utilisateur

La qualité de la requête utilisateur représente les besoins de l'utilisateur sous forme de critères de qualité. Elle a une représentation formelle similaire à celle utilisée pour la qualité des catégories des SWG en ajoutant un autre élément important (seuil) à travers lequel la dimension qualité sera acceptée ou non. Donc la qualité de la requête utilisateur est formalisée comme suit :

- $Q = Q_d \cup Q_s \cup Q_u$: représente la qualité globale de la catégorie
 - Q_d : représente les critères de qualité des données tel que
 $Q_d = \{D_i (\text{Nom}, \text{Poids}, \text{Seuil}) \quad i:1..N \}$
 - Q_s : représente les critères de qualité des services tel que
 $Q_s = \{D_i (\text{Nom}, \text{Poids}, \text{Seuil}) \quad i:1..M \}$
 - Q_u : représente les critères de qualité des données tel que
 $Q_u = \{D_i (\text{Nom}, \text{Poids}, \text{Seuil}) \quad i:1..P \}$
le poids attribué à D représente le degré d'importance de la dimension, sa valeur est entre 0 et 1, le seuil représente la valeur au delà cette dimension sera acceptée.

- Qd, Qs et Qu sont disjoints
- La relation sémantique entre les différentes dimensions d'une qualité est assurée à travers une ontologie de concepts de qualité (même ontologie utilisée pour la qualité des catégories des SWG).

5.3.3 Méthodes d'évaluation de la qualité

Pour évaluer la qualité globale nous faisons appel aux méthodes d'évaluation de qualité inspirée des travaux de la recherche d'information. Parmi les méthodes d'analyse multicritères les plus connues nous pouvons citer [HAR05] :

- la méthode SAW (Simple Additive Weighting): est une méthode multicritère caractérisée par une agrégation additive des critères par sommation pondérée.
- la méthode TOPSIS (Technique for Order Preference by Similarity to Ideal Solution): est une méthode multicritère développée par Hwang et Yoon (1981) se basant sur une relation de dominance qui résulte de la distance par rapport à la solution idéale. Elle se caractérise par une agrégation compensatoire entre les critères
- la méthode AHP : l'analyse procédurale hiérarchique AHP (Analytic Hierarchy Process), développée par Thomas Saaty (1971), est une méthode de prise de décision qui décompose le problème en structure hiérarchique de choix ou objectifs, par la suite en réduisant les décisions complexes en séries de simples paires de comparaison et en synthétisant les résultats.
- la méthode DEA (Data Envelopment Analysis): est une méthode introduite par Thomas Charnes, Cooper et Rhodes (1978). Cette méthode détermine le degré d'efficacité d'une solution en résolvant un programme linéaire, pour trouver la valeur optimale. Dans la méthode DEA les variables du programme sont les poids des critères qui ne sont pas spécifiées par l'utilisateur.

Dans notre travail les préférences utilisateur sont exprimées essentiellement par une affectation de poids aux dimensions de la qualité et la stratégie d'évaluation de la qualité repose essentiellement sur le calcul de score de la qualité des dimensions à partir des scores des sous dimensions. En se basant sur la comparaison de Naumann [NAU98] et l'étude de complexité de Burgess [BUR03], le choix de la méthode de calcul se fait selon les propriétés suivantes :

- interaction : l'utilisateur peut exprimer des préférences sur les critères (affectation d'un poids numérique aux critères de qualité).
- dominance : une source d'information domine l'autre si elle est égale ou meilleure pour tous les critères et meilleure au moins pour un critère.

- type de résultat : on distingue deux types de résultats : ordonnancement (score de qualité) ou classement.
- critères positifs et négatifs : la méthode distingue entre les critères négatifs (prix, début réseau, temps de réponse...) et les critères positifs (disponibilité, complétude ...).
- complexité : la complexité de l'algorithme de calcul de scores de qualité.

Propriété	SAW	TOPIS	AHP	DEA
Interaction	Poids	Poids	Décisions	-
Dominance	Oui	Oui	Oui	Oui
Type de résultat	Ordonnancement	Ordonnancement	Ordonnancement	Classement
Critères positifs et négatifs	Oui	Non	Non	Non
Complexité	O(n,m)	O(n,m)	Exponentielle	polynomiale

Tableau 14: Comparaison des méthodes d'analyses multicritères [HAR05].

5.3.3.1 La méthode SAW (Simple Additive Weighting)

L'algorithme de calcul de score dans la méthode SAW se décompose en trois étapes :

Étape 1 : création de la matrice de décision

La première étape est la création de la matrice de décision qui représente les données disponibles et les critères de qualité (tableau 15).

	Critère			
Alternatives	C1	C2	...	Cn
Poids Alternatives	W1	W2	...	Wn
A1	D11	D12	...	D1n
A2	D21	D22	...	D2n
...
Am	Dm1	Dm2	...	Dmn

Tableau 15: Matrice de décision de la méthode SAW

A_i est la i ème alternative, C_j est le j ème critère et d_{ij} est la mesure de la performance de la i ème alternative pour le j ème critère. Dans notre travail, les critères sont les dimensions, et les alternatives sont les catégories des SWG.

Étape 2 : création de la matrice de décision normalisée

Cette étape consiste à normaliser les poids et les mesures de performance afin d'obtenir des valeurs entre 0 et 1. La normalisation se fait selon les formules suivantes :

$$\text{Pour les poids : } w_j' = \frac{w_j}{\sum_{j=1}^m w_j}$$

Pour les mesures :

$$v_{ij} = \frac{d_{ij} - d_j^{\min}}{d_j^{\max} - d_j^{\min}} \quad \text{Pour les critères positifs}$$

$$v_{ij} = \frac{d_j^{\max} - d_{ij}}{d_j^{\max} - d_j^{\min}} \quad \text{Pour les critères négatifs}$$

$$\text{Où } d_j^{\max} \equiv \max_i [d_{ij}] \text{ et } d_j^{\min} = \min_i [d_{ij}]$$

Étape 3 : calcul du score de chaque alternative.

Le score de qualité de chaque alternative est donné par cette formule :

$$\text{Score}(A_i) = \sum w_j v_{ij} \quad 0 \leq \text{Score}(A_i) \leq 1$$

5.4 Algorithme de « matching »

Dans la phase de découverte dans cette approche nous avons essayé d'améliorer la pertinence de la liste des SWG choisis à travers l'utilisation d'un algorithme de « matching » de LARKS [SYC02] (détaillé dans 5.4.1) au niveau métadonnées. Cet algorithme sera adapté en appliquant les trois premiers filtres (contexte, profil, similarité) sur les critères de qualités figurant dans la requête et les descriptions des classes de SWG disponibles dans le catalogue de métadonnées. Cette solution nous permet d'avoir un résultat de type «relaxed match». L'algorithme de « matching » niveau catalogue exploite une ontologie de qualité des SWG. Le « matching » au niveau registre UDDI est le même algorithme de l'approche précédente (présenté 4.2.3.3). Il est basé sur le mécanisme d'interrogation du registre UDDI en faisant appel aux différentes pages (pages blanches, pages jaunes, pages vertes).

5.4.1 L'algorithme de « matching » LARKS

L'algorithme de « matching » LARKS (Language for Advertisement and Request for Knowledge Sharing)[SYC02] se base non seulement sur la syntaxe du service mais aussi sur sa sémantique. Cet algorithme de « matching » est composé des étapes suivantes :

- agent fournisseur : son rôle consiste à fournir des types spécifiques de services comme les services de recherche d'information;
- agent demandeur : il demande au agent fournisseur de lui assurer certains services;

- agent médiateur : cet agent stocke les différentes annonces de l'agent fournisseur et assure la médiation entre les deux autres agents.

Le « matching » dans LARKS consiste à trouver le fournisseur le plus approprié pour un demandeur à travers le médiateur. Il est composé des étapes suivantes :

- l'agent fournisseur publie ces capacités à l'agent médiateur ;
- l'agent médiateur stocke ces publications ;
- l'agent demandeur demande à l'agent médiateur de lui indiquer le fournisseur le plus appropriés à sa demande;
- l'agent médiateur fait le « matching » en se basant sur les publications stockées et présente les résultats.

Ce processus apparaît simple, mais il devient très compliqué à cause de l'hétérogénéité des différents fournisseurs qui sont incapable de se comprendre.

Dans le système LARKS, les services sont modélisés à travers des frames composés de :

- Context : le contexte de la spécification du service ;
- Types : définitions des types de données utilisées dans la spécification ;
- Input and output : déclarations des variables d'entrées et de sorties de la spécification ;
- Inconstraints and Outconstraints : les contraintes logiques sur les variables d'entrées et les variables de sorties.
- ConsDescription : description de la signification des mots utilisés dans la spécification. Elle s'appuie sur une ontologie locale.
- TextDescription : texte facultatif de la signification de la spécification.

le processus de « matching » est conçue en tenant compte des critères suivants [SYS02]:

- Le rapprochement ne devrait pas être fondé sur les mots clé seulement. Au lieu de cela, contrairement à l'habitude des moteurs de recherche en texte libre, la sémantique des requêtes et les annonces doivent être prises en considération.
- Le processus de « matching » doit être automatisé. Une grande quantité d'agents apparaissent et disparaissent dans l'Internet. Il est presque impossible pour un utilisateur de rechercher manuellement ou parcourir toutes les capacités des agents
- Le processus de « matching » doit être exact. Par exemple, si les résultats retournés par le moteur de match sont prétendument correspondance exacte ou plug-in de match, ces matches doivent répondre aux définitions d'appariement exact et plug-in correspondant
- Le processus de « matching » doit être efficace, en termes de rapidité.
- Le processus de « matching » doit être efficace, l'ensemble des matchs ne devrait pas être trop grande. Il est préférable d'utiliser une petite série de matches très bien cotées à une

requête donnée.

Le mécanisme de recherche est basé sur cinq filtres. Ces filtres se basent sur différents aspects de correspondance entre requête et annonce :

1. le contexte: la requête et l'annonce doivent partager le même domaine. Le filtre calcule les distances entre les différents mots-clés d'annonce et de requête. Cette distance est basée sur les liens entre les mots tels que la généralisation, la spécialisation et l'association positive.
2. le profil: la requête et l'annonce doivent partager un certain nombre de mots dans leurs spécifications. Ce filtre compare les occurrences des mots de l'annonce et de la requête.
3. la similarité: la structure de la requête et la structure de l'annonce doivent être similaires, c'est à dire les concepts utilisés, par exemple dans la partie input, correspondent dans les deux cas. Au lieu de faire une comparaison de l'ensemble du document, on compare les différentes sections (input, output, précondition, ...) avec le filtre de contexte.
4. la signature: la signature de l'annonce doit ressembler à la signature de la requête. On compare les différents termes (chaque paramètre individuellement) grâce au filtre de similarité. De plus, pour les paramètres input et output, on vérifie également que les types sont compatibles (équivalent, sous-classe, sur-classe).
5. les contraintes : l'utilisateur voulant utiliser un service doit vérifier les pré-conditions de ce service. De même, le service doit répondre aux attentes de l'utilisateur.

Ces filtres sont indépendants les uns des autres. Le filtre contexte permet de sélectionner l'espace de recherche qui peut être comparé avec la requête dans le même contexte ou similaire. Le filtre profil, similarité et signature sont appliqués seulement sur l'espace sélectionné par le filtre contexte [SYC02].

La composition de ces filtres permet d'établir différents degrés de correspondance :

- exact match: le service correspond parfaitement à la requête. On compare la requête avec l'annonce au moyen des cinq filtres.
- plug in match: le service annoncé peut correspondre (être branché) à la requête. Par exemple, l'utilisateur désire un service triant une liste d'entiers. Le médiateur propose un service de tri de liste d'entiers et de strings. On compare la requête avec l'annonce au moyen des deux derniers filtres : signature et contraintes.
- relaxed match: la relation relaxed match ne signifie pas que l'annonce correspond à la requête mais détermine que leur distance sémantique est inférieure à un seuil déterminé. Par exemple, un utilisateur cherche l'adresse d'un magasin vendant des ordinateurs HP et un service fournissant le prix et le numéro de téléphone d'un tel magasin. On compare la requête avec l'annonce au moyen des trois premiers filtres. Il est donc le moins consommateur de ressource.

En toute logique, les résultats du relaxed match devraient inclure les exact et plug in match.

5.4.2 Pourquoi LARKS

Parmi les algorithmes de «maching» cités [PAO02], nous utilisons l'algorithme de «matching» LARKS au niveau métadonnées pour les raisons suivantes:

- Le formalisme de spécification du LARKS correspond à la description proposée dans notre travail. Cette correspondance est indiquée dans le tableau suivant :

LARKS	Description des catégories SWG
Context	Domaine d'application Critère de qualité
Types	Les types d'attributs de la dimension qualité
Input	Les attributs de la dimension qualité
Output	Liste des catégories des SWG
Incontraint	Les valeurs de préférences des attributs de qualité
Outcontraint	
ConDescription	Ontologie

Tableau 16: Tableau de correspondance entre LARKS et QGWSR-Model

- l'utilisation des trois premiers filtres (contexte, profil, similarité) permet d'assurer un «relaxed matchig » qui englobe les deux autres type de matching (exact, plug-in) et consomme le moins de ressources. Le rôle de chaque filtre dans le cadre de notre travail est comme suit :
 - Filtre de contexte : dans le cadre général, le rôle du filtre de contexte est de spécifier le domaine cible de la requête utilisateur. dans notre cas nous utilisons le filtre de contexte pour spécifier la catégorie (sous catégorie) du SWG correspondante à la requête utilisateur en se basant sur l'ontologie de qualité de SWG. Ce qui permet de restreindre l'espace de recherche en déterminant les entrées concernées dans le catalogue des métadonnées. Ce filtre est basé sur le filtrage multidimensionnel selon les critères de qualité et le calcul des distances entre les différentes dimensions de qualité d'annonce et de requête. Cette distance est basée sur les liens entre les dimensions de qualité tel que la généralisation, la spécialisation et l'association positive, qui sont indiqué dans une ontologie de qualité des catégories des SWG en se basant sur la méthode d'évaluation de qualité SAW. Les autres filtres seront appliqués sur l'espace choisi par le filtre contexte.
 - Filtre de profil: ce filtre consiste à calculer la distance entre les mots de la requête et ceux des entrées du catalogue de métadonnées choisi par le filtre de contexte afin de spécifier les éléments du catalogue qui partage le plus des mots clés avec la requête

utilisateur. Ce calcul est basé sur l'utilisation du modèle vectoriel exploité dans le domaine de la recherche d'information.

- Filtre de similarité: ce filtre permet d'assurer la comparaison entre les différentes sections de la requête utilisateur et celles choisi par le filtre de profil. Dans notre cas la similarité est assurée à travers le formalisme proposé pour spécifier la classe requête utilisateur.

Nous remarquons que l'application de ces trois filtres assure une restriction incrémentale de l'espace de recherche, tel que le filtre de contexte spécifie les entrées concernées dans le catalogue des métadonnées ce qui élimine les autres entrées, le filtre de profil spécifie les entrées du catalogue de métadonnées en se basant sur le partage des mots clés ce qui élimine les autres entrées de l'espace choisi par le filtre de contexte en acceptant seulement les entrées dont la similarité dépasse un seuil donné. Enfin le filtre de similarité exploite les résultats des deux autres filtres en se basant sur la similarité de la requête et les entrées des catalogues (dans notre cas les requêtes et les annonces sont similaires).

5.4.3 Notions utilisées dans l'algorithme de « matching » de notre approche

Dans cette section nous présentons les différentes notions utilisées ainsi que l'algorithme de « matching » correspond aux deux niveaux (catalogue et registre UDDI).

5.4.3.1 Filtrage multidimensionnel selon les critères de qualité

Le filtre utilisé dans cas est inspiré des travaux de Burgess et al [BUR02]. Ils proposent une hiérarchie pour caractériser la qualité des données. Les principales dimensions de qualité sont coût, utilité et temps. Chaque dimension se décompose en sous dimensions et critères de qualité. Dans cette approche les préférences de l'utilisateur en terme de qualité sont exprimées par :

- Le choix des dimensions ou critères de qualité (représenté par des poids).
- Le choix du seuil de qualité.

Les auteurs définissent une taxonomie de la qualité afin de filtrer les informations. Les informations qui ont une qualité inférieure au seuil sont éliminées.

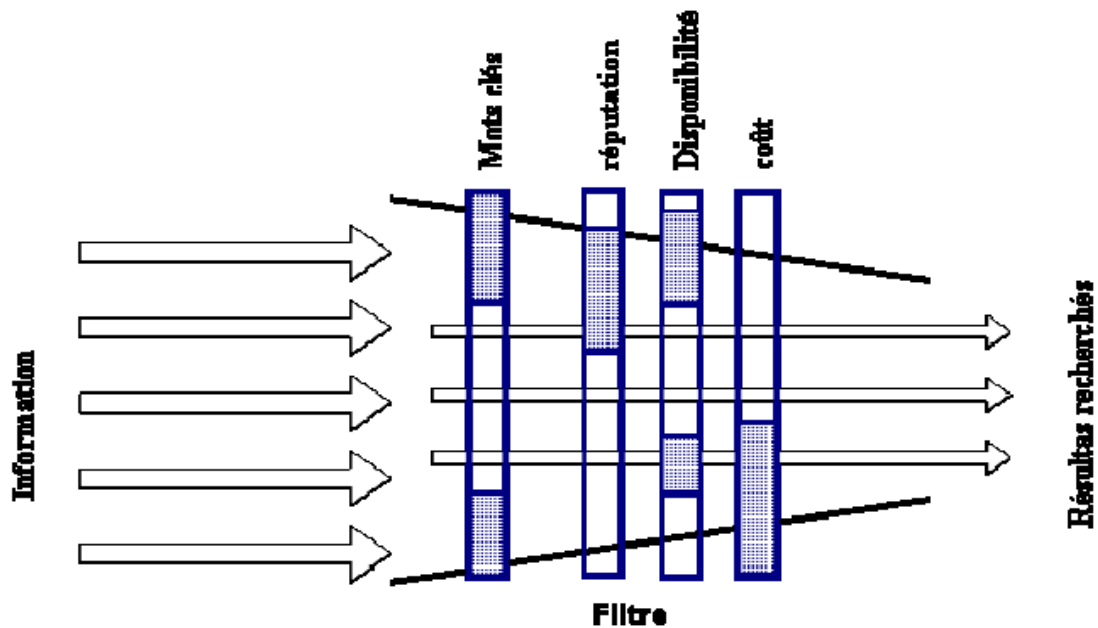


Figure 53 Filtrage d'information en utilisant des facteurs qualité [BUR02].

Dans notre travail, d'une part l'utilisateur exprime ses préférences en termes de qualité en fonction d'un poids et un seuil pour chaque dimension, et d'autre part le fournisseur des SWG exprime l'importance de chaque dimension en fonction d'un poids. Le filtrage multidimensionnel selon les critères de qualité consiste à filtrer l'ensemble des dimensions de qualité pour chaque annonce en fonction de la requête utilisateur. Après le filtrage nous obtenons seulement la liste des catégories des SWG candidats avec les dimensions de qualité acceptées pour chaque catégorie.

5.4.3.2 La distance sémantique entre deux concepts dans une ontologie

La détermination du degré de similarité entre deux concepts reliés à des termes d'un document est un problème qui se pose dans beaucoup d'applications : désambiguïsation, résumé automatique, extraction d'information, indexation automatique, etc.

Rada et al. [RAD89] ont suggéré que la similarité dans un réseau sémantique peut être calculée en se basant sur les liens taxonomiques « is-a ». Plus généralement, le calcul de similarité entre concepts peut être basé sur les liens hiérarchiques de spécialisation/généralisation. Un moyen des plus évidents pour évaluer la similarité sémantique dans une taxonomie est de calculer la distance entre les concepts par le chemin le plus court [ZAR04].

il existe une bonne dizaine de mesures permettant de calculer, dans une hiérarchie de concepts, une valeur de similarité sémantique entre le deux concepts à comparer. Parmi ces mesures nous citons [ZAR04] :

5.4.3.2.1 La mesure de Resnik

La notion de contenu informationnel (CI) a été la première fois introduite par [RES95]. Elle utilise conjointement l'ontologie et le corpus. Le contenu informationnel d'un concept traduit la pertinence d'un concept dans le corpus en tenant compte de sa spécificité ou généralité. Nous disons qu'un concept général *subsume* un concept plus spécifique. La fréquence de concepts dans le corpus est calculée pour retrouver le contenu informationnel. Cette fréquence regroupe la fréquence d'apparition du concept lui-même ainsi que des concepts qu'il subsume. La formule est la suivante :

$$\mathbf{CI(c) = -\log(P(c))}$$

Où $P(c)$ est la probabilité de retrouver une instance du concept c . Ces probabilités sont calculées par : $\text{frequence}(c)/N$ où N est le nombre total de concepts.

Resnik définit la similarité sémantique entre deux concepts par la quantité d'information qu'ils partagent. Cette information partagée est égale au contenu informationnel du plus petit généralisant (PPG) – le concept le plus spécifique qui subsume les deux concepts dans l'ontologie.

$$\mathbf{Sim(c1, c2) = CI(ppg(c1, c2))}$$

Cette mesure ne dépend que du PPG et est de ce fait un peu sommaire car nous pouvons avoir $\text{ppg}(a,b) = \text{ppg}(d,e)$ même si d et e sont plus proches du PPG que a et b .

5.4.3.2.2 La mesure de Jiang-Conrath

La mesure de ([JIA97]) pallie aux limites de la mesure de Resnik en combinant le contenu informationnel du PPG à ceux des concepts. Elle prend en considération aussi le nombre d'arcs. Ainsi une distance est définie :

$$\mathbf{distance(c1, c2) = CI(c1) + CI(c2) - (2 \cdot CI(ppg(c1, c2)))}$$

La mesure de similarité devient donc : $\mathbf{Sim(c1, c2) = 1/distance(c1, c2)}$

5.4.3.2.3 La mesure de Hirst-St.Onge

La mesure de [HIR98] prend en considération toutes les relations dans WordNet. Les liens sont classés comme *haut* (eg. partie-de), *bas* (eg. sous-classe), *horizontal* (eg. antonyme). La similarité est calculée entre mots par le poids du chemin le plus court qui mène d'un terme à un autre. Il est calculé en fonctions de ces classifications qui indiquent les changements de direction :

$$\mathbf{Sim(c1, c2) = T - \text{chemin} - K \times d \quad (5)}$$

Tels que T et K sont des constantes, *chemin* est la longueur du chemin le plus court en nombre d'arcs et d est le nombre de changements de direction. L'idée est que deux concepts sont proches sémantiquement si leurs synsets sont connectés par un chemin qui n'est pas très long et qui ne change pas souvent de direction. S'il n'y a pas de chemin, le poids est égal à zéro.

5.4.3.2.4 La mesure de Wu-Palmer

Dans un domaine de concepts, la similarité est définie par rapport à la distance qui sépare deux concepts dans la hiérarchie et également par leur position par rapport à la racine. La similarité

entre C1 et C2 est :

$$\text{ConSim}(C1, C2) = (2*N3)/(N1+N2+2*N3)$$

Plus formellement cette mesure devient :

$$\text{ConSim}(C1, C2) = (2*\text{depth}(C))/(\text{depth}_c(C1)+\text{depth}_c(C2))$$

Où C est le PPG de C1 et C2 (en nombre d'arcs), depth(C) est le nombre d'arcs qui sépare C de la racine et depth_c(C_i) avec i le nombre d'arcs qui séparent C_i de la racine en passant par C.

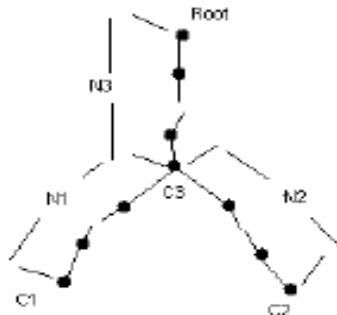


Figure 54: Les relations conceptuelles [WU94]

Dans notre travail nous faisons appel à cette dernière mesure. Elle a l'avantage d'être simple à implémenter et d'avoir d'aussi bonnes performances que les autres mesures de similarité [Lin98].

5.4.4 Mesure sémantique utilisée dans le filtre de profil

Le filtre de profil représente l'un des filtres utilisés dans notre solution. Il consiste à calculer la similarité entre les mots clés de la requête et ceux de l'annonce. Il est basé sur les modèles de recherche d'information dont le rôle est de fournir une formalisation du processus de recherche d'information. Il doit accomplir plusieurs rôles dont le plus important est de fournir un cadre théorique pour la modélisation de cette mesure de pertinence. De façon générale, les modèles de RI peuvent être classés en trois principales classes ou modèles qui sont [BAZ05]:

- les modèles booléens : où l'on peut distinguer le modèle booléen pur (boolean model), le modèle booléen étendu (extended boolean model) et le modèle basé sur les ensembles flous (fuzzy set model).
- les modèles vectoriels : ils englobent le modèle vectoriel (vector model), le modèle vectoriel généralisé (generalized vector model), Latent Semantic Indexing (LSI) et le modèle connexionniste.
- les modèles probabilistes : ces modèles comprennent le modèle probabiliste général, le modèle de réseau de document ou d'inférence (Document Network) et le modèle de langages.

Nous constatons que le modèle booléen ne permet pas de calculer une valeur de score et les autres modèles imposent que chaque mot doit avoir un poids.

la similarité des deux documents D et C . Chaque document est définis par une liste de mots clés et chaque mot a un poids tel que

$D = \{(d1,W1),(d2,W2), \dots, (dn,Wn)\}$, $C = \{(c1,W1),(c2,W2), \dots, (cm,Wm)\}$, la similarité des deux documents D et C est calculée à travers la formule suivante [HAR10] :

$$\text{Sim}(D,C) = \frac{\sum_{i=1}^N \sum_{j=1}^M (W_i * W_j * \text{Sim}(d_i,c_j))}{N * M}$$

Dans le cas ou chaque mot est pondéré nous faisons appel au modèle vectoriel [BAZ05]. Dans ce modèle, les requêtes et les documents sont représentés dans l'espace vectoriel engendré par les termes d'indexation [SAL83]. L'espace est de dimension N (N étant le nombre de termes d'indexation de la collection de documents).

Chaque document est représenté par un vecteur $D_j = (d1j, d2j, d3j, \dots, dNj)$,

Chaque requête est représentée par un vecteur $Q = (q1, q2, q3, \dots, qN)$,

avec dij : poids du terme ti dans le document Dj , qi : poids du terme ti dans la requête Q .

La pondération des composantes de la requête est soit la même que celle utilisée pour les documents, soit donnée par l'utilisateur lors de sa formulation.

Le mécanisme de recherche consiste à retrouver les vecteurs documents qui s'approchent le plus du vecteur requête.

5.4.5 Algorithme de « matching »

L'objectif de notre algorithme de « matching » est de présenter une liste triée des SWG appropriés. Pour cela, nous supposons que la structure de la requête et la structure de l'annonce sont similaires.

5.4.5.1 Algorithme de « matching » au niveau catalogue des catégories des SWG

Une catégorie de SWG (CSWG) est présentée par la catégorie, dimensions de qualité et la liste des mots clés : CSWG (Categorie, DQc(Nom,P) , Cc) ou

- Catégorie : représente l'un de type de SWG suivants : Affichage, Analyse, Abstraction, Archivage, Acquisition.
- DQc : représente un critère qualité qui définie par un nom (D) et un poids (P)
- Cc: représente la liste des concepts de la catégorie ou chaque concept a un nom (C) et un poids (P)

Une requête utilisateur est présentée par la catégorie, dimensions de qualité et la liste des mots clés : $R(\text{Catégorie}, \text{DQu}(\text{Nom}, \text{P}, \text{S}), \text{Cu})$ ou

- Catégorie : représente l'un de type de SWG suivants : Affichage, Analyse, Abstraction, Archivage, Acquisition.
- DQu: représente un critère qualité qui définie par un Nom (D), un poids (P) et seuil (S) qui permet d'indiquer que le service vérifie cette dimension de qualité ou non.
- Cu: représente la liste des concepts de la requête ou chaque concept a un nom (C) et un poids (P)

Notre algorithme est présenté comme suit :

1. Déterminer les critères de qualité de la requête utilisateur DQu et les mots clés Cu ou

$\text{DQu} = \{ (D_i, P_i, S_i) \text{ tel que } i=1..M \}$

$\text{Cu} = \{ (C_{ui}, P_{ui}) \text{ tel que } i=1..K \}$

2. Filtrage des catégories des SWG selon les dimensions de qualité :

Pour chaque catégorie

nous avons les dimensions de qualité de la catégorie DQc ou

$\text{DQc} = \{ (D_j, P_j) \text{ tel que } j=1..N \}$

DA: représenté les concepts acceptés

DA = { };

Pour chaque C_j de la catégorie

S'il existe un D_i de la requête correspond à D_j alors

Si $P_j \geq S_i$ alors insérer D_j à DA finsi

Sinon Trouver D_f telle que $\text{Sim}(D_j, D_f)$ est max

Si $P_j \geq S_f$ alors insérer D_j à DA finsi

Finsi

Finpour

{ Dans la phase de filtrage multidimensionnel, le calcul de similarité est basé sur l'ontologie de la qualité des SWG }

Calcul de la qualité globale de la catégorie Qcat en utilisant la méthode SAW en fonction des dimensions de qualité acceptées

Finpour

3. Classement selon les critères de qualités : consiste à classer les catégories des SWG suivant leur qualité globale Qcat.

4. Filtrage selon le profil :

Pour chaque catégorie retenue

Pour chaque concept Cui de la catégorie

trouver le concept Ccj de la catégorie correspond. C'est à dire trouver le concept Ccj dont la similarité $Sim(Cui, Ccj)$ est maximale

finpour

calculer le score de la catégorie Scat en utilisant la formule présentée dans 5.4.4

Finpour

{Dans la phase de filtrage selon le profil, le calcul de similarité est basé sur l'ontologie de domaine}

5. Classement selon les deux filtrages en utilisant la formule $Qcat * Scat$.

Figure 55 : Algorithme de « matching » au niveau catalogue des catégories des SWG

5.4.5.2 Algorithme de « matching » niveau registre UDDI

Initialement, L'Algorithme de « matching » utilisé au niveau registre UDDI permet l'interrogation de la partie spécifiée dans l'algorithme de « matching » niveau catalogue, de registre UDDI. Il effectue une recherche par mot-clé, par tModel et donc, par voie de conséquence, une recherche par spécification (ex : classification, service utilisant WSDL,...).

Il est basé sur la comparaison de l'information disponible à la requête utilisateur à celle fournie par le registre UDDI en utilisant les pages suivantes :

- les pages blanches : liste alphabétique des différentes compagnies, d'un numéro de contact ainsi que des différents services proposés;
- les pages jaunes : classification des services et des compagnies suivant une taxonomie standardisée ou définie par l'utilisateur;
- les pages vertes : indication techniques pour faire appel au service, joindre le fournisseur.

Afin d'améliorer les résultats de « matching » au niveau UDDI nous proposons le renforcement de l'aspect sémantique au niveau registre UDDI en se basant sur l'extraction des différents éléments de la description de SWG à partir des fichiers WSDL [LOP08] et faire un calcul de similarité sémantiques entre les éléments de la requête et ceux de l'annonce ce qui me permet d'avoir une classification des SWG proposés comme SWG pertinents.

5.5 Conclusion

Le travail présenté dans ce chapitre représente la continuité de celui proposé dans le chapitre précédent en :

- ajoutant des critères de qualité à la description du SWG et à la requête utilisateur d'une part et le détail du modèle de description des différents éléments de notre approche à l'image des catégories des SWG, les SWG, le domaine d'application et le profil utilisateur.

- Appliquant l'algorithme de « matching » LARKS au niveau catalogue de métadonnées, ce qui permet d'assurer un « matching » efficace à travers l'application de ces différents filtres.

Donc dans l'objectif de rendre la découverte des SWG plus pertinente, nous avons proposé une phase de publication dont le but est d'enrichir la description des SWG. Cette dernière est effectuée à deux niveaux de description. Le premier niveau concerne le catalogue des classes des SWG en se basant sur la norme ISO19119 et les critères de qualité, le deuxième niveau concerne le registre UDDI à travers le WSDL.

Cet enrichissement est consolidé à travers l'utilisation d'un ensemble d'information à propos de la catégorie du SWG en se basant sur la norme ISO 19119 et les critères de qualité, le service lui-même à travers l'exploitation du WSDL, le domaine d'application dont le SWG peut être utilisé et le profil utilisateur basé sur les travaux de D Kostadinov [KOS07].

Dans la phase de découverte, cet enrichissement permet de restreindre l'espace de recherche et d'augmenter le nombre de services pertinents à travers l'application de l'algorithme de « matching » LARKS au niveau métadonnées. Cet algorithme sera adapté en appliquant les trois premiers filtres (contexte, profil, similarité) sur les critères de qualités figurant dans la requête et les descriptions des classes de SWG disponibles dans le catalogue de métadonnées. A travers cet algorithme nous avons fait appel au domaine de la recherche d'informations où nous avons utilisé la méthode SAW pour l'évaluation de la qualité globale et l'algorithme de mesure de similarité de Wu_Polmer pour le calcul de similarité sémantique entre deux concepts dans une ontologie et une mesure sémantique basé sur la similarité sémantique pour le filtre de profil. Cette solution nous permet d'avoir un résultat de type «relaxed match». Le résultat de ce premier « matching » permet de restreindre l'espace de recherche au niveau du registre UDDI ce qui permet d'avoir de bons résultats pour le « matching » basé sur l'aspect fonctionnel des SWG en matière temps.

Pour la mise en œuvre de notre approche, un prototype de découverte sémantique, à base de qualité, dont le champ d'application est le domaine de tourisme, fait l'objet du prochain chapitre.

***Chapitre 6: Prototype de la découverte
sémantique des services Web
géographiques à base de qualité***

Chapitre 6 : Prototype de découverte sémantique des SWG à base de qualité

A travers le quatrième et le cinquième chapitre, nous avons présenté deux approches de découverte sémantique des services Web géographique (SWG). La première approche est basée sur les métadonnées exploitants la norme ISO 19119 à travers un système multi-agents. La deuxième représente une approche optimisée de la première. Elle est basée sur l'utilisation des métadonnées basées sur la norme ISO 19119 et les critères de qualités des SWG. Elle est concrétisée à travers deux modèles, le premier concerne les SWG et sera adapté aux classes de SWG dans l'objectif de la construction du catalogue de métadonnées. Le deuxième concerne le profil de l'utilisateur qui sera utilisé pour enrichir et compléter les requêtes de l'utilisateur. La correspondance entre les requêtes utilisateurs et les annonces des SWG publiés est assurée à travers l'utilisation d'un algorithme de « matching » inspiré du travaux du « LARKS » en se basant sur une ontologie de qualité des SWG au niveau catalogue des métadonnées, et un algorithme de « matching » basé sur l'interrogation des pages (blanches, jaunes et vertes) en se basant sur une ontologie de domaine (dans notre cas nous utilisons une ontologie de tourisme).

Pour la validation de notre deuxième approche nous avons opté à réaliser un prototype de l'approche optimisée en faisant appel à un ensemble d'outils (Protégé, XML/XML Schemas, PHP,...), en prenant comme domaine d'application le domaine du tourisme. Ce prototype est une application Web en adaptant une architecture Web classique “n-tiers”.

Ce chapitre est organisé comme suit. La première section présente un survol des différents outils exploités dans notre mise en œuvre. La deuxième section décrit les deux ontologies exploitées dans notre prototype. , la description du prototype en détail fait l'objet de la troisième section. Elle consiste à présenter :

- l'architecture générale de l'application ;
- la structure du catalogue, ainsi que ses rôles et ses fonctionnalités ;
- le sous système de profil utilisateur en présentant les interfaces utilisées et les fonctionnalités assurées;
- le sous système de publication en indiquant les interfaces de description des SWG et leurs catégories ainsi que le fonctionnement de ce sous système;
- le sous système de découverte des SWG en présentant l'interface de la requête utilisateur ainsi que les différentes phases de « matching ».

Enfin une quatrième section est réservée à la conclusion du travail réalisé.

6.1 Outils de développement

Pour mettre en œuvre notre modèle de découverte sémantique à base de qualité nous faisons appel à un ensemble d'outils décrits comme suit :

6.1.1 XML/XML Schémas :

Le langage XML (eXtensible Markup Language) connaît depuis son apparition en 1996 un succès considérable. Recommandation W3C en 1998, XML est devenu un standard pour les documents semi structurés. XML est un sous-ensemble de SGML. Un document XML est structuré en éléments dont les balises marquent le début et la fin de chacun. Les éléments peuvent contenir du texte et éventuellement d'autres éléments. L'ensemble des données du document XML est contenu dans un élément unique appelé racine qui contient tous les autres éléments.

XML est un méta-langage servant de base aux langages qui définissent des vocabulaires et des grammaires qui leur sont propres. Ainsi, par exemple, le langage XHTML définit les éléments légaux (head, title, body, etc.) et la façon dont ils peuvent s'agencer. Le langage HTML, qui est avec XML l'autre très célèbre descendant de SGML, offre une plus grande liberté dans l'utilisation des balises définies, ce qui rend certains documents HTML mal formés au regard de la syntaxe XML.

Le langage XML Schema est une recommandation du W3C depuis 1998. Les schémas XML sont eux-mêmes des documents XML. Ils permettent de définir la structure d'un document XML, de vérifier la validité de ce document. Une instance d'un XML Schema est un peu l'équivalent d'une définition de type de document (DTD) avec un peu de différence.

Nous optons à utiliser XML et XML Schema pour les raisons suivantes [YAN96]:

- XML permet l'utilisation d'espaces de nommage (namespace) ;
- XML n'est lié à aucun système d'exploitation particulier. Les documents XML sont de simples fichiers texte lisibles pour un lecteur humain ;
- XML Schema est un langage bien adapté pour traduire notre modèle conceptuel ;
- XML Schema intègre le principe d'héritage.
- XML Schema permet l'expression de contraintes fines sur le contenu des éléments.

6.1.2 OWL

OWL (Web Ontology Language) est le langage standard pour la représentation d'ontologies, proposé par le consortium W3C. OWL utilise les primitives de base modélisées par les schémas RDF. OWL permet de définir des classes qui représentent des ensembles d'individus, et des propriétés qui représentent les caractéristiques des individus ou les relations possibles entre ces derniers. Les connaissances sont prises en compte selon deux niveaux d'abstraction : la représentation des concepts et des relations relève du niveau *terminologique*, tandis que la représentation des individus et les assertions dans lesquelles ils interviennent relèvent du niveau assertionnel [MIR09].

Pour répondre à des besoins de représentation et d'inférence variés, exprimés à travers

différentes applications, le consortium W3C propose trois sous langages : OWL Lite, OWL DL et OWL Full, énumérés ici par ordre décroissant d'expressivité [MIR09] comme suit :

- OWL Full: est la version la plus complexe de OWL, celle qui offre le plus haut niveau d'expressivité. Elle autorise l'utilisation de toutes les primitives OWL combinées d'une façon arbitraire aux structures RDF et RDF Schema.
- OWL DL: il s'adresse aux utilisateurs qui souhaitent une expressivité maximale sans perdre la complétude et la décidabilité du calcul de subsomption.
- OWL Lite: est le langage le plus simple (il est moins expressif que les deux autres), il est décidable et complet. Il s'adresse aux utilisateurs qui ont principalement besoin de hiérarchies de classification et de constructeurs élémentaires.

De nombreux outils pour la construction des ontologies ont été développés ces dernières années. Les plus largement utilisés par la communauté ingénierie des ontologies sont cités dans [MIR09] à l'image d'OILEd, KAON, Jena et Protégé.

Pour l'inférence, il existe plusieurs moteurs d'inférence, la plupart conçus pour raisonner sur les Logiques de Description, mais qui acceptent en entrée des fichiers OWL/RDF(S). Parmi ceux-ci, nous citons RacerPro, Pellet, Fact , Fact++, Surnia, F-OWL, Hoolet, etc. Certains moteurs d'inférence ne peuvent raisonner qu'au niveau terminologique, alors que des moteurs comme Pellet et RacerPro permettent de raisonner aussi sur les instances de concepts [MIR09].

6.1.2.1 Protégé :

Parmi les outils de construction des ontologies cités dans [MIR09], nous utilisons Protégé qui représente une plate-forme de développement d'ontologies et d'acquisition de connaissances développée par Stanford Medical Informatics, Standford University School of Medicine. Logiciel open-source, Protégé est muni d'une interface graphique qui permet aux développeurs d'ontologies de se concentrer sur leur tâche de modélisation conceptuelle, sans connaître, ni penser à la syntaxe d'un langage de représentation.

Le modèle de connaissances de Protégé est flexible et son architecture est facilement extensible à travers des plugins. [MIR09]

La plate-forme Protégé offre deux possibilités de modélisation d'ontologies :

- en utilisant l'éditeur Protégé-Frames, qui permet à l'utilisateur de créer et peupler des ontologies définies selon le modèle de frames, et conformes au protocole Open Knowledge Base Connectivity (OKBC).
- à travers l'éditeur Protégé-OWL. Comme son nom l'indique, Protégé-OWL est une extension de l'éditeur d'ontologie Protégé pour manipuler le format OWL. Il permet

notamment de visualiser, d'éditer classes et propriétés OWL, de communiquer avec des raisonneurs logiques ou de peupler l'ontologie d'instances trouvées dans des documents. Visuellement, le logiciel se présente sous la forme de plusieurs onglets permettant l'accès à différents types d'information : nous pouvons citer, entre autres, les onglets Classes, Properties (avec une subdivision supplémentaire pour chaque type de propriété OWL) et Individuals.



Figure 56 : interface du logiciel protégé 3.4

6.1.3 PHP

PHP (officiellement "PHP: Hypertext Preprocessor") est un langage de script qui est principalement utilisé pour être exécuté par un serveur HTTP, mais il peut fonctionner comme n'importe quel langage interprété en utilisant les scripts et son interpréteur sur un ordinateur. Il permet aux développeurs Web d'écrire des pages dynamiques rapidement.

PHP n'est pas un langage compilé, c'est un langage interprété par le serveur : le serveur lit le code PHP, le transforme et génère la page HTML. Pour fonctionner, il a donc besoin d'un serveur Web. De ce fait une plateforme minimale de base pour l'exécution d'un site Web développé en PHP comprend [PHP02] :

- l'interpréteur PHP (serveur PHP)
- un serveur Web (Apache, IIS, ...)

Le langage PHP possède les mêmes fonctionnalités que les autres langages permettant d'écrire des scripts CGI, comme collecter des données, générer dynamiquement des pages Web ou bien envoyer et recevoir des cookies,

Il a les avantages suivants [PHP02] :

- la plus grande qualité et le plus important avantage du langage PHP est le support d'un grand nombre de bases de données et la simplicité d'interfaçage avec eux.
- PHP est utilisable sur la majorité des systèmes d'exploitation, comme Linux, de nombreuses variantes Unix (incluant HP-UX, Solaris et OpenBSD), Microsoft Windows, Mac OS X, RISC OS et d'autres encore.
- PHP supporte aussi la plupart des serveurs Web actuels : Apache, Microsoft Internet Information Server, Personal Web Server, Netscape et iPlanet servers, Oreilly Website Pro server, Caudium, Xitami, OmniHTTPd et beaucoup d'autres encore.
- la gratuité et la disponibilité du code source.
- la simplicité d'écriture de scripts (apprentissage rapide)

Malgré ses avantages il a les limitations suivantes [PHP02] :

- Langage interprété ;
- L'orienté objet reste limité ;
- Pas adéquat en termes de rapidité et de maintenabilité pour des projets de grandes envergures

6.1.4 MySQL

MySQL dérive directement de SQL (Structured Query Language) qui est un langage de requête vers les bases de données exploitant le modèle relationnel. MySQL est très souvent utilisé avec le langage de création de pages Web dynamiques : PHP. Il vient d'ajouter une couche supplémentaire de stockage des données qui est plus commode, rapide et puissante d'utilisation : le serveur de base de données [MYS03]

Lorsque le serveur reçoit une demande d'un client de consultation d'une page en PHP qui fait appel à des données stockées sous MySQL:

1. Le serveur Web envoie le nom de la page PHP demandée à l'interpréteur PHP.
2. PHP exécute le script existant dans la page. Si des instructions relatives à la connexion à une base de données trouvées, PHP se charge d'envoyer les requêtes d'exécution à MySQL.
3. MySQL exécute la requête et renvoie à PHP le jeu de données résultat.
4. PHP termine son traitement et renvoie la page HTML générée au serveur Web qui la transmet à l'internaute.

6.2 Ontologies exploités

Pour la mise en œuvre de notre prototype nous faisons appel à deux ontologies, la première est une ontologie générale qui consiste à présenter les différents concepts de la qualité des Services Web Géographique. La deuxième concerne le domaine d'application qui est dans notre cas le domaine de tourisme. L'utilisation de ces deux ontologies nous permet de calculer la similarité

sémantique entre les différents concepts en utilisant la mesure de calcul de Wu-Palmer (5.4.3.2.4)

6.2.1 Ontologie de la qualité des SWG

La qualité des catégories des SWG est représentée sous forme ontologique à travers un ensemble de classes et de sous classes. Cette ontologie permet d’assurer une bonne couverture de la requête utilisateur recherchée. Dans le cadre général la qualité des catégories des SWG est composée de trois sous classes à savoir :

1. La qualité des données : cette sous classe contient plusieurs sous classes, elle est spécifiés selon la norme ISO 19113.
2. La qualité des services : cette dernière est composé de plusieurs sous classes suivant les taxonomies indiqués dans le chapitre 2.
3. La qualité d’usage : cette sous classe contient les sous classes spécifiant les dimensions de qualité dont l’objectif est de décrire l’environnement désiré par l’utilisateur.

Dans notre mise en œuvre nous utilisons une partie de cette ontologie générale, représentée un graphe comme indique dans la figure 57 et par un fichier qualite.OWL.

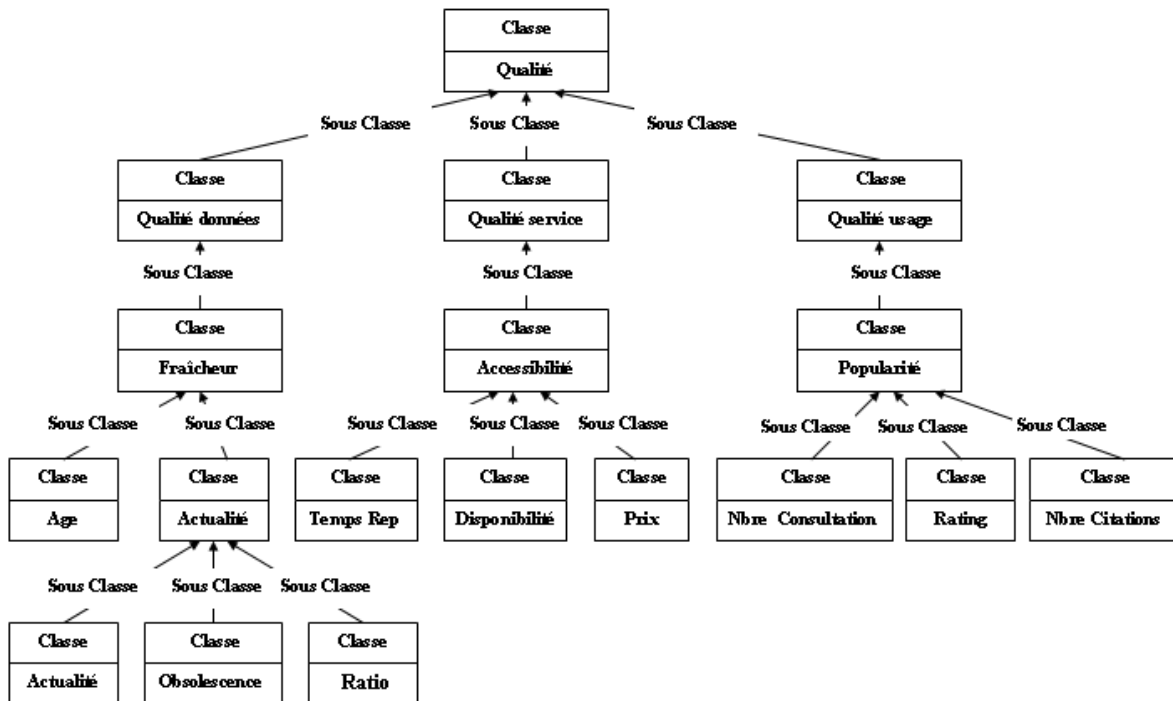


Figure 57 : graphe d’une partie de l’ontologie qualite.OWL


```

<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY swrl "http://www.w3.org/2003/11/swrl#" >
  <!ENTITY swrlb "http://www.w3.org/2003/11/swrlb#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <!ENTITY protege "http://protege.stanford.edu/plugins/owl/protege#" >
  <!ENTITY xsp "http://www.owl-ontologies.com/2005/08/07/xsp.owl#" >
]>
<rdf:RDF xmlns="http://www.owl-ontologies.com/Qualite.owl#"
  xml:base="http://www.owl-ontologies.com/Qualite.owl"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="accessibilite">
    <rdfs:subClassOf rdf:resource="#qualite_service"/>
  </owl:Class>
  <owl:Class rdf:ID="actualite">
    <rdfs:subClassOf rdf:resource="#fraicheur"/>
  </owl:Class>
  <owl:Class rdf:ID="age">
    <rdfs:subClassOf rdf:resource="#fraicheur"/>
  </owl:Class>
  <owl:Class rdf:ID="disponibilite">
    <rdfs:subClassOf rdf:resource="#accessibilite"/>
  </owl:Class>
  <owl:Class rdf:ID="fraicheur">
    <rdfs:subClassOf rdf:resource="#Qualite_donnees"/>
  </owl:Class>
  <owl:Class rdf:ID="nb_citation">
    <rdfs:subClassOf rdf:resource="#popularite"/>
  </owl:Class>
  <owl:Class rdf:ID="nb_occurrences">
    <rdfs:subClassOf rdf:resource="#popularite"/>
  </owl:Class>
  <owl:Class rdf:ID="popularite">
    <rdfs:subClassOf rdf:resource="#qualite_usage"/>
  </owl:Class>
  <owl:Class rdf:ID="prix">
    <rdfs:subClassOf rdf:resource="#accessibilite"/>
  </owl:Class>
  <owl:Class rdf:ID="Qualite_donnees"/>
  <owl:Class rdf:ID="qualite_service"/>
  <owl:Class rdf:ID="qualite_usage"/>
  <owl:Class rdf:ID="Ratting">
    <rdfs:subClassOf rdf:resource="#popularite"/>
  </owl:Class>
  <owl:Class rdf:ID="temps_reponse">
    <rdfs:subClassOf rdf:resource="#accessibilite"/>
  </owl:Class>
</rdf:RDF>

```

Figure 58 : La description OWL de l'ontologie qualité SWG Qualite.OWL

6.2.2 Ontologie du domaine du tourisme

La deuxième ontologie exploitée dans notre mise en œuvre est celle du domaine de tourisme. Cette dernière est créée à travers le logiciel protégé comme indique dans la figure 59.

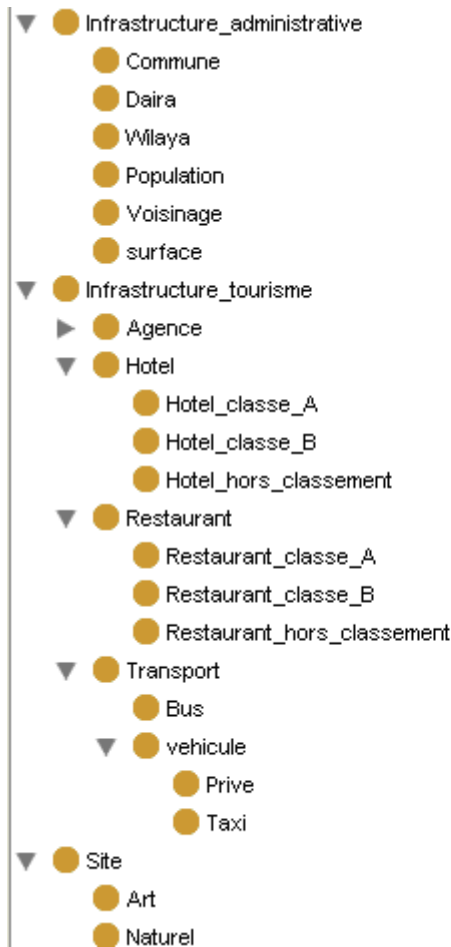


Figure 59 : Une prise d'écran de l'ontologie TOURISME à travers protégé 3.4.4

6.3 Description du prototype

La mise en œuvre de notre approche est assurée par un prototype QGWSSD (Quality Geospatial Web Service Semantic Discovery). Ce prototype exploite un ensemble d'outils (Protégé, XML/XML Schemas, PHP,...) et prend comme domaine d'application le domaine du tourisme. QGWSSD est une application Web adaptée d'une architecture Web classique "n-tiers". Il est basé sur l'exploitation des métadonnées et l'algorithme de «matching». Pour rendre ces métadonnées persistantes, nous avons choisi le couple XML/XML Schema, basé sur le principe des langages à balises de la famille SGML. Un ensemble d'interfaces est mis à la disposition de l'utilisateur (Fournisseur/client / Utilisateur), ces interfaces sont réalisées en utilisant le PHP qui assure une simplicité d'interfaçage avec des bases de données.

6.3.1 Architecture de l'application

Pour le développement de notre prototype. Nous adoptons une architecture Web classique "n-tiers":

A- Utilisateur : l'utilisateur dispose d'un ordinateur équipé d'un navigateur Web standard et connecté à l'internet. IL accède aux pages HTML via le protocole HTTP. Il peut être :

- 1 **fournisseur:** il fait la saisie de la description du SWG (Niveau classe et niveau service),
- 2 **Client :** il soumet de requêtes et transmet des données au serveur Web, et fait la saisie du profil utilisateur

B- Le serveur Web reçoit les requêtes de l'utilisateur. Il les reformule et les transmet au serveur d'application. Le serveur d'application renvoie la réponse au format XML. Le serveur Web effectue la mise en forme en transformant le XML en HTML qui est alors envoyé au client;

C- Le serveur d'application effectue toutes les opérations sur les métadonnées autres que leur mise en forme : vérification de la conformité de la base de métadonnées au modèle durant la phase de publication ainsi que l'application de l'algorithme de « matching » durant la phase de découverte.

D- Les serveurs de données stockent les bases de métadonnées.

L'architecture générale de notre application est présentée dans la figure suivante :

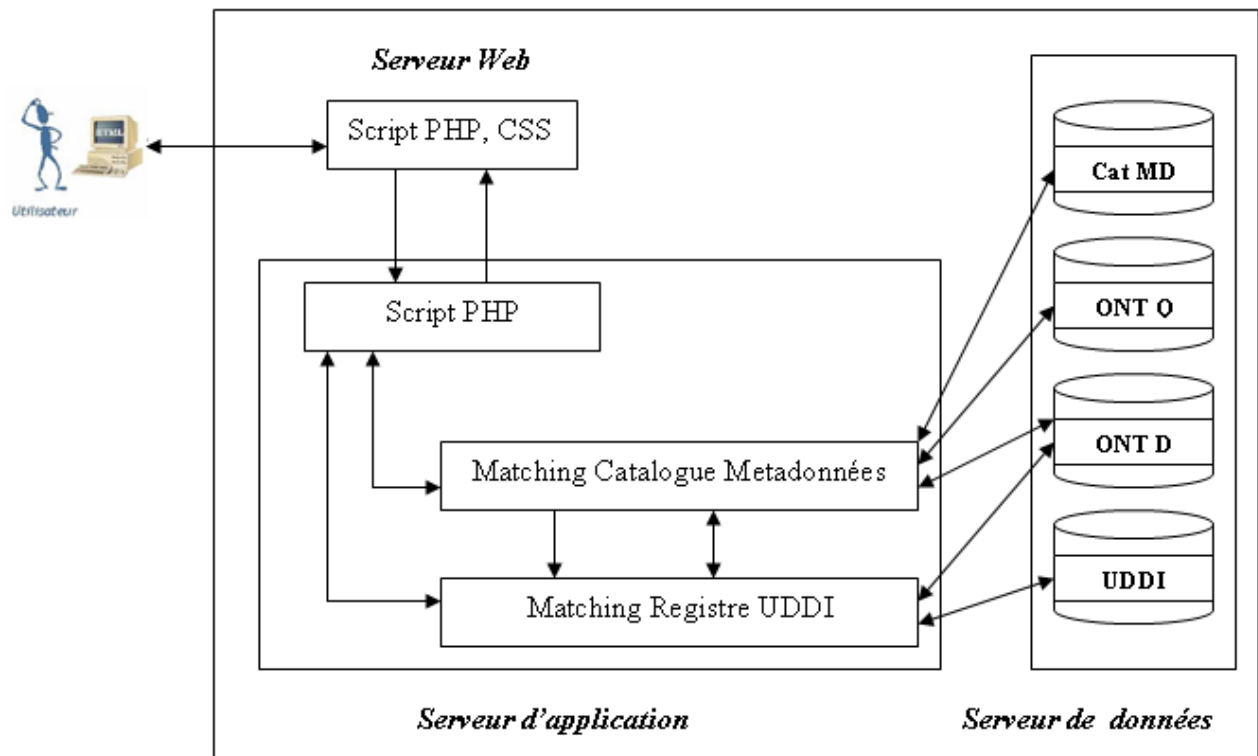


Figure 60 : Architecture générale de l'application proposée

6.3.2 Les composants du QGWSSD

Notre prototype QGWSSD est basé sur un registre de description de SWG qui défini dans le domaine du service Web comme suit [CUR02]:

- un registre **stocke la représentation** des services publiés ;
- un registre est **un moyen d'interroger** et de mettre à jour les informations qui décrivent les services.

La fonctionnalité d'un registre est donc double et concerne deux types d'utilisateur : les **fournisseurs** qui publient leurs services, et les **clients** qui interrogent les bases de représentations de services.

Dans notre cas ce registre est structuré en deux niveaux: le premier concerne les métadonnées des catégories des SWG et le deuxième concerne la description WSDL des SWG.

Notre prototype se présente à l'utilisateur (fournisseur / client) à travers des sites Web. Il est composé des sous systèmes suivants:

6.3.2.1 Le sous système de profil utilisateur

Pour la gestion du profil utilisateur nous avons besoin de deux interfaces à savoir :

6.4.2.1.1 Interface groupe utilisateurs

Elle concerne les groupes d'utilisateurs pour saisir d'une part les données de domaine d'intérêt qui est destiné à exprimer les centres d'intérêts de l'utilisateur. Dans notre cas nous spécifions le domaine d'intérêt à travers une liste de mots clés. Dans le cas général chaque mot a un poids qui représente son importance pour l'utilisateur. D'autre part, elle permet à l'utilisateur de spécifier les dimensions de qualité désirée en spécifiant la dimension de qualité, son poids et son seuil. La liste des dimensions de qualité ne doit pas contenir des dimensions de qualité en double. Pour assurer cette contrainte la saisie des dimensions de qualité doit être contrôlée en faisant appel à l'ontologie de qualité des SWG. Cette interface est présentée à travers la figure suivante :

The screenshot shows a web browser window with the URL `http://localhost/www/projet_these/Groupe_Utilisateur.html`. The page content is as follows:

- Identification du Groupe Utilisateur**:
 - Identificateur du groupe utilisateur
 - Désignation du groupe utilisateur
 - Description du groupe utilisateur
 - Enregistrer Groupe Utilisateur
- Domaine d'intérêt**:
 - Mot clé
 - Poids
 - Explication du mot de clé
 - Inserer un mot clé
- Données de qualité**:
 - Dimension qualité
 - Poids
 - Seuil
 - Description de la dimension qualité
 - Inserer dimension qualité

Figure 61 : Interface groupe utilisateurs

6.4.2.1.2 Interface des données personnelles d'utilisateur

Elle est utilisée pour saisir les données personnelles de l'utilisateur. Ces dernières représentent la partie statique du profil. Elles contiennent des informations qui décrivent l'utilisateur et ne dépendent pas du système à interroger. Ces informations sont classées selon trois catégories: (i) données personnelles, (ii) données démographiques et (iii) données des contacts

Cette interface est présentée à travers la figure suivante :



Figure 62 : Interface données personnelles utilisateur

Après cette étape, les données utilisateur seront représentées à travers le fichier XML schemas suivant :

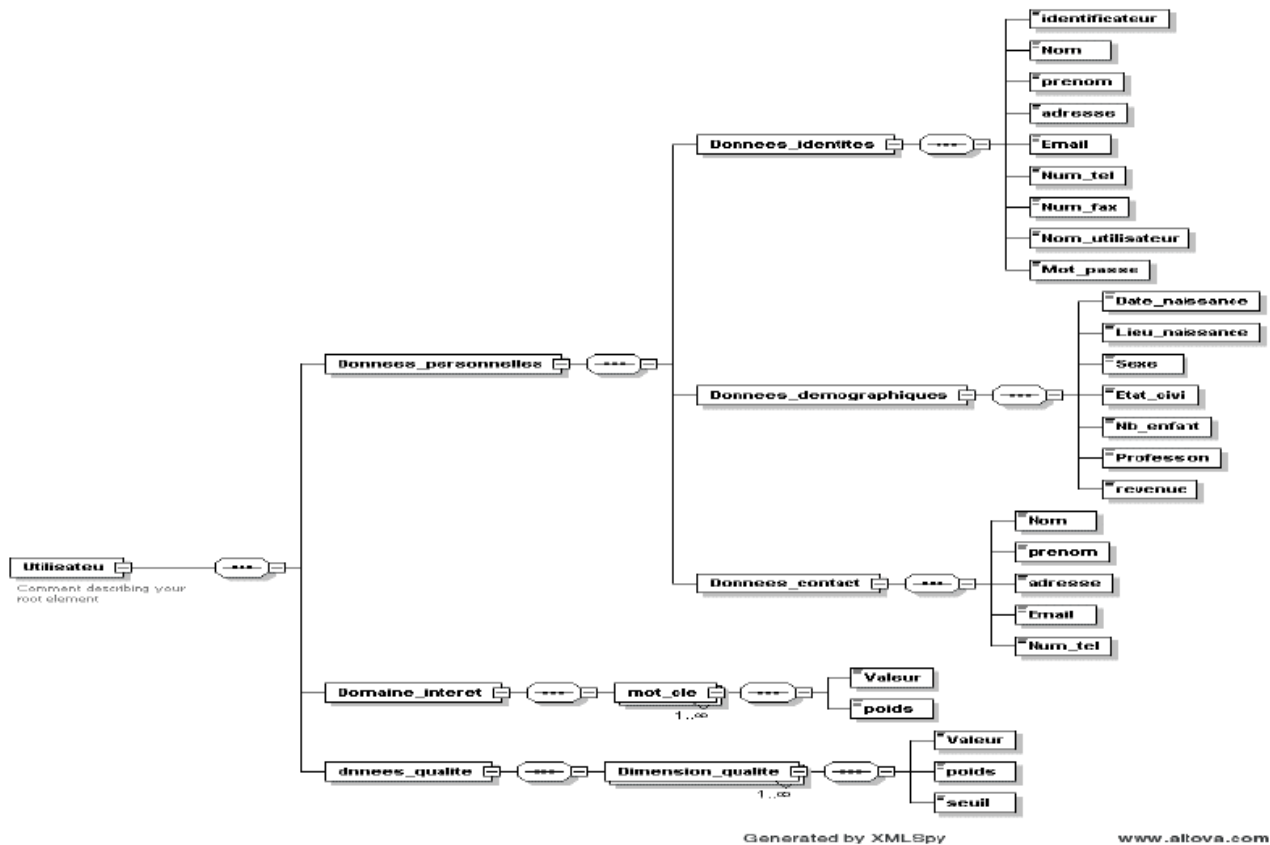


Figure 63 : Extrait du fichier XML Schema de la classe utilisateur (XML Spy 2005)

6.3.2.2 Le sous système de publication

Ce sous système consiste à assurer la mise à jour de la publication des SWG en adoptant la solution présentée dans [YAN06] qui consiste à éclater les métadonnées au format XML en plusieurs fichiers identifiés par leur id, et stockés dans les champs BLOB (Binary Large Object) d'une base de données relationnelle. Cette solution a pour objectif la conciliation des avantages d'un SGBD et du format XML. La publication des SWG consiste à enregistrer la description des catégories des SWG dans le catalogue des métadonnées et à enregistrer la description des SWG dans le registre UDDI. En plus de la publication, nous présentons d'autres fonctionnalités pour permettre aux fournisseurs et aux clients de prendre une idée de la base des SWG disponibles à travers :

- l'indexation : permet d'indexer les ressources représentées par les métadonnées.
- les statistiques : permettent de présenter les statistiques des ressources suivant différents critères par exemple : les fournisseurs, les catégories, les critères de qualités.

La phase de publication est assurée à travers deux interfaces à savoir :

6.3.2.2.1 Interface de description des catégories des SWG

Cette interface sert à introduire une entrée dans la catalogue des métadonnées. Elle est composée de trois parties (i) la description des catégories des SWG selon la norme ISO 19119, (ii) domaine d'intérêt à travers un ensemble de mots clés et leurs poids et (iii) les dimensions de qualité qualifiant la catégorie du SWG et leur poids. La liste des dimensions de qualité doit être contrôlée en faisant appel à l'ontologie de qualité des SWG (même ontologie utilisé dans la description du groupe utilisateurs). Cette interface est présentée à travers la figure suivante :

The screenshot shows a web browser window with the URL `http://localhost/www/projet_these/categorie_SWG.html`. The page content is as follows:

- Description des catégories des SWG selon la norme ISO 19119**
 - Identifiant:
 - Titre:
 - Langue du métadonnées:
 - Date de création:
 - Date de mise à jour:
 - Jeu de caractères de métadonnées:
 - Parcourir:
 - Résumé de la catégorie du SWG:
 - Catégorie:
 - Date de publication du métadonnée:
 - Niveau hiérarchique:
 - Niveau de sécurité:
 - Points de contact:
 - Type de service:
 - Parametres 1:
 - Parametres 2:
 - Enregistrer catégorie de SWG:
- Domaine d'intérêt**
 - Mot clé:
 - Poids:
 - Explication du mot de clé:
 - Inserer un mot clé:
- Données de qualité**
 - (This section is partially obscured by a grey bar at the bottom of the screenshot)

Figure 64 : Interface de publication des catégories des SWG

Le résultat de la publication des catégories des SWG peut être représenté par un fichier XML schema comme indiqué dans la figure suivante :

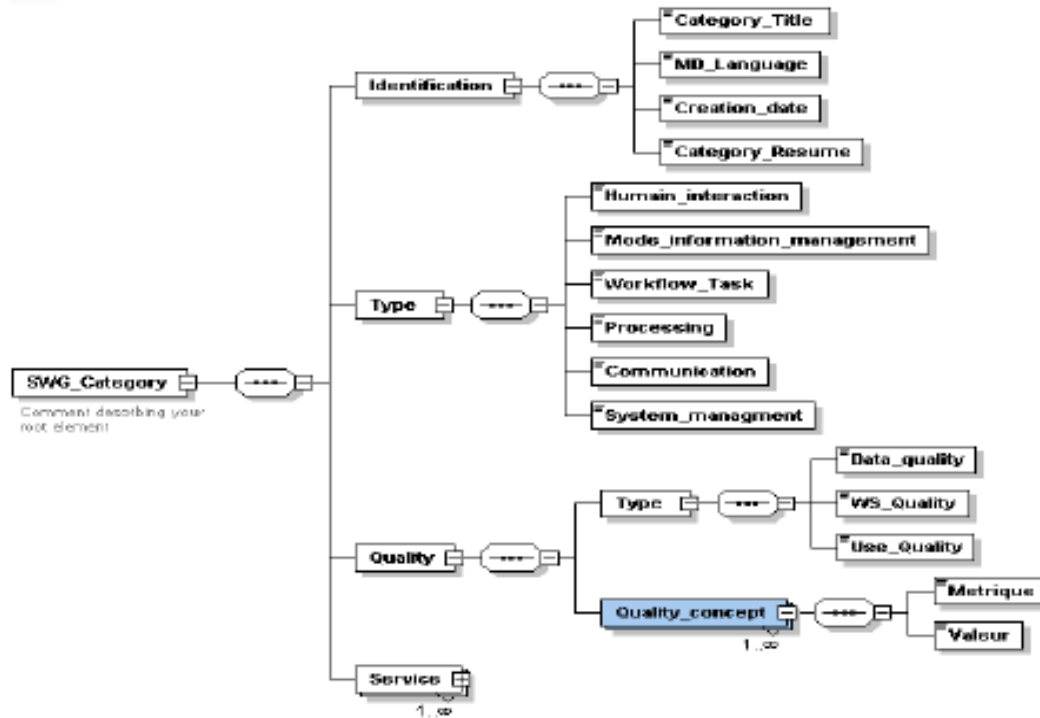


Figure 65 : Extrait du fichier XML Schema de la catégorie du SWG (XML Spy 2005)

6.3.2.2.1 Interface de description des SWG

La publication des SWG est assurée en spécifiant un fichier de description WSDL du SWG ainsi que la catégorie dont il appartient pour assurer la liaison entre les SWG et leur catégorie.



Découverte sémantique des services Web Géographiques à base de qualité

Service Web Géographique

L'interface Service Web Géographique le fichier de description WSDL du SWG ainsi que la catégorie dont il appartient pour assurer la liaison entre les SWG et leur catégorie.



Figure 66 : interface de publication des SWG

Le résultat de la publication des catégories des SWG est représenté par des fichiers XML dont chacun est enregistré au niveau de la base de données en exploitant son identificateur. Ces fichiers permettent d'accéder aux différents SWG appartenant à cette catégorie.

6.3.2.3 Le sous système de découverte

Ce sous système représente le noyau de notre travail. Il consiste à permettre à l'utilisateur de spécifier ces besoins à travers une requête. Cette dernière est composée de deux parties à savoir :

- La partie fonctionnelle : cette partie est composée de la fonction de la requête (l'une des fonctions des 5A) et ses paramètres (paramètres d'entrées et paramètres de sorties) qui seront utilisés dans la phase de « matching » niveau registre UDDI.
- La partie non fonctionnelle : cette partie doit contenir : (i) les données spécifiant le domaine d'intérêt représenté par une liste de mot clés et (ii) les données de qualité représentées par une liste de dimension de qualité et ses paramètres (poids, seuil). Les données de cette partie seront exploitées dans la « matching » niveau catalogue.

La deuxième partie de la requête sera proposée en fonction du profil utilisateur en lui donnant la possibilité de changer ses préférences d'une manière temporaire (seulement pour cette requête).

Ce sous système assure à travers un ensemble d'interface les opérations suivantes :

- la connexion de l'utilisateur à notre système ;
- la saisie et l'enrichissement de la requête utilisateur en fonction du profil utilisateur
- la recherche des SWG candidats ;
- la proposition d'une liste ordonnée de SWG pertinents.

6.3.2.3.1 Connexion d'utilisateur

Cette interface représente le premier contact de l'utilisateur avec notre système dans la phase de découverte. Elle a la forme classique d'une interface d'authentification. Elle permet de charger le profil utilisateur pour proposer la partie non fonctionnelle de la requête utilisateur. Dans ce cas (utilisateur inscrit) l'utilisateur doit compléter la partie fonctionnelle et valider la partie non fonctionnelle de la requête. Dans le cas où l'utilisateur est non inscrit, il doit créer sa propre requête qui peut être le point de départ pour la création de son profil.

Figure 67 : Interface de connexion d'utilisateur

6.3.2.3.2 Construction la requête utilisateur

Une fois la connexion de l'utilisateur est établie, nous avons une requête initiale dont la

saisie est assurée travers un langage permettant de spécifier tous les éléments nécessaires de la classe requête utilisateur mentionnée dans 5.2.2.2.4. Cette requête initiale contient initialement l'ensemble des mots clés définissant le domaine d'intérêt de l'utilisateur ainsi que les dimensions de qualité. Dans cette étape l'utilisateur doit :

- compléter la partie fonctionnelle de la requête en spécifiant la fonction demandée ainsi que ses paramètres ;
- Valider la partie non fonctionnelle de la requête : cette phase consiste à confirmer la requête par l'utilisateur, en lui proposant de mettre à jour sa requête en fonction de ses besoins actuels. cette mise à jour de la requête peut être utile à la mise à jour de profil utilisateur (actuellement n'est pas prise en compte).

Cette étape est effectuée à travers l'interface suivante :



Figure 68 : Interface de construction de requête utilisateur

La requête utilisateur peut être représentée par un fichier XML schema comme indiqué dans la figure suivante :

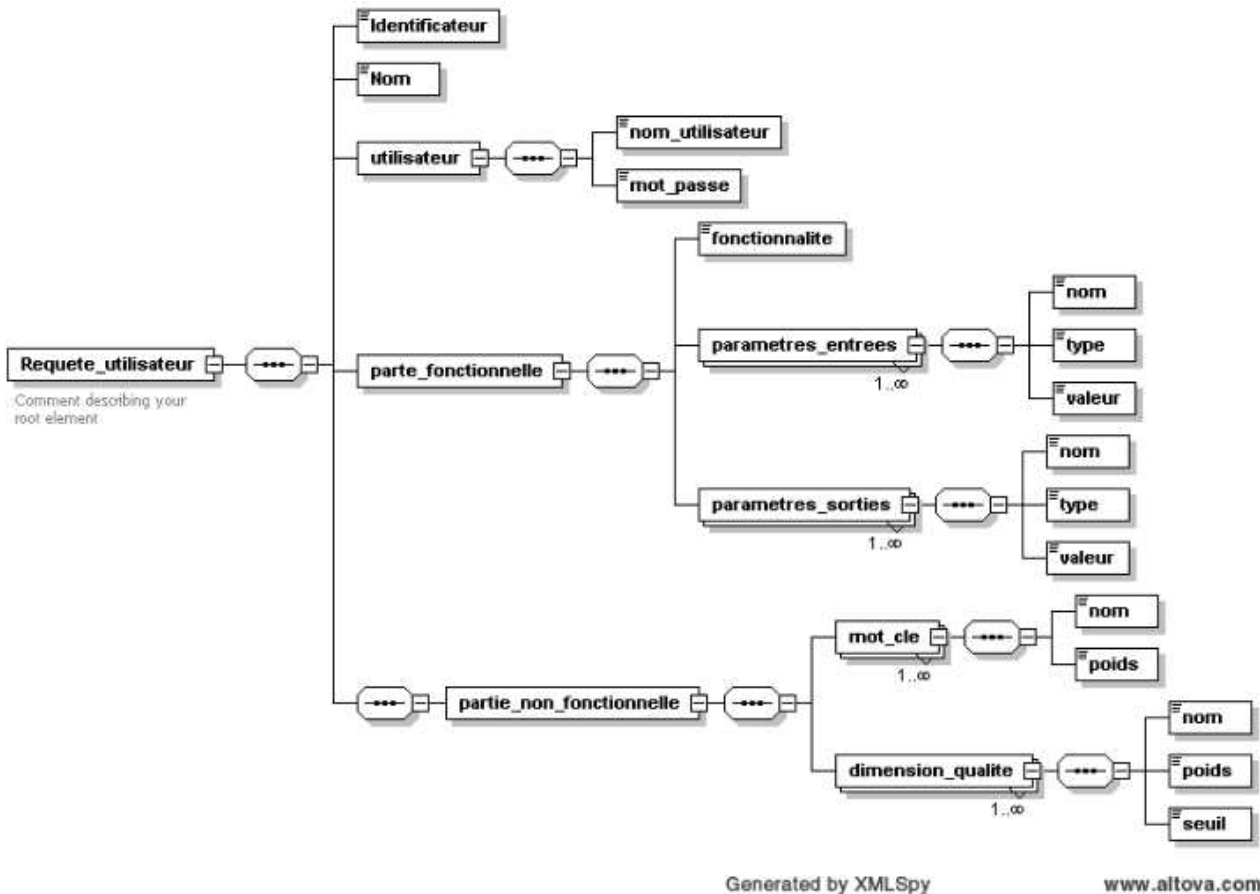


Figure 69 : Extrait du fichier XML Schema de la requête utilisateur (XML Spy 2005)

6.4.2.3.3 Recherche et classement des SWG

La phase de recherche est une phase primordiale de la découverte de SWG. Elle est composée de :

- **La recherche au niveau du catalogue de métadonnées :** cette phase consiste à appliquer l'algorithme de «matching» inespéré de l'algorithme de LARKS en appliquant le filtre de contexte en se basant sur les critères de qualités et le filtre de profil en se basant sur les mots clés. Cette étape consiste à appliquer notre algorithme de classement des catégories des SWG pertinents par ordre de mérite dont la qualité dépasse un seuil donné. Cet algorithme est effectué à travers les étapes suivantes :
 1. Déterminer les critères de qualité et les mots clés de la requête utilisateur ;
 2. Filtrage des catégories des SWG selon les dimensions de qualité (filtrage de contexte): cette étape permet de construire l'ensemble des catégories candidats pour satisfaire la requête utilisateur ;
 3. Construction et normalisation de la matrice de décision et le calcul du score de chaque catégorie : cette phase permet de calculer de la qualité globale de la catégorie en utilisant la méthode SAW (5.3.3.1) en fonction des dimensions de qualité acceptées ;

4. Classement selon les critères de qualités : consiste à classer les catégories des SWG suivant leur qualité globale.
5. Filtrage selon le profil : cette phase concerne les catégories des SWG retenues le filtrage de contexte. Elle consiste à calculer le score de la catégorie Scat en utilisant la formule présentée dans 5.4.4.
6. Classement selon les deux filtrages en utilisant la formule de la qualité globale et le score de la catégorie.

Le résultat de l'application de cet algorithme sera présenté sous forme de table précisant les catégories concernées par ordre de mérite dont la qualité dépasse un seuil donné. Ces résultats sont présentés à travers l'interface des résultats de l'algorithme de « matching » au niveau catalogue (figure 70) pour permettre à l'utilisateur de choisir une catégorie donnée. Ce choix permet de restreindre l'espace de recherche au niveau du registre UDDI.



Découverte sémantique des services Web Géographiques à base de qualité
Recherche et classement des catégories des Services Web Géographiques

Liste des catégories pertinents

Figure 70 : Interface des résultats de « matching » niveau catalogue

- **La recherche au niveau registre UDDI** : pour cette phase, initialement nous optons à appliquer l'algorithme de « matching » en se basant sur l'aspect syntaxique en utilisant l'interrogation des différentes pages (blanches, jaunes et vertes) basée sur le tmodel (en utilisant par exemple le nom du service ou bien le nom du fournisseur) ce qui nous donne comme résultat un ensemble de SWG mais sans aucune classification. Pour enrichir cette

phase de recherche nous optons à renforcer l'aspect sémantique même au niveau registre UDDI en se basant sur l'extraction des différents éléments de la description de SWG à partir des fichiers WSDL [LOP08] et faire un calcul de similarité sémantiques entre les éléments de la requête et ceux de l'annonce ce qui me permet d'avoir une classification des SWG proposés comme SWG pertinents. les résultats de la recherche syntaxique au niveau du registre UDDI sont présentés à travers l'interface des résultats de l'algorithme de «matching» au niveau UDDI (figure 71).

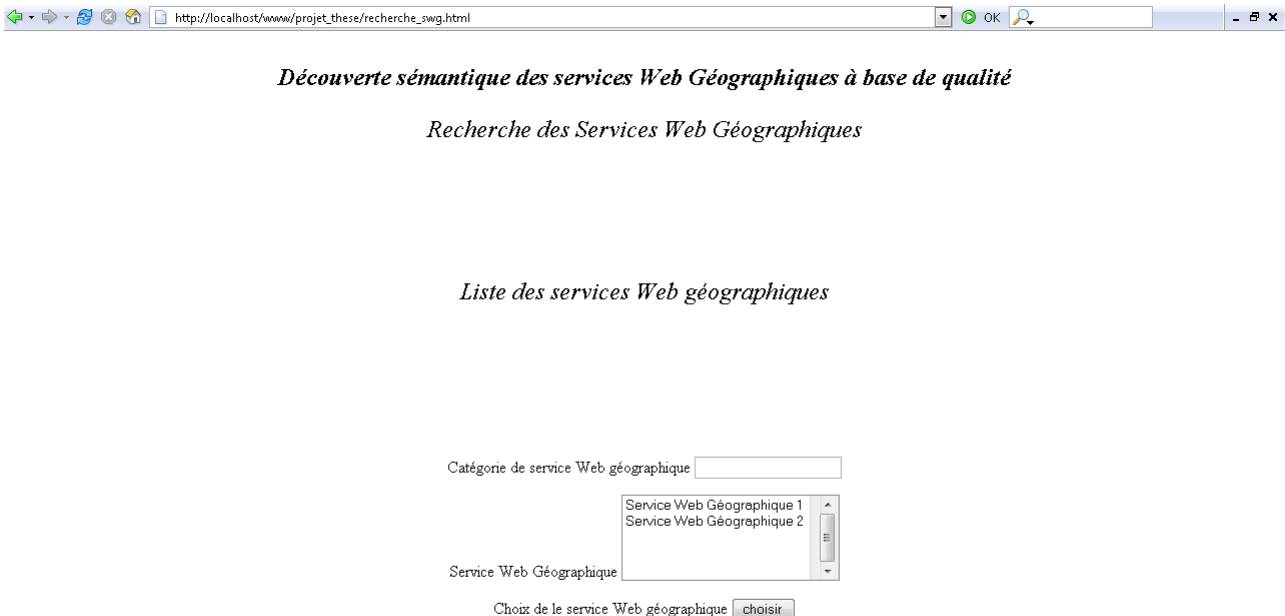


Figure 71: Interface des résultats de « matching » niveau Registre UDDI

6.3.2.3.3.1 Analyse de la description WSDL du service Web

Dans le cas de renforcement du «matching » au niveau UDDI par l'aspect sémantique nous optons le processus d'extraction des éléments de la description WSDL proposé par [LOP08] détaillé comme suit:

Le processus d'analyse de la description WSDL des services publiés dans le registre UDDI est illustré par la figure 72. Ce processus repose sur deux étapes : l'extraction et l'analyse des éléments de la description WSDL.

- **L'extraction.** L'extraction des éléments du document WSDL a été implémentée par le biais de l'API `wsdl4j` (L'API `wsdl4j` permet de faire l'extraction des éléments des fichiers WSDL issus de la version 1.1.) proposée par le *Java Community Process*⁷⁸. Cette extraction permet d'obtenir le contenu de quatre éléments ou groupe d'éléments (type, portType et operation, binding, et service) que nous devons analyser lors de la publication.
- **L'analyse.** Une fois que les éléments nécessaires à l'instanciation du modèle de

représentation de services Web sont extraits, le module d'analyse crée les instances de la Base de Connaissances du Domaine Service AROM correspondantes :

- le contenu de l'élément type permet la création des instances *InParameter* et *OutParameter*.
- le contenu de l'élément *portType* et *operation* permet la création des instances *Method*.
- le contenu de l'élément *binding* permet la création des instances *Binding*.
- le contenu de l'élément *service* permet la création des instances *WebService*.

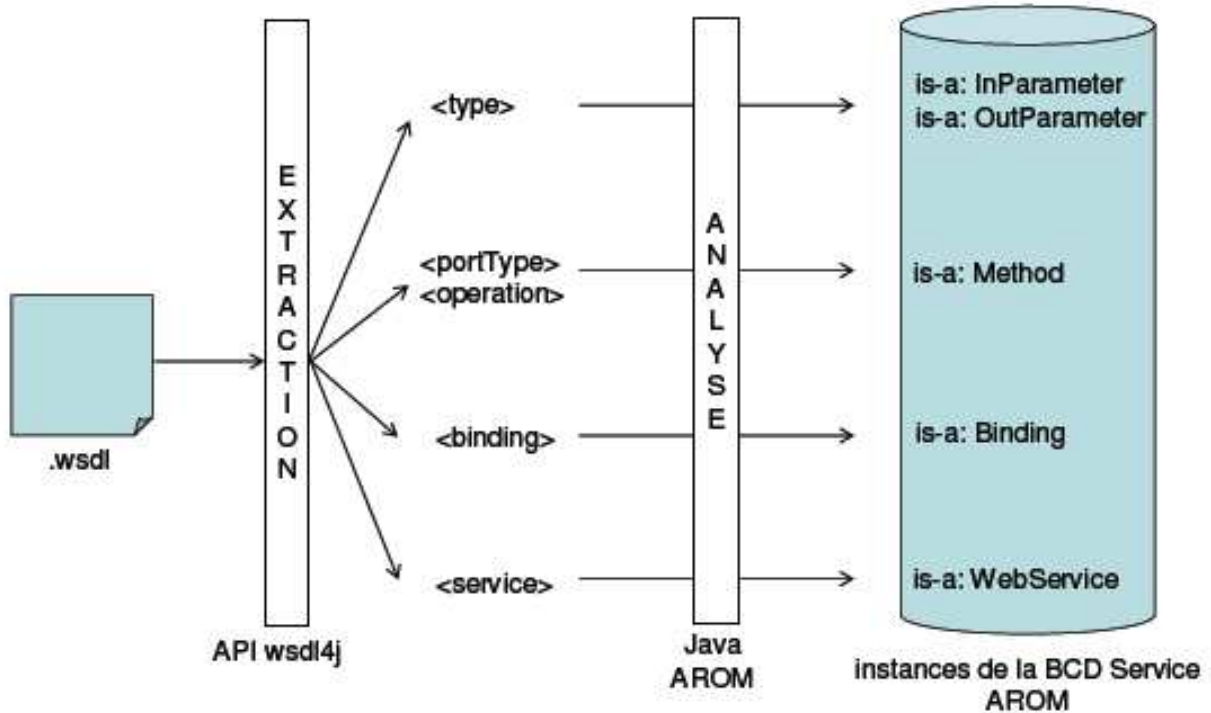


Figure 72: Processus d'extraction d'une description WSDL vers les instances d'une base de connaissances AROM selon le modèle de représentation de services Web.

Dans notre travail nous avons besoin de la phase d'extraction pour arriver à calculer la similarité sémantique entre les constituants de la partie fonctionnelle de la requête utilisateur et ceux des SWG publiés dans le registre UDDI et décrits en utilisant le WSDL.

6.4 Conclusion

Nous avons présenté notre prototype QGWSSD (Quality Geospatial Web Service Semantic Discovery) qui concrétise le modèle proposé dans le cinquième chapitre. Le prototype développé nous a permis d'expérimenter notre modèle de métadonnées représenté par le catalogue de métadonnées des catégories des SWG ainsi que la liaison avec le registre UDDI. QGWSSD assure la découverte sémantique des SWG en se basant sur la norme ISO 19119 et les critères de qualité en utilisant l'ontologie de qualité des SWG et l'ontologie de domaine de tourisme. Pour la réalisation nous avons fait appel à un ensemble d'outils permettant de manipuler les connaissances au niveau

catalogue de métadonnées, registre UDDI et le profil utilisateur sous forme de fichiers d'XML organisé en base de données relationnel.

Notre prototype permet d'une part d'enrichir la description des SWG en renforçant la description WSDL des SWG par la couche de la description de la catégorie des SWG. D'autre part rendre la découverte des SWG plus pertinente.

Notre prototype est basé sur une base de connaissances composée d'un catalogue de métadonnées et un registre d'UDDI.

Notre prototype représente une application Web qui propose aux utilisateurs humains d'utiliser les fonctionnalités de publication et de recherche. Pour le rendre plus utilisé nous proposons comme perspective ou complément de ce travail d'ajouter la partie de construction de requête effectuée par un service Web.

Conclusion générale

Conclusion générale

Les systèmes d'information géographiques actuels donnent accès à un grand nombre de sources hétérogènes et distribuées. La multiplication des ressources et l'accroissement des données nécessitent l'utilisation des services Web qui apparaissent comme solution pour assurer l'interopérabilité entre les différentes ressources. Ils sont capables de collaborer et de partager le contrôle des processus de données et d'informations entre les applications sur différentes plateformes. L'exploitation des données géographiques à travers les services Web fait apparaître les Services Web Géographiques (Geospatial Web Services). Un nombre croissant de Services Web Géographiques (SWG) est conçu pour interroger et faire interopérer des informations spatiales sur le Web. Les SWG sont en train de changer la manière dont les informations spatiales, les systèmes et les applications sont conçus, développés et déployés. Au fur et à mesure que les SWG se multiplient, la difficulté de la découverte de service s'accroît. La maîtrise de cette phase nécessite la personnalisation de la sélection des services Web géographiques désirés par l'utilisateur.

Notre thèse concerne particulièrement la proposition pour la découverte des services web géographique en se basant sur la norme ISO 19119 et les critères de qualité. Il est lié à quatre domaines en pleine évolution: les SMA, les SIG, le Web sémantique et les services Web. Ce couplage permet de profiter pleinement de la puissance des SMA dans la modélisation, de la richesse des SIG comme outils de traitement de données géographiques, de la prise en charge de l'aspect sémantique par le Web sémantique et surtout l'interopérabilité par les services Web grâce aux standards WSDL, SOAP, UDDI.

Nous avons débuté notre thèse par un état mettant en relief les principaux domaines et les concepts de base de notre recherche. Cet état de l'art s'articule sur, d'une part, les systèmes d'information géographiques et les services web géographiques, les normes (Norme ISO 19119), les services Web sémantiques, les métadonnées ; et d'autre part les systèmes multi-agents et les critères de qualités des SWG.

Pour la concrétisation de notre objectif de thèse nous avons proposé une première approche [REZ09] basée sur les métadonnées exploitant la norme ISO 19119 à travers un système multi-agents. Cette approche repose sur la coopération d'un ensemble d'agents utilisant les capacités sémantiques des services web géographiques à travers les catalogues basés sur les métadonnées selon la norme ISO19119. Les agents de notre approche intègrent les propriétés générales des agents à savoir la portabilité, l'autonomie, la stabilité, la persistance,... etc. L'idée de base de notre approche est de classer les services web géographiques selon la fonctionnalité (les 5 A) en exploitant des modèles de métadonnées basés sur la norme ISO19119.

Afin de renforcer l'approche précédente nous avons fait appel à la notion de qualité et de

personnalisation. La qualité joue un rôle primordial pour satisfaire les besoins explicites ou implicites de l'utilisateur, alors que la personnalisation facilite l'expression des besoins de l'utilisateur et rend l'information sélectionnée intelligible à l'utilisateur et exploitable. Dans ce cadre, une optimisation de l'approche précédente en intégrant la qualité et la personnalisation fait l'objet d'une deuxième proposition [REZ10]. Cette dernière est basée sur l'utilisation des métadonnées basées sur la norme ISO 19119 et les critères de qualité des SWG. Elle est concrétisée à travers deux modèles, le premier concerne les services web géographique et est adapté aux classes de services web géographiques dans l'objectif de la construction du catalogue de métadonnées. Le deuxième concerne le profil de l'utilisateur qui est utilisé pour enrichir et compléter les requêtes de l'utilisateur. La correspondance entre les requêtes utilisateurs et les annonces des SWG publiés est assurée à travers l'utilisation d'un algorithme de « matching » inspiré du travaux du « LARKS » en se basant sur une ontologie de qualité des SWG au niveau catalogue des métadonnées, et un algorithme de « matching » basé sur l'interrogation des pages (blanches, jaunes et vertes) en se basant sur une ontologie de domaine (dans notre cas nous utilisons une ontologie de tourisme).

. Notre deuxième contribution est validée à travers la mise en œuvre d'un prototype QGWSSD (Quality Geospatial Web Service Semantic Discovery). Pour la réalisation de ce prototype nous avons exploité un ensemble d'outils (Protégé, XML/XML Schemas, PHP,...) en prenant comme domaine d'application le domaine du tourisme. Ce prototype est une application Web adaptée d'une architecture Web classique «n-tiers».

La particularité de notre thèse se résume aux points suivants:

- La richesse de la description des SWG : la description des SWG est basée sur deux niveaux: le premier concerne les catégories des services en se basant sur la norme ISO19119 et les critères de qualité à travers l'utilisation de catalogue de métadonnées. Le deuxième concerne le service lui même en utilisant le WSDL. La couche sémantique dans notre cas est représentée par le catalogue de métadonnées.
- La pertinence de la découverte : la phase de découverte permet de restreindre l'espace de recherche et d'augmenter le nombre de services pertinents à travers l'application de l'algorithme de « matching » LARKS au niveau métadonnées. Cet algorithme sera adapté en appliquant les trois premiers filtres (contexte, profil, similarité) sur les critères de qualité figurant dans la requête et les descriptions des classes de SWG disponibles dans le catalogue de métadonnées. Cette solution nous a permis d'avoir un résultat de type «relaxed match». Le résultat de ce premier « matching » permet de restreindre l'espace de recherche au niveau du registre UDDI ce qui permet d'avoir de bons résultats pour le « matching » basé sur l'aspect fonctionnel des SWG en matière temps.

- L'utilisation des ontologies: deux ontologies sont utilisées dans notre thèse. La première est une ontologie de qualité de SWG, elle représente l'aspect sémantique des dimensions de qualité des SWG. Elle est utilisée au « matching » niveau catalogue de métadonnées dans le filtrage de qualité. Elle peut être exploitée quelque soit le domaine (domaine de tourisme ou autre). La deuxième concerne le domaine d'application (domaine de tourisme dans notre cas). Elle est utilisée au « matching » niveau catalogue dans le filtre de profil et peut être utilisée dans le « matching » au niveau registre UDDI.

Dans le cadre de notre objectif global qui consiste à permettre à l'utilisateur de profiter du service Web géographique le plus approprié, nous proposons d'enrichir notre modèle de découverte sémantiques en s'intéressant aux points suivants:

- l'expérimentation de notre prototype dans plusieurs domaines : cette perspective permet d'enrichir les métadonnées des catégories des SWG par d'autres dimensions de qualité. Cet enrichissement permet, d'une part, d'avoir une ontologie de qualité des SWG englobant la plupart des concepts de qualité, d'autre part de cela faciliter la découverte sémantique des SWG.
- la composition de services Web géographiques: nous optons d'une part à l'enrichissement des métadonnées pour prendre en charge la découverte des services Web géographiques composables, en assurant les conditions nécessaires pour compléter la composition des services découverts. D'autre part une vision répartie de l'UDDI central qui consiste à le dispatcher en sous registres UDDI selon les classes des services Web géographiques (UDDIC : UDDI Classe).
- Les services Web géographiques localisés : les dernières avancées en termes de réseaux sans fil et de capacités techniques des dispositifs mobiles permettent aux Systèmes d'Information Géographique (SIG) d'être mobiles. Cette nouvelle génération de SIG repose sur des services localisés (LBS –*Location-Based services*) [FRA04], services exploitant la localisation de l'utilisateur à travers différents capteurs (tels qu'un GPS, un compas, un lecteur RFID). Notre perspective dans ce cadre consiste à proposer une extension de notre modèle de découverte sémantique en prenant en charge des difficultés en terme de mémoire nécessaire et de vitesse d'exécution.

Références bibliographiques

Références bibliographiques

- [ACH08] M Acher ,«Vers une ligne de services pour la grille :application à l'imagerie médicale», Rapport de stage Master recherche PLMT, Université de Nice Sophia-Antipolis, 2008
- [AKT05] M.S. Aktas, G. Aydin, G.C. Fox, H.Gadgil, M.Pierce, A.Sayar «Information Services for Grid/Web Service Oriented Architecture (SOA) Based Geospatial Applications», Scientificcommons, 2005
- [ALA06] A. Alam, G. Subbiah, B. Thuraisingam, L.Khan , « Reasoning with Semantics-aware Access Control Policies for Geospatial Web Services», Proceedings of the 3rd ACM workshop on Secure web, 2006
- [AMA99] G. Amato, U. Straccia, «User Profile Modeling and Applications to Digital Libraries», In: Processing of the Third European Conference on Research and Advanced Technology for Digital Librairies, Paris, France, 1999
- [BAA05] F. Baader, C. Lutz, M. Milicic, U. Sattker, F. Wolter, « A Description Logic Based Approach to Reasoning about Web Services ». In: Procs of the International Conference WWW, Workshop on Web Service Semantics: Towards Dynamic Business Integration (WSS2005), , Chiba Japan., 2005
- [BAD03] J. Barde, J. Divol, « Projet Syscolog : Métadonnées et ontologie », 2003
- [BAG03] J-F. Baget, E. Canaud, J. Euzenat, et M. Saïd-Hacid. « Les langages du Web sémantique ». In [CLR03], pp. 9-24, 2003.
- [BAH06] D. Bahloul, « Une approche hybride de gestion des connaissances basée sur les ontologies, application aux incidents informatiques », Thèse de doctorat, INSA de Lyon, 2006
- [BAR03] K. Barry Douglas, «The Savvy Manager's Guide to Web Service and Service Oriented Architectures», Morgan Kaufman Publishers Inc, Francisco CA, USA, 2003.
- [BAR04] S. Bard, « Méthode d'évaluation de la qualité de données géographiques généralisées Application aux données urbaines », Université PARIS 6, 2004
- [BAR05] F. Badra, «Utilisation de techniques de fouille de données pour enrichir une ontologie d'objets géographiques », mémoire de Master, Université Lyon 2, 2005
- [BAZ05] M. Baziz, « Indexation conceptuelle guidée par ontologie pour la recherche d'information », thèse de doctorat, IRIT, 2005

- [BEN04] S.Benmokhtar, «Synthèse ad hoc de services Web dans les environnements de l'informatique diffuse»; Rapport de stage DEA Systèmes Informatiques Répartis (SIR), Université Pierre et Marie Curie 2003-2004
- [BEN08] A. Benna, N.Boudjlida, H. Talantikite, «SAWSDL, Mediation and XQUERY for Web Services Discovery»,NOTERE'08: 8ème Conférence Internationale sur les NOuvelles TEchnologies de la REpartition , Lyon France, 23-27 juin 2008
- [BEN09] R. Benhalima, « Conception, implantation et expérimentation d'une architecture en bus pour l'auto-réparation des applications distribuées à base de services Web », Thèse de doctorat, Université de Toulouse, 2009
- [BET05] L.Berti-Équille, "Assurer la qualité des données: un défi permanent pour les systèmes d'information, bases et entrepôts de données", Genie Logiciel, 74, pp. 13-20, Septembre, 2005.
- [BOC01] G. Bochmann, B. Kerherve, H. Lutfiyya, M.M. Salem, H. Ye, « Introducing QoS to Electronic Commerce Applications », Springer-Verlag Berlin Heidelberg, pp.138-147, 2001
- [BOG07] L. Bourgois, « Représentation et comparaison de Web services complexes avec des logiques dynamiques »,thèse de doctorat ; Université PARIS 13-VILLETANEUSE, 2007
- [BON00] M. Bonin, C. Le page, « SIG, SMA, connaissances et gestion de l'espace », Revue internationale de géomatique, Vol 10 n° 1/2000.
- [BOO04] D. Booth, H. Haas,F. McCabe, E. Newcomer C. Ferris, D. Orchard , »Web Service Architectures », Technical report, W3C, Web Service Architectures group working, 2004
- [BOR01] S. Borzsonyi, D. Kossmann, K. Stocker, « The Skyline Operator »,IEEE Conf. On Data Engineering, 2001
- [BOU00] M.Boutry , «Qualité, Construction d'indicateurs », University nancy 2, 2000
- [BOU04] M Bouzeghoub, D Kostadinov, «Une approche multidimensionnelle pour la personnalisation de l'information», RapportPRiSM, Versailles, France, 2004
- [BOU05] M. Bouzeghoub, D. Kostadinov: Personnalisation de l'information: aperçu de l'état de l'art et définition d'un modèle flexible de profils. CORIA 2005: 201-218

- [BOU06] L. Bouzguenda, R. Bouaziz, E. Andonoff, « Utilisation d'Ontologies pour la Coordination dans le Workflow », Inter-Organisationnel Lâche, pp 10-11, 2006
- [BOU06a] M. Bouzeghoub, D. Kostadinov, « Data Personalization: a Taxonomy of User Profiles Knowledge and a Profile Management Tool », rapport technique, PRISM, 2006
- [BOU07] M. Bouzeghoub, S. Calabretto, N. Denos, R. Harrathi, D. Kostadinov, A. Nguyen, V. Peralta, « Accès personnalisé aux informations: approche dirigée par la qualité », INFORSID, pp 105-120, 2007
- [BOU08] M. Boutry, « Construction d'indicateurs », Université Nancy, 2008
- [BRE03] G. Bressoles, « La qualité de service traditionnelle versus la qualité de service électronique : similarités, différences et voies futures de recherche », Actes des 2èmes Journées Normandes de Recherche sur la consommation « Sociétés et Consommation » Rouen, 2003
- [BRE07] K.K. Breitman; M. A. Casanova; W. Truszkowski, (2007) Semantic Web: Concepts, Technologies and Applications. Londres: Springer, v.1. pp.337, 2007.
- [BUR02] M. Burgess, W. Alex Gray, N. Fiddian, « Establishing Taxonomy of Quality for Use in Information Filtering », Proceedings of the 19th British National Conference on Databases (BNCOD 2002), Lecture Notes in Computer Science: Advances in Databases (LNCS 2405), Sheffield, UK, July, p.103-113, 2002.
- [BUR03] M. Burgess, « Using Multiple Quality Criteria to Focus Information Search Results », PhD Thesis Cardiff University, UK, September, 2003.
- [CHA08] B. Chaib draa, « Agents et systèmes multiagents », Cours IFT-7001, Université Laval, 2008
- [CHO02] J. Chomicki, « Querying with Intrinsic Preferences », EDBT, 2002
- [CHR01] E. Christensen, F. Curbera, G. Meridith, S. Weerawarana, « Web Service Description Language (WSDL) Version 1.2. W3C », <http://www.w3.org/TR/wsdl.html>, 2001
- [CLR03] J. Charlet, P. Laublet, et C. Reynaud, « Le Web sémantique », rapport final de l'action spécifique 32 CNRS/STIC (version 3 de décembre 2003), publié chez Cépaduès (Hors-série de la collection Information interaction intelligence), 2003.
- [CNI07] Métadonnées et catalogages fiche du CNIG N° 106- 207, 2007
- [CNI94] CNIG, Éléments d'un vocabulaire géomatique, 1ère édition, 1994.

- [COR03] O. Corcho, M. Fernandez Lopez, A. Gomez-Pérez, « Methodologies, tools and language for building ontologies », *Data and knowledge Engineering* 46:41-64. 2003.
- [CRA03] J.B. Crampes, « Génie logiciel - Méthode orientée-objet intégrale MACAO, Démarche participative pour l'analyse, la conception et la réalisation de logiciels », Ellipses (Eds.), 2003.
- [CRA98] E. Crawley, R. Nair, B. Rajagopalan, H. Sandick, « A Framework for QoS-based Routing in the Internet », IETF RFC 2386, 1998
- [CUR02] F. Curbera, M. Duftler, R. Khalaf, W. Nagy, N. Mukhi, S. Weerawarana, «Unraveling the Web services Web: An introduction to SOAP, WSDL, and UDDI. IEEE», *Internet Computing*, vol.6, n°2, pp.86-93, 2002
- [DAM02] DAML Services Coalition, «DAML-S: Web Service Description for the Semantic Web». In *The First International Semantic Web Conference (ISWC)*, pp 348–363, 2002
- [DEN96] J. Denègre et F. Salgé. *Les systèmes d'information géographique*. collection « Que sais-je ? », PUF, Paris, pp.12, 62, 1996.
- [DEV05] R. Devillers, K. Beard “Communication et utilisation de l'information sur la qualité dans les SIG”. *Qualité de l'information géographique*. Lavoisier : Paris 2005. ISBN 2746210975, pp 275-290, 2005.
- [DON08] A. Donaubauer, T.Kutzner, F. Straub «Towards a Quality Aware Web Processing Service», *GI-Days*, 2008
- [DOR03] D. Gorga « Métadonnées et XML – une initiation » TECFA, 2002 – 2003, Juin 2003
- [DUM07] R. Dumitri, E. Klien, «SWING – A Semantic Framework for Geospatial Services », *The Geospatial Web* , Springer London, 2007
- [EGE02] M. J. Egenhofer, «Toward the Semantic Geospatial Web»,*GIS'02*, November 8-9, 2002, McLean, Virginia, USA, ACM, 2002.
- [EGE94] M.J. Egenhofer, « Spatial SQL: A Query and Presentation Language », *IEEE Transactions on Knowledge and DataEngineering*, vol. 6, no. 1, pp. 86-95, 1994
- [ELF04] A. El Fallah-Seghrouchni, S. Haddad, T. Melitti, Alexandru Suna « Interopérabilité des systèmes multi-agents à l'aide des services Web » , *JFSMA*, 2004
- [FDC03] F.T.Fonseca, C.A. Davis, G. Câmara, « Bridging ontologies and conceptuel schemas in geographic information integration », *Geoinformatica*, 7 (4) : pp 355-378, 2003

- [FEN02] D. Fensel, C. Bussler, A. Maedche, «Semantic Web Enabled Web Services ». In International Semantic Web Conference, Sardinia, Italy, pages 1–2.2000
- [FER95] J. Ferber. « Les systèmes multi-agents, vers une intelligence collective. », InterEditions, 1995.
- [FER99] M. F. Lopez. “Overview of methodologies for building ontologies”, IJCAI-99 Workshop on Ontologies and problem-solving Methods: Lessons learned and future Trends. 1999
- [FLU03] FLUKE NETWORKS, «Contrôle du temps de réponse, des applications Composantes et analyse des différentes méthodes», Fluke Corporation, 2003
- [FOR05] M. Fornefeld, P. Oefinger, «Modèles de compensation financière pour les géoservices web – une contribution à la mise en place d’une infrastructure des géodonnées en Suisse», MICUS Management Consulting GmbH Stadttor 1, Düsseldorf, 2005
- [FRA04] C. Frank, D. Caduff, M. Wuersch, M. « From GIS to LBS, An Intelligent MobileGIS. », IfGI prints, Münster, vol.22, 2004.
- [FRA08] S. Fraisse, H. Pronon, « Les metadonnées corvée ou nécessité ? », Géomatique Expert N° 63, 2008
- [FRI05] A. Friesen, M. Altenhofen, «SEMANTIC DISCOVERY OPTIMIZATION Matching composed semantic web services at publishing time», WEBIST pp 347-350,2005
- [GAN02] F. Gandon, « Ingénierie d’ontologie : une synthèse un retour d’expérience », Rapport de recherche INRIA RR-4396, mai 2002.
- [GEO07] «Annexe Informations techniques géoportail « administrations » » ANNEXE : INFORMATIONS TECHNIQUES GEOPORTAIL, 2007
- [GON04] A. Gonçalves, A Rodrigues, L Correia, "Multi-agent simulation within geographic information system", Proceedings of the 5th Workshop on Agent-Based Simulation, Lisbon, May 2004.
- [GOO01] Goodrum A. A.; McCain K. W., Lawrence S., Giles C. L., “Scholarly Publishing in the Internet Age: A Citation Analysis of Computer Science Literature”, Information Processing & Management, vol. 37 n° 5, pp. 661-75, 2001
- [GRO82] C. Gronroos, « Strategic Management and Marketing in the Service Sector », Helsingfors: Swedish School of Economics and Racine.- Administration,(cité dans [BRE03]), 1982

- [GUN98] N. J. GUNTHER, « The Practical Performance Analyst », published by McGraw-Hill, 1998.
- [GUT07] C. Gutiérrez, S. Servigne « Métadonnées et Qualité pour les Systèmes de Surveillance en Temps-Réel », SAGEO, 2007
- [GUT07a] C. GUTIERREZ, S. SERVIGNE, « Métadonnées spatiotemporelles Temps-réel », Revue des sciences et technologies de l'information : Ingénierie des Systèmes d'Information 12(2):pp. 97-119, Lavoisier, ISBN 978-2-7462-1913-, ISSN 1633-1311. 2007
- [HAR05] R. Harrati, S Calabretto, « Facteurs de qualité et personnalisation de l'information », mémoire de master, INSA de lyon, 2005
- [HAR10] R. Harrati, « Recherche d'information conceptuelle dans les documents semi structurés » Thèse de doctorat LIRIS, INSA de lyon, France, 2010
- [HIR98] G. Hirst, D. St Onge, « Lexical chains as representations of context for the detection and correction of malapropisms. In Christiane Fellbaum (editor), WordNet: An electronic lexical database », Cambridge, MA: The MIT Press .1998
- [HUB02] M. F. HUBERT, « Modèle de Traduction des Besoins d'un Utilisateur pour la Dérivation de Données Géographiques et leur Symbolisation par le Web », Thèse de doctorat, IGN FRANCE, Université de CEAN, 2003
- [ISO03] International Organisation for Standardization – Technical Committee 211. Geographic information – Metadata. ISO, 2003, 148 p. ISO 19115 : 2003
- [JAR05] I. Jars, « Contribution des sciences sociales dans le domaine de l'intelligence artificielle distribuée ALONE, Un modèle hybride d'agent apprenant », Thèse de doctorat université Claude Bernard Lyon I, 2005
- [JEL08] I. Jellouli, M. El Mohajir et E. Zimanyi. « Classification conceptuelle et ontologie de domaine pour l'intégration sémantique des données ». e-TI - la revue électronique des technologies d'information, Numéro 5, 5 novembre 2008
- [JEN98] N. R. Jennings, M. Wooldridge, and K. Sycara., « A roadmap of agent research and development ». Int Journal of Autonomous Agents and Multi-Agent Systems, 1(1) :7-38, 1998.

- [JIA97] J. Jiang, D. Conrath, « Semantic similarity based on corpus statistics and lexical taxonomy. » In Proceedings on International Conference on Research in Computational Linguistics, Taiwan, 1997.
- [KAL03] S.Kalepu, S. Krishnaswamy, S.W Loke, « Verity: A QoS Metric for Selecting Web Services and Providers ». In: Procs of the 4th International Conference on Web Information Systems Engineering Workshops (WISEW'03), Italy. pp.131-139, 2003
- [KEL03] P. Kellert et F. Toumani « Les web services sémantiques » spécifique 32 CNRS / STIC Web sémantique Rapport final, décembre 2003
- [KEL92] N. Belkin and W. Croft. «Information ltering and information retrieval : Two sides of the same coin ?» Communication of the ACM, 35(12) :29.38, 1992.
- [KOS04] D Kostadinov, « Personnalisation de l'information et gestion des profils utilisateurs », rapport, 2004
- [KOS07] D Kostadinov, «Personnalisation de l'information : une approche de gestion de profils et de reformulation de requêtes»; thèse de doctorat, Université de Versailles Saint-Quentin-En-Yvelines, 19 Décembre 2007
- [LAC87] M. Lacroix, P. Lavency, « Preference:Putting More Knowledge into Queries », Proceeding of the 13th VLDB Conference, Brighton, 1987
- [LAR06] J. Larson, M. Antonia, O. Siliceo, M. Pereira dos Santos Silva, E. Klien, S. Schade, «Are Geospatial Catalogues Reaching their Goals? », in AGILE Conference on Geographic Information Science, 2006
- [LI-05] Li-Li He; Jian-Gang Yang; Chao Deng; Heng-Nian Qi, « Multi-Agent Framework for Service-Oriented Geospatial Computing », Machine Learning and Cybernetics, 2005.
- [LIN06] LIN Hui, YE Lei, «A Mash-up of the Tracking System and Real-time Monitoring System Based on MAS »,IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.12, December 2006
- [LIN98] D. Lin., «An information-theoretic definition of similarity ». In Proceedings of 15th International Conference On Machine Learning, 1998.
- [LOP08] C. Lopez-Velasco, « Sélection et composition de services Web pour la génération d'applications adaptées au contexte d'utilisation », thèse de doctorat, École Doctorale Mathématiques, Sciences et Technologie de l'Information, Informatique, Grenoble, 2008.

- [LUD09] A. Ludovic, «Analyse des diagrammes de l'apprenant dans un EIAH de la modélisation orientée objet Le système ACDC», Thèse de doctorat, Université du Maine, France, 2009
- [MAN02] A. Mani, A. Nagarajan «Understanding quality of service for Web services», <http://www.ibm.com/developerworks/library/ws-quality.html>, 01 Jan 2002
- [MAR01] J. Marttinen, « Quality, Quality Standards and Spatial Data Quality », GI Norden, 2001
- [MAS07] E. Al-Masri, Q.H. Mahmoud, « Discovering the Best Web Service ». **In:** Procs of the 16th International Conference on World Wide Web (WWW2007), 7, Banff, Alberta, Canada. ACM, pp.1257-1258., 2007
- [MAZ01] H. Mazouzi, "Ingénierie des protocoles d'interaction : des systèmes distribués aux systèmes multi-agents», Thèse de doctorat université paris IX, 2001
- [MCI01] S. MCILRAITH, T.C. SON, & H. ZENG (2001). Semantic Web Services. IEEE Intelligent Systems. Special Issue on the Semantic Web, 16(2):46–53, 2001
- [MED05] L. Médini, «Systèmes d'Information Avancés (et répartis), Introduction aux services Web» Université Lyon 1 MIAGE, mars 2005.
- [MEL04] T. Melliti, « Interopérabilité des services Web complexes. Application aux systèmes multi-agents », Thèse de doctorat, Université Paris IX Dauphine, 2004.
- [MER05] R. Meraihi, « Gestion de la qualité de service et contrôle de topologie dans les réseaux ad hoc », l'Ecole nationale supérieure des télécommunications PARIS, 2005
- [MIR09] A. D. Miron, «Découverte d'associations sémantiques pour le Web Sémantique Géospatial: le framework ONTOAST », Thèse de doctorat, Université JOSEPH FOURIER, France, 2009.
- [MOS03] S.K. Mostefaoui, G. K. Mostefaoui, « Towards A Contextualisation of Service Discovery and Composition for Pervasive Environments », Workshop on Web services And Agent-based engineering, Melbourne, Australia, 2003.
- [MRI07] M. Mrissa, « Médiation Sémantique Orientée Contexte pour la Composition de Services Web », Thèse de doctorat, Université Claude Bernard Lyon I, 2007
- [MYS03] <http://cyberzoide.developpez.com/php4/mysql/>, 2003
- [NAU98] F. Naumann, « Data Fusion and Data Quality, Seminar on New Techniques et Technologies for Statistics, », 1998.

- [NIS04] National Information Standards Organisation (NISO), Understanding Metadata, 2004, <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- [OAS02] OASIS. UDDI Version 3.0. UDDI Spec Technical Committee Specification, <http://www.oasis-open.org/committees/uddi-spec/>, 2002
- [OSU02] J. O'sullivan, D;Edmond, A Ter Hofstede, « What's in a service ? », Distributed and parallel databases 12, P 117-133, Kluwer Academic Publishers, 2002
- [OWL04] “OWL-S: Semantic Markup for Web Services”, <http://www.w3.org/Submission/2004/SUBM-OWL-S-20041122/>
- [PAO02] M. Paolucci, T. Kawamura, T.R. Payne, Katia Sycara1 « Semantic Matching of Web Services Capabilities »,The Semantic Web — ISWC 2002, Springer Berlin / Heidelberg, 2002
- [PAR88] A. Parasuraman, A. Zeithaml Valarie, L. Berry Leonard , « SERVQUAL : A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality », *Journal of Retailing*, 64, 1, 12-40, 1988 (Cité dans [BRE03])
- [PER02] G. Percivall, «ISO 19119 and OGC Geographic Information Service Architecture », NASA Geographic Interoperability Office Global Science and Technology, Inc,2002
- [PER06] V. Peralta, « Data Quality Evaluation in Data Integration Systems », Thèse de doctorat Université de Versailles, France and Universidad de la República, Uruguay, 2006
- [PHP02] Le langage PHP, <http://cerig.efpg.inpg.fr/Note/2002/php.htm>
- [PIE06] L. Piechocki, «UML en français », <http://uml.free.fr/>, 2006.
- [POI06] A. Pointet, « Requêtes spatiales » , LSNE , École Polytechnique Fédérales de LAUSANE, 2006
- [PRI03] Y. Prié, Serge Garlatti « Méta-données et annotations dans le Web sémantique » Action spécifique 32 CNRS / STIC, Web sémantique Rapport final décembre 2003
- [RAD89] R. Rada, H. Mili, E.Bicknell, M.Blettner, « Development and application of a metric on semantic nets ». *IEEE Transaction on Systems, Man, and Cybernetics*, 19(1): 17-30, 1989
- [RAN03] S.Ran, “A Model for Web Services Discovery With QoS.” ACM SIGecom Exchanges, vol.4, n°1, pp.1-10, 2003

- [RES95] P. Resnik P. « Using information content to evaluate semantic similarity in a taxonomy ». In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal., 1995
- [REZ09] K. Rezeg, M.T. Laskri, O. Kazar, S. Servigne, « Geospatial Web Services Discovery Approach Using Metadata and Multi-Agents System», International Computers and Software (IRECOS) ISSN 1828-6003, Vol 4 N.1 January 2009 PP 26-33
- [REZ10] Geospatial Web Services Discovery Approach Using Quality, REZEG Khaled, LASKRI Mohamed Tayeb, Sylvie SERVIGNE, Journal of Convergence Information Technology (JCIT) ISSN: 1975-9320 Volume 5, Number 2, April 30, 2010 PP 28-35.
- [ROB76] S. Robertson and K. Sparck Jones. «Relevance weighting for search terms. Journal of The American Society for Information Science », 27(3) :129.146, 1976.
- [SAC09] E. R. Sacramento, « Automatic Discovery and Composition of OGC Web Services», thèse de doctorat, université fédérale de Cesra, Brésil, 2009
- [SAL71] G. Salton. «The SMART Retrieval System. Experiments in Automatic Document» Processing. Prentice-Hall Inc, NJ, 1971
- [SAL83] G. Salton, E.A. Fox, H. Wu. « Extended Boolean information retrieval system ». CACM 26(11), pp. 1022-1036, 1983.
- [SCH01] J. B. Schafer, J. A. Konstan, J. Riedl J., «E-Commerce Recommendation Applications », Data Min. Knowl. Discov, vol. 5, n° 1-2, pp. 115-153, 2001
- [SEG90] Segev A., Weiping F., “Currency-Based Updates to Distributed Materialized Views”, Actes de la 6th Int. Conf. on Data Engineering (ICDE), Los Angeles, USA, 1990.
- [SEM07] “Semantic Annotations for WSDL and XML Schema”,
<http://www.w3.org/TR/2007/REC-sawSDL-20070828/>
- [SER05] S. Servigne, « Requêtes spatio-temporelle », Cours département informatique, INSA de Lyon, 2005
- [SER06] S. Servigne, T. Libourel, « Fondements des bases de données spatiales », édition Hermes, Lavoisier, 2006
- [SER06b] S. Servigne, N. Lesage, T. Libourel. « [Spatial data quality components, standards and metadata](#). Fundamentals of Spatial Data Quality ». International scientific and technical encyclopedia. Wiley, ISBN 1905209568. Pp 179-208, 2006

- [SHE03] M. Sheshagiri, M. Desjardins, T. Finin « A Planner for Composing Services Described in DAML-S », Workshop on Web services And Agent-based engineering, Melbourne, Australia, 2003.
- [SHE03] M. Sheshagiri, M. Desjardins, T. Finin « A Planner for Composing Services Described in DAML-S », Workshop on Web services And Agent-based engineering, Melbourne, Australia, 2003.
- [SHU03] R Shuping « A Model for Web Services Discovery With QoS » ACM, 2003
- [SMI01] B. Smith et C. Welty, «Ontology: Towards a new synthesis». In Actes de la 2^{ème} conférence internationale FOIS'01 (Formal Ontology in Information Systems), Ogunquit, Maine, USA, 2001.
- [SOA07] SOAP Version 1.2 Part 1: Messaging Framework (Second Edition) W3C Recommendation , <http://www.w3.org/TR/2007/REC-soap12-part1-20070427/>, 27 April 2007
- [SRI04] N. Srinivasan, M. Paolucci , K. Sycara, «Adding OWL-S to UDDI, implementation and throughput », First International Workshop on Semantic Web Services and Web Process Composition, 2004
- [SYC02] K. Sycara, S. Widoff, M. Klusch, M., J. Lu, « Larks: Dynamic matchmaking among heterogeneous software agent in cyberspace. Autonomous Agents and Multi-Agent Systems 5», pp :173–203, 2002
- [TRI07] G, TROISPOUX « La qualité des données géographiques au service des utilisateurs », CERTU (centre d'études sur les réseaux, les transports, l'urbanisme et les constructions publiques, ministère de l'écologie, du développement et du management durables), 2007
- [TRO04] R. Troncy. « Formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies : application à la description de documents audiovisuels ». Thèse de doctorat d'informatique de l'université Joseph Fourier – Grenoble I, 2004.
- [URB06] D. URBANI, « Elaboration d'une approche hybride SMA-SIG pour la définition d'un système d'aide à la décision; application à la gestion de l'eau », thèse de doctorat université de CORSE – PASQUALE PAOLI UFR SCIENCES ET TECHNIQUES, 2006

- [VEN86] C. Van Rijsbergen. « A non-classical logic for information retrieval ». The computer journal, 29(6) :481-485, 1986.
- [VER99] H. Veregin, "Data quality parameters". In Geographical Information Systems, Vol. Principles and Technical Issues (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, Inc., pp. 177-189. 1999.
- [W3C04] W3C Working Group. W3c working group note, February 2004. <http://www.w3.org/TR/ws-gloss/>.
- [WAN96] Wang R., Strong D., "Beyond accuracy: What data quality means to data consumers", Journal on Management of Information Systems, vol. 12, n° 4, 1996, pp. 5-34.
- [WOO02] M.Wooldridge, «An Introduction to multiagent systems », Université de Liverpool UK, 2002
- [WU94] Z. Wu, M. Palmer, « Verb Semantics and Lexical Selection », Proceedings of the 32nd Annual Meetings of the Associations for Computational Linguistics, pp 133-138, 1994.
- [YAN06] A. Yann, «Conception et exploitation d'une base de méta données de traitements informatiques, représentation opérationnelle des connaissances d'expert, application au domaine géographique », Thèse de doctorat, IGN FRANCE, Université de CEAN, 2006
- [YUE06] P. Yue, L. Di., W. Yang, G. Yu, P. Zhao « Semantics-based automatic composition of geospatial Web service chains » ,Elsevier, 2006
- [YUQ08] Yuqi Bai ,Liping Di, Yaxing Wei, « A taxonomy of geospatial services for globalservice discovery and interoperability » Elsevier Computers & Geosciences ,2008
- [ZAR04] H. Zargayouna, S. Salotti « Mesure de similarité sémantique pour l'indexation de documents semi-structurés » dans [12ème Atelier de Raisonnement à Partir de Cas](#), Mars 2004
- [ZAY07] C. A. Zayani, I. Amous, "Approche pour l'adaptation de l'intégration de documents semi structures" RSTI-IS 12/2007 Métadonnées et nouveaux SI, pages 61 à 73, 2007
- [ZEI00] K. Zeitouni et L. Yeh. « Le Data Mining Spatial et les bases de données spatiales». In Actes des Journées Data Mining Spatial et Analyse du Risque, Versailles, 2000.

- [ZEM08] W N Zemirli, «Modèle d'accès personnalisé à l'information basé sur les Diagrammes d'Influence intégrant un profil utilisateur évolutif», thèse de doctorat, l'Université Paul Sabatier de Toulouse III, 12 Juin 2008
- [ZET00] A; Zeithaml Valarie, A. Parasuraman, A. Malhotra, «A Conceptual Framework for Understanding e-Service Quality: Implications for Future Research and Managerial Practice », *MSI Working Paper Series*, Report No. 00-115, 2000,(Cité dans [BRE03])
- [ZHA07] P. Zhao, G. Yu, L. Di, « Geospatial.Web.Services », Idea Group Inc., 2007