

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ SAAD DAHLAB, BLIDA

FACULTÉ DES SCIENCES

DÉPARTEMENT INFORMATIQUE



MÉMOIRE POUR L'OBTENTION D'UN DIPLÔME DE MASTER EN INFORMATIQUE

OPTION : GÉNIE DES SYSTÈMES INFORMATIQUE

SUJET :

---

**Conception et réalisation d'un portail documentaire  
sémantique avec la technologie des web des données**

---

Présenté par : ABDELBAKI ELFODIL

BENDJEBBAS ABDESSELEM

Encadreur de la recherche : Dr.HASSINA ALIANE

Organisme d'accueil : CERIST

Membres de jury composé de :

Président de jury : • ..Derrar Hacene .

Examinateur 1 : Rezzougue, N .

## Remerciements

Nous tenons à remercier tout d'abord Allah le tout puissant pour nous avoir donné la force d'accomplir ce travail ;

Notre encadreur Madame Hassina Alliane pour le sujet qu'elle nous a proposé. Merci d'avoir accepté de suivre la réalisation du travail, pour les conseils et l'encouragement.

Les membres de jury qui ont accepté de juger ce travail et d'y apporter leur caution.

Et nous remercions tous ceux qui ont aidé de loin ou de près dans la réalisation de ce mémoire.

## Dedicace

Je dédie ce travail pour toute personnes qui ma était d'aide de prés ou de loin pour obtenir mon diplôme, toute d'abord a ma maman qui a était la avec ses encouragement et ses prières, ensuite pour ma deuxième maman, ma tante qui n'as pas cessé de croire en moi, toute ma famille, mes amis, spécialement, mon binôme El fodil Abdelbaki, mes frères El merrerbi Youcef, Ababsia Abdellah, Reguieg Khalil et Stambouli Abdelkarim.

Bendjabbas Abdesselem.

Je dédie ce travail à mes très chers parents, ma mère pour ses encouragements, mon père pour tout ce qu'il a fait pour avoir ce résultat, que dieu les protèges. Je le dédie à mon frère ainsi que toute la famille. A mon ami et binôme Bendjebbas Abdesselem

El fodil Abdelbaki.

## Résumé

Le but de notre travail dans ce mémoire, est de réaliser un portail documentaire sémantique, . Deux travaux complémentaires ont été réalisés en vue d'atteindre cet objectif :

- Construire une ontologie selon le modèle FRBR pour l'aspect sémantique du travail.
- Relier les données qui proviennent de la bibliothèque nationale de France avec les données de notre ontologie pour des résultats plus pertinentes

## Abstract

The purpose of our work in this master thesis is to provide a semantic document portal for managing digital library documents. Two complementary works were done to achieve this goal :

- Building an ontology with FRBR model for the semantic aspect of the work.
- Link the data coming from the national library of France with our ontology data for more relevant results

# Table des matières

<b>I</b>	<b>Introduction Générale</b>	<b>1</b>
<b>II</b>	<b>Etat de l'Art</b>	<b>4</b>
<b>1</b>	<b>Généralité sur Le Web sémantique et les ontologies</b>	<b>5</b>
1.1	Introduction . . . . .	6
1.2	Bref historique du web actuel . . . . .	6
1.3	Introduction aux Web Sémantique . . . . .	7
1.3.1	Force et faiblesse du web actuel . . . . .	8
1.3.2	Solution aux problèmes du web classique . . . . .	8
1.3.3	Le web sémantique . . . . .	9
1.4	Technologie du web sémantique . . . . .	10
1.4.1	Standards du web sémantique . . . . .	10
1.4.2	Vocabulaire . . . . .	15
1.5	Ontologies . . . . .	19
1.5.1	Qu'est-ce qu'une ontologie ? . . . . .	19
1.5.2	Éléments(Composante) d'une ontologie . . . . .	21
1.5.3	Type d'ontologie . . . . .	21
1.5.4	Cycle de vie d'une ontologie . . . . .	23
1.5.5	Méthode et Méthodologies d'ingénierie ontologique . . . . .	26
1.5.6	La méthode METHONTOLOGY . . . . .	27
1.5.7	Outils et langage de construction d'ontologie . . . . .	31

1.6	Linked data . . . . .	33
1.6.1	Linked data, le web des donnée . . . . .	33
1.6.2	Le principe du Linked data . . . . .	33
1.6.3	Les avantages de Linked data . . . . .	36
1.6.4	les caractéristiques qu'il vau mieux éviter . . . . .	36
1.7	conclusion . . . . .	37
<b>2</b>	<b>Le portail documentaire</b>	<b>38</b>
2.1	Introduction . . . . .	39
2.2	Définitions . . . . .	39
2.2.1	Le terme portail . . . . .	39
2.2.2	Le portail documentaire . . . . .	40
2.3	Moteur de recherche fédérée . . . . .	40
2.3.1	Fonctionnement . . . . .	40
2.3.2	Exemples de moteur de recherches fédérée . . . . .	41
2.4	Les enjeux du portail documentaire . . . . .	42
2.4.1	L'hétérogénéité des ressources . . . . .	42
2.4.2	Adaptation aux usages . . . . .	42
2.4.3	La simplicité de la recherche . . . . .	43
2.4.4	Architecture du portail documentaire . . . . .	44
2.5	Le Portail sémantique . . . . .	45
2.5.1	Les enjeux du portail sémantique . . . . .	45
2.5.2	La collecte d'informations . . . . .	45
2.6	Exemple de portails web sémantiques . . . . .	46
2.6.1	Le portail sémantique SEAL . . . . .	46
2.6.2	presentation de fonctionnement de SEAL . . . . .	46
2.6.3	Le portail sémantique Kasimir . . . . .	47
2.6.4	Le Portail sémantique data.BNF . . . . .	48
2.7	Conclusion . . . . .	49

<b>III</b>	<b>Conception Et Réalisation</b>	<b>50</b>
<b>3</b>	<b>Conception de notre système</b>	<b>51</b>
3.1	Introduction . . . . .	52
3.2	Conception de l'ontologie . . . . .	52
3.2.1	Le modèle FRBR . . . . .	53
3.2.2	FRBR, Illustration ! . . . . .	59
3.2.3	Construction de notre ontologie . . . . .	61
3.2.4	Exemple illustratif . . . . .	75
3.3	Linked data . . . . .	78
3.3.1	Publication du Linked Data . . . . .	78
3.3.2	Relier nos données avec ceux de la BNF . . . . .	80
3.4	Conclusion . . . . .	82
<b>4</b>	<b>Réalisation du portail sémantique</b>	<b>83</b>
4.1	Introduction . . . . .	84
4.2	Environnement de travail : . . . . .	84
4.2.1	Langage de Programmation : . . . . .	84
4.2.2	Éditeurs . . . . .	84
4.2.3	Appache Jena Framework . . . . .	85
4.2.4	Serveurs : . . . . .	88
4.2.5	SparQl . . . . .	88
4.3	le portail Cerist : . . . . .	89
4.4	Conclusion . . . . .	91
<b>IV</b>	<b>Conclusion Générale</b>	<b>92</b>

# Table des figures

1.1	Architecture pyramidale du web sémantique [W3C] . . . . .	11
1.2	Représentation d'un triplet RDF. . . . .	12
1.3	Représentation schématique d'un graphe RDF. . . . .	13
1.4	Ontologie sous FOAF au format RDF/XML . . . . .	14
1.5	Typologie des ontologies selon l'objet de conceptualisation . . . . .	22
1.6	Cycle de vie d'une ontologie . . . . .	23
1.7	Figure Processus de développement et le cycle de vie de METHONTOLOGY. . . . .	28
1.8	Graphe RDF pour person :Richard Cyganiak . . . . .	34
1.9	Graphe RDF pour Berlin . . . . .	35
1.10	Principe de Linked data . . . . .	35
1.11	Langage OWL des trois graphes . . . . .	36
2.1	Architecture de Seal . . . . .	46
2.2	Architecture du portail kasimir . . . . .	48
3.1	Les éléments de l'entité du "Groupe1" . . . . .	55
3.2	Relations entre l'entité "Groupe2" et "Groupe1" . . . . .	55
3.3	Entité "Sujet" dans FRBR . . . . .	56
3.4	Les classes les plus générales de notre ontologie . . . . .	66
3.5	les sous classe des classes générales de notre ontologie . . . . .	66
3.6	Procédé de Haut en Bas . . . . .	66
3.7	Diagramme des relations binaires . . . . .	67

3.8	Extrait de notre ontologie . . . . .	74
3.9	Extrait des relation de l'ontologie . . . . .	74
3.10	Extrait du code source de l'ontologie . . . . .	75
3.11	Création des instance de l entité Œuvre . . . . .	76
3.12	Création des relation entre les instance de l'élément Œuvre . . . . .	76
3.13	L'expression de nos données et leur matérialisation . . . . .	77
3.14	Relier les documents a leur auteurs . . . . .	78
3.15	Configuration de fuseki server . . . . .	79
3.16	Interface du fuseki sparql endpoint . . . . .	79
3.17	Conversion de notre ontologie en donnée rdf . . . . .	80
3.18	Création d'une nouvelle ressource et sont URI avec protégé . . . . .	81
3.19	Generation du model rdf de la ressource . . . . .	81
3.20	Exemple de requête sur notre ressource . . . . .	81
3.21	Résultat finale . . . . .	81
4.1	Jena Downloads . . . . .	85
4.2	Architecture Jena Framework . . . . .	86
4.3	instanciation du model de l'ontologie . . . . .	87
4.4	JenaFramework Eclipse . . . . .	88
4.5	Prise du portail documentaire sémantique . . . . .	89
4.6	Prise du portail documentaire sémantique : Recherche avancée . . . . .	90
4.7	Prise du portail documentaire sémantique : Inscription . . . . .	90
4.8	Prise du portail documentaire sémantique : connexion . . . . .	91

# Liste des tableaux

1.2	Vocabulaire de description de données . . . . .	17
3.2	Construction du glossaire de termes . . . . .	65
3.4	Construction de dictionnaire des concepts . . . . .	71
3.6	Construction de la table des relations binaires . . . . .	72
3.8	Construction de la table d'attribut . . . . .	73

# **Première partie**

## **Introduction Générale**

---

## Introduction Générale

Dans le domaine du web, la richesse joue un rôle très important par rapport à son évolution, une richesse qui se définit en quantité d'informations que reçoit un utilisateur après avoir exécuter sa requête, mais aussi le nombre d'utilisateurs qu'ils peuvent avoir besoin de ces informations. On parle d'utilisateurs humain ou systèmes, ces derniers sont des agents web, des petits processus s'exécutant dans des ordinateurs utilisant le web.

Pour les bibliothèques, utiliser le web pour partager les documents était un objectif à réaliser, alors des portails documentaire on vite vue le jour, on parle de portails web qui proposent des documents de bibliothèques, c'est la que les utilisateurs humains trouvent généralement leurs documents souhaités tandis que pour les agents web, ils ne comprenaient pas le contenu dans les sites des bibliothèques à cause de la mal structuration des données.

La solution pour que les machines peuvent lire et comprendre le contenu des sites web sans avoir besoin d'un être humain comme intermédiaire est de repenser le web, et aller vers un web plus significatif qu'on appelle le web sémantique. C'est un web où les utilisateurs systèmes peuvent séparer le contenu de la présentation dans les sites internet. Ainsi les bibliothèques cherchent à réaliser des portails documentaire sémantique à savoir des portails qui regroupent les documents de bibliothèques et qui sont accessibles et compréhensibles par les agents web. Ce document consiste en une étude d'une réalisation d'un portail documentaire sémantique, le travail présenté a été élaboré au sein du Centre de Recherche Sur L'information Scientifique Et Technique (CERIST).

Le document est organisé en quatre chapitres :

- **Chapitre 1** : Où nous présentons les définitions et les concepts liés aux web sémantique, ainsi que ses technologies comme la base du web sémantique qui est l'ontologie.

- 
- **Chapitre 2** : Où nous présentons un élément de base de notre travail, à savoir, le portail documentaire, nous verrons ce qui est un portail documentaire, ces fonctionnalités de base ainsi que quelques exemples sur des portails documentaire sémantique.
  - **Chapitre 3** : Où nous donnons la description et la spécification de la conception de notre système à travers la construction de notre ontologie selon le modèle FRBR et l'apport des Linked data dans l'enrichissement et la publication de nos données.
  - **Chapitre 4** : Où nous exposons d'une part l'implémentation du système avec une vue général sur notre portail documentaire et ces principaux fonctionnalité et d'autre part l'environnement de notre travail et les outils que nous avons utilisé pour achever ce dernier.

## **Deuxième partie**

### **Etat de l'Art**

# Chapitre 1

## Généralité sur Le Web sémantique et les ontologies

### Contents

---

1.1	Introduction . . . . .	6
1.2	Bref historique du web actuel . . . . .	6
1.3	Introduction aux Web Sémantique . . . . .	7
1.4	Technologie du web sémantique . . . . .	10
1.5	Ontologies . . . . .	19
1.6	Linked data . . . . .	33
1.7	conclusion . . . . .	37

---

### 1.2 Bref historique du web actuel

En 1989, Tim Berners-Lee, travaillant au CERN (Conseil européen pour la Recherche nucléaire) invente le Web (World Wide Web) en liant le principe d'hypertexte avec Internet. Surfer sur le net, partager des documents, échangés avec des millions de personnes à travers le monde... Tout ceci n'aurait pas été possible sans l'invention du World Wide Web. Le Web a été pendant les premières années de son apparition un outil de diffusion, où un ensemble de producteurs de contenu diffusent l'information aux internautes, via des pages HTML simples liées entre elles par des hyperliens. C'est ce qu'on appelle le Web 1.0 auquel succède le Web 2.0.

Outre les évolutions techniques, le Web 2.0 se traduit par une évolution des usages. Grâce aux interfaces simplifiées, l'internaute s'approprie plus facilement le contenu du Web et ses nouvelles fonctionnalités. Avec les nouvelles applications (blogs, wikis, réseaux sociaux), les utilisateurs peuvent participer à la création du contenu.

L'éditeur Tim O'reilly rappelle certainement le mieux d'où vient la formule : à son origine, le terme web 2.0 capturait le sentiment commun selon lequel il se passait quelque chose de qualitativement différent sur le web d'aujourd'hui. Pour lui, la clef du succès dans cette nouvelle étape de l'évolution du web réside dans l'intelligence collective. "Le web 2.0 repose sur un ensemble de modèles de conception : des systèmes architecturaux plus intelligents qui permettent aux gens de les utiliser, des modèles d'affaires légers qui rendent possible la syndication et la coopération des données et des services... Le web 2.0 c'est le moment où les gens réalisent que ce n'est pas le logiciel qui fait le web, mais les services!", de son côté l'ancien rédacteur en chef de « Wired », Kevin Kelly partagé le même avis de ce qui est le web 2.0 "Le web 2.0 repose sur un ensemble de modèles de conception : des systèmes architecturaux plus intelligents qui permettent aux gens de les utiliser, des modèles d'affaires légers qui rendent possible la syndication et la coopération des données et des services... Le web 2.0 c'est le moment où les gens réalisent que ce n'est pas le logiciel qui fait le web, mais les services!" [Tip11]

Le Web 2.0 est marqué par l'interactivité entre les utilisateurs, l'échange et le partage d'informations, c'est ce qu'on appelle le Web social. Mais le Web ne cesse d'évoluer. De nos jours, la question qui se pose, c'est quel serait le web de demain ?

L'évolution future du Web 2.0 est souvent désignée par le terme Web 3.0. Deux évolutions majeures sont pressenties. Certains spécialistes voient le Web sémantique comme la future évolution du Web, d'autres pensent que le Web 3.0 sera l'internet des objets[i]. Cependant, le Web 3.0 pourrait être la conjonction des deux thèses précédentes. Dans l'internet des objets, les objets physiques sont identifiés de manière unique, avec des codes-barres, des étiquettes RFID etc. les données sur les

comportements de ces objets sont collectées via des dispositifs tels que les capteurs. Ainsi, le Web sémantique permettrait de structurer et de mieux exploiter les données collectées.[Piq12]

### 1.3 Introduction aux Web Sémantique

De nos jours, le web est chargé d'un grand nombre d'informations, il fut un temps ou on considérait le fait qu'un utilisateur fait ses achats sur internet comme un véritable exploit, qu'on est il si une machine fait ses propres achats sur internet ?

#### 1.3.1 Force et faiblesse du web actuel

Malgré que le web classique est une forte technologie qui est indispensable pour la vie de l'être humain sauf qu'à force de se reposé sur lui, le web s'affaiblit et ne répond plus aux exigences cet être humain

##### **Force du web actuel**

Le web de nos jours connaît de plus en plus de portails, qui dit trop de portails dit trop d'informations et qui dit trop d'informations dit aussi trop de recherches d'informations, alors des moteurs de recherche ont vue le jour et des méta-moteurs ont vite suivis, cela représente une force incontestable pour le web actuel, de plus les ressources sont hétérogènes et distribuées. Avec un web statique et dynamique dans les serveurs, ces derniers profitent des services mise en place par internet email, FTP, Chat, E-commerce etc. ce qui fait que le web devient simple à utiliser mais seulement par des lecteurs humains, les données sont organisées principalement pour être simple à transmettre.

##### **Faiblesse du web actuel**

Le web actuel, un grand volume de ressources hétérogènes distribuées, simple à transmettre mais destiner à des humains et seulement pour des humains. Internet est de plus en plus utilisé par des machines, des agents web intelligents etc, qui ont accès à l'information la traitent, qu'ils peuvent l'extraire et l'interpréter tandis que la majorité des données qui sont sur le web sont sous une forme qui ne permette pas leur utilisations par des machines, on a besoin d'un web ou une machine peut extraire de l'information mais aussi la générer et ce n'est qu'en franchissant cette porte que nous pouvons vraiment dire que des machines se communiquent entre elles. Le web actuel fais face a une

importante faiblesse qui fait que l'information n'est ni accessible ni interprétable par des machines les informations sont caché dans l'ombre du code HTML.

### 1.3.2 Solution aux problèmes du web classique

A imaginer les avantages que peut offrir une véritable communication entre deux machines, les développeurs ont vite essayé de remédier aux problèmes du web, les technologies d'XML étaient parmi les solutions qui ont aidé à séparer la présentation du contenu, sauf que XML n'est pas un langage mais un métalangage sinon comment comparer deux documents XML écrits avec des DTD différents ? Le web sémantique, est une extension du web classique facilitant l'automatisation du traitement des connaissances disponibles, les connaissances ne sont pas représentées dans une langue naturelle mais plus tôt formalisées grâce à de langages pouvant être interprétés par des machines En effet, et selon la vision de son créateur, Tim Berners-Lee, l'évolution du web est marquée par deux grandes étapes. La première consiste simplement à relier les internautes entre eux pour partager des données et travailler en coopération. A cet effet, le HTML est largement suffisant, notamment à travers les liens hypertextes qui permettent de naviguer d'une ressource à l'autre. Toutefois, le défaut majeur du HTML est qu'il ne sépare pas le contenu de la présentation, ce qui pose un problème d'interopérabilité. Le second volet de cette évolution, quant à lui, a porté sur l'extension des fonctionnalités du Web, favorisant ainsi la coopération homme/machine. Pour cela, cette seconde génération de Web doit être apte à interpréter la sémantique des ressources utilisées. Ce principe repose déjà sur la séparation du contenu et de la présentation. Des standards mis au point par le W3C permettent déjà de structurer ce premier et de lui donner un sens, en lui affectant des méta-données. En se basant sur ces informations et les connaissances du monde réel, un système peut alors effectuer des inférences au nom des utilisateurs. Cette technologie permet d'assurer un plus grand partage et une réutilisation des données, tout en favorisant la coopération homme/machine.

### 1.3.3 Le web sémantique

Le web sémantique est un moyen de rendre des contenus accessibles aux ordinateurs sur internet, les informations vont être inter-connectées entre elles, pour but de laisser les ordinateurs raisonner sur ces informations. Tim Berners-Lee, directeur du W3C évoque le web sémantique comme suite : Certains ont pensé qu'il s'agissait d'un Web qui permettrait par exemple d'effectuer des recherches sur Internet en posant des questions sous forme de phrases, en langage naturel. Cependant ce n'est pas le vrai but du web sémantique. En effet la philosophie de ce dernier est d'avoir des

connaissances structurés par des langages de structuration et aussi de crée un lien automatique pour relier les données qui sont stockées dans les différents fichiers et bases de données de nos ordinateurs. Cela va permettre une interopérabilité inégalée et une capacité de faire partager l'accès aux données avec des machines qui n'existe pas aujourd'hui [r06] Le Web sémantique repose sur les ontologies qui permettraient aux humains et aux machines de partager les connaissances du domaine et de collaborer ensemble mais aussi pour permettre aux machines de raisonner et d'interpréter les informations ainsi que d'améliorer la pertinence des recherches. Le principe consiste à définir une interprétation commune d'une partie du monde réel, et modéliser les concepts et des relations entre les concepts par des classes et des relations entre les classes. Une ontologie est souvent gérée comme un wiki, dans le sens où chacun peut intervenir pour modifier certains sens ou ajouter des nouvelles définitions. L'ontologie doit donc être entretenue et mise à jour au fur à mesure de l'évolution du monde qu'elle modélise. [Pal01]

« Avec la notion de Web sémantique, vous définissez un espace virtuel où les liens hypertexte pointerait non plus sur des documents (textes ou images), mais sur des concepts » et « Le futur Web exprimera l'intelligence collective de l'humanité » [Pal01] Étant donné le caractère très large et désordonné des données disponibles sur le réseau du net, on attend du web sémantique une meilleure structuration de ces dernières et une affectation optimale des méta-données qui puisse simplifier leurs interprétations par les machines.

### 1.4 Technologie du web sémantique

Le web sémantique consiste en l'application d'un ensemble de standards promus par le W3C, permettant d'identifier, modéliser, encoder et interroger des données. Des vocabulaires viennent compléter ces standards en donnant un ensemble de termes concrets que quiconque peut réutiliser pour décrire ses propres ressources dont la manipulation de ces derniers par des machines requiert l'expression ou la description de ces ressources. Plusieurs langages sont donc définis à cet effet, ils doivent permettre d'exprimer données et métadonnées (RDF, Cartes Topiques), de décrire les services et leur fonctionnement (UDDI, WSDL, DAML-S, etc.) et de disposer d'un modèle abstrait de ce qui est décrit grâce à l'expression d'ontologies (RDFS, OWL). Dans cette section nous essayerons de donné un éclairage sur ces standards et vocabulaires qui constituent les deux piliers du web sémantique.[w3c16]

### 1.4.1 Standards du web sémantique

L'ensemble des technologies du Web sémantique est organisé dans une architecture en couches. C'est ce qu'on appelle la « Semantic Web Stack » ou « Pile du Web Sémantique » (voir figure ci-dessous). Elle constitue la vision du W3C sur l'architecture du Web Sémantique. Ainsi l'ensemble des technologies (langages et protocoles) qu'elle comporte, sont standardisées par le W3C, et font toujours l'objet de recherches et de travaux d'amélioration et de normalisation au sein du W3C. Certaines couches sont déjà implémentées. En effet, les langages et protocoles qui permettent de remplir leurs fonctions existaient déjà ou ont été conçus et créés pour répondre aux spécifications de chaque couche. Le Web sémantique est encore un grand chantier, les couches supérieures ne sont pas encore très évoluées, mais le rôle qu'elles doivent occuper est globalement défini. Chaque couche utilise la couche du dessous, et a une fonction bien déterminée dans l'architecture.[w3c16] Dans ce

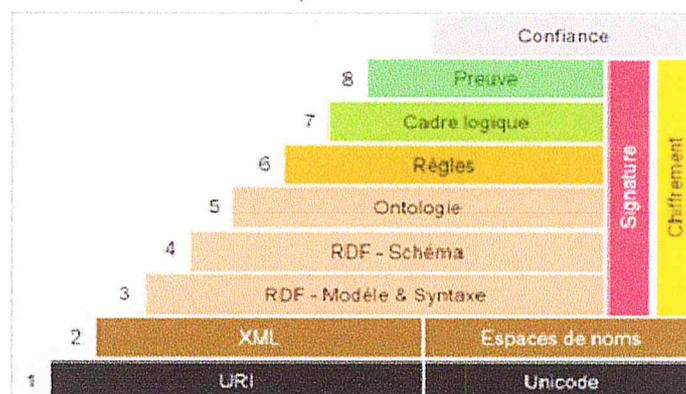


FIGURE 1.1: Architecture pyramidale du web sémantique [W3C]

qui suit, nous présenterons plus en détail les différents niveaux de cette technologie.

#### La couche URI et Unicode

L'URI (Uniform Resource Identifier) est justement un protocole simple et extensible pour identifier, d'une manière unique et uniforme, toute ressource sur le web (document électronique, une page HTML, un fichier, une image, une vidéo, un service) ayant une identité. La notion d'uniformité, Elle permet d'abord à différents types d'identificateurs de ressources d'être utilisés dans le même contexte, même si le mécanisme qui leur permet d'y accéder est différent. Ensuite elle assure une interprétation sémantique uniforme des conventions syntaxiques communes pour les différents identificateurs. Par ailleurs, elle permet d'introduire de nouveaux types d'identificateurs de ressources sans interférer avec ceux existants. Enfin, cette uniformité permet la réutilisation de ces identificateurs dans différents contextes. [Tip11] Il existe plusieurs types d'identificateurs de res-

sources. Un URI peut être classé, en effet, en 2 catégories, selon qu'il soit destiné à la localisation ou au nommage :

- URL (Uniform Resource Locator) désigne un sous-ensemble d'URI qui identifie les ressources via une représentation de leur mécanisme d'accès, plutôt que par le nom ou autre attribut de cette dernière.
- URN (Uniform Resource Name). désigne un sous-ensemble d'URI qui identifie les ressources via une syntaxe de chaîne de caractères durant toute son existence, indépendamment de sa localisation ou de son accessibilité par Internet. L'URL et l'URN sont donc des cas particuliers d'URI.

Les données seront encodées avec le jeu de caractères Unicode pour un maximum d'interopérabilité.

[Lav16]

### La couche XML, NS, XML Schéma

À ce niveau d'architecture, nous ne sommes toujours pas au point d'affecter une sémantique à l'information, c'est-à-dire de la décrire et lui donner un sens. Il s'agit seulement d'une couche syntaxique, de bas niveaux, qui permet de structurer les données et organiser selon un format de message standard, et ce, grâce au langage de balisage extensible XML [Lav16]

XML est un langage proposant une syntaxe de description de la structure d'un document, mais aussi de création et de manipulation de documents. Il utilise un espace de nommage (name space) permettant d'identifier des balises utilisées dans les documents Xml. Le document est validé et définit par un schéma. Cependant, XML n'impose aucune contrainte sémantique à la signification de ces documents. Xml ne suffit donc pas pour assurer aux logiciels la compréhension du contenu des données incluses dans le document : l'interopérabilité syntaxique n'est pas suffisante. [Lav16]

### La couche RDF et RDF Schéma

Après avoir référencé les ressources avec le protocole URI et structuré les informations avec le XML, l'étape suivante consiste à les annoter, afin de les doter d'un sens interprétable par la machine. C'est justement le rôle de la couche RDF et RDF-S dans l'architecture du Web sémantique. Le terme "annotation" est défini comme un commentaire ou une explication d'un document ou de son contenu, de son côté le RDF est la norme qui spécifie comment définir des liens sémantiques, les relations entre les nœuds de données, les ressources. Comment associer les métadonnées aux ressources et entre les ressources. La construction de base en RDF consiste en un triplé d'éléments (Ressource, Propriété,

Valeur); qu'on appelle déclaration RDF. Par analogie, un triplé RDF est similaire à la déclaration (sujet , prédicat , objet)



FIGURE 1.2: Représentation d'un triplé RDF.

- Ressource (Sujet) : Cela peut être n'importe quel objet référencé par une URL, qu'il concerne le web (Page HTML, document PDF, fichier multimédia...), ou non (Personne, Région, Etc.).
- Propriété (prédicat) : Critère, caractéristique, attribut ou relation qui peut décrire la ressource (titre, couleur, taille, auteur, etc.). Une propriété ses valeurs permises et ses relations avec les autres propriétés.
- Valeur (objet) : C'est la valeur qui sera affectée à la propriété de la ressource. Cette affectation peut être soumise à certaines restrictions.[w3c16]

Une déclaration RDF est souvent schématisée par un graphe RDF, où les sujets et les objets sont représentés par des ellipses, qu'on appelle nœuds, et le prédicat par une flèche qui va du sujet vers l'objet. Les rectangles, quant à eux, représentent des instances d'une ressource.

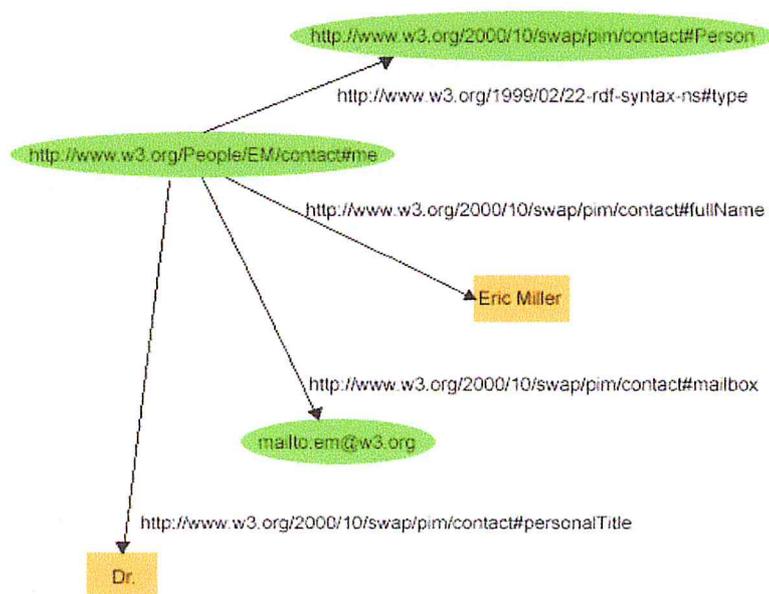


FIGURE 1.3: Représentation schématisée d'un graphe RDF.

- RDF-S (RDF Schema) est un vocabulaire de base pour décrire les déclarations RDF, au même titre que le XML-S pour le langage XML. Il ajoute à RDF la possibilité de définir des hiérarchies

de classes et de définir les genres et les propriétés des ressources, d'assigner des contraintes spécifiques sur la nature des documents et de fournir des informations sur l'interprétation des déclarations RDF. Les schémas RDF permettent donc de garantir qu'un document RDF est sémantiquement consistant.

### Couche ontologique

Une ontologie est un modèle de donnée représentatif d'un ensemble de concept dans un domaine. Cette couche utilise le langage OWL, qui est basé sur RDF. C'est une extension de RDF Schémas. Il définit un vocabulaire riche pour décrire les ontologies. Une des ontologies les plus utilisées est le projet FOAF (Friend Of A Friend) qui permet de décrire des personnes et les relations qu'elles entretiennent entre elles. Un exemple concret consisterait par exemple à décrire Mr Dan Brickley, avec son adresse mail, son site web ainsi qu'une de ses connaissances. La figure suivante quant à elle montre un exemple d'une ontologie sous FOAF

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <foaf:Person>
    <foaf:name>Dan Brickley</foaf:name>
    <foaf:title>Mr.</foaf:title>
    <foaf:givenName>Dan</foaf:givenName>
    <foaf:familyName>Brickley</foaf:familyName>
    <foaf:mbox rdf:resource="mailto:webmaster@foaf-project.org"/>
    <foaf:homepage rdf:resource="http://danbri.org"/>
    <foaf:knows>
      <foaf:Person">
        <foaf:name>Libby Miller</foaf:name>
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>
```

FIGURE 1.4: Ontologie sous FOAF au format RDF/XML

**SPARQL, SPARQL Protocol and RDF Query Language** Comme son nom l'indique, SPARQL est un protocole, un langage et aussi un formalisme pour l'expression des résultats qui permet d'interroger dynamiquement les données en RDF, sans télécharger l'ensemble des données brutes. SPARQL est l'équivalent de SQL pour le Web des données. L'un des avantages de SPARQL est le fait qu'il se base directement sur les métadonnées RDF. Cela permet aux machines ou aux humains d'interroger des bases de données sur le Web, sans forcément en connaître le schéma au préalable, ce qui permettrait un accès aux données sans intermédiaire. Il est à noter qu'on reviendra plus tard avec plus de détails sur cette couche qui est le noyau du web sémantique

### Les Couches Logique, Confiance et preuve

Si les premières couches ont déjà été standardisées et recommandées par le W3C, dans le cadre de l'architecture du web sémantique, il en est tout autrement pour les couches de haut niveau, que très peu d'auteurs évoquent dans leurs littératures.

En effet, Ces trois couches n'ont pas encore été implémentées. Elles s'appuieront les unes sur les autres afin de permettre l'identification et la validation des informations récoltées grâce aux données RDF. La couche Logique est la première de ces couches. Son nom est assez transparent : elle donne la possibilité de construire des moteurs d'inférence logique afin de faire des liens entre les entités RDF sans que ceux-ci soient explicitement exprimés. Elle se baserait pour ça sur les règles définies par la couche Rules. La couche Logic sera à la couche Rules ce que la couche OWL est à la couche RDF c'est-à-dire qu'on cherche à regrouper les règles pour leurs donner un sens et les rendre utilisable. La couche Preuve utilisera la couche logique afin d'assurer la vérification des déclarations effectuées dans le Web sémantique, et de prouver la pertinence de l'information retournée par les couches de plus bas niveau et des déductions obtenues à partir des inférences Enfin la couche Confiance vient s'appuyer sur les deux autres et c'est elle qui permettra de contrôler la véracité d'une information en attribuant plus ou moins de confiance aux sources de données rencontrées lors de la recherche ainsi que d'évaluer la fiabilité de l'information et des raisonnements. Cette couche repose sur les signatures numériques, le cryptage des données et sur la fiabilité des sources d'information.

### 1.4.2 Vocabulaire

Dans ce qui suit on va présenter les principaux vocabulaires du web sémantique qui sont utilisés pour exprimer les données dans un modèle RDF.

### Qu'est-ce qu'un vocabulaire du web sémantique ?

Lorsqu'on exprime des données en RDF, il est préférable d'utiliser autant que possible des classes et des propriétés déjà définies dans des vocabulaires existants. En RDF, les ressources appartiennent généralement à des classes qui les regroupent par types (documents, concepts, personnes, etc.). Elles sont qualifiées grâce à des propriétés (prédicats) qui définissent un aspect, une caractéristique, un attribut, ou une relation spécifique de ces ressources. Les classes et les propriétés sont décrites dans des vocabulaires RDF qui permettent ensuite aux machines de les comprendre et de les exploiter. Ces vocabulaires affectent des URI à chaque classe et chaque propriété qu'ils définissent.

### Comment choisir le vocabulaire le plus adapté ?

Il existe plusieurs type de vocabulaire qu'on peut employer pour étiqueter, décrire des données. Le choix du vocabulaire dépend du type de donnée manipulé afin d'assurer la richesse des termes et leur cohérence, nous avons récapitulé les principaux vocabulaires existant dans le tableau ci-dessous :

Vocabulaire	Gouvernance	Communauté d'usage	Cas d'usage
RDFS	Norme W3C	Web	Décrire une ontologie
OWL	Norme W3C	Web	Décrire une ontologie
SKOS	Norme W3C	Web	Exprimer un concept, thésaurus Décrire des concepts
FOAF	Réseaux sociaux	Web	Décrire des personnes
Dublin Core	Dublin Core	Web et bibliothèques	Décrire des documents de manière simple DC Elements : 15 éléments descriptifs DC Terms : ontologie
ISBD	IFLA (Open Metadata Registry)	Bibliothèques	Décrire un document selon les éléments et les règles de l'ISBD

FRBR, FRBER, FRAD, FRSAD	IFLA (Open Metadata Registry)	Bibliothèques	Décrire les entités et les relations du modèle FRBR pour les données bibliographiques, les données d'autorité et les sujets
RDA	Communauté RDA	Bibliothèques	Décrire des informations catalographiques avec une granularité fine et les relations FRBR
OAI ORE	Communauté OAI	Bibliothèques, archives, musées	Décrire des relations entre des objets, notamment de type agrégation d'objets
Schema.org	Moteurs de recherche	Web	Favoriser l'indexation par les moteurs Intégration de métadonnées dans les pages web (html)

TABLE 1.2: Vocabulaire de description de données

- **RDFS (RDF Schema) et OWL (Web Ontology Language)** : RDFS et OWL sont deux vocabulaires RDF qui permettent de définir d'autres vocabulaires, c'est-à-dire de définir des classes, des propriétés, des hiérarchies de classes et de propriétés, et leurs comportement, OWL permet en outre de décrire une logique formelle facilitant l'exploitation des données par des machines : on parle alors d'ontologies. Il existe aussi dans OWL et RDFS des propriétés et classes très génériques qu'on utilise dans de très nombreux contextes, comme la propriété owl:sameAs qui permet d'exprimer une équivalence exacte entre deux ressources.[Lav16]
- **SKOS (Simple Knowledge Organisation System)** :SKOS est un vocabulaire RDF permettant de décrire des référentiels de type thésaurus. Il permet de décrire des concepts (en uti-

lisant la classe principale, SKOS :Concept) et d'exprimer les relations entre ces concepts (relations hiérarchiques – termes plus spécifiques ou génériques -ou autres – termes en relation). [Lav16] SKOS comporte également des propriétés appropriées pour décrire des résultats d'alignements automatiques ou manuels entre des concepts issus de thésaurus distincts (SKOS :closeMatch, SKOS :exactMatch).

- **FOAF (Friend of a friend)** :FOAF est un vocabulaire RDF destiné à la description des personnes et des liens qu'elles entretiennent entre elles. Principalement conçu pour exprimer les données de type réseaux sociaux, FOAF est un vocabulaire adapté et très répandu pour décrire les personnes d'une façon plus générale. [Lav16]
- **Dublin Core Metadata Terms** :Le Dublin Core est un vocabulaire de métadonnées développé par le DCMI (Dublin Core Metadata Initiative) pour exprimer des descriptions de documents. Le Dublin Core comprend deux ensembles :
  - le «Dublin Core element set » qui comprend les 15 propriétés de base (Title, Subject, Description, Source, Format, Type, Creator, Contributor, Publisher, Rights, Relation, Coverage, Language, Date et Identifier) ;
  - le « Dublin Core metadata terms » ou DC Terms qui reprend les 15 éléments de base, plus un certain nombre de classes et de propriétés supplémentaires. qui traitent les 15 éléments du Dublin Core Element Set comme des propriétés de la ressource décrite, et qui par ailleurs définissent, entre autres : (a) un certain nombre de nouvelles propriétés comme des sous-propriétés de ces propriétés de base, (b) un certain nombre de classes qui peuvent être associées à la ressource décrite par l'une de ces propriétés, et (c) une liste fermée de types de ressources.

Chaque classe et propriété du DC Terms possède une URI. Des informations sur l'utilisation des classes et propriétés sont également fournies, et leur hiérarchie est exprimée en RDFS.[Lav16]

- **ISBD(International standard for bibliographic description)** :La version consolidée de l'ISBD (2011) est publiée en RDF. C'est un ensemble normatif de règles validées au niveau international, pour la description bibliographique de toute ressource publiée existant dans les bibliothèques, quel qu'en soit le support. La version actuelle de référence est l'ISBD International standard bibliographic description : consolidated edition, publiée en 2011.
- **RDA (Resource Description and Access)** :RDA est un nouveau code de catalogage qui prend acte du modèle FRBR. Il est également utilisé comme vocabulaire pour exprimer des notices bibliographiques en RDF. Un premier travail a été effectué pour attribuer des URI et

créer une ontologie en OWL et déclarer les principales entités des FRBR.[dF16]

- **La famille des modèles FRBR (Functional Requirements for Bibliographic Records) :**Le modèle FRBR est un modèle conceptuel de données bibliographiques. Les relations entre les entités sont exprimées dans le vocabulaire FRBRER (FRBR Entité-Relation). Ce modèle a été étendu aux données d'autorité dans FRAD (Fonctionnalités requises pour les données d'autorité) et à la notion de sujet dans FRSAD (Fonctionnalités requises pour les données d'autorité matières).[dF16]

### 1.5 Ontologies

Cela fait un bon moment que nous parlons de l'aspect sémantique, et combien ce dernier est-il nécessaire pour la séparation du contenu de la présentation dans les pages web, ainsi même des machines pourront communiquer entre elles, l'ontologie est la base de toute sémantique dans le domaine du web. nous allons prêter attention à la façon dans laquelle elle opère.

#### 1.5.1 Qu'est-ce qu'une ontologie ?

L'histoire de l'intelligence artificielle a montré que la connaissance est un élément majeur pour les systèmes intelligents. Dans de nombreux cas, une meilleure connaissance peut être plus pertinente qu'un algorithme pour résoudre une tâche. Afin d'avoir des systèmes vraiment intelligents, la connaissance doit être capturée, transformée, réutilisée, et communiquée. Les Ontologies prennent en charge toutes ces tâches. Mais qu'est-ce qu'une ontologie ?

Le terme ontologie est emprunté de la philosophie où il fait référence à la science qui « étudie l'être en tant qu'être ». Par la suite ce terme a été pris par des chercheurs en intelligence artificielle, qui l'utilisent pour désigner les blocs qui construisent notre modèle du monde. Un agent d'un modèle particulier (robot par exemple) ne sera capable de percevoir que ce que l'ontologie est en mesure de représenter de ce dernier. Dans ce sens seuls les choses dans son ontologie peuvent exister pour cet agent. De cette façon, une ontologie devient le niveau le plus bas d'un système de représentation des connaissances.

Avec l'émergence de l'ingénierie des connaissances et du web sémantique, ce terme a pris une toute autre tournure pour désigner la problématique de représentation et de manipulation des connaissances dans un système informatique. [y08] Neches et ses collègues ont été les premiers à en proposer une définition : « une ontologie définit les termes et les relations de base du vocabulaire d'un

*domaine ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire* ». Cette définition descriptive donne un premier aperçu sur la manière de construire une ontologie, à savoir l'identification des termes de bases d'un domaine et les relations entre ces termes ainsi que les règles pouvant s'appliquer sur ces derniers.

Par la suite les ontologies ont été définies comme « *une spécification explicite d'une conceptualisation* » Ce que signifie qu'une ontologie est une description (comme une spécification formelle d'un programme) des concepts et des relations qui peuvent exister pour un agent ou d'une communauté d'agents dans un système donné.

En mettant l'accent sur l'importance du partage et la réutilisation des connaissances, cette définition a été reprise et modifiée : « *Une ontologie est une spécification formelle d'une conceptualisation partagée* » cette dernière a été expliquée comme suit :

**Conceptualisation** : réfère à un modèle abstrait d'un phénomène dans le monde, en ayant identifié les concepts appropriés à ce phénomène.

**Explicite** : signifie que le type de concepts utilisés et les contraintes liées à leur usage sont définis explicitement.

**Formel** : réfère au fait que l'ontologie doit être traduite en langage interprétable par une machine.

**Partagé** : réfère au fait qu'une ontologie capture la connaissance consensuelle, c'est-à-dire non réservée à quelques individus, mais partagée par un groupe ou une communauté.

Cette définition, qui reste la plus citée dans le domaine de l'informatique, permet d'une part, de voir l'ontologie comme une représentation abstraite des connaissances, et d'autre part, précise le fait que l'ontologie doit être formelle, exprimé sous forme d'une logique pouvant être manipulée par la machine.

On peut concevoir une ontologie comme une structuration de divers concepts spécifiques à un domaine. Ces concepts sont rassemblés pour fournir les éléments de base et exprimer les connaissances dont on dispose dans ce domaine. Ces concepts sont reliés entre eux par ce que l'on appelle des relations, qui sont en général des actions. Par exemple, dans « un auteur écrit un livre », les deux concepts sont « auteur » et « livre » et la relation est « écrit ». On peut aussi voir ça dans l'autre sens : « un livre est écrit par un auteur », les deux concepts sont toujours les mêmes mais la relation est maintenant « est écrit ». On dit alors que la relation est symétrique. [Lav16] Les ontologies font partie intégrante des normes du W3C pour le Web sémantique, dans lequel elles sont utilisées pour spécifier la norme des vocabulaires conceptuels dans lequel l'échange de données entre différents systèmes a pour but de fournir :

- Des services pour répondre aux requêtes ;
- De publier des bases de connaissances réutilisables ;

- D'offrir des services pour faciliter l'interopérabilité entre plusieurs systèmes hétérogènes et bases de données.

Le rôle clé des ontologies par rapport aux systèmes de base de données est de spécifier une représentation de modélisation des données à un niveau d'abstraction au-dessus des schémas d'une base de données spécifique (logique ou physique), afin que les données puissent être exportées, traduites, interrogées et unifiées pour tous les systèmes développés de manière indépendante.

### 1.5.2 Éléments(Composante) d'une ontologie

L'ontologie n'est qu'une modélisation du monde réel en concept et relation entre ces concepts.

Les principales composantes qu'on peut distinguer sont donc les suivantes :

- **Concept** : C'est la représentation abstraite des éléments du domaine. On peut également les appeler termes ou classes. Ces concepts peuvent être classés selon différents critères dans la taxonomie (niveau d'abstraction, atomicité, etc.)
- **Relation** : traduisent les associations existant entre les concepts présents dans le segment analysé de la réalité, ces relations incluent les associations suivantes :
  - Relations
  - Sous-classe-de (généralisation – spécialisation)
  - Partie-de (agrégation ou composition)
  - Associée-à
  - Instance-de

Ces relations nous permettent d'apercevoir la structuration et l'interrelation des concepts, les uns par rapport aux autres.

- **Les fonctions** : constituent des cas particuliers de relations, dans laquelle un élément de la relation, le nième est défini en fonction des n-1 éléments précédents.
- **Axiomes** : Constituent des assertions considérées toujours comme vraies.
- **Instances** : Ce sont des exemples particuliers de concepts.

### 1.5.3 Type d'ontologie

Les éléments de l'ontologie (concepts, relations, axiomes et instances) sont définis explicitement avec un langage dont la sémantique est plus ou moins formelle, dépendant du degré d'abstraction voulu. Par conséquent, les ontologies présentent des degrés d'abstraction différents dépendant du domaine auquel elles sont dédiées. On identifie plusieurs types d'ontologies selon leur niveau d'abs-

traction, on parle ici d'une Typologie selon l'objet de conceptualisation.

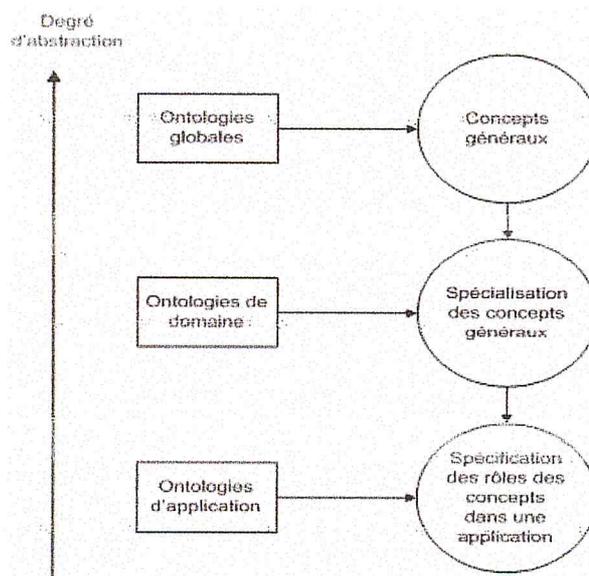


FIGURE 1.5: Typologie des ontologies selon l'objet de conceptualisation

**Les ontologies de représentation :** Ce type d'ontologies regroupe les concepts utilisés pour formaliser les connaissances. Parmi les ontologies de représentation, on trouve des ontologies qui vont décrire les notions utilisées dans toutes les ontologies pour spécifier les connaissances, telles que les substances, les concepts, les relations etc. Par exemple, la « Frame-Ontology » est une ontologie de représentation. Elle définit de manière formelle les concepts utilisés principalement dans les langages à base de frames : classes, sous-classes, attributs, valeurs, relations et axiomes. Les ontologies de représentation sont indépendantes des différents domaines de connaissances, puisqu'elles décrivent des primitives cognitives communes aux différents domaines.

**Les ontologies globales (ontologies de haut niveau) :** c'est-à-dire celles qui présentent le plus haut niveau d'abstraction et de généralité, sont les ontologies formelles, car elles sont issues d'un développement systématique, rigoureux et axiomatique de la logique de toutes les formes et modes d'existence. L'adoption de principes rigoureux dans la conception de l'ontologie formelle répond au besoin de disposer de connaissances pouvant être partagées et transférées d'un contexte à l'autre. Elles sont dédiées à des utilisations générales (ex : Word Net). Une ontologie formelle est donc une théorie des distinctions formelles entre les éléments d'un domaine, indépendamment de leur réalité.

**Les ontologies de domaine :** qui sont limitées à la représentation de concepts dans des domaines donnés (géographie, médecine, écologie, etc.) et qui spécialise les concepts de l'ontologie globale.

**Les Ontologies de Tâches :** Ce type d'ontologies est utilisé pour conceptualiser des tâches spécifiques dans les systèmes, telles que les tâches de diagnostic, de planification, de conception, de confi-

guration, de tutorat, soit tout ce qui concerne la résolution de problèmes. Elle régit un ensemble de vocabulaires et de concepts qui décrivent une structure de résolution des problèmes inhérente aux tâches et indépendante du domaine.

**Les ontologies d'applications :** offrent le plus fin niveau de spécificité, c'est-à-dire qu'elles sont dédiées à un champ d'application précis à l'intérieur d'un domaine et décrivent le rôle particulier des entités de l'ontologie de domaine dans ce champ. Par exemple, l'ensemble des spécifications sur la forêt de Montmorency constitue une ontologie d'application qui spécifie les concepts généraux pouvant provenir d'une ontologie de domaine forestier générale.

### 1.5.4 Cycle de vie d'une ontologie

La construction d'une ontologie doit passer par de différentes phases, à commencer par l'analyse des besoins, jusqu'à la validation et l'évaluation, et ce, comme le montre la figure ci-dessous.

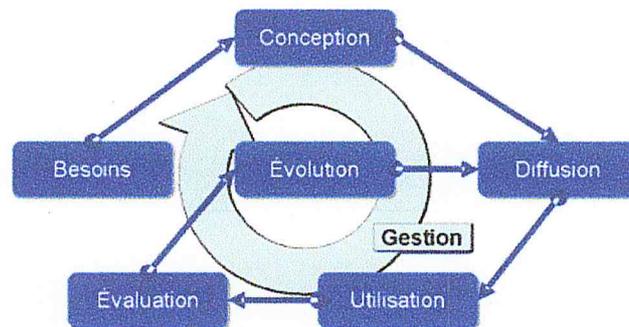


FIGURE 1.6: Cycle de vie d'une ontologie

Cette élaboration doit passer par une phase d'ingénierie qui détermine son cycle de vie. Étant donné que ces ontologies seront utilisées comme composantes de systèmes logiciels, leur développement doit s'appuyer sur les mêmes principes que ceux utilisés en génie logiciel. On peut alors faire émerger chaque étape de son cycle de vie : chacune d'elle pose des problèmes de recherche, il convient donc d'expliquer à quoi elles correspondent. Ce cycle de vie rassemble sept activités : détection des besoins, conception, gestion et planification, évolution, diffusion, utilisation, évaluation qu'on peut rassembler en quatre Phases principale :

#### A/ Phase d'analyse des besoins :

L'activité de cette phase se résume en l'identification des objectifs, des limites et des contraintes, ainsi qu'à l'étude du domaine d'application autour duquel s'articule le système en cours d'élaboration. Cette étape est pour l'ingénierie de l'ontologie ce qu'est l'analyse des besoins pour le

génie logiciel et se doit de donner une réponse aux différentes questions qui tournent autour du domaine d'application, de l'objet, de l'objectif, de la durée de vie et de la langue utilisée pour cette ontologie.

Par ailleurs, un rapport récapitulatif, dit rapport d'analyse, sera élaboré à la fin de cette phase. Celui-ci servira d'appui pour la conception de l'ontologie au cours de la phase suivante.

### **B/ Phase de conception :**

Cette phase fait suite à celle de l'analyse des besoins et consiste à concevoir l'ontologie qui servira comme base de raisonnement du système en cours d'élaboration. Cette tâche n'est pas du seul rôle du développeur. En effet, il est extrêmement recommandé de faire intervenir des experts du domaine pour garantir la validité et la complétude de l'ontologie en question. La phase de conception se déroule en général sur deux étapes à savoir, la collecte d'informations et l'identification des termes :

#### **– Collecte d'information**

Il s'agit de collecter le maximum d'information sur le domaine d'application du système en cours d'élaboration. Ces données sont alors appelées connaissances du domaine. Cette tâche est souvent guidée par un expert du domaine, des documentations ou d'autres ontologies. Il faut noter, par ailleurs que l'on fait souvent appel aux techniques de génie logiciel pour assurer cette tâche, notamment, les interviews, les documentations formelles ou informelles et autres.

#### **– Identification des termes**

Le développeur, en collaboration avec un expert du domaine, entame cette tâche aussitôt que les informations sur le domaine d'application seront réunies. Cette tâche consiste plus exactement à retracer une abstraction du monde réel, sous forme de concepts et de relations entre ces concepts. Pour des fins de simplification et de clarté, il est vivement recommandé de représenter cette abstraction en langage naturel, avant de commencer à la coder de façon formelle.

### **C/ Phase de développement :**

Dans la phase précédente (Conception) les connaissances du monde réel étant représentées en concepts et relations entre ces concepts, il ne reste plus à présent qu'à les coder d'une façon formelle. Une fois le développement réalisé, Le code source résultant de cette phase doit être bien entendu enregistré dans un document qui portera l'extension « .OWL ». Celui-ci servira de référence pour le partage et la réutilisation des connaissances sur le monde réel, d'une part, et comme base de raisonnement pour les moteurs d'inférence du système d'autre part.

Il existe une multitude d'outils pour le développement d'ontologies qui diffèrent selon leurs faiblesses et leurs points forts. Ces derniers sont appelés dans le jargon informatique EDI (Environnements de développement intégrés). En d'autres termes, il s'agit d'un environnement permettant d'assister le développeur dans sa tâche de programmation ou de génération du code source. Parmi les EDI les plus utilisés pour l'élaboration des ontologies, nous citons : CHIMAERA, DAG-EDIT, OILED, PROTEGE C'est d'ailleurs ce dernier qui est le plus utilisé.

### D/ Phase de validation et évaluation de l'ontologie :

La validation de l'ontologie, de façon générale, doit se faire au regard des tâches qui lui sont attribuées, dans la mesure où cette étape de l'ingénierie ontologique ne peut se concevoir que par l'usage. Cependant, et comme il est absurde de demander aux utilisateurs de valider un à un tous les termes définis dans cette ontologie, nous avons jugé plus judicieux d'effectuer une série de tests. Il s'agit en effet de vérifier pour chaque test, si les termes retournés comme résultats répondent à la requête de l'utilisateur ou non.

Par ailleurs, monsieur Yang propose une autre technique d'évaluation de l'ontologie du domaine. Celle-ci est plus conforme à la méthode de méta-modèle utilisée pour l'élaboration de l'ontologie et se distingue des méthodes classiques par des critères spécifiques d'évaluation. Il est à rappeler que cet auteur a lui-même adopté la même méthodologie dans l'élaboration de l'ontologie du domaine pour son système de recommandation de cartes de crédit, basé sur la technologie du web sémantique. Parmi ces critères, nous citons :

#### – Déterminer les limites du domaine d'application :

Il est très difficile de cerner les limites du domaine et d'affirmer le niveau d'abstraction que doit couvrir cette ontologie. L'essentiel est que cette ontologie puisse être utilisée par tous les utilisateurs et qu'elle définisse suffisamment de concepts et de relations entre ces concepts afin de limiter la dépendance vis à vis de l'expert du domaine.

#### – Souligner l'importance de l'analyse des besoins :

Il est très important de vérifier si l'ontologie satisfait tous les besoins cités dans le rapport d'analyse. Si tel est le cas, l'ontologie répond alors aux attentes des usagers, sinon il faudrait revoir la conception.

#### – Être attentif à l'évolution de l'ontologie :

Que ce soit en génie logiciel, ou en ingénierie des connaissances, l'évolution du système est un facteur d'une importance capitale à étudier de près. Il en est autant pour l'ontologie dans le cadre de l'ingénierie ontologique.

#### – Assurer la mise à jour de la documentation :

La documentation doit évoluer dans le même sens et à la même vitesse que l'élaboration de

l'ontologie. Il s'agit de donner des comptes-rendus tout au long des différentes phases. Le rapport d'analyse des besoins, le rapport de conception et le rapport de développement font tous partie de cette documentation.

– **Assurer le partage et la réutilisation des connaissances :**

Il est à rappeler que le principal objectif de l'ontologie est de décrire un domaine de façon abstraite et formelle en concepts et relations entre ces concepts, afin de permettre un meilleur partage et une plus large réutilisation des connaissances. Il s'agit de deux idées clés dans la philosophie de la mise en place de l'ontologie. Au cours de la validation, le développeur doit donc s'assurer que son ontologie couvre tout le domaine, dans la mesure du possible, afin de ne plus dépendre d'un expert du domaine.[y08]

### 1.5.5 Méthode et Méthodologies d'ingénierie ontologique

Une méthodologie est une succession, compréhensible et intégrée, des méthodes et des techniques qui créent ensemble un système théorique explicatif pour l'accomplissement d'une tâche cognitivement complexe.

Il faut savoir que Jusqu'en 1996, les premières ontologies ont été développées de façon complètement artisanale, sans suivre de méthode prédéfinie, c'est après une suite de projet (Mikrokosmos, Enterprise Ontology, TOVE, MENELAS) que des listes de recommandations constituant des ébauches de méthodes, ou cadres méthodologiques ont vu le jour.

La conception d'ontologies est une tâche difficile nécessitant la mise en place de procédés élaborés afin d'extraire la connaissance d'un domaine, manipulable par les systèmes informatiques et interprétable par les êtres humains. Deux types de conception existent :

1. **La construction manuelle :** Ces principes se basent sur des fondements philosophiques et suivent des procédés de modélisation collaboratifs. Ils mènent à la conception d'ontologies dites légères et d'ontologies dites lourdes (ces ontologies se distinguent par la présence ou non d'axiomes). Cependant, ce procédé de génération est très coûteux en temps et pose surtout des problèmes de maintenance et de mise à jour, parmi ces méthodes on peut citer les suivantes :
  - **La méthode d'Uschold et King**
  - **La méthode de Grüninger et Fox**
  - **La méthode basée sur SENSUS**
  - **La méthode METHONTOLOGY**
  - **La méthode de KACTUS**

2. **La construction automatique** : c'est une conception qui repose sur des apprentissages, elle commence à émerger comme un sous-domaine de l'ingénierie des connaissances. Face à la masse croissante de documents présents sur le Web et aux avancées technologiques dans le domaine de la recherche d'information, de l'apprentissage automatique et du traitement automatique des langues, de nouveaux travaux portent sur la recherche de procédés plus automatiques de génération d'ontologies. Ces mécanismes mènent généralement à la conception d'ontologies dites légères. Différents types d'approches sont distingués en fonction du support sur lequel elles se basent : à partir de textes, de dictionnaires, de bases de connaissance, de schémas semi-structurés et de schémas relationnels, parmi ces méthodes on peut citer les suivantes :

- La méthode ARCHONTE
- La méthode TERMINAE

### 1.5.6 La méthode METHONTOLOGY

Nous avons vu dans la section précédente qu'avant de se lancer dans la construction d'une ontologie, il faut d'abord tracer le chemin que nous devons suivre pour le faire, ou ce qu'on appelle le processus de construction de notre ontologie, pour faire cela nous avons vu qu'il n'existe pas qu'une seule méthodologie « correcte » pour développer des ontologies, et que chaque méthode répond à des besoins spécifiques et a ses propres avantages, dans ce qui va suivre et à titre d'exemple nous allons essayer de présenter les points essentiels de la méthodologie « METHONTOLOGY ».

Notre choix s'est porté sur cette méthodologie parmi plusieurs autres que nous avons vues pour plusieurs raisons, tout d'abord cette méthodologie est utilisée pour la construction des ontologies à partir de zéro sans être basée sur des ontologies déjà existantes avec une conception réalisée manuellement selon nos besoins ce qui est le cas dans notre projet, en plus de cela cette méthodologie présente des avantages majeurs qui la rendent complète du point de vue du processus de développement ou elle présente un certain nombre de phases spécifiées de manière très détaillée entre autres la phase de conceptualisation, ainsi que la phase d'évaluation qui permet de vérifier la cohérence de notre ontologie avec le monde qu'elle modélise et ça grâce à une liste de critères de validation et de vérification englobés dans l'outil de construction d'ontologie « Protégé » que nous allons voir par la suite.

Parler de ces deux phases nous pousse à mettre en lumière le cycle de vie de METHONTOLOGY, qui est basé sur des prototypes évolutifs et apporte un plus lors de la construction de l'ontologie du moment qu'on peut passer d'une phase à une autre à n'importe quel moment de notre conception si on est amené à faire des modifications ou des définitions incorrectes ou incomplètes du coup de

cycle de vie permet l'inclusion, la suppression ou la modification des définitions de l'ontologie à tout moment comme il est présenté dans la figure ci-dessous :

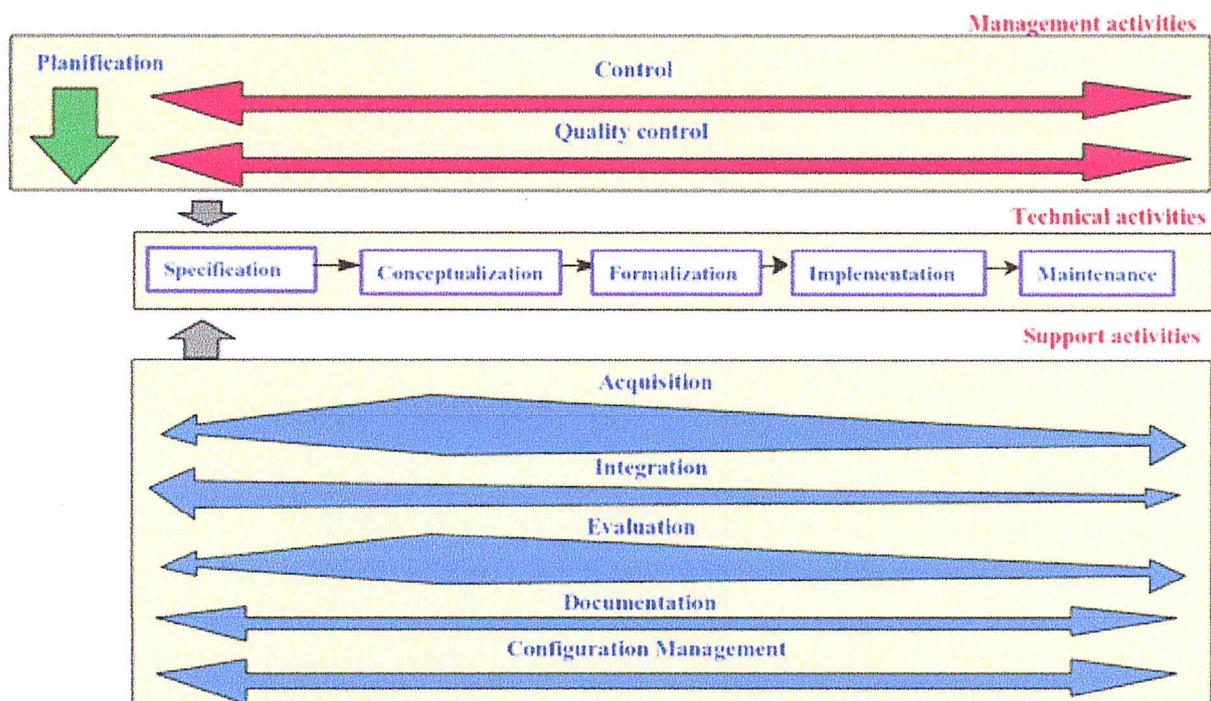


FIGURE 1.7: Figure Processus de développement et le cycle de vie de METHONTOLOGY.

De ce cycle de vie on peut dégager trois axes majeurs sur lesquelles notre méthodologie se repose. Le premier axe « **Management activities** » ou activités de gestion de projet consiste à étudier en manière profonde l'étendu de notre projet et les probables problèmes qu'on peut faire face par la suite, afin d'assurer une qualité à notre travail, cette partie on peut la résumer en trois mots : (prévision, contrôle, assurance de qualité).

Le deuxième axe et pas le moindre « **Development activities** » ou activités orientées-développement consiste à présenter les différentes phases de conception de notre ontologie (spécification, conceptualisation, formalisation, implémentation, maintenance) de manière très détaillée. Ces phases sont le cœur de notre méthode et ont combinaison avec les règles et les conseils de l'ingénierie ontologique on a réussi à dégager un processus efficace en matière de construction d'ontologies à partir de zéro dans un domaine donnée, l'importance de cette partie nous pousse à voir en détail les différentes étapes et phases dans ce qui va suivre surtout que la construction de notre ontologie repose majoritairement sur ces derniers.

1. **Spécification** : Pour commencer la conception de notre ontologie, il est primordial de définir et de spécifier les besoins de celle-ci ainsi que son domaine et sa portée, ça revient à dire qu'il faut répondre à des questions de base telle que :

- Quel est le domaine que va couvrir l'ontologie ?
- Dans quel but utiliserons-nous l'ontologie
- A quels types de questions l'ontologie devra-t-elle fournir des réponses ?
- Qui va utiliser et maintenir l'ontologie ?

Les réponses à ces questions peuvent varier au cours du processus de la conception de l'ontologie, mais à chaque moment donné, elles aident à limiter la portée du modèle et se rapproché de nos besoins.

2. **Conceptualisation** : Une des plus importantes étapes de la conception si ce n'est pas la plus importante, ou on vise à structurer les connaissances du domaine en un modèle conceptuel, en d'autres termes cette étape nous permet le passage de l'informelle aux structures formelles. Les principales tâches suivantes sont réalisées :

- **Tâche1** : Enumérer les termes importants dans l'ontologie sans se soucier des relations et les propriétés qu'il peut y avoir entre ces différents termes pour avoir au final ce qu'on appelle un glossaire de termes ;
- **Tâche2** : définir les classes et la hiérarchie des classes ou ce qu'on appelle « Taxonomie » ce qui facilite la recherche d'un terme en fonction de ses relations hiérarchiques avec d'autres termes, il faut noter qu'il existe un certain nombre d'approches hiérarchiques :
  - (a) Approche ascendante (bottom-up strategy) : les concepts les plus spécifiques sont identifiés, par la suite, ils sont généralisés en concept plus abstraits.
  - (b) Approche descendante (top-down strategy) : les concepts les plus généraux sont identifiés, par la suite, ils sont spécialisés en plus spécifiques.
  - (c) Approche centrifuge (middle-out strategy) : les concepts les plus importants sont identifiés (centraux), par la suite, ils sont généralisés et spécialisés comme il est nécessaire.
- **Tâche3** : construire le diagramme de relations binaires où les relations entre les concepts de l'ontologie sont identifiées ;
- **Tâche4** : construire le dictionnaire de concepts qui inclut principalement les instances de chaque concept, leurs attributs et leurs relations ;
- **Tâche5** : décrire en détail dans une table de relations binaires les relations apparues dans le diagramme de relation binaires ;
- **Tâche6** : décrire les attributs apparus dans le dictionnaire de concepts en spécifiant leur contraintes dans une table d'attributs ;
- **Tâche7** : spécifier les axiomes sur les concepts dans une table d'axiomes logiques ;
- **Tâche8** : décrire les instances des concepts dans une table d'instances ; Il faut noter que

### 1.5.7 Outils et langage de construction d'ontologie

La réalisation des ontologies s'effectue au moyen de langages et plusieurs types de langages peuvent jouer le rôle de langage de spécification, allant des langues naturelles jusqu'aux langages de représentation opérationnels (exécutables).

Dans un projet de construction d'ontologie, il est courant d'utiliser plusieurs langages, adaptés aux différentes étapes de la construction.

- Pour acquérir les connaissances ontologiques, on peut faire usage de la langue naturelle ou d'un langage de modélisation informel.
- Si l'ontologie doit devenir un composant d'une application, on aura recours à des langages de représentation formels et/ou exécutables.

Jusqu'au milieu des années 90, seul existait le langage formel « Ontolingua » comme langage d'échange d'ontologies.

Depuis la fin des années 90, une nouvelle génération de langages voit le jour. Leur développement répond à plusieurs objectifs :

- améliorer le processus de construction des ontologies (ex : OCML, Mdos),
- échanger les ontologies sur le Web (ex : RDF(S), SHOE, XOL, OIL),
- intégrer différents paradigmes de représentation, comme les langages de Frames et les Logiques de description (ex : OIL, DefOnto),
- doter les langages de services inférentiels performants (ex : OIL, PowerLOOM, DefOnto).

[w3c16] Dans ce qui va suivre on va voir un aperçu sur différents outils qui sont fréquemment utilisés par la communauté :

- **ODE «ONTOLOGY DESIGN ENVIRONMENT»** : construction des ontologies au niveau connaissances (Méthodologie METHONTOLOGY).
- **ONTOEDIT «ONTOLOGY EDITOR»** : est indépendant de tout formalisme. Il permet l'édition des hiérarchies de concepts et de relations et l'expression d'axiomes algébriques.
- **PROTEGE2000** : Il permet l'édition, le contrôle, la visualisation et l'extraction d'ontologie à partir des textes.
- **ONTOLINGUA** : qui constitue une extension du langage KIF (KNOWLEDGE INTERCHANGE FORMAT). Ce serveur d'édition permet la fusion d'ontologie. L'ontologie est immédiatement représentée dans un formalisme.
- **OILED «OIL EDITOR»** : est un éditeur d'ontologie s'inspirant du formalisme OIL. Un éditeur de petites ontologies avec un moteur d'inférence pour tester la cohérence de l'ontologie construite.

- **RDF et RDF Schéma « Resource Description Framework »** : est un formalisme graphique pour représenter des méta-données. Il est basé sur la notion de triplet (sujet, prédicat, objet). Le sujet et l'objet sont des ressources liées par le prédicat.

RDF utilise la syntaxe XML, mais il ne donne aucune signification spécifique pour le vocabulaire comme sous classe de, ou le type. Les primitives de modélisation offertes par RDF sont très basiques.

RDF Schéma est un langage qui étend RDF avec un vocabulaire de termes et les relations entre ces termes, par exemple : Class, Property, type, subclassOf, subPropertyOf, range et domain.

RDFS est reconnu comme un langage d'ontologie qui définit :

- Des classes et des propriétés.
- Les sous-classes, les superclasses, les sous-propriétés, et les super-propriétés.
- Le domaine de définition et le domaine image des propriétés.

[w3c16]

- **OWL (Web Ontology Language)** : est le standard actuellement recommandé par W3C pour représenter les ontologies. C'est une extension du vocabulaire de RDF(S). Il est dérivé du langage d'ontologie DAML+OIL.

OWL a été fractionné en fait en trois sous-langages distincts offrant une complexité croissante et destinés à des communautés différentes d'utilisateurs. Chacun est une extension par rapport à son prédécesseur plus simple. Ils sont caractérisés par l'expressivité croissante (d'OWL Lite à OWL FULL) et la calculabilité des inférences décroissantes (d'OWL Lite à OWL FULL).

- **OWL LITE** : est d'expressivité faible par rapport aux autres sous langages, mais qui reste quand même suffisant pour des utilisateurs qui ont principalement besoin d'une hiérarchie de classification et de contraintes simples. Par exemple, malgré qu'il permet d'exprimer des contraintes de cardinalité, il limite leurs valeurs à 0 ou 1. Une cardinalité 0 ou 1 correspond à des relations fonctionnelles, par exemple, une personne a une adresse. Toutefois, cette personne peut avoir un ou plusieurs prénoms, OWL Lite ne suffit donc pas pour cette situation. Ainsi, le langage offre une calculabilité maximale (temps de calcul acceptables), ce qui compense sa faible expressivité.
- **OWL DL** : nommé DL car il correspond à la logique descriptive. Il est d'expressivité maximale sans perte de calculabilité. Il convient aux utilisateurs qui veulent une expressivité maximale tout en maintenant la complétude de calcul (toutes les inférences sont garanties calculables) et la décidabilité (tous les calculs s'effectuent dans un temps fini). OWL DL inclut tous les constructeurs du langage OWL, mais ils sont utilisables seulement sous cer-

taines restrictions pour garantir la décidabilité des calculs. Par exemple, lorsqu'une classe est sous classe de plusieurs classes, elle ne peut pas être instance d'une autre classe.

- **OWL FULL** : offre un maximum d'expressivité. Il a l'avantage de la compatibilité complète avec RDF/RDFS, mais aucune garantie concernant la calculabilité (la complétude et la décidabilité des calculs liés à l'ontologie) n'est offerte par ce langage.

### 1.6 Linked data

Jusqu'au la nous avons parlé des technologies qui permettent la révolution du web classique, tandis qu'il reste une dernière, qui figure comme une technologie majeure et qui promet un avenir meilleur pour le web sémantique. Supposons qu'on cherche à faire connecter deux ontologies pour étendre l'espace de données, d'interroger non seulement la donnée elle-même mais aussi les relations entre ces données, les avantages de cette manipulation seront plus que bénéfiques pour les utilisateurs du web sémantiques. Voyons comment cela peut arriver ? Et qu'est ce qui nous permet de mettre en œuvre le web de données

#### 1.6.1 Linked data, le web des données

Linked data, cette technologie devient le cœur du web sémantique, on l'utilise pour l'intégration à grande échelle de données, on parle du web de données contenant un nombre de données très important et qui sont liées entre eux, sa permet de laisser de côté le fait de collecter de petits ensembles de données. Pour que le web de données devienne une réalité il faut non seulement avoir un très grand nombre de données mais aussi ces derniers doivent être dans un format standard qui soient accessibles et gérables par les outils du web sémantiques. Le but de cette technologie est de permettre au gens de partager de données structurées aussi facilement que comme ils le font avec le partage de document. [N.P14]

#### 1.6.2 Le principe du Linked data

L'objectif du web sémantique est relier les données sur le web et de les rendre interopérables. Pour obtenir des données liées ou des Linked Data, quatre principes doivent être respectés [af11] :

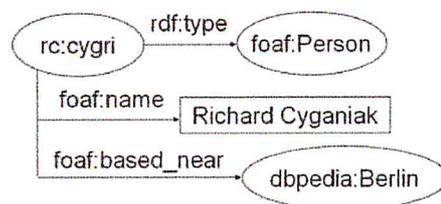
1. Utiliser des adresses URI pour identifier les choses ;
2. Utiliser des adresses URI HTTP pour que l'on puisse consulter ces identifications ;

3. Fournir des informations utiles sous forme de standards (RDF, SPARQL) lors d'une recherche d'adresse URI
4. Inclure des liens vers d'autres adresses URI qui permettent de découvrir d'autres informations.

En d'autres termes, il faut donc attribuer des identifiants aux ressources sous forme d'adresses URI HTTP, les décrire en utilisant des standards du web sémantique et créer des liens vers des données externes pour créer un réseau. Si plusieurs acteurs du web utilisent les mêmes standards et les mêmes modèles, il devient plus facile d'échanger, de réutiliser et de lier les données. Mais ce scénario n'est possible qu'à condition que les données soient disponibles sous licence ouverte. C'est pour cela que les Linked Data à elles seules ne suffisent pas dans l'optique du web sémantique ; elles doivent également être des Open Data, ou données ouvertes en français. Ce second concept est défini ainsi par l'Open Knowledge Foundation (2012) : « Une donnée ouverte est une donnée qui peut être librement utilisée, réutilisée et redistribuée par quiconque - sujette seulement, au plus, à une exigence d'attribution et de partage à l'identique. » Par déduction, cette définition stipule donc qu'une donnée ouverte doit pouvoir être réutilisée à des fins commerciales également. [Fou16] De manière générale, pour faciliter la réutilisation des données, trois aspects doivent être pris en compte (Chignard 2012 : p. 13-14) :

1. Aspect technique : un format le plus ouvert possible
2. Aspect juridique : une licence ouverte
3. Aspect économique : des pertes nulles ou limitées

Maintenant que nous savons de quoi s'agit le web de données, nous allons expliquer son principe en exemple et bien montré son avantage. Supposons que nous avons deux graphes RDF, le premier est pour une personne qui s'appelle Richard Cyganiak et qu'elle habite à Berlin, un autre pour Berlin qui a comme informations le nombre d'habitants et le pays dont elle appartient qui est l'Allemagne. Linked data permet d'assurer la relation de ces deux graphes à l'aide des URI depuis des sources de données différentes. Les figures qui suivent illustrent le principe de Linked data. [af11]



GET /resource/Berlin HTTP/1.0  
Accept: application/rdf+xml

FIGURE 1.8: Graphe RDF pour person :Richard Cyganiak

La figure précédente montre une personne qui s'appel Richard Cyganik et qui habite a Berlin

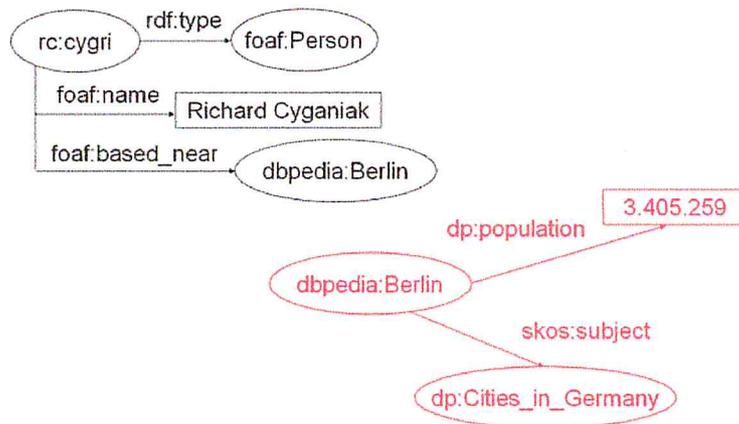


FIGURE 1.9: Graphe RDF pour Berlin

La figure précédente montre un autre graphe de Berlin, Berlin est une ville de Germany et elle a 3.409.259 habitants, Berlin est aussi la ville de Richard, alors on peut relier ou bien aligner ces deux graphes.

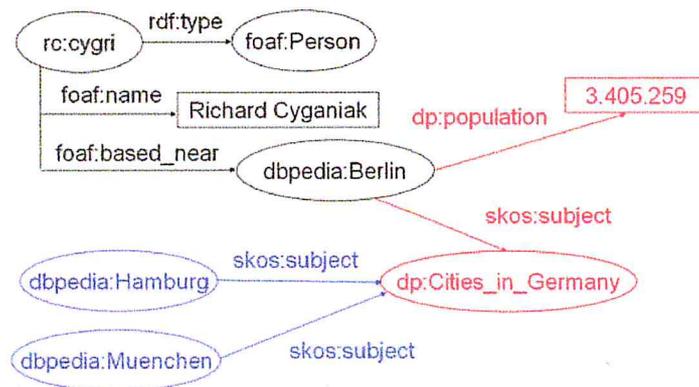


FIGURE 1.10: Principe de Linked data

Bien évidemment Germany n'a pas qu'une seule ville, la figure montre la force de Linked data, ou on peut interroger et les données et les associations entre les données. mais aussi pourquoi aligner des données permet de dire bienvenue à un web de donnée.

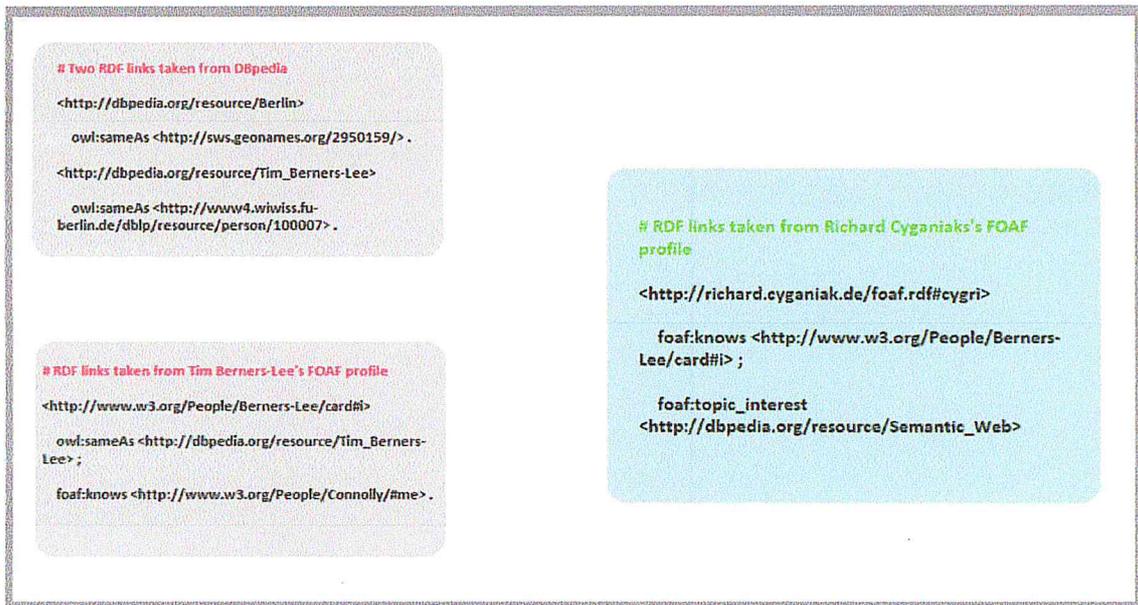


FIGURE 1.11: Langage OWL des trois graphes

### 1.6.3 Les avantages de Linked data

Après avoir montré le principe de données liées ou bien alignées, voilà les avantages que nous apporte cette technologie :

- Les clients peuvent consulter tous les URI dans un graphe RDF sur le Web pour récupérer des informations supplémentaires.
- Informations provenant de différentes sources fusionnent naturellement.
- Le modèle de données vous permet de définir des liens RDF entre les données provenant de différentes sources
- Le modèle de données vous permet de représenter l'information qui est exprimée à l'aide des schémas différents dans un seul modèle.[dF16]

### 1.6.4 les caractéristiques qu'il vaudrait mieux éviter

La technologie Linked-data, est très attirante et promet beaucoup pour l'avenir du web, en effet il existe quelques caractéristiques en RDF qu'il vaudrait mieux éviter et ce sont :

- l'utilisation de nœuds anonymes. Il est impossible d'établir des liens externes RDF à un nœud vide, et la fusion de données provenant de différentes sources devient beaucoup plus difficile lorsque les nœuds vides sont utilisés . Par conséquent , toutes les ressources d'une certaine importance doivent être nommées en utilisant les références URI
- l'utilisation de triplet avec le même prédicat rends l'utilisation de Sparql pas simple à réaliser

### 1.7 conclusion

Nous voilà arriver maintenant a la fin de ce deuxième chapitre, nous avons vue de diverses technologie, expliquer en détail les couche mises en place par le modèle W3C, et démontrer les technologies que prenait chaque couche, nous avons aussi profiter de ce chapitre pour donner une brève idée de ce qu'est l'alignement de donnée ou bien la notion de Linked data et pourquoi c'est si nécessaire d'utiliser le web de donnée. dans le chapitre suivant, nous allons montrer une conception de notre ontologie et une présentation des technologies annoncées précédemment

# Chapitre 2

## Le portail documentaire

### Contents

---

<b>2.1 Introduction</b> . . . . .	<b>39</b>
<b>2.2 Définitions</b> . . . . .	<b>39</b>
<b>2.3 Moteur de recherche fédérée</b> . . . . .	<b>40</b>
<b>2.4 Les enjeux du portail documentaire</b> . . . . .	<b>42</b>
<b>2.5 Le Portail sémantique</b> . . . . .	<b>45</b>
<b>2.6 Exemple de portails web sémantiques</b> . . . . .	<b>46</b>
<b>2.7 Conclusion</b> . . . . .	<b>49</b>

---

### 2.1 Introduction

Dans le domaine du web, le volume d'informations numériques n'arrête pas de nous impressionner, la richesse des documents numériques joue un rôle très important par rapport à l'évolution du web, une richesse qui se définit en quantité d'informations que reçoit un utilisateur après avoir exécuté sa requête et aussi le niveau de précision de sa recherche d'informations. Une fois qu'une information numérique est introduite, on doit s'attendre à ce que celle-ci soit demandée par un ou plusieurs utilisateurs, ou chacun ne cherche pas forcément les mêmes détails que l'autre. Comment gérer tout ça, et quelle sont les technologies utilisées par les développeurs de nos jours ?

### 2.2 Définitions

Lorsqu'on parle du web, on est souvent amené à utiliser des termes techniques comme portail web, moteur de recherche, agent web intelligents etc. Voici quelques définitions qui s'imposent pour mieux encadrer l'objectif de ce travail.

#### 2.2.1 Le terme portail

Avant d'avancer sur ce qu'est un portail documentaire, voyons d'abord ce que veut dire le mot portail, en effet ce terme est flou et très général ce qui nous oblige à lui ajouter de la précision. Une définition du mot portail a été proposée par le dictionnaire en ligne Business-Dictionary : «... Web site that serves as a gateway or a main entry point ('cyber door') on the internet to a specific field-of-interest or an industry. A portal provides at least four essential services, 1 : search engine(s), 2 : email, 3 : links to other related sites, and 4 : personalized content, ... Also called portal site or web portal. » [bus16] « Un site web qui sert de passerelle ou d'un point d'entrée principale sur internet pour spécifier les champs d'intérêts ou une industrie. Un portail a au moins quatre essentiels services, un moteur de recherche, la gestion des emails, les liens vers d'autre site connexes, et un contenu personnalisé, ... aussi appelé le site portail ou bien le portail web. » [bus16] Enfin, en une seule phrase, le terme portail web est une porte en ligne unique qu'une entreprise utilise pour proposer son produit. "Indeed, 'Portal' is one of the least helpful words we have developed in recent years as we come to terms with the changing information environment in which we research, learn and work" [BRE09]

### 2.2.2 Le portail documentaire

Initialement le portail documentaire est tout simplement un portail web avec des fonctionnalités spécifiques où les principaux services que peut offrir ce dernier à ses utilisateurs sont par exemple proposer des informations sur la vie d'une bibliothèque ou un ensemble de bibliothèques, des informations comme horaires d'ouverture / fermeture, ou peut être conditions de prêt, des actualités ou de proposer les services offerte par la bibliothèque elle-même. Ce type de portail se base sur l'interrogation simultanée des ressources hétérogènes alias la recherche fédérée, sur la gestion des contenu du web et la gestion du profil qui est tous ce qui est mis en avant selon les caractéristique du profil de l'utilisateur, mais aussi il se base sur l'utilisation des applications en ligne par exemple la réservation en ligne, acheter des billets, pour faire simple un portail documentaire est un portail web qui permet un accès à toutes sortes de ressources documentaire qui diffèrent dans leur thèmes et cela par une porte d'entrée unique. En fin, un portail documentaire est un portail qui permet de faire tomber les frontière géographique et de réduire les contraintes temporelles [Tai13]

## 2.3 Moteur de recherche fédérée

Un moteur de recherche fédérée est un outil de recherche qui envoie des requêtes à une ou plusieurs bases de données distantes et cela de manière simultanée, cet outil propose à l'utilisateur un formulaire de recherche unique, transmet la requête et enfin l'affiche pour l'utilisateur sur une page unique. Cette recherche fait partie des composant de bases d'un portail documentaire.

### 2.3.1 Fonctionnement

Le moteur de recherche fédérée utilise des fichiers pour traduire son information appelé Connecteurs, ces derniers font la traduction après avoir changé la syntaxe. La figure 1 est un petit schéma qui montre le déroulement d'une recherche fédérée faite par un utilisateur.

1. le connecteur récupère la requête de l'utilisateur selon la syntaxe du moteur de recherche.
2. le connecteur envoie la requête de l'utilisateur au moteur disant après avoir fait une traduction et modification de syntaxe.
3. le connecteur récupère les résultats des bases distantes dans une liste unique.
4. le connecteur affiche les résultats dans une page unique selon les règles de l'affichage du moteur de la recherche fédérée.

Le moteur de recherche fédérée fait souvent face à trois types de bases de données

- **Bases de données où il y accède par un Protocole web ;**
- **Bases de données où il y accède par le Protocole Z39.50 ;** cette norme concerne la recherche et la récupération d'informations au sein d'un ou plusieurs catalogues de bibliothèques, ou plusieurs bases de données. Protocole créé par des bibliothécaires américains et maintenue par la Library of Congress, elle permet une recherche précise sur un grand volume de données contenues dans des bases bibliographiques, elle permet aussi une gestion partagée des catalogues bibliographiques : une notice peut être récupérée sans que le catalogueur ait à connaître le mode d'accès à la base. La BnF offre ainsi la récupération des notices de son catalogue BnOpale Plus grâce au protocole clientserveur Z39.50 et enfin elle permet la mise en place d'OPAC et de catalogues collectifs (Sudoc, Catalogue collectif de France) dans lesquels l'utilisateur récupère un résultat global et homogène. [GD12]
- **Bases de données qui ne proposent pas un accès par Protocole :** Elle fonctionne à base d'une sorte de connecteur. Le connecteur décompose le codage, récupère une information et lui donne un sens en par une restructuration et en donnant un sens aux différents composants (titre, auteur, etc.).

Nous avons vu ce qu'est un connecteur et son rôle dans la communication d'un moteur de recherche fédérée, il faut noter qu'il est nécessaire de mettre en place autant de connecteur que de bases de données interrogeables, un moteur de recherche fédérée est dit performant si c'est un moteur avec un bon fonctionnement de ses connecteurs. [BRE09]

### 2.3.2 Exemples de moteur de recherches fédérée

**Le moteur de recherche Qwam Federator (Qwam-QES) :** Il est européen, c'est une solution de la recherche fédérée la plus complète du marché en raison de son intégration de tous les types de ressources informationnelles. La recherche s'applique à tous les bases de données, ainsi qu'aux applications du web. Noter que toutes les solutions Qwam-QES sont disponibles en mode hébergé sur la plateforme d'hébergement de Qwam Content Intelligence ou bien installés dans l'environnement informatique du client.

**Le moteur de recherche Dbwiz :** C'est un moteur de recherche fédérée libre développé par l'Université Canadienne Simon Fraser, DbWiz a été initialement développé pour aider les élèves à déterminer le meilleur point de départ pour leur recherche. On craignait que les utilisateurs novices de la bibliothèque soient submergés par le nombre croissant de ressources en ligne, et exigé un outil

en ligne pour les aider à l'absence d'un bibliothécaire. Ce moteur de recherche travail avec le protocole Z39.50. **Le moteur de recherche 360 Search** : Il est développé et commercialisé par la société Serials Solutions, implantée sur tous les continents. **Le moteur de recherche PazPar2** : C'est un moteur de recherche fédérée libre développé par Index Data, un groupe danois, implanté à Copenhague ainsi qu'aux États-Unis. [GD12]

## 2.4 Les enjeux du portail documentaire

Nous allons étudier les facteurs qui ont favorisé l'éclosion et le développement des portails documentaires.

### 2.4.1 L'hétérogénéité des ressources

Les bibliothèques proposent de nombreuses ressources hétérogènes, en effet sans compter les ressources possédées par la bibliothèque qui sont gérés par son système de gestion, on peut citer d'autres qui peuvent être classé entre ressources payantes (revues, journaux, e-books...), ou gratuites (catalogues, recherche d'étudiants, articles...), celles qui sont gratuites sont diffusées à une échelle plus grande tandis que celles qui sont payantes sont d'ordre plus important que les gratuites. Les ressources peuvent aussi être classées par types en considérant ce qu'elles présentent comme intérêt. Marc Maisonneuve ajoute une définition du portail documentaire, « Pour devenir un portail d'information documentaire, il faut que l'outil intègre quelques fonctions supplémentaires : choix de bases à solliciter dans une liste fermée, choix des accès à utiliser, consultation simultanée des bases choisies, fusion des résultats de la recherche et si possible affichage de ces résultats par ordre de pertinence » [Mai03]

### 2.4.2 Adaptation aux usages

Les bibliothèques universitaires, les instituts de formation et tous ce qui détient un ensemble de document numérisées à partager, doivent s'adapter à leur public, en sachant que les enjeux diffèrent de l'un à l'autre nous permet de dire qu'une connaissance de public ciblé doit être considérée. En effet si on a des étudiants de formation paramédicale on ne peut pas se permettre de cadrer les ressources numériques en ressources purement municipales. Or, savoir comment le public ciblé cherche son information revient à le connaître lui-même et à mieux opérer avec lui. Le public de la génération d'aujourd'hui est particulièrement issu d'une génération qui voit les bibliothèques comme un

lieu d'étude, une génération ou les ressources papier sont considérées comme de l'antiquité, Marc Maisonneuve démontre cela en faisant une enquête sur les pratiques culturelles et médiatiques générationnelles en 2007, cette enquête montre que notre public ciblé est en fait très vaste allant de celui qui veut se renseigner sur quelque chose dans un domaine précis vers celui qui fait de la vraie recherche pour faire évoluer son monde de pratique, ce qui fait que connaître son public cible est important pour réaliser son portail documentaire, une possibilité d'échange par conférences est la bienvenue d'où la raison du développement actuel de Learning center. L'idée est de permettre de lire une revue, un livre ou s'informer de quoi que ce soit tout en buvant du café de manière parallèle. [BRE09]

Comme mentionné précédemment, la génération d'aujourd'hui se base sur internet pour la majorité de ses services, le principe de résultat immédiat et dématérialisation de l'information fait en sorte de permettre au usage de consulter son fond en ligne (des articles visités, des documents, vidéos, etc) et cela de manière immédiate ce qui implique l'investissement dans des outils de web qui offrent ce genre de services, cela entre dans l'habitude de l'usage donc du public ciblé, désormais les nouvelles technologies détenues dans le web d'aujourd'hui rendent les utilisateurs plus exigeants, tout simplement ils veulent tout et tout de suite. L'accès à distance, des ressources disponibles et une époque de moteurs de recherche performants fait en sorte que l'usage se libère du temps d'une part mais aussi ajoute des contraintes aux développeurs de nouveaux portails documentaires d'une autre part, Ainsi comme le résume dans un commentaire un usager de la bibliothèque d'HEC lors de l'enquête LibQual de 2004, « We need easy access to all databases in English and in French from both school and home within two clicks of the library home page and three clicks from the intranet home page PERIOD. Anything less is unacceptable ». Telle est aujourd'hui l'attente extrême des usagers vis-à-vis des bibliothèques (Organiser l'accès électronique à des périodiques : Analyses des outils et proposition pour la mise en place d'une solution d'accès à la bibliothèque d'HEC). [GD12]

L'amélioration des services pour les utilisateurs est l'un des objectifs principaux de développement du portail documentaire ou sa conception fait en sorte que l'usage peut avoir des résultats qui diffèrent selon son profil. Cela est parmi les bases du portail documentaire SCD de l'université d'Artois bibliography.

### 2.4.3 La simplicité de la recherche

Cette notion est d'une difficulté majeure pour les développeurs de portails web documentaires, c'est ce qui permet à l'utilisateur d'avoir un maximum de résultats souhaités de n'importe quel endroit et le plus rapidement possible, une statistique du portail de l'université "Artois" montre que

les étudiants persistent à utiliser des recherche simples et des outils basics malgré que l'offre du portail et générale ainsi que ses ressources sont spécifiques et répondent à un maximum de besoin. « Les usagers se sont appropriés le portail documentaire dans sa fonction d'OPAC (90% des usagers se concentrent sur cet outil qui ne représente que 10% des outils mis en place). La plus part des recherches effectuées passent par l'encart « recherche simple » [...]» [Leb09]

Dans le portail documentaire JUBIL de l'université Pierre et Marie Curie les statistiques d'interrogation montrent les ressources accessibles sur la page d'accueil sont demandé fortement par rapport aux ressources qui ne sont qu'en recherche multi-ressources, enfin une recherche simple que comporte un portail documentaire le rend plus visité et apporte un aide à sa communautés d'utilisateurs et au public ciblé. Jusqu'au là nous avons vu ce qu'est un portail documentaire par une définition native, les enjeux du portail documentaire qu'on a résumé sur l'hétérogénéité des ressources, le choix du public ciblé, savoir opérer avec lui et trouver le moyen de mettre en œuvre une recherche simple et consistante. Avec de telles propriétés, un portail documentaire peut s'avérer complet mais il ne faut pas oublier que ce dernier pourra apparaitre dans plus qu'un seul endroit, les développeurs pensent en une architecture de briques métiers modulables qui se regroupe dans de différent environnements évolutifs.[Pap11]

### 2.4.4 Architecture du portail documentaire

En 1990, le noyau de l'informatique des bibliothèques était le SIGB, on n'avait aucun problème pour gérer au mieux les données de cette époque, toute étaient physique Or avec l'évolution des technologies informatiques et l'explosion des ressources documentaires l'architecture des bibliothèques est devenue inadaptée surtout lorsqu'il s'agit d'interroger des sources externes ou d'en émettre donc on a évolué vers un système plus performant. Dix ans plus tard les bibliothèques on évoluer, leurs système de gestion est plus ouvert en terme de ressources externe qu'interne cette évolution est l'évènement déclencheur de la conception du portail documentaire. Actuellement les développeurs cherchent à améliorer encore le portail documentaires en fait le portail de nos jours manque plus d'interopérabilité car il faut qu'il ait une plus grande capacité à fonctionner et à communiquer avec d'autres systèmes, aussi plus de portabilité cela revient à dire que le portail doit pouvoir évoluer dans des environnements différents et évolutif et enfin il manque de plus d'évolutivité, en effet ce le portail doit pouvoir être modifié de manière facile et rapide. [BRE09]

Conclusion, une architecture de brique est le plus grand pas que fait la bibliothèque , en effet, ce passage à une telle architecture qui est du à internet a permis au portail de voir le jour, ce n'est plus un système de gestion isolé d'une bibliothèque l'empêchant d'interagir avec d'autres bibliothèques

distantes mais plus tôt une architecture de brique de composant qui permette une évolution documentaire en facilitant la vie de l'utilisateur qui n'a pas besoin de se déplacer vers une bibliothèque pour se documenter. [GD12]

### 2.5 Le Portail sémantique

Le portail sémantique est tout d'abord un portail web comme tout autre portail, un login, gestion d'email des fonctionnalités selon son public ciblé et autres, sauf que ce dernier s'opère sur une autre catégorie de web, comme son nom l'indique, ce genre de portail utilise ce qu'on appelle le web sémantique c'est un web révolutionnaire plus satisfaisant. Pour le moment nous allons donner une brève idée de ce qu'est ce genre de portail.

Pour avoir un portail sémantique, il faut utiliser un ensemble de technologies et de standards, certains sont présents dans le web actuel et d'autres non.

#### 2.5.1 Les enjeux du portail sémantique

De plus que c'est un portail comme tout autre, le portail sémantique utilise quelques outils et technologies qui le rendent très satisfaisant aux besoins des utilisateurs. Aussi comme un portail classique, un portail sémantique se voit fixer ses propres défis dans le domaine du web. [yu11]

#### 2.5.2 La collecte d'informations

Supposons une ressource, quel que soit son type (une personne, une idée, un événement, un produit ou autre). Lorsqu'on utilise un portail sémantique on doit savoir tout ce qu'il est possible à savoir à propos de cette ressource. Donc il doit y avoir une collecte d'information de la ressource comme il doit y avoir une interrogation de ces informations collectées.

Pour réaliser une idée pareille, le web traditionnel n'aide pas beaucoup, supposons un agent web, un petit programme qui donne des instructions à un ordinateur quelconque pour récolter des informations sur une personne, il va donc falloir une URL pour accomplir cette tâche sauf que cet agent web dans un web classique n'aura pas une collection d'informations très intéressante puisque il va parcourir des tags comme `<b>` `<href>` `<table>` etc. Tandis qu'il existe des informations précieuses et intéressantes sur cette personne, à savoir son nom sa fonction, les personnes qui travaillent avec lui etc. Le web sémantique est révolutionnaire car il permet une bonne collecte d'informations ainsi même de simples agents web pourront faire communiquer des machines entre elles. [yu11]

## 2.6 Exemple de portails web sémantiques

Il existe déjà un ensemble de portails sémantiques publié sur internet, nous avons trouvé quelques exemples qui semblaient être intéressants et parmi ces portails, on a choisi de vous présenter les suivants.

### 2.6.1 Le portail sémantique SEAL

Comme son nom l'indique Seal (semantic portal) est un portail sémantique qui permet le partage et la structuration des informations entre les agents web et les humains, il est vu aussi comme une approche du web sémantique. En effet c'est une approche générique pour le développement de portails sémantiques, avoir cette approche permet d'avoir une sorte de modèle de base réutilisable a chaque fois quand cherche a développé un portail sémantique dans un cadre donné.

Parmi tant d'ontologies, Seal repose sur Onto-broker, une ontologie qui a pour but de rechercher les connaissances ainsi que de les partager. Elle représente le principal mécanisme d'inférence pour cette approche.[q01]

### 2.6.2 présentation de fonctionnement de SEAL

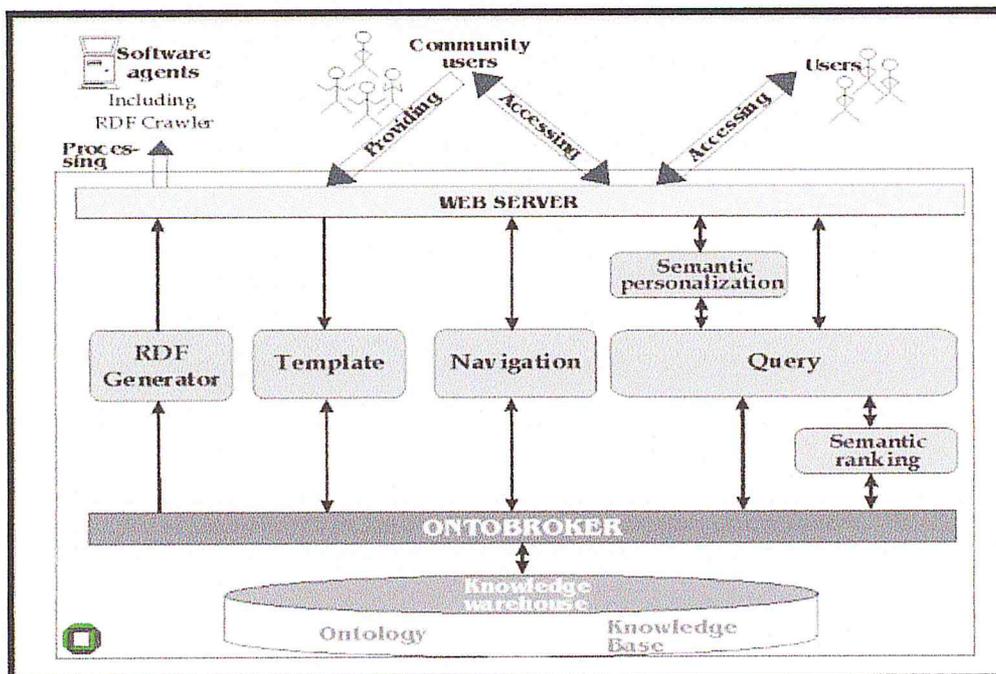


FIGURE 2.1: Architecture de Seal

Un Agent web, communauté d'utilisateurs ou des utilisateurs généraux sont trois modes qui interagissent avec le système à travers le serveur web. Un agent web intelligent « RDF-Crawler » peut traiter les informations stockées dans le portail, pour cela il existe un générateur RDF qui expose les informations structurées via le portail. L'application distante peut donc collecter les informations et aura un accès au stockage des connaissances sur le portail sémantique.

La communauté d'utilisateurs et les utilisateurs généraux peuvent accéder aux informations stockées dans le portail. Il existe deux formes d'accès, une Navigation simple en exploitant les liens HyperText et cherchant les informations en postant des requêtes, La structure de lien HyperText est en partie donnée par le constructeur du portail, mais il peut être prolongé avec l'aide de le module de navigation, celui-ci exploite les capacités d'inférence du moteur d'inférence a fin de construire des structure des lien HyperText conceptuels.

La Recherche et l'interrogation est effectuée par l'intermédiaire du module Query, de plus l'utilisateur du portail peut personnaliser sa recherche a travers le module Semantic Personalization comme il peut récupérer les informations grâce a la similarité sémantique avec le Module Sémantic Ranking, Les requête profitent de l'avantage qu'offre l'ontologie Ontobroker. Pour la communauté d'utilisateurs, ils peuvent introduire des données personnelles, ils profitent d'un accès administrateur où - à titre d'exemple - ils gèrent les domaines de publication mais le plus important c'est qu'il gère le composant principal du portail sémantique qui est l'ontologie.

L'administrateur remplit des formulaires et des modèles dans le portail et le module Template stocke les données dans le knowledge warehouse qui est une base de connaissances. [q01]

### 2.6.3 Le portail sémantique Kasimir

D'abord Kasimir est un portail sémantique en sorte de plateforme toujours en développement basé sur l'extraction, représentation et gestion de connaissances en cancérologie il est la pour but de faciliter la recherche et aider à la décision.

Au début le projet Kasimir visait un appui sur un ensemble de prototypes comme l'approche SEAL du web sémantique, aujourd'hui, sur Kasimir on cherche à poursuivre et enrichir ces prototypes, ainsi que de permettre une amélioration dans le développement à grande échelle du portail lui-même en terme de taille de données.

C'est un portail où les connaissances étudiées et traités sont dans des référentiels médicaux servant à des guides de bonnes pratiques médicales. Gérer ces connaissances revient à formaliser le contenu des référentiels et d'en permettre un accès et une manipulation intelligente par des utilisateurs se trouvant dans de différents endroits, et aussi gérer la maintenance et l'évolution des connaissances

dans les référentiels. [F.B12]

Le portail Kasimir est basé sur des technologies sémantiques qui lui ont permis de devenir une plateforme pour soutenir et valider la recherche dans son domaine. Aujourd'hui il sert aussi de référence dans le réseau européen « knowledge web » pour la mise en place du Web.

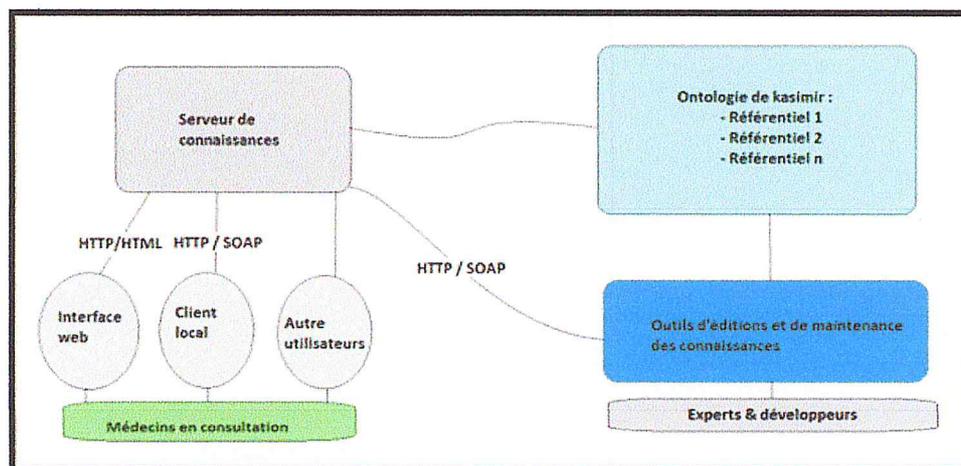


FIGURE 2.2: Architecture du portail kasimir

### 2.6.4 Le Portail sémantique data.BNF

BNF, un portail sémantique pour la bibliothèque nationale de France, un portail où on trouve toutes ses ressources en regroupant sur une même page toutes les informations issues de ses différents catalogues, comme Kasimir, BNF repose aussi sur l'approche du web sémantique SEAL, aujourd'hui on cherche à faciliter l'exploitation des données structurées pour que les machines arrivent à les comprendre.

BNF utilise le web de données qui consiste à publier des données sur le web et les relier entre elles pour accroître leur visibilité et leur réutilisation c'est un projet en cours de développement, le portail français utilise un standard d'ontologie qui s'appelle FRBR une sorte de modèle d'ontologie qui sort du standard W3C.

Le projet data.bnf.fr s'inscrit dans une démarche d'ouverture des données sur le web, y compris hors de la communauté des bibliothèques, les outils qu'il utilise répondent à des problèmes anciens des bibliothèques en particulier autour de la gestion de formats divers et de l'échange de données.

Dans ce projet on cherche à accroître la visibilité des données de la bibliothèque nationale de France et ce par une meilleure exposition sur le Web, Fédérer les données de la BNF, mais aussi à mettre en œuvre un bon fonctionnement du web de données et cela en assurant l'échange de métadonnées par création de liens entre des ressources de confiance structurées et faciliter la réutilisation

des métadonnées sous licence ouverte. En fin le projet BNF se voit fixer des objectifs comme développé son sémantique en utilisant un modèle d'ontologie qui est le FRBR. BNF est en ligne sur <http://data.bnf.fr/> [dF16]

### 2.7 Conclusion

Nous voilà déjà arrivé à la fin du premier chapitre, nous avons vue ce qu'est un portail documentaire et quelque aspects qui le caractérisent, nous avons vue aussi ce qu'est un portail sémantique les enjeux du portail et surtout que ce dernier opère sur un nouveau web, c'est un web repensé, c'est le web sémantique. Dans le chapitre suivant nous allons aller plus loin dans cette nouvelle catégorie de web, essayer de bien montrer ses forces, et surtout quels sont les outils utilisé pour ses technologies.

## **Troisième partie**

### **Conception Et Réalisation**

# Chapitre 3

## Conception de notre système

### Contents

---

<b>3.1 Introduction</b> . . . . .	<b>52</b>
<b>3.2 Conception de l'ontologie</b> . . . . .	<b>52</b>
<b>3.3 Linked data</b> . . . . .	<b>78</b>
<b>3.4 Conclusion</b> . . . . .	<b>82</b>

---

### 3.1 Introduction

Nous arrivons dans le chapitre 3 ou nous allons essayer de répondre à la problématique posée dans notre mémoire à savoir comment concevoir un portail documentaire sémantique mais avant d'arriver à la notion du portail et comment ce dernier a été conçu, il faut tout d'abord voir en profondeur comment ce dernier fonctionne réellement et quel sont les technologies sur lesquelles il se base ?

Dans ce qui va suivre nous allons essayer de voir par étape comment notre système a été conçu, en premier lieu nous allons voir l'aspect sémantique qui part du principe de créer des données structurées et d'exprimer des vocabulaires et des règles qui les décrivent afin d'assurer une bonne interopérabilité de ces derniers et faciliter leur manipulation par la machine, ceci est traduit dans notre travail par la conception de notre ontologie à travers le modèle FRBR qui est un modèle conceptuel pour les données bibliographiques et que nous avons adopté pour la structuration de nos données dans un schéma standardisé et que nous allons voir plus en détail. Pour ce qui on est de la construction elle-même nous avons suivi la méthode « Methontology » qui est une méthode référence dans le domaine d'ingénierie ontologique.

La seconde partie de notre travail va nous amener au Linked data, ou comment publier nos donnée sur le web, par la suite nous allons montrer comment enrichir nos données en les reliant avec les données de la bibliothèque nationale de France afin de montrer l'apport que peut avoir le web des donnée et sa contribution pour la réalisation d'une des objectif du web sémantique à savoir faciliter l'exploitation des données par les machines de manière indépendante de l'humain pour des recherches plus pertinentes.

### 3.2 Conception de l'ontologie

Dans cette section nous allons voir en détail les étapes de la conception de notre ontologie ainsi que les méthodes et les standards que nous avons employée pour, en commençons par avoir un aperçu sur ce qui le modèle conceptuel FRBR ? Pourquoi l'avoir choisi ? Ainsi que son fonctionnement, par la suite nous allons aborder la méthode que nous avons adoptée pour la construction de notre ontologie à savoir « Methontology », définir ces étapes et ce qui nous a encouragé a travaillé avec et pour finir voir les étapes de construction de l'ontologie à partir de notre méthode et avoir

une vue globale.

### 3.2.1 Le modèle FRBR

Une notice bibliographique est tout d'abord une fiche qui permet la description d'un document dans une bibliothèque. On utilise une notice pour un fichier afin de l'identifier, le localiser ou le décrire. Pour cela, la notice bibliographique comporte plusieurs types d'informations qui sont relatives au document lui-même, par exemple elle peut comporter le titre et le nom de l'auteur du document qui facilitera son identification lors de la lecture. La notice peut aussi comporter un résumé pour but de donner une vision sur le contenu du document, elle peut même donner une clé d'accès pour retrouver le document dans les moteurs de recherches, au final bien que la notice bibliographique détient un rôle très important lorsqu'il s'agit d'un grand nombre de documents mais aussi il faut mettre en œuvre des normes à respecter comme ISBD (standard international des description bibliographique) ou Unimarc (le standard marc universel).

FRBR ou bien les spécifications fonctionnelles des notices bibliographiques (en anglais : Functional Requirements for Bibliographic Records) est une modélisation des informations contenues dans les notices bibliographiques. Ces spécifications ne sont ni standards ni normes, elles sont nées d'un certain groupe de travail de l'IFLA (en anglais International Fédération of Library Associations and Institutions) dans les années 1990, ce dernier cherchait à mettre en œuvre une nouvelle manière de concevoir l'information bibliographique. Plus le temps passe plus il y a des spécifications héritées de FRBR comme le FRBRoo de l'orienté objet. Enfin le FRBR est publié en 1998.

#### Objectif de la norme FRBR

FRBR aujourd'hui prend de plus en plus d'importance, il est devenu une sorte de clé pour adapter les catalogues bibliothèques pour aller sur le web et on le trouve dans l'évolution des règles de catalogage pour pouvoir décrire les documents possédées par les bibliothèques selon cette approche. Utiliser FRBR est donc tout un travail du côté des systèmes informatiques, ainsi pour structurer les informations selon cette méthode on doit passer par deux manières :

- Traiter l'existant avec des programmes pour pouvoir faire un affichage frbairisé de l'information bibliographique [data.bnf.fr](http://data.bnf.fr) est un exemple
- repenser complètement l'organisation des informations et aller vers une structuration en RDF et selon les standards du web sémantique.

Le groupe de travail IFLA met en œuvre le FRBR pour aider à mieux comprendre ce que nous renseignent les notices bibliographiques, et savoir si c'est en adéquation avec les besoins de l'utilisateur. Pour cela, il doit être un langage parlant que comprend tout le monde. Alors IFLA a choisi le modèle d'entité-association qui permet de représenter l'information sous forme de diagrammes, et cela en attribuant pour chaque entité un terme avec un sens précis. De plus le FRBR couvre un très grand nombre de types de documents ce qui fait de lui un candidat intéressant pour que les futures RDA qui se présente comme une nouvelle norme pour la description des ressources et l'accès à celles-ci, conçue pour le monde numérique, s'appuieront dessus.

### Fonctionnement de la norme FRBR

Nous allons maintenant découvrir comment opère ce modèle, d'abord un document en FRBR se compose de trois parties qu'on va appeler entités.

– **Entité du groupe 1** : Les entités de ce groupe sont subdivisées en quatre éléments.

1. **Entités œuvres** : Ce sont des créations intellectuelles ou artistique déterminées, ces entités sont abstraites qu'on reconnait à travers des réalisations individuelles, il est difficile de définir précisément les frontières de ces entités à cause de leurs caractère abstrait. définir Œuvre comme entités du modèle permet de nommer la création intellectuelle abstraite qui englobe toutes les expressions. [IFL12]
2. **Entités expression** : Elles représentent la réalisation intellectuelle ou artistique des œuvres et qui peut être sous un ou plusieurs formes ex : musicale, visuelle, notation alphanumérique, ou bien une combinaison des formes. Les limites de ces entités peuvent apparaitre du point de vue que les aspects purement matériels qui ne font pas partie intégrante de la réalisation intellectuelle ou artistique de l'œuvre en tant que telle. ex : la typographie et La mise en page des œuvres de documents sont exclus de l'expression. [IFL12]
3. **Entités manifestations** : Elle représentent les manifestations de l'une des expressions de l'œuvre, il s'agit d'une matérialisation des expressions, ces entités couvre un grand intervalle de supports, des manuscrits, des livres, des périodiques, film, cartographie, etc. C'est dans ces entités qu'on trouve tous les objets matériels qui présentent les même caractéristiques sur le plan intellectuel que matériel. [IFL12]
4. **Entités Item** : UUn Item représente un exemplaire isolé d'une manifestation, mon exemplaire, ton exemplaire, son exemplaire de la manifestation. [IFL12]

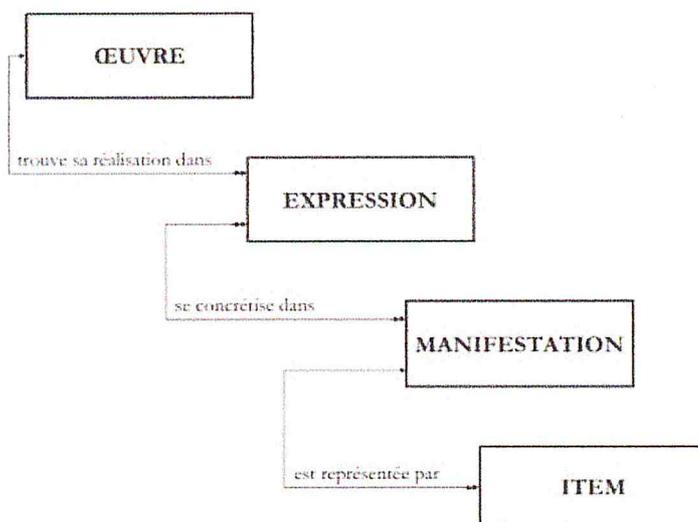


FIGURE 3.1: Les éléments de l'entité du "Groupe1"

- **Entité du groupe 2** : Sera dédié pour spécifier l'auteur du document, cette entité peut représenter soit une personne ou bien une organisation (un groupe d'individus), c'est l'entité la plus simple du modèle FRBR. [IFL12]

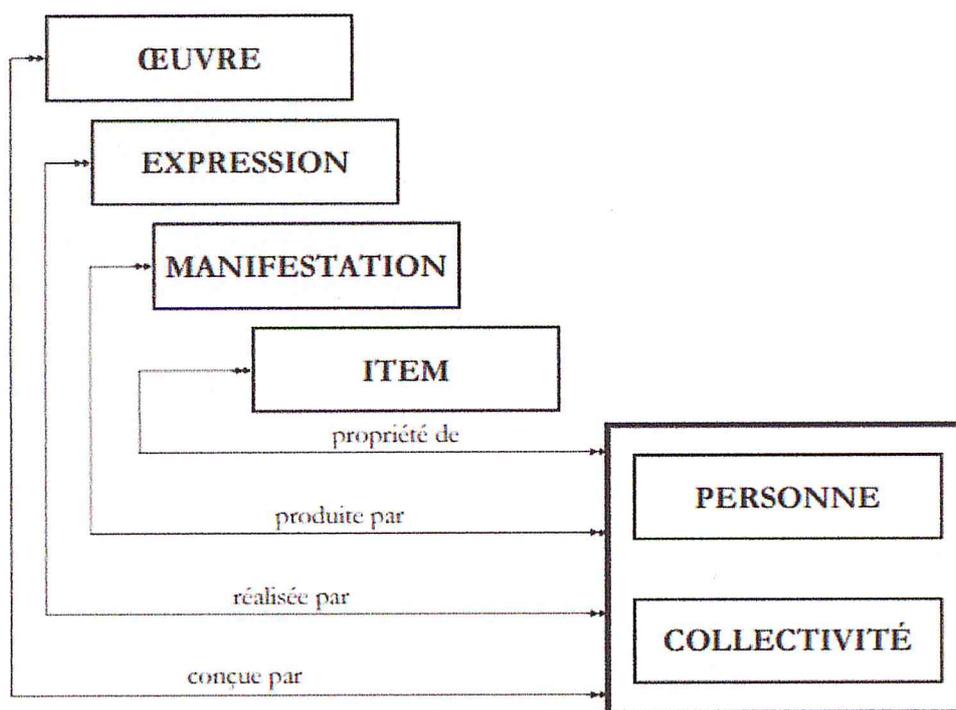


FIGURE 3.2: Relations entre l'entité "Groupe2" et "Groupe1"

- **Entité du groupe 3** : Ces entités représentent un nouvel ensemble pour déterminer le sujet

de l'œuvre, il est subdivisé en quatre parties :

1. **la partie concept** : Un groupement d'entité dédié a une notion ou une idée abstraite.
2. **la partie objet** : Un ensemble d'entités qui représente une réalité matérielle.
3. **la partie événement** : C'est une partie qui englobe des entité qui représentent une action ou un fait.
4. **la partie lieu** : C'est les des données topographiques.

Dans la figure suivante l'entité du groupe 3 est encadrée d'un trait gras [IFL12]

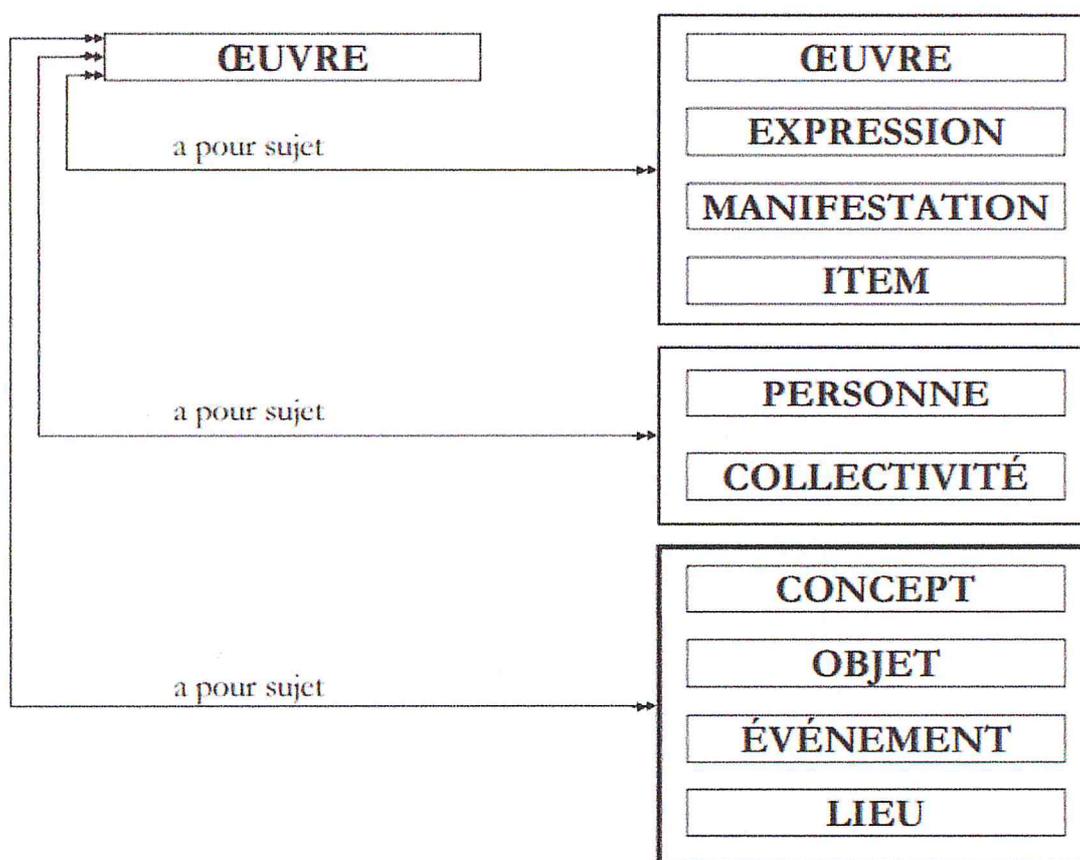


FIGURE 3.3: Entité "Sujet" dans FRBR

Le modèle FRBR que nous avons vue jusqu'au là, répond presque déjà aux besoins des notices bibliographique, ses entité aide à mieux renseigner ce que proposent les notices, et il est parfait pour un grand nombre de document variés, sauf que le FRBR nous propose encore plus. En effet, dans ce modèle d'entité association chacune des entités est doter d'un ensemble de caractéristiques ou attributs, c'est grâce à ces attributs que les utilisateurs du portail exécutent leurs requêtes et interprètent leurs réponses lorsqu'il recherche une entité déterminée. [IFL12]

### Les attribut du modèle FRBR

Les attributs du modèle FRBR sont divisés en deux grandes catégories majeures, On trouve d'un côté les attributs inhérents à une entité et autres qui sont forgés. Les attributs de la première catégorie sont des attributs de caractéristique physique, ex : la matière et les dimensions d'un objet, mais aussi sont des éléments qui nous permette de distinguer ex : distinguer deux objet par leurs couverture, boitier etc. Les attributs de la seconde catégorie comprennent des identifiants affectés à une entité ex : un numéro de catalogue thématique pour une composition musicale, mais aussi des informations contextuelles ex : le contexte politique dans lequel une Œuvre a été conçue. En général on peut établir les attributs inhérents à une entité par un examen de l'entité elle-même tandis que les attributs forgés font souvent appel à une source extérieure. [IFL12] Nous allons maintenant détailler les attributs de chaque entité, ceci va nous permettre de bien comprendre le vocabulaire de ce modèle [IFL12] Les quatre entités "œuvre", "expression", "manifestation" et "item" comprennent un ensemble de vocabulaire, nous allons essayer de citer quelques attributs généraux :

#### Les attributs de "œuvre"

- Titre de l'œuvre
- Forme de l'œuvre
- Date de l'œuvre
- Contexte de l'œuvre : C'est le contexte historique, social, intellectuel, artistique ou autre au sein duquel l'œuvre a été conçu à l'origine. l'œuvre a été conçue à l'origine.
- Autre caractéristique distinctive :C'est toute caractéristique peut nous aider à distinguer une "œuvre" d'une autre "œuvre" portant le même titre ex : c'est la zone géographique d'origine qui sert à distinguer entre les deux mystères médiévaux intitulés L'adoration des bergers, l'un provenant de Coventry et l'autre de Chester [IFL12]

#### Les attributs de "Expression"

- titre de l'expression
- forme de l'expression
- date de l'expression
- langue de l'expression
- contexte de l'expression
- Autre caractéristique distinctive [IFL12]

### les attribut de "Manifestation"

- titre de la manifestation
- mention de responsabilité : C'est le nom d'un ou plusieurs personnes qui ont une responsabilité dans la réalisation ou la création du contenu intellectuelle concrétisé par la manifestation.  
ex : Auteur d'un livre, un traducteur, un compositeur etc.
- lieu d'édition/diffusion
- date d'édition/diffusion
- éditeur/diffuseur de la manifestation :
- identifiant de la manifestation : il représente un numéro ou code associé de manière unique à la manifestation. [IFL12]

### les attribut de "Item"

- identifiant de l'item
- empreinte : C'est pour signaler des différences entre plusieurs exemplaires de livres anciens
- provenance de l'item : C'est une liste des précédents possesseurs ou dépositaires de "Item"
- annotations/inscriptions : C'est toutes signatures, numérotations, notes, etc. qui est propre à l'item
- état matériel de l'item : C'est une description de l'état physique de l'item ex : Des pages manquante dans un livre.
- restrictions de communication : toute limitation quant à l'accès physique à l'exemplaire ex : Consultation sur place et sous contrôle. [IFL12]

Nous avons vu maintenant ce qu'est le modèle FRBR, l'architecture de ce dernier facilite la structuration de données pour les bibliothèques, nous allons voir maintenant comment utiliser les avantages du web sémantique sur ce modèle. Pour une première partie nous allons voir une conception de l'ontologie en mentionnant la méthodologie suivie, ensuite nous allons voir la jonction entre cette ontologie et le modèle FRBR. *Rappel* : le web sémantique est un web révolutionnaire grâce à sa force qui se résume en :

- Il permet la séparation du contenu de la présentation dans les pages web, ce qui donne la possibilité pour des machines de comprendre d'autres machines.
- Il a donné vite naissance à ce qu'on appelle le web de données ou bien les données alignées.

### 3.2.2 FRBR, Illustration !

Après avoir vue l'ensemble de définitions de FRBR, nous allons maintenant suivre un exemple de A à Z afin d'illustrer ce modèle et voir comment sa se passe en réalité.

- Meriem, une femme qui va au cinéma pour voir un film qui s'appelle "**Ne touchez pas a la hache**" de Jacques Rivette
- Meriem, s'aperçoit dans le générique que le film est inspiré d'un Roman qui s'appelle "**Honoré de Balzac**" et souhaite l'obtenir. ce dernier appartient a un cycle romanesque qui est " **l'Histoire des XIII** "
- Meriem donc peut obtenir une version textuel ou bien une version audio si elle souhaite l'entendre dans sa voiture, cette dernière est aussi sous différentes versions, une avec un lecteur anonyme qui est juste un donneur de voix , ou bien une lecture avec une qualité meilleure par une interprète qui s'appelle par ex : La Fanny Ardant qui a fait la lecture de la duchesse de Langeais de Honoré de Balzac
- Meriem a une fille "Ahlem" qui doit faire un exposé sur un Roman de Balzac et elle lui suggère Son roman pour faire sa présentation.
- Ahlem va donc avoir besoin d'édition du roman qui sont accompagné de commentaires et donc elle va chercher des éditions élaborées et éventuellement des document qui présente l'œuvre du Roman pour les étudiant.
- Ahlem a une amie Sihem qui étudie l'allemand, et qui peut aussi obtenir le document élaboré puisque il est traduit en Allemand.

Voici tout un parcours recherche qui pourra se faire dans le web nous allons voir maintenant comment organiser les entités utilisées dans la recherche selon le modèle FRBR.

**Rappel** : *une œuvre est une création intellectuelle ou artistique déterminées de manière abstraite.*

Les entités "**œuvres**" appliqué pour cette exemple sont :

- Le film "Ne Touchez pas a la hache - 2006".
- Le Roman "La duchesse de longeis - 1833".
- Le cycle Romanesque " Histoire des XIII "
- La version élaboré du roman "L'étude critique de la Duchesse de longeis".

Ces entité œuvres ont des auteurs qui sont caractérisés par leur noms, date de naissances etc, c'est un autre groupe d'entités, c'est les "**entités du groupe2**"

**Rappel :** Une expression est la réalisation des œuvres abstraites. Donc c'est la réalisation de cette œuvre dans une forme donnée ex soit textuelle, soit lu.

Les entités "Expression" appliquées pour cet exemple sont :

- Le Roman "La duchesse de longeais - 1833" Format papier - langue Française.
- Le Roman "La duchesse de longeais - 1833" Format papier - langue Allemande.
- Le Roman "La duchesse de longeais - 1833" Format audio - Donneur de voix anonyme.
- Le Roman "La duchesse de longeais - 1833" Format audio - De La Fanny Ardant.

**Rappel :** Une Manifestation est la matérialisation des expressions, c'est le support sur lequel repose l'expression, c'est ce qu'on peut manipuler physiquement dans une bibliothèque.

Les entités "Manifestation" appliquées pour cet exemple sont :

- Le Roman "La duchesse de longeais - 1833" Format papier - langue Française de la maison d'édition de Gallimard.
- Le Roman "La duchesse de longeais - 1833" Format papier - langue Française de la maison d'édition de Livre de poche.
- Le Roman "La duchesse de longeais - 1833" Format papier - langue Française de la maison d'édition de Flammarion.
- Le Roman "La duchesse de longeais - 1833" Format audio - De La Fanny Ardant - Fichier mp3
- Le Roman "La duchesse de longeais - 1833" Format audio - De La Fanny Ardant - CD audio

**Rappel :** Un Item est les diverses exemplaires de la manifestation qu'on peut posséder, mon exemplaire, ton exemplaire, son exemplaire.

Les entités "Item" appliquées pour cet exemple sont :

- Mon exemplaire du Roman de "La duchesse de longeais - 1833" auquel j'ai mis mes propres commentaires
- L'exemplaire du Roman de "La duchesse de longeais - 1833" de mon ami qu'il a eu hier

De cet exemple nous constatons un ensemble de relations entre entités par exemple :

- Une œuvre peut faire partie d'une autre œuvre, l'œuvre du roman "La duchesse de longeais" fait partie de l'œuvre du cycle romanesque de Balzac
- Des relations entre groupes d'entités, une relation de responsabilité, Jacques Rivet (entités du groupe 2) est le réalisateur du film ne touchez pas à la hache.

l'entité du groupe 3 de FRBR renseigne sur les informations comme le lieu où se passe l'histoire du roman.

Maintenant que nous avons illustré cet exemple dans le modèle FBRR, nous allons voir comment implémenter cet exemple avec notre ontologie une fois expliquer comment nous avons construit notre ontologie à travers la méthode Methontology.

### 3.2.3 Construction de notre ontologie

La procédure que nous avons choisie pour construire notre ontologie est la méthodologie « METHONTOLOGY ». Notre choix s'est porté sur cette méthodologie parmi plusieurs d'autre que nous avons vu dans la partie de l'état de l'art pour plusieurs raison, dont principalement le faite que cette méthodologie est utilisée pour la construction des ontologies à partir de zéro sans se baser sur des ontologies déjà existante avec une conception réalisée manuellement selon nos besoin ce qui est le cas dans notre projet. Dans cette section nous allons passer du coté théorique a l'implémentation de notre ontologie toute en respectant les standard du model FRBR en premier lieu ainsi que les étapes de notre méthodologie, ces étapes que nous développer dans ce qui suit jusqu' à arriver à une ontologie opérationnel.

1. **Spécification** : Une ontologie ne peut être construite qu'après la phase de spécification. Il s'agit d'établir un document informel de spécification de besoins. Au niveau de ce document, nous décrivons l'ontologie à construire à travers les questions qu'on a vues précédemment et qu'on peut englober dans les aspects-suivant : **Domaine de connaissance** : Domaine Bibliographique (gestion de document bibliographique) **Objectif** : Construire une ontologie qui modélise les données Bibliographiques selon un modèle définis appelé FRBR, cette ontologie sera utilisé par la suite dans le cadre d'un portail documentaire sémantique afin de permettre aux utilisateurs de notre portail d'avoir un vocabulaire conceptuel commun pour utiliser et partager les données bibliographique facilement. **Utilisateurs** : Etudiants, Professionnels, bibliothécaires, éditeurs, usagers. . . **Sources d'informations** : Catalogues bibliographiques, notices bibliographiques. **Portée de l'ontologie** : L'ontologie va fournir un ensemble de données et d'informations sur des documents numériques.
2. **Conceptualisation** : Cette partie du travail aborde l'étape de conceptualisation du domaine en question (création et formalisation des informations que regroupe la branche « bibliothèque »
  - **a. Construction du glossaire de termes** : Construire un glossaire de termes est la première tâche à effectuer dans l'étape de conceptualisation. Il recueille et décrit tous les termes (concepts, instances, attributs, relations entre les concepts, etc.) qui sont utiles et potentiellement utilisables dans le domaine qu'on essaye de modélisé. Le tableau qui suit représente

une partie des concepts regroupés dans le glossaire. (Un extrait seulement de ce qu'on peut trouver dans le domaine bibliographique !)

Nom	Description
Annuaire	Ensemble d'adresses (de personnes, d'organismes , de sources d'information, de sites web, etc.) pouvant être complétées par d'autres informations associées (chiffre d'affaires d'un organisme, rôle et fonction d'une personne, nom du dirigeant d'un organisme ect...) sous une forme papier, il peut paraître annuellement.
Article	En documentation, texte indépendant constituant une partie d'une publication périodique, d'un dictionnaire, d'une encyclopédie.
Document administratif	Document produit ou reçu dans le cadre des opérations administratives générales et des opérations spécifiques au mandat et aux fonctions de toute organisation (gestion des ressources et des activités humaines, financières et matérielles), et susceptible d'être utilisé pour action, étude ou preuve.
Document audiovisuel	Terme générique désignant des documents dont le contenu est soit sonore, soit visuel, soit un combiné des deux, quel qu'en soit le support. Il regroupe qui sont : les images fixes, les phonogrammes et les documents images animées. Il comporte un numéro international normalisé d'identification (ISAN).
Document électronique	Document (fichier de données et/ou de programmes) sur support électronique nécessitant un matériel pour être consulté. Il bénéficie des possibilités de l'informatique qui sont : reproductibilité (sur plusieurs écrans en même temps), modification selon le type de code, transmissibilité via les réseaux de communication... On distingue trois modes principaux de représentation de l'information de ces documents électroniques qui sont : mode caractère, mode image, mode vectoriel
Extrait	Copie conforme de certaines parties d'un document, y compris la pagination d'origine.

Film	Bande transparente de pellicule photographique portant une séquence d'images distinctes qui, projetées l'une après l'autre, donnent l'illusion d'un mouvement continu. Le son peut être enregistré par des procédés magnétiques ou optiques sur la même bande que l'image ou sur une bande séparée.
Journal	Publication en série contenant des informations sur l'actualité, d'intérêt particulier ou général, ayant des formes et une périodicité différentes suivant son mode de diffusion qui sont : papier, réseau de télécommunication (Internet ou téléphone par exemple), réseau radiophonique ou télévisuel.
Livre	Ensemble imprimé, illustré ou non, publié sous un titre, ayant pour objet la représentation d'une œuvre de l'esprit d'un ou de plusieurs auteurs.
Manuscrit	œuvre dictée ou écrite et corrigée par l'auteur antérieurement à son impression.
Multimédia	Ensemble de documents dont un ou plusieurs composants sont de nature audiovisuelle, dont les différents éléments ou supports, quelle qu'en soit la nature, ne peuvent être dissociés par la commercialisation, et qui ont une parité d'importance entre eux. Leur consultation nécessite au moins un appareil de lecture
Norme	Spécification technique (ou autre document de référence accessible au public), établie par consensus qui fournit, pour des usages communs et répétés, des solutions optimales à des problèmes techniques et commerciaux concernant les produits, biens et services. Les normes intéressent trois domaines qui sont domaine général, domaine électronique, domaine des télécommunications. Elles sont validées et publiées par un organisme qualifié sur le plan régional, national et international.

Périodique	Catégorie de publications en série, à auteurs multiples, dotée d'un titre unique, dont les fascicules, généralement composés de plusieurs contributions répertoriées dans un sommaire, se succèdent chronologiquement à des intervalles en principe réguliers (journal, magazine, lettre, numéro spécial, revue), pendant une durée non limitée a priori. Les publications annuelles sont comprises dans cette définition ; les journaux et les collections de monographies en sont exclus
Publication annuel	Catégorie de publications en série dont la périodicité est en principe annuelle et dont l'objet reste constant d'un volume à l'autre. Les publications annuelles comprennent les rapports d'activité, les annuaires, les répertoires, les chronologies, etc.
Rapport	Document (écrit) qui décrit les résultats de recherches, d'enquêtes, d'observation ou d'autres études et qui est soumis en premier lieu à la personne ou à la collectivité pour laquelle l'étude a été réalisée ou par qui elle était commanditée. Un rapport peut traiter n'importe quel sujet. On identifie différents types de rapports qui sont : rapport de recherche, rapport d'étude méthodologique, rapport d'activités, rapport de conjoncture, de mission, etc. En règle générale, les rapports ne sont ni commercialisés, ni disponibles en librairie. Publié le plus souvent irrégulièrement, un rapport peut être une monographie isolée ou appartenir à une collection. Un rapport est identifié par un numéro international normalisé d'identification (ISRN).

Ressource électronique	Document (données ou logiciels) encodé afin d'être traité par un ordinateur et considéré comme une unité bibliographique. Les ressources électroniques comprennent d'une part des ressources d'information stockées en local, d'autre part celles qui nécessitent l'utilisation d'un périphérique relié directement à l'ordinateur (par exemple, un disque dur, un lecteur de cédérom), et enfin les services en ligne (par exemple, les forums ou les listes de discussions, des sites web). Une ressource électronique peut comporter soit du texte, soit de l'image fixe ou animée, soit du son. Elle peut être aussi multimédia.
Supplément	Publication généralement éditée de façon distincte, qui complète la publication principale en la mettant à jour, en la continuant d'une autre manière ou en traitant un sujet particulier qui n'y figure pas. Un supplément peut porter ou non un titre dépendant de celui de la publication principale.
Thèse	Document réalisé par un chercheur, rendant compte de ses travaux de recherche et de leurs résultats sous la responsabilité d'un directeur de thèse, soutenu devant un jury universitaire pour l'obtention d'un diplôme universitaire (doctorat).
Compte rendu	Texte ou exposé, oral ou écrit, qui présente et/ou analyse des faits, des débats, une réunion, le contenu d'un ouvrage, une visite, une mission, etc., pour informer, faire savoir, expliquer ou justifier.

TABLE 3.2: Construction du glossaire de termes

- **b. Définir les classes et la hiérarchie des classes** : Dans cette partie nous nous sommes inspirés du modèle FRBR pour définir les classes les plus générales de notre ontologie, en premier lieu, ces classes représentent les entités du modèle FRBR que nous avons vu dans la section précédente, le résultat est définis dans la figure ci-dessous :



FIGURE 3.4: Les classes les plus générales de notre ontologie

Dans cette figure on peut voir clairement les 3 entités du modèle FRBR, dans notre ontologie sous 3 groupes que nous avons nommée « Document » pour œuvre, « Responsable » pour responsable et « Subject » pour le sujet de l'œuvre, pour ce qui est de la deuxième couche elle représente les sous classe des classes générales définies par une relation taxonomique proposée par METHONTOLOGY qui est sous-classe-de (Subclass-Of) et qu'on peut voir clairement dans la figure ci-dessous ou Item, Expression, Work, Manifestation sont des sous classes de Document.

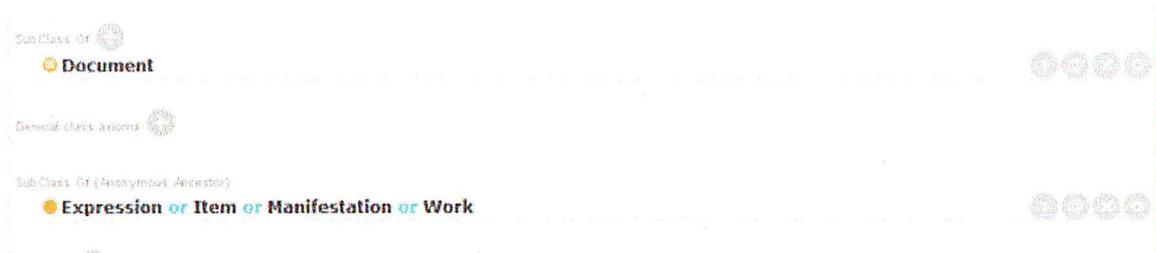


FIGURE 3.5: Les sous classe des classes générales de notre ontologie

Cette relation taxonomique nous permet d'abord la seconde partie de cette étape qui est d'établir la hiérarchie des classes dans notre ontologie ou on a vu précédemment qu'il existe trois approches pour cela, la première approche que nous avons utilisée et celle de « Procédé de Haut en Bas » dans notre définition du troisième groupe du modèle FRBR qui est « Sujet ».

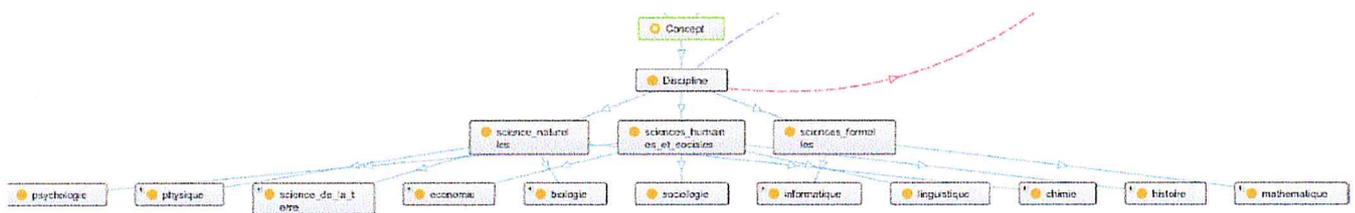


FIGURE 3.6: Procédé de Haut en Bas

Nous avons, donc, commencé par les classes les plus générales, à savoir, «Discipline » qui est une sous classe de « Sujet » et par la suite prendre une des discipline qu'on peut trouver par exemple « Science Humaine » et ensuite affiné cette filière « Histoire , Philosophie... etc. » La deuxième approche qu'on a utilisée et celle de « Procédé du Bas en Haut » et qu'on peut voir clairement dans notre définition de l'élément « Expression » du premier groupe du modèle FRBR ou on précise comment notre œuvre a été exprimée et sous quel moyen, prenons par exemple le concept « movie » (film) nous avons à une œuvre exprimée dans un travail cinématographique qu'on situe par la suite dans la catégorie Document audiovisuel et voilà comment on a procédé d'un concept spécifique qui représente la feuille de la hiérarchie pour le regrouper par la suite dans des classes plus générales

- **c. Construction du diagramme de relation binaire :** Dans cette tâche nous avons essayé de récapitulé les différentes relations qui peut y avoir dans notre ontologie, des relations qui répondent à la norme FRBR.

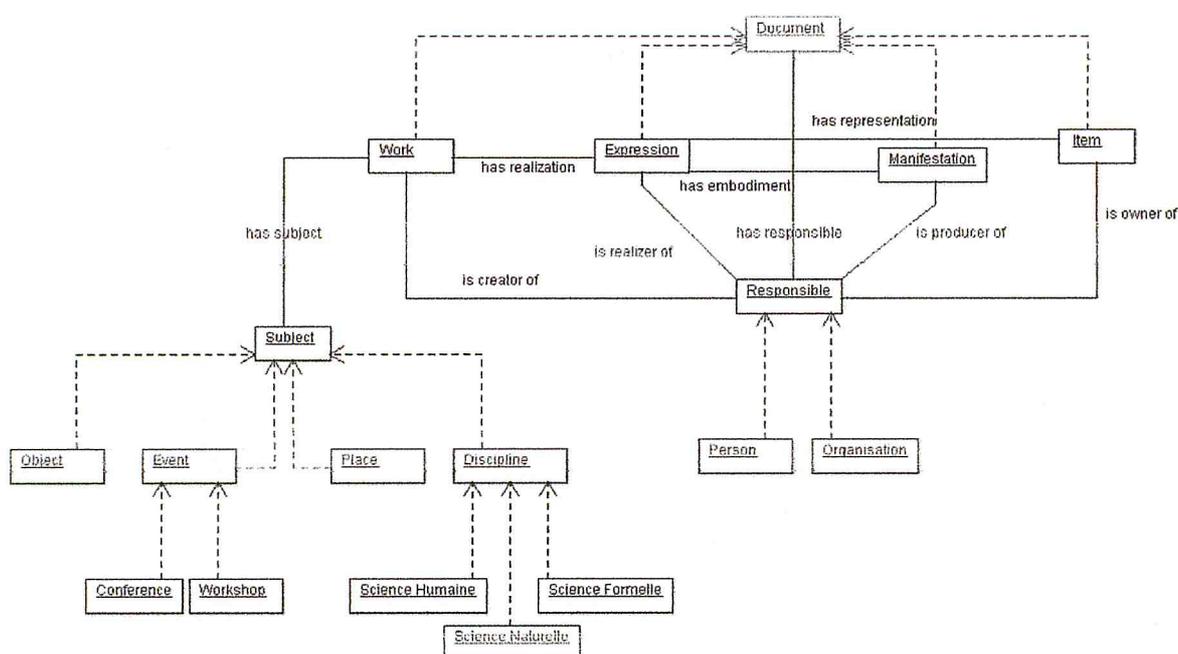


FIGURE 3.7: Diagramme des relations binaires

- **d. Construction de dictionnaire des concepts :** Une fois la taxonomie de concepts et le diagramme de relations binaires sont effectués, il faut donc spécifier les propriétés qui décrivent chaque concept de la hiérarchie dans un dictionnaire de concepts. Un dictionnaire

de concepts contient tous les concepts du domaine, leurs attributs et leurs relations. Les relations spécifiées pour chaque concept sont celle où leur domaine est le concept lui-même

Nom du concept	Description	Attribut des classes	Relation
Responsible	entités responsables du contenu, ou de la production matérielle, ou de la gestion juridique de ces œuvres		Is owner of Is producer of Is creator of Is realizer of
Person	Une personne physique auteur d'une création intellectuelle	nom de la personne dates de la personne titre de la personne	
Corporate Body	un organisme ou un groupe d'individus et/ou d'organismes responsable de la création ou la distribution d'un œuvre intellectuelle ou artistique	nom de la collectivité numérotation localisation date	
Document	Produits d'une activité intellectuelle ou artistique nommés ou décrits dans les notices bibliographiques		Has format ; Has license ; Has creator ; Has discipline ; Has manifestation ; Has realization ; Has exemplar ; Has subject ;

Work	une création intellectuelle ou artistique qui sont publiées ou potentiellement publiables, et qui contiennent ou sont désignés par des références bibliographiques, ou les entités utilisées pour définir des références bibliographiques.	titre de l'œuvre ; forme de l'œuvre ; date de l'œuvre ;	
Expression	la réalisation intellectuelle ou artistique d'une œuvre sous la forme d'une notation alphanumérique, musicale ou chorégraphique, ou sous une forme sonore, visuelle, objective, cinétique, etc., ou bien sous l'aspect d'une combinaison de ces formes.	titre de l'expression ; forme de l'expression ; date de l'expression ; langue de l'expression ; volume de l'expression ; résumé du contenu ; contexte de l'expression	Is Realization of
Manifestation	la matérialisation de l'expression d'une œuvre, recouvre un large éventail de supports, parmi lesquels : manuscrits, livres, périodiques, documents cartographiques, affiches, phonogrammes, films	Titre propre ; Mentions de responsabilité ; lieu de publication ; éditeur ; type de support ; date de publication. ISBN.	Is Manifestation of

Item	un exemplaire isolé d'une manifestation. il s'agit d'un seul objet matériel (par exemple, un exemplaire d'une monographie en un volume, une seule cassette audio, etc.)	identifiant de l'item ; provenance de l'item ; annotations/inscriptions ; état matériel de l'item ; opérations de sauvegarde ; effectuées opérations de sauvegarde à effectuer restrictions de communication ;	Is Exempler of
Concept	une notion abstraite ou une idée.		
Discipline	L'entité Discipline recouvre un large champ d'abstractions pouvant constituer le sujet d'une œuvre : domaines de la connaissance, disciplines, écoles de pensée (en philosophie, en religion, en idéologie politique, etc.)		Is Discipline of
Subject	entités envisagées comme sujet d'une œuvre		

Event	un fait accompli ou survenu. L'entité événement recouvre un large champ de faits accomplis ou survenus pouvant constituer le sujet d'une œuvre : événements historiques, époques, périodes chronologiques.		
Place	L'entité lieu recouvre un large champ de coordonnées topographiques : sur la Terre et dans l'espace ; dans l'histoire et dans le monde d'aujourd'hui ; selon des critères purement géographiques ou géopolitiques		
Object	C'est une réalité matérielle. L'entité objet recouvre un large champ de réalités matérielles pouvant constituer le sujet d'une œuvre : objets animés et inanimés présents dans la nature, objets mobiles, immobiles et déplaçables créés par l'homme, objets qui ont cessé d'exister		

TABLE 3.4: Construction de dictionnaire des concepts

- **e. Construction de la table des relations binaires** : Le but de cette tâche consiste à construire une table de relations binaires décrites en détail. Pour chaque relation utilisé dans le diagramme des relations binaires, nous définissons le nom de la relation, le nom des concepts sources et cibles, le nom de la relation inverse et les cardinalités

Ces relations sont tirées du modelé FRBR qu'on a vu précédemment.

Nom de la relation	Concept source	Concept cibles	Card	Card Cible	Relation inverse
Œuvre (Work)	Work	Expression	1...n	1...1	Is realization of
Has embodiedment	Expression	Manifestation	1...n	1...n	Is embodiedment of
Has exemplar	Manifestation	Item	1...n	1...1	Is exemplar of
Has creator	Work	Responsible	1...n	1...n	Is creator of
Has owner	Item	Responsible	1...n	1...n	Is owner of
Has producer	Manifestation	Responsible	1...n	1...n	Is producer of
Has realizer	Expression	Responsible	1...n	1...n	Is realizer of
Has discipline	Work	Discipline	1...n	1...n	Is Discipline of
Has subject	Work	Subject	1...n	1...n	Is subject of
Has location	Work	Place	1...n	1...n	Is location of
Has event	Work	Event	1...1	1...1	Is event of

TABLE 3.6: Construction de la table des relations binaires

- **f. Construction de la table d'attribut** : La table des attributs comporte une description détaillée des attributs inclus dans le dictionnaire de concepts, et l'ensemble de contraintes et de restrictions sur ces valeurs. tableau modif

Concept concerné	Nom d'attribut	type	Card
Œuvre (Work)	Titre	string	1...1
	Forme	string	1...n
	DateDeCreation	datetime	1...1
Expression	Titre	string	1...n
	FormeExpression	string	1...n
	DateExpression	datetime	1...1
	Langue	string	1...1
	Volume	string	1...1
	Contexte	string	1...1
	Resume	string	1...1
Manifestation	Editeur	string	1...n
	Date D Edition	datetime	1...1
	Type de support	string	1...n

	Identifiant	string	1...1
	Lieu D Edition	string	1...1
Item	IdentifiantItem	string	1...1
	localisation	string	1...n
Person	nomP	string	1...1
	Dateofbirth	string	1...1
Corporate body	nomC	string	1...1
	Dateoffondation	string	1...1
	localisation	string	1...1

TABLE 3.8: Construction de la table d'attribut

3. **Formalisation** : Cette étape consiste en la représentation des connaissances récoltées dans un langage formel tel que les graphes conceptuelles. Cette partie stipule la création d'un model formel avant de générer le model computable mais grâce aux éditeurs d'ontologie cette phase se fait implicitement et c'est au logiciel de crée et vérifier le model formel.
4. **Implémentation** : La construction de l'ontologie est une tâche complexe, même pour implémenter une petite ontologie, celle-ci va prendre plusieurs lignes de code et nécessite un grand effort et du temps. Notamment, si cette ontologie est codée directement par le développeur en langage d'ontologie sans faire recours à aucun outil. Pour cela, plusieurs outils sont conçus pour libérer « l'ontologiste » de cette complexité et générer automatiquement la structure de l'ontologie créée en OWL-DL. Pour développer notre ontologie nous avons choisi l'éditeur d'ontologie Protégé 2000 pour les raisons suivant :
  - (a) Protégé prend en charge la méthode que nous avons utilisé dans la partie conception.
  - (b) C'est l'un des meilleurs éditeurs d'ontologie sur recommandation des experts du l'ingénierie ontologique.
  - (c) L'éditeur de classe Protégé permet la création et la visualisation des classes et des hiérarchies OWL (Ontology Web Language).
  - (d) De même, il permet la création et la gestion des relations en termes de propriétés et de hiérarchies.

Nous avons choisi le langage de représentation des connaissances OWL (Ontology Web Language). Qu'est le standard W3C Pour le développement des ontologies [W3C05] Nous don-

nous dans ce qui suit des figures qui représentent un extrait de notre ontologie, un extrait de relations de l'ontologie ainsi qu'un extrait de code source.

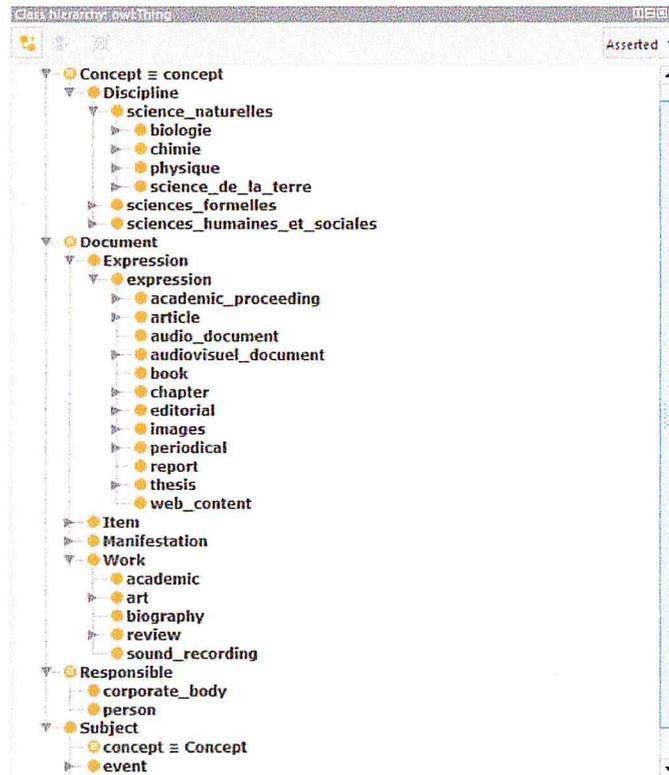


FIGURE 3.8: Extrait de notre ontologie

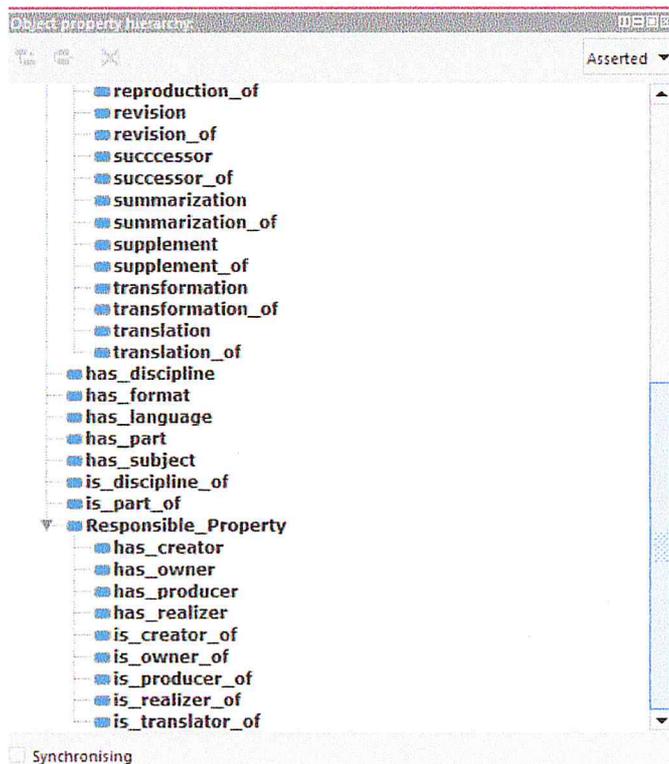


FIGURE 3.9: Extrait des relation de l'ontologie

```

<owl:Class rdf:about="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#manifestation">
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#Manifestation"/>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Restriction>
          <owl:onProperty rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#has_part"/>
          <owl:allValuesFrom rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#Manifestation"/>
        </owl:Restriction>
        <owl:Restriction>
          <owl:onProperty rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#is_part_of"/>
          <owl:allValuesFrom rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#Manifestation"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#has_exemplar"/>
      <owl:allValuesFrom rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#item"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#is_embodiment_of"/>
      <owl:allValuesFrom rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#expression"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

FIGURE 3.10: Extrait du code source de l'ontologie

### 3.2.4 Exemple illustratif

Maintenant que nous avons une ontologie opérationnelle, nous allons montrer le fonctionnement de cette dernière et l'avantage qu'elle présente dans la structuration des données et d'informations sur les documents numérique et ceux par l'intermédiaire d'un exemple que nous avons vue dans la section [3.2.2]. Pour commencer on va récapituler les éléments qui constitue le premier groupe du modèle FRBR (œuvre, Expression, Manifestation, Item) et voir comment passé d'un élément a un autre dans notre présentation de l'exemple. Tout d'abord l'élément œuvre qui représente les différentes création intellectuel dans notre exemple et que nous récapitulons comme suit :

- Le film "Ne Touchez pas a la hache - 2006".
- Le Roman "La duchesse de longeais - 1833".
- Le cycle Romanesque " Histoire des XIII ".
- La version élaboré du roman "L'étude critique de la Duchesse de longeais".

Dans notre ontologie nous avons crée les instances de l'entité **Work(œuvre)** comme le montre la figure suivante :

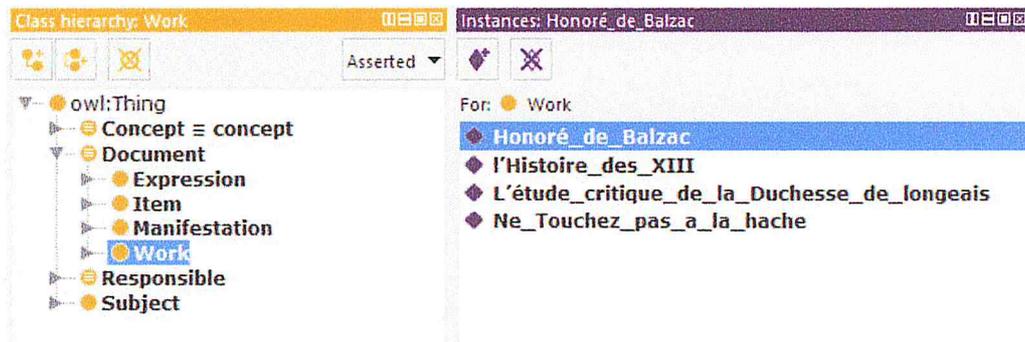


FIGURE 3.11: Création des instance de l'entité Œuvre

une fois ces instances crée nous avons définis les relation entre ces dernier selon l'énoncée de l'exemple que nous pouvons récapitulé comme suit

- Le film "Ne Touchez pas a la hache" est une adaptation du roman "La duchesse de longeais".
- Le Roman "La duchesse de longeais " fait partie du cycle romanesque " Histoire des XIII ".
- La version élaboré du roman "L'étude critique de la Duchesse de longeais" a pour sujet le roman "La duchesse de longeais".

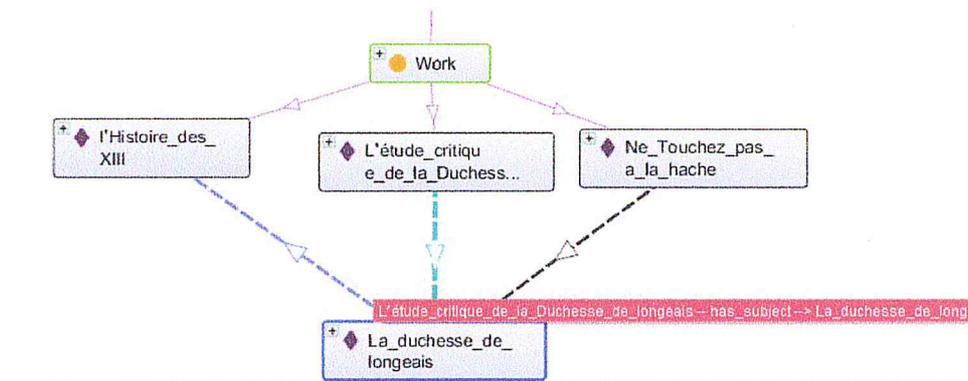


FIGURE 3.12: Création des relation entre les instance de l'élément Œuvre

Jusqu'à présent nous avons créé les œuvres qui sont à notre possession ainsi que les relations qui peuvent avoir entre eux maintenant nous allons voir comment ils sont exprimés et c'est ce qui va nous amener au deuxième élément de notre entité qui est *Expression*, dans cette étape nous avons ce qui suit :

- Le Roman "La duchesse de longeais " est réalisé sous format texte dans la langue française avec deux langues différentes en français et en allemand.
- Le Roman "La duchesse de longeais " est réalisé sous format audio mais avec deux interprètes différents donc deux versions différentes.

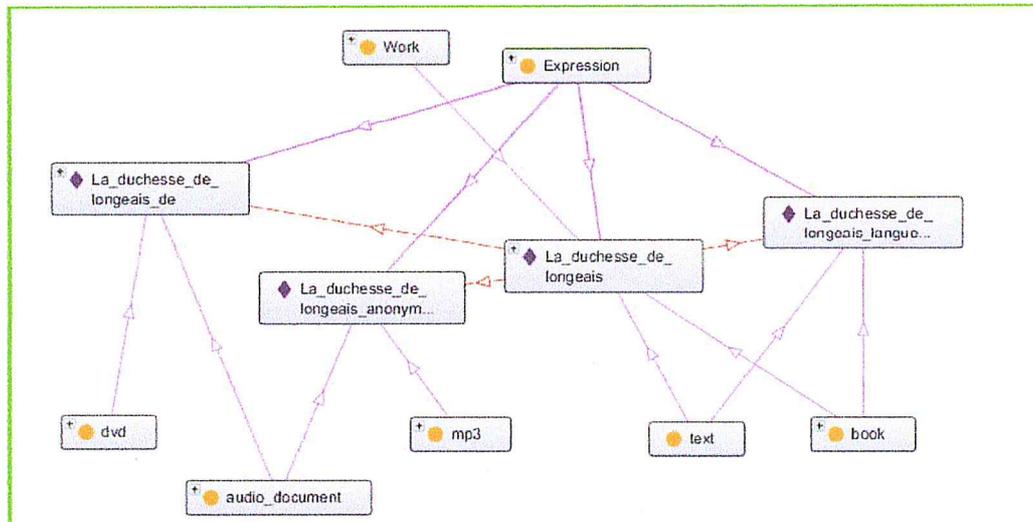


FIGURE 3.13: L'expression de nos données et leur matérialisation

Dans la figure ci-dessus nous pouvons par ailleurs que les deux document audio son englobé ou matérialisé soit sous forme numérique ou physique, la forme numérique se concrétise dans la version mp3 ou DVD dans les document audio dans notre exemple et pour ce qui est de la forme physique elle se concretise dans le livre qui matérialise les deux version du roman soit la version française ou la version traduite en allemand, ceux ci et notre troisième élément du 1er groupe du modèle frbr a savoir l'élément **Manifestation**. Le dernier élément du groupe1 **Item** est l'exemplaire lui même qu'une personne peut posséder et que nous pouvons trouver dans une bibliothèque, dans notre ontologie on peut le définir simplement par deux attribut comme un identifiant et le lieu ou le document est disponible.

Une fois le travail avec le groupe 1 terminé, nous passons au deuxième groupe qui représente l'ensemble de personne responsable de la création des œuvre intellectuel que nous avons vue dans le groupe précédent et que nous pouvons récapitule dans la figure suivante comme suit :

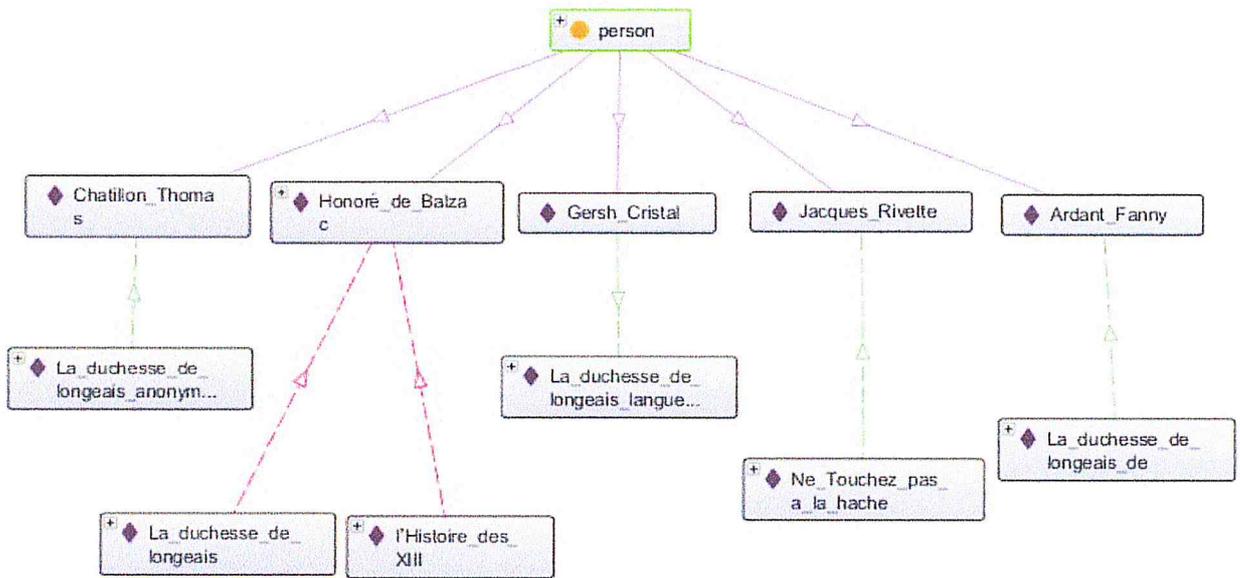


FIGURE 3.14: Relier les documents a leur auteurs

Pour finir il nous suffit de définir les disciplines pour qui nos documents appartiennent pour achever notre travail, dans cet exemple les documents que nous avons manipulés sont des romans et du coup ça appartient à la science humaine et sociale et plus précisément au domaine linguistique ceci représentera notre 3ème groupe qui englobe l'ensemble des sujets et disciplines de nos documents.

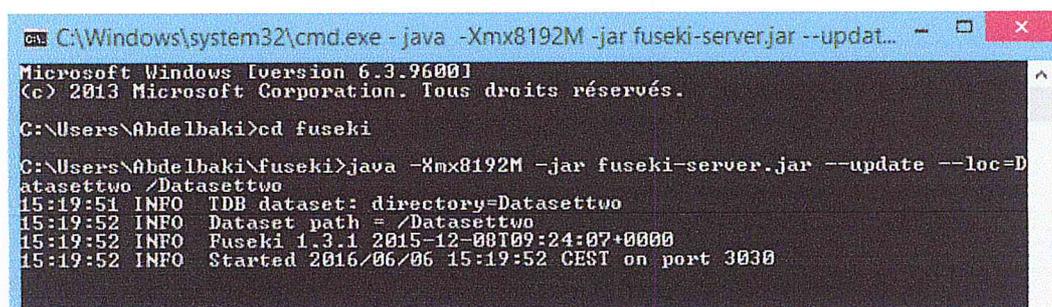
### 3.3 Linked data

Nous avons vu dans la section précédente comment nous avons créé notre propre ontologie et comment nous avons structuré nos données, dans la section qui va suivre nous allons voir comment nous avons publié notre ontologie afin de pouvoir répondre au principe du web des données et pouvoir ainsi relier notre ontologie et du coup nos données avec d'autres données sur le web.

#### 3.3.1 Publication du Linked Data

Le processus de publication commence par la conversion de nos données en triplets RDF afin de contourner la complexité des technologies (XSLT, XPATH..) surtout que protégé génère l'ontologie sous forme XML/RDF ce qui est le but de la conversion des données en triplets RDF est simple mais primordiale pour la suite, qui est de pouvoir rendre chaque ressource dans notre ontologie sous forme de triplet ou chaque triplet est constitué par une association du type < sujet > < prédicat > < objet > ;

l'ensemble des triplets constitue une base de données et stockés dans un triplestore pour former donc un graphe qu'il est possible de requêter grâce au langage Sparql (assez similaire au SQL, le langage de requête des bases de données relationnelles) via un point d'accès web, un « Sparql Endpoint ». Voilà de manière générale la procédure de publication de Linked data, voyons maintenant comment nous avons procédé, pour cela nous avons utilisé l'outil « Apache Jena Fuseki » qui permet le stockage et la manipulation de nos données une fois qu'il génère les triplets RDF, et pour finir nous avons juste besoin de le configurer pour jouer le rôle d'un Sparql Endpoint et afficher nos données, c'est ce que nous allons voir en détail dans ce qui va suivre. Tout d'abord il faut configurer Fuseki et le lancer comme étant un serveur local autonome



```
Microsoft Windows [version 6.3.9600]
(c) 2013 Microsoft Corporation. Tous droits réservés.

C:\Users\Abdelbaki>cd fuseki

C:\Users\Abdelbaki\fuseki>java -Xmx8192M -jar fuseki-server.jar --update --loc=D
atasettwo /Datasettwo
15:19:51 INFO TDB dataset: directory=Datasettwo
15:19:52 INFO Dataset path = /Datasettwo
15:19:52 INFO Fuseki 1.3.1 2015-12-08T09:24:07+0000
15:19:52 INFO Started 2016/06/06 15:19:52 CEST on port 3030
```

FIGURE 3.15: Configuration de fuseki server

Une fois la configuration terminée il nous reste plus qu'à ramener notre ontologie et la publier dans le serveur Fuseki qui va générer de manière automatique les triplets RDF.

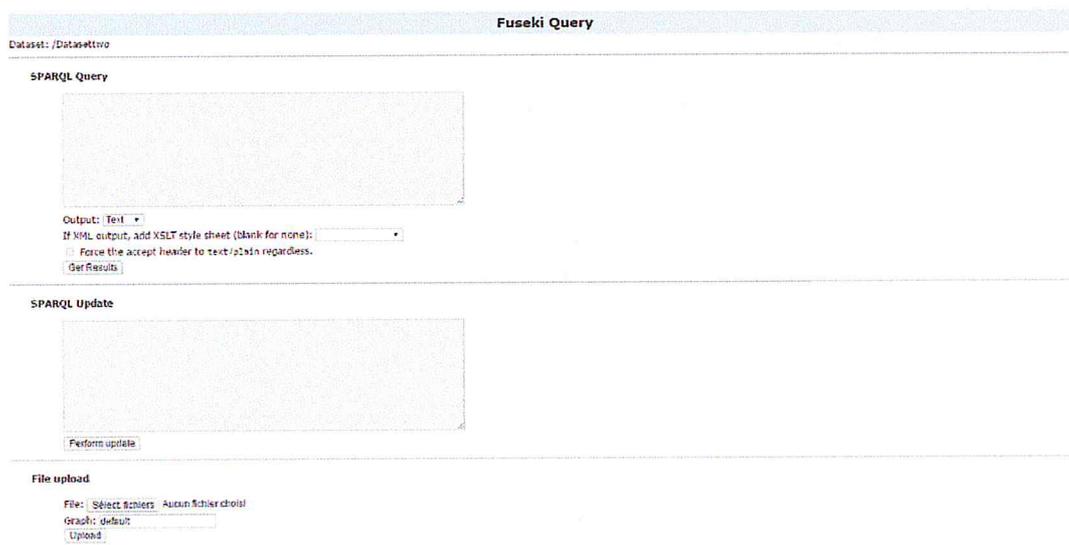


FIGURE 3.16: Interface du fuseki sparql endpoint

Ceci est l'interface de Fuseki ou ce qu'on appelle « Sparql Endpoint » ou on peut apercevoir ces fonctionnalités comme le Sparql Query ou on peut faire des requêtes sur nos données.

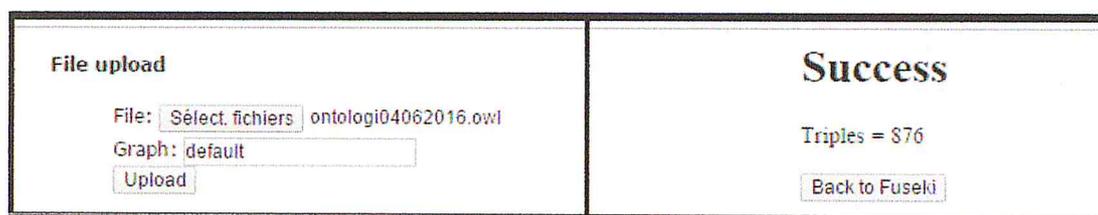


FIGURE 3.17: Conversion de notre ontologie en donnée rdf

Nous avons un message qui confirme la réussite de l'opération de conversion de nos données en triplets rdf ainsi que le nombre de triplet généré

### 3.3.2 Relier nos données avec ceux de la BNF

Nous avons vu dans la section précédente comment nous avons publié nos données mais l'intérêt du linked data n'est pas la publication des données seulement loin de là, mais l'enjeu majeur est de comment enrichir nos données avec des référentiel externes pour des résultats plus pertinents et plus riche en information, notre approche pour montrer cela dans notre travail est d'essayer de lier nos données avec ceux de la bibliothèque nationale de France « BNF », mais avant de parler de notre travail dans cette partie il faut avant tout parlé de la difficulté majeur qu'on a dû faire face, à savoir, l'énorme quantité des données que contient la BNF et qui nous était difficile de manipuler localement afin de les relier avec celles présentes dans notre ontologie et précédemment publiées dans notre serveur Fuseki, pour cela nous avons récupéré les données à travers des dumps par types de donnée ce qui revient à dire que nous avons pris seulement les informations qu'on est susceptible d'utiliser le plus dans notre travail et qui sont : Liste des Œuvres et leur Informations, Liste des auteurs et leur informations, Liste des sujets. Ces données sont récupérées sous forme de RDF/XML et nous avons dû procéder de la même sorte pour les publier dans notre serveur Fuseki. À cet instant nous nous retrouvons avec deux ensembles de données de différentes sources dans notre Dataset intégré dans notre interface de requête ou « Sparql Endpoint ». Pour la suite de notre travail nous allons revenir au principe du linked data afin d'illustrer en cas pratique comment exploiter les données de la BNF avec nos propres données, pour réaliser cela on va rappeler que les linked data se basent sur deux axes majeurs le premier est le « RDF data model » ou comment publier des données structurées dans le web et c'est ce que nous allons expliquer dans la section précédente, le second c'est le « RDF links » ou comment relier deux ensembles de données de différentes sources pour cela on utilise des URI pour identifier nos ressources. Pour illustrer cela, nous allons créer une ressource dans notre ontologie qu'on va nommer « Pierre Lamiral » qui fait référence à un auteur cette ressource on va lui attribuer un URI :

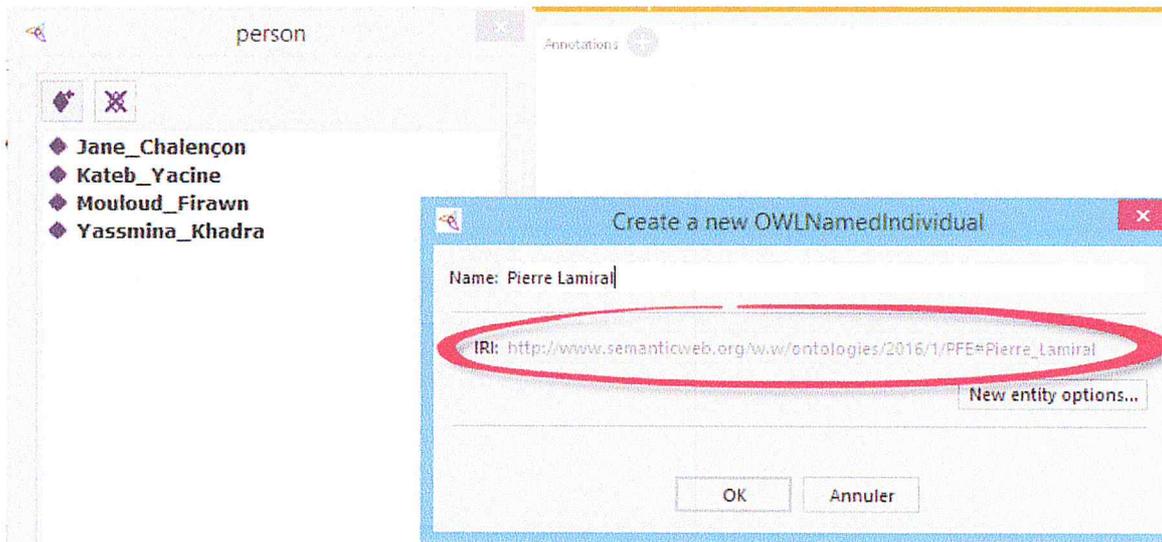


FIGURE 3.18: Création d'une nouvelle ressource et sont URI avec protegé

```
<!-- http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#Pierre_Lamiral -->
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#Pierre_Lamiral">
  <rdf:type rdf:resource="http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#person"/>
  <perso rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Pierre Lamiral</perso>
</owl:NamedIndividual>
```

FIGURE 3.19: Generation du model rdf de la ressource

Maintenant à partir de cette ressource, notre but est de l'enrichir avec des informations supplémentaires d'une autre source, dans notre cas ça sera la BNF et pour cela on va utiliser l'URI pour identifier notre ressource et revenir à notre Fuseki Sparql Endpoint pour faire la requête qui récupéra les données et les affichera dans notre point d'accès, la requête est la suivante :

```
PREFIX pfe:<http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX bio:<http://vocab.org/bio/0.1/>
SELECT ?name ?source ?DateOfBirth ?DateOfDeath
WHERE {
  <http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#Pierre_Lamiral>
  foaf:knows ?name.
  ?source foaf:name "Pierre Lamiral".
  ?source bio:birth ?DateOfBirth.
  ?source bio:death ?DateOfDeath. }
```

FIGURE 3.20: Exemple de requête sur notre ressource

Dans la première ligne on voit clairement notre URI qui définit l'ensemble des ressources de notre ontologie et pour définir une ressource précise il suffit juste de donner le nom après le "#", on peut traduire cette requête par « donnez-moi la date de naissance et de mort de Pierre Lamiral » or que dans notre ontologie nous avons que le nom de l'auteur, le résultat de la requête est le suivant :

name	Source	DateOfBirth	DateOfDeath
"Pierre Lamiral"	<http://data.bnf.fr/ark:/12148/cb169470858#foaf:Person>	"1898"	"1979-09-26"

FIGURE 3.21: Résultat finale

Le lien afficher dans la source nous renvoie dans le profil de l'auteur sur le site de la BNF « <http://data.bnf.fr/> », ainsi on peut faire une multitude de requête à partir de là comme l'ensemble des œuvre de l'auteur ou bien tous les œuvres qui ont une relation avec une de ces œuvre que nous avons dans notre ontologie ainsi de suite.

### 3.4 Conclusion

Dans ce chapitre, nous avons vu de manière détaillée comment nous avons construit le noyau de notre système qui est notre ontologie à travers la norme FRBR qui modélise le domaine de bibliothèque, nous avons aussi vu comment nous avons publié notre ontologie et relier avec les données de la bibliothèque nationale de France pour des résultats plus pertinents et plus riches en informations.

# Chapitre 4

## Réalisation du portail sémantique

### Contents

---

4.1	Introduction . . . . .	84
4.2	Environnement de travail : . . . . .	84
4.3	le portail Cerist : . . . . .	89
4.4	Conclusion . . . . .	91

---

### 4.1 Introduction

Nous arrivons au dernier chapitre de notre mémoire ou nous avons vue précédemment comment nous avons réussi à structuré nos données ainsi que l'intégration de donnée externes pour enrichir la recherche ou ce qu'on appelle les Linked data, dans ce chapitre nous allons arriver à notre dernière étape qui est l'affichage de ces données par le biais du portail documentaire que nous avons conçu selon les critères que nous avons vu dans le chapitre des portails documentaire, mais avant de voir notre portail en détails nous allons tout d'abord présenter notre environnement de travail et les différents outils qui nous ont aidé à achever notre travail.

### 4.2 Environnement de travail :

Cette section consiste à mentionner les différents outils utilisés pour la réalisation de notre étude.

#### 4.2.1 Langage de Programmation :

Pour le développement de l'interface du portail documentaire sémantique nous avons choisi le langage **java** en utilisant sa technologie JSP dans sa version d'entreprise "j2ee", le choix s'est porté sur ce langage à cause du Framework Jena pour la manipulation sémantiques des informations et qui propose une riche documentation gratuite et Open Source.

#### 4.2.2 Éditeurs

Nous allons présenter les deux éditeurs que nous avons utilisé pour la réalisation du portail documentaire sémantique

- **Eclipse** : Pour réaliser notre portail avec Java, nous avons utilisé l'IDE eclipse, c'est aussi une communauté open source dont les projets sont axés sur la construction de plate-forme de développement extensible
- **Protégé** : Protégé est un éditeur qui permet de construire une ontologie pour un domaine donné, de définir des formulaires d'entrée de données et d'acquérir des données à l'aide de ces formulaires sous forme d'instances de cette ontologie. Protégé est également une librairie Java qui peut être étendue pour créer de véritables applications à bases de connaissances en utilisant un moteur d'inférence pour raisonner et déduire de nouveaux faits par application de règles d'inférence aux instances de l'ontologie.

### 4.2.3 Apache Jena Framework

Jena est un Framework Open source pour le langage JAVA qui nous permet d'exercer dans le web sémantique ainsi dans le web de données, ce Framework est constitué d'un certain nombre d'API pour le traitement des données RDF. Ce Framework est la principale raison d'utilisation du langage java, une bonne documentation est mise en disposition dans le site officiel de Jena.

#### Installation et Configuration de Jena Framework

Jena Framework est mis en disposition dans le site officiel de Apache Jena, la figure 4.1 montre l'interface du téléchargement de Jena Framework.

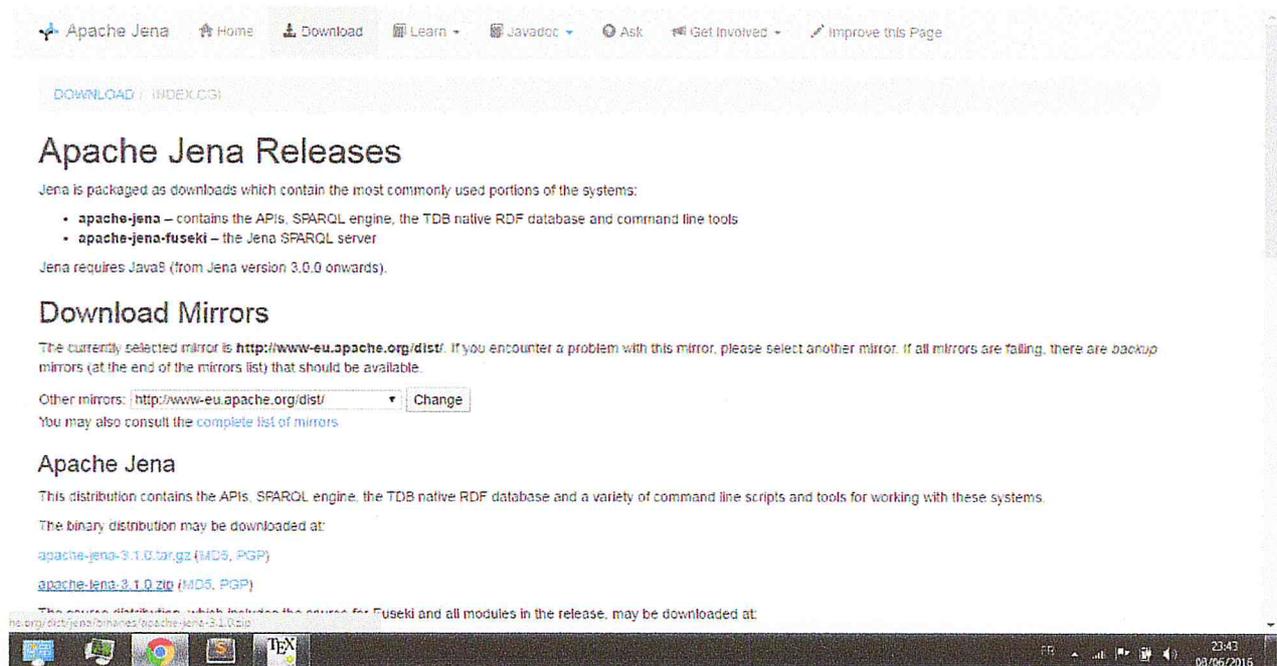


FIGURE 4.1: Jena Downloads

#### Architecture de Jena Framework

Après avoir parler de l'environnement de Jena framework, nous allons montrer son architecture et ses différents interactions avec quelque API sémantiques. La figure 4.4 montre une architecture générale du Framework Jean :

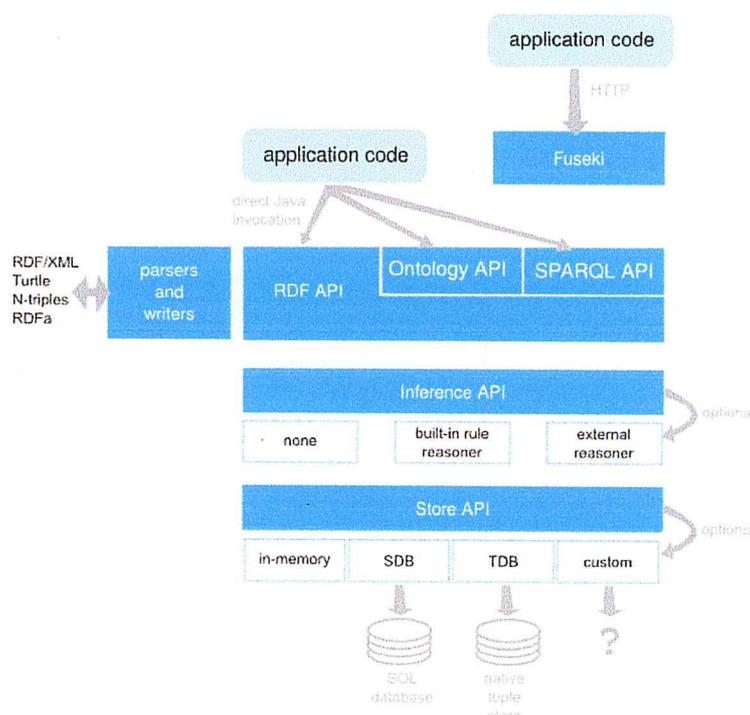


FIGURE 4.2: Architecture Jena Framework

### API RDF

**Rappel :** *RDF est un Framework de Description de Ressource que la W3C le détient parmi ses standards, il permet de description et la structuration des données a fin de permettre aux machines des les comprendre.*

Avec Jéna, les informations qui sont introduit sont sous format de triplet RDF, ces dernier vont être mise dans une conteneur qu'on va appeler "**Model**", ce modèl est une collection de noeuds RDF attachés les uns au autres par des relations définits et unidirectionnelles. En Java, nous avons utiliser la classe Model, elle fournis plusieurs méthodes pour rendre le conteneur des donnée RDF facile a manipulé. En fin, il existe trois différents concepts de conteneurs RDF à Jena :

- **graphique** : Une vue mathématique des relations dirigées entre les noeuds dans une structure connectée
- **Modèle** : Une riche API Java avec de nombreuses méthodes pratiques pour les développeurs d'applications Java
- **Graphique** : Une API simple Java destinée à étendre la fonctionnalité de Jena [App16]

Pour utiliser Jena, on doit donc passer par trois étapes :

- Créer un **Model** de donnée qui va contenir notre ontologie en format RDF/XML.
- Traduire une requête **sparQL** en une requette HTTP(request/response).
- Afficher les résultat a partir de l'ontologie.

1. **Création du Model** : Dans un premier temps, il faut créer le modèle de notre ontologie en format RDF/XML, ceci en la stockant dans un fichier à part qu'on va appeler mode. Ensuite nous allons l'instancier à l'aide de l'instruction suivante :

```
//Connexion
static String OpenConnectOWL(String serviceURI){
    Model mode = ModelFactory.createDefaultModel();
    DatasetAccessorFactory factory;
    DatasetAccessor accessor;
    accessor = DatasetAccessorFactory.createHTTP(serviceURI);

    accessor.putModel(mode);
    return (serviceURI);
}

static String GetResultAsString(String serviceURI, String Query){
    String s=null;

    try {
        com.hp.hpl.jena.query.Query query = QueryFactory.create(Query);
        QueryExecution qe = QueryExecutionFactory.sparqlService(serviceURI,
            query);
        com.hp.hpl.jena.query.ResultSet results = qe.execSelect();
        if(results.hasNext()){
```

FIGURE 4.3: instanciation du modèle de l'ontologie

2. **Interrogation de l'ontologie** : Nous avons maintenant la possibilité d'interroger l'ontologie, et cela par traduire les requête sparQl en requête HTTP avec les instruction **queryString** dans la figure 4.2, le cadre vert montre la translation de la requête SparQl
3. **Récupération des données** : Il ne reste qu'à récupérer les données voulue de l'ontologie a travers **GetResultAsString** dans la figure 4.2, le cadre rouge montre la récupération des informations de l'ensemble de données.

```
1 package pfe;
2
3
4 //Imports
5 import com.hp.hpl.jena.ontology.OntModel;
6
7
8
9
10
11
12
13
14
15
16
17
18 public class Main {
19
20     public static void main(String[] args) {
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35     String queryString = "PREFIX pfe:<http://www.semanticweb.org/w.w/ontologies/2016/1/PFE#> "+
36         "PREFIX foaf:<http://xmlns.com/foaf/0.1/>"+
37         "PREFIX bio:<http://vocab.org/bio/0.1/>"
38         + "SELECT ?Nom ?auteur ?jour ?dateOfDeath ?dateOfBirth "
39         + "WHERE { ?name pfe:Nom ?Nom. "
40
41         + "}" ;
42
43
44     String a = GetResultAsString("http://localhost:3030/Datasettwo/query",queryString);
45     System.out.println(a);
46
47
48
49
50 }
```

FIGURE 4.4: JenaFramework Eclipse

### 4.2.4 Serveurs :

- **Fuseki** : est un serveur SPARQL qui permet la manipulation et la requête des graph RDF via le protocole HTTP et c'est le meilleur choix pour une architecture client/serveur, il est considéré aussi comme un module du framework jena et déployant un point d'accès SPARQL pour stocker des données dans le triple store TDB.
- **Apache Tomcat** : est un conteneur libre de servlets et JSP Java EE. Issu du projet Jakarta, c'est un projet principal de l'Apache Software Foundation. Il implémente les spécifications des servlets et des JSP du Java Community Process, est paramétrable par des fichiers XML et de propriétés, et inclut des outils pour la configuration et la gestion. Il comporte également un serveur HTTP.

### 4.2.5 SparQL

SparQL est un langage de requête pour les données qui sont structurées par RDF, ce langage interroge uniquement les informations contenues dans le modèle, son travail consiste à prendre la description de ce que l'application veut, sous la forme d'une requête, et renvoie cette information, sous la forme d'un ensemble de liaisons ou d'un graphe RDF. [App16]

### 4.3 le portail Cerist :

Maintenant nous allons voir une revue de l'interface de notre portail "*Cerist*", c'est un portail documentaire sémantique dédié à la recherche sémantique des informations des bibliothèques, conçu sur le modèle FRBR pour les notices bibliographique et adopte le principe des linked data. le portail est concidéré comme résultat de notre travail sur ce document.

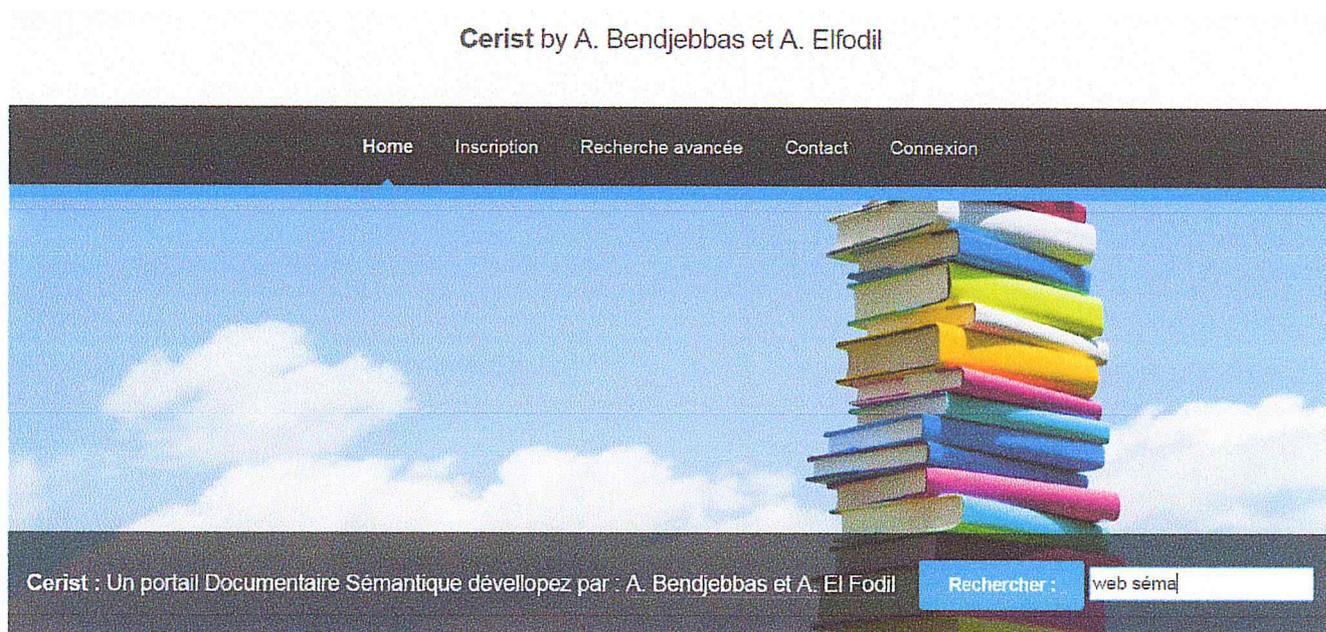


FIGURE 4.5: Prise du portail documentaire sémantique

Comme toute portail,"*Cerist*" propose un ensemble de fonctionnalité de base que nous avons vu toute au long de notre recherche sur les portail documentaire a savoir les notions de Recherche et d'authentification. Dans notre portail deux type de recherche sont proposé la première est une recherche simple par mot clé, la deuxième c est ce qu'on appelle recherche avancée qui permet de proposer de la précision, ainsi un utilisateur peut chercher par mot clé ou bien en cochant les cases souhaitée, la figure suivante est une prise de vue pour la recherche avancée du portail.

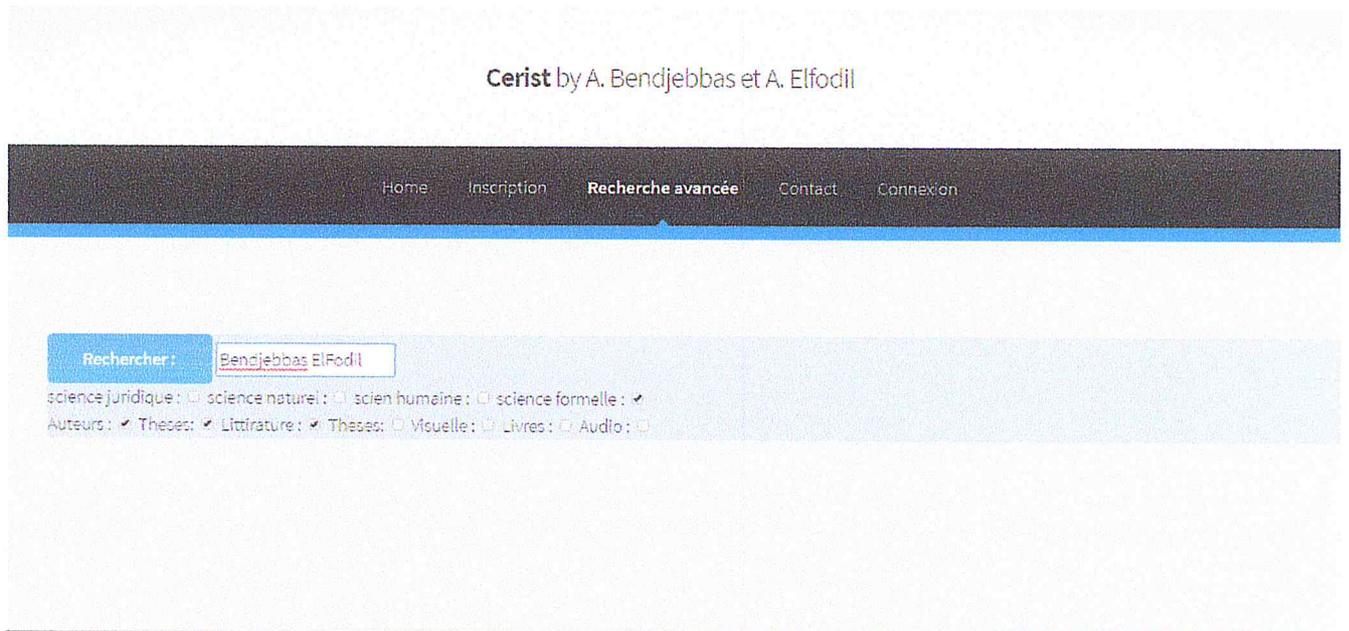
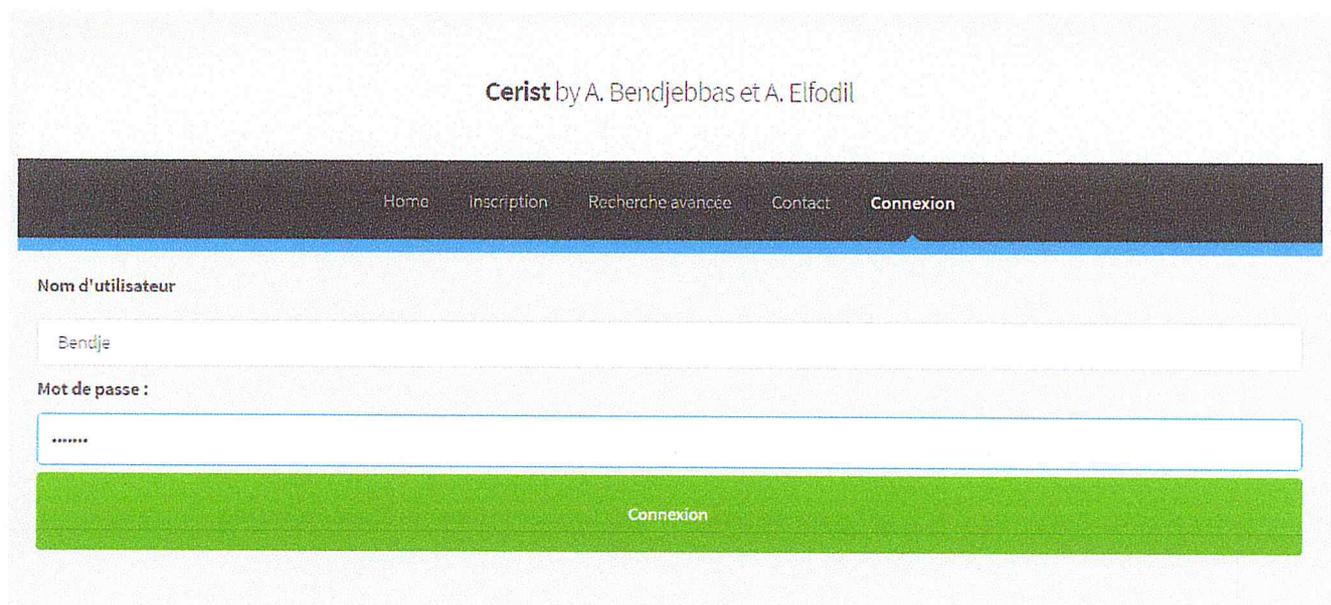


FIGURE 4.6: Prise du portail documentaire sémantique : Recherche avancée

D'autres fonctionnalités sont présentes dans notre portail comme l'inscription et la connexion à notre portail.

The screenshot shows the 'Inscription' (Registration) page of the Cerist portal. At the top, the same dark navigation bar is visible with links for 'Home', 'Inscription', 'Recherche avancée', 'Contact', and 'Connexion'. The main content area contains a registration form with the following fields: 'Nom', 'Prenom', 'Nom d'utilisateur', 'Mot de passe', and 'Confirmation du mot de passe'. Each field is represented by a text input box.

FIGURE 4.7: Prise du portail documentaire sémantique : Inscription



Cerist by A. Bendjebbas et A. Elfodil

Home Inscription Recherche avancée Contact **Connexion**

Nom d'utilisateur

Bendje

Mot de passe :

\*\*\*\*\*

Connexion

FIGURE 4.8: Prise du portail documentaire sémantique : connexion

### 4.4 Conclusion

Nous avons vu tout au long de ce chapitre les différents outils et technologies nécessaires pour la réalisation de notre système tels que l'environnement de développement, les langages de programmations, les bibliothèques open source utilisée. Nous avons montré par la suite les différente fonctionnalité que propose notre portail documentaire.

## **Quatrième partie**

### **Conclusion Générale**

---

## Conclusion Générale

Nous arrivons à la fin de notre mémoire, une conclusion générale est pour nous une occasion de revenir sur les principales étapes de la recherche que nous venons de présenter.

Le travail prévoyait dans ce document était de traiter une thématique qui révèle d'être très intéressante pour la recherche sur le web d'aujourd'hui. En effet grâce à ce sujet nous avons pu prendre connaissance et mettre en pratique des notions nécessaires pour la mise en place d'un portail documentaire sémantique avec les technologies du web de données, notamment les ontologies, le modèle FRBR, et l'alignement des données avec celles de la Bibliothèque Nationale de France.

Pour cela :

- Nous avons créé une ontologie pour le modèle FRBR, pour standardiser la recherche documentaire des bibliothèques avec des technologies sémantiques.
- Nous avons relié les données de la bibliothèque nationale de France qui utilise aussi le modèle FRBR avec les principes du web de données.
- Nous avons réalisé un portail documentaire sémantique pour le centre de recherche CERIST, qui est une interface web qui fait la recherche de document de manière sémantique.

# Bibliographie

- [af11] *Linked Data : Evolving the Web into a Global Data Space*. MorganClaypool, 2011.
- [App16] Apache. Apache jena framewok, 2016.
- [BRE09] Sophie BREZEL. Concevoir un portail documentaire académique des pistes pour la bibliothèque de sciences po. *Diplôme de conservateur de bibliothèque*, 2009.
- [bus16] businessdictionary. Bdictio, 2016.
- [dF16] Bibliothèque National de France. Présentation générale du projet data.bnf.fr, 2016.
- [F.B12] A.Napoli F.Badrav, J.Lieber. Etude et conception d'un portail semantique dans le cadre du systeme kasimir de gestion des connaissances et d'aide a la décision en cancérologie, 2012.
- [Fou16] OpenKnowledge Foundation. Opk, 2016.
- [GD12] Marion Grand-Démery. Évolution d'un moteur de recherche fédérée dans un contexte multiinstitutionnel : La bibliothèque ouverte montpellier - languedocroussillon, 2012.
- [IFL12] IFLA. Fonctionnalités requises des notices bibliographiques, 2012.
- [Lav16] LavalUniversity. Collection mémoires et thèses électroniques, 2016.
- [Leb09] Corinne Leblond. L'expérience du visual catalog : comment mettre les usagers au coeur du dispositif de médiation documentaire". 2009. <http://slideplayer.fr/slide/2265680/>.
- [Mai03] Marc Maisonneuve. "recherche multibases : de nouveaux outils pour accroître l'autonomie des usagers". 2003. [http://www.toscaconsultants.fr/articles/recherche\\_multibases.pdf](http://www.toscaconsultants.fr/articles/recherche_multibases.pdf).
- [N.P14] J.HÜGI N.PRONGUÉ. Les bibliothèques face aux linked open data : de nouvelles applications web et de nouvelles compétences professionnelles, 2014.
- [Pal01] Sean B. Palmer. The semantic web : An introduction, 2001.
- [Pap11] Fabrice Papy. Bibliothèque numérique et recherche : Les freins au usages, 2011.
- [Piq12] Julien Piquot. Petite introduction au web sémantique : les technologies du web sémantique, 2012.

## BIBLIOGRAPHIE

---

- [q01] *SEAL : a framework for developing SEmantic PortALs*. Proceeding, 2001.
- [r06] *The Semantic Web Revisited*. IEEE, 2006.
- [Tai13] Anne-Laure Taillant. Les enjeux d'un portail documentaire dans la mise en place des formations paramédicales, 2013.
- [Tip11] Chris Tipping. Web sémantique, 2011.
- [w3c16] w3c. Vocabularies, 2016.
- [y08] *Applied Ontology An Introduction*. ontos verlag, 2008.
- [yu11] Liyang yu. *A Developer's Guideto the Semantic Web*. Springer, 2011.