Université Abou Bekr Belkaid
Tlemcen Algérie
جامعة أبي بكر بلقايد
تلمسان الجزائر

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

Mémoire de fin d'études

Pour l'obtention du diplôme de Master en Informatique

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

Thème

Sémantification du Dossier de Patient

Réalisé par :

- ***BELAIDOUNI Nesrine***
- ***MOSTEFAI Ilheme Fatima Zohra***

Présenté le 22 Juin 2017 devant le jury composé de :

- *M^r BELABED Amine* *(Président)*
- *M^r CHIKH Azzedine* *(Encadreur)*
- *M^r TADLAOUI Mohammed* *(Examinateur)*

Année universitaire: 2016-2017

REMERCIEMENTS

Nous remercions Dieu tout puissant pour la volonté, la santé et la patience pour accomplir ce modeste travail.

Au terme de ce modeste travail nous tenons un vif remerciement et gratitude à notre promoteur ***Mr. CHIKH Azzedine*** *pour ces conseils et ses orientations judicieuses et précieuses qui nous ont été infiniment utiles le long de ce travail.*

Nos remerciements s'adressent aussi à :

- ***Mr. BELABED Amine*** *qui a en la bienveillance d'accepter de juger ce travail et de présider le jury de soutenance.*
- ***Mr. TADLAOUI Mohamed*** *pour avoir accepté d'examiner ce travail et honorer le jury ;*

A tous ceux qui nous ont aidé de près ou de loin, à tous ceux qui nous ont encouragé, à tous ceux qui nous ont accordé leur confiance.

Enfin, nous saisissions cette même occasion pour remercier tous nos amis et toutes les personnes ayant contribué, de près ou de loin, à l'élaboration de ce travail.

Merci.

Dédicace

Je dédier ce modeste travail :

A

Mes très chers parents ma mère et mon père, pour leur patience, leur amour, leur soutien, et qui m'ont toujours poussé et motivé dans mes études.

A

Mes chers frères pour leur encouragement, leur conseil et leur aide pendant toute ces années d'études.

A

Mes cousins, mes cousines, mes ami « e »s ainsi tous mes camarades de la promotion SIC.

A

Toute personne qui nous aidés du près ou du loin à terminer ce modeste travail.

✍ *BELAIDOUNI Nesrine*

Dédicace

Je dédie ce modeste travail à l'âme de ma mère.

✍ MOSTEFAI Ilheme Fatima Zahra

Table des matières

Glossaire 11

Introduction générale 1

Partie I : 1

Background 1

Chapitre I : 2

Dossier Electronique de Patient 2

1. Introduction 4

2. Historique du dossier Médical 4

3. Types de dossier Médical 5

a. Dossier Santé Electronique 5

b. Dossier Médical Electronique 5

c. Dossier Electronique de Patient 5

4. Bénéfice du dossier électronique de patient 5

Chapitre II : 4

Web Sémantique 4

1. Introduction 6

2. Web sémantique 6

3. Historique du web 6

3.1. Web traditionnel 6

3.2. Web structuré 7

3.3. Web sémantique 7

3.4. Différences entre le WEB traditionnel et le WEB sémantique 7

4. Objectif du web sémantique 8

5. Architecture du web sémantique 9

5.1. La couche URI 9

5.2. La couche XML 9

5.3. La couche RDF 10
5.3.1. Type de RDF 11
5.4. La couche Ontologie 13
5.5. Les couches restantes 13
Chapitre III : 7
Les Ontologies 7
1. Introduction 14
2. Définition 14
3. Composantes d'une ontologie 14
4. Classification des ontologies 15
4.1 Typologie selon l'objet de conceptualisation 15
4.2 Typologie selon le niveau de formalisation utilisé 16
5 langage ..... **Error! Bookmark not defined.**
5.1 OWl 16
5.2 SPARQL 17
Exemple 18
6 Editeurs d'ontologie 19
6.1 *OntoEdit* 19
6.2 *Ontolingua* 19
6.3 *Protégé2000* 19
6.4 INA 19
Chapitre IV : 15
Annotations 15
1. Introduction 20
I. Annotation classique 20
1. Définition 20
2. Méta-donnée ou Annotation 20

3. Cycle de vie des annotations classiques 21
4. Acteurs interagissant avec l'annotation classique 21
5. Quelques outils pour l'annotation classique 22
II. Annotation sémantique 22
1. Définition 22
2. Annotation et méta-donnée sur le web sémantique 23
3. Caractéristiques des annotations sémantiques 24
3.1. Processus des annotations sémantiques 24
3.2. Exploitation des annotations sémantiques 24
3.3. Méthodes et éléments de base pour l'annotation sémantique 25
4. Les outils d'annotation sémantique 25
Partie II- Contribution : Sémantification 21
Du Dossier Electronique Médical 21
1. Introduction 27
2. Composition du DEP 28
a. Renseignements personnels 28
b. Renseignements médicaux 29
3. Architecture du système 30
a. Module 1 : « Editeur d'ontologie » 30
b. Module 2 : « Editeur DTD/XSD » 33
c. Module 3 : « Editeur DEP » 34
d. Module 4 : « Générateur d'Annotations Sémantiques » 36
e. Module 5 : « Annotateur sémantique » 37
f. Module 6 : « Recherche Sémantique » 38
4. Annotateur 39
4.2 L'Annotateur Sémantique « semi-automatique » 39
4.3 Les annotateurs sémantiques automatiques 40

Conclusion générale 41

Références bibliographiques 27

Webographies ..... **Error! Bookmark not defined.**

Table des Figures

Figure 1: Architecture du web sémantique ‘layer cake’ 9
Figure 2: Exemple d'un model RDF 10
Figure 3: Classification des ontologies Selon l’objet de conceptualisation 15
Figure 4: Relation entre l’annotation et la Méta-donnée 21
Figure 5: Annotation sémantique 23
Figure 6: Interface du DEP 28
Figure 7: Partie renseignements personnels 29
Figure 8: Extrait du renseignement médical 29
Figure 9: Ontologie: Structure du DEP 30
Figure 10: Ontologie de vie privée 31
Figure 11: Ontologie de la Famille 31
Figure 12: Ontologie Médicale "DMTO" 33
Figure 13: Module "Editeur DTD/XSD" 33
Figure 14: Extrait du DTD du DEP 34
Figure 15: Module "Editeur DEP" 34
Figure 16: Extrait du fichier XML DEP 35
Figure 17: Validité du document xml 35
Figure 18: Module "Générateur d'annotation sémantique" 36
Figure 19: Module "Annotateur Sémantique" 37
Figure 20: Module "Recherche sémantique" 38
Figure 21: Architecture du système global 38
Figure 22: Extrait du code généré par l'API Jena 39
Figure 23: Extrait du code de transformation via XSLT 40
Figure 24: Extrait du fichier annoté en RDF 40

Glossaire

DEP: Dossier Electronique de Patient

DTD: Document Type Description.

IHM: Interface Homme Machine.

NS :name space.

OWL: Web Ontology Language.

RDF: Ressource Description and Framework.

RDFS: Ressource Description and Framework Schéma.

SPAQL: Protocol and RDF Query Language.

URI: UNIFORM RESOURCE IDENTIFIER.

W3C: World Wide Web Consortium.

XML: eXtensible Markup Language.

XSLT: Extensible Stylesheet Language Transformations.

Introduction générale

La mémoire humaine est limitée et le médecin ne peut avoir l'ensemble des connaissances médicales dont il a besoin pour sa pratique quotidienne. Avec l'augmentation rapide de la population l'enregistrement numérique des données de patients est devenu indispensable.

Une solution peut être trouvée dans le Web sémantique qui vise à enrichir les structures syntaxiques du Web actuel avec leur contenu sémantique. Par conséquent, l'exploitation de ces technologies dans le cadre médical électronique semble être bénéfique. Surtout pour assurer un rapprochement automatique entre le patient et l'ensemble médical (Médecin, Pharmacien, Assurance). Le principe consiste à expliciter les informations du patient à l'aide d'éléments ontologiques formant un référentiel commun (ontologie), inspiré des parties communes entre le Dossier Electronique du Patient (DEP) et l'entourage médical. Ce référentiel peut être utilisé par un médecin (ou infirmier) a fin d'enlever l'ambiguïté, pour annoter son document avec des éléments pertinents. Les annotations résultantes peuvent être exploitées pour rendre le DEP intelligible.

Les nouvelles générations de Technologies de l'Information pour la Santé sont en train de passer d'une problématique de gestion des données à une problématique de gestion des informations. Ce passage impose une modélisation complexe des données et des connaissances médicales. Parmi les problèmes rencontrés, nous citons :

* Avoir une terminologie commune du domaine de la santé ;
* Avoir une interopérabilité des systèmes d'information de santé ;
* Avoir un consentement des patients avant le partage de leurs données médicales sensibles.

Plusieurs travaux concernant la sémantification du dossier médical ont vu le jour. Bataille et al. se focalisent dans [Bat, 2002] sur la modélisation et l'exploitation de dossiers et de documents numériques. Ils ajoutent une valeur à la version numérique par rapport à la version papier. Le lecteur, doit disposer d'outils sémantiques pour naviguer plus efficacement vers ce qui l'intéresse en fonction de sa tâche de lecture.

Alors que l'auteur doit pouvoir produire de nouvelles ressources par réutilisation de certaines déjà existantes et être assuré d'une gestion efficace de l'évolution de ses productions. Pour atteindre ces buts, ils proposent une approche ontologique et orientée objet fondée sur le concept central de relation sémantique.

Après un constat qui affirme que le document est le support le plus approprié pour manipuler les connaissances médicales pendant les soins, les auteurs proposent l'utilisation des annotations dans les documents sachant que cette fonctionnalité existe déjà au niveau des logiciels de traitement de texte et de collaboration. Barry et al. [Bar,2005] montrent l'intérêt d'une sémantique hypertextuelle annotationnelle pour l'appliquer sur le DEP et les fonctions qui en découlent.

Après plusieurs années de proposition des approches et mécanismes pour gérer un DEP sémantique, d'autres alors ont proposé d'utilisé les ontologies, tel que Charlet et al. [Cha, 2009] qui présentent dans leur article des exemples dans trois spécialités différentes dont ils font une coordination entre les thésaurus médicaux et les ontologies. Ils prennent en considération l'expression linguistique des connaissances acquises par les professionnels de santé durant leurs activités.

En 2012, il y a eu la naissance du projet (ontolurgences) développé par Declerck et al. [Dec ,2012] .Ce projet est basé sur la méta modélisation pour modéliser l'ensemble du système organisationnel et structuré des connaissances et de mettre le lien entre les concepts et la façon dont ils sont nommés dans le DEP.
Ontolurgences est une Ressource Termino-ontologique (RTO). Elle assure :

1- le rôle du modèle du domaine répertoriant tous les concepts pertinents

2- le lien entre les concepts et la façon dont ils sont nommés dans les documents du DEP

Le développement d'ONTOLURGENCES a été passé par 6 étapes :

1- construction du squelette ontologique grâce à une méthode à base d'analyse de corpus

2- utilisation de ressources terminologiques et ontologiques existantes pour compléter manuellement le système de concepts de la RTO,

3- enrichissement automatique de la RTO au niveau des termes

4- enrichissement semi-manuel de la RTO au niveau des termes

5-enrichissement de la RTO en concepts en rapport avec les médicaments

6-mise en œuvre de procédures de validation et assurance qualité.

Le présent travail de master rentre dans le cadre du Web sémantique et son application au domaine médical. Il consiste à faire une sémantification du dossier électronique du patient (XML) à base d'ontologies (OWL). Quatre ontologies (OWL) ont été choisies : (1) une ontologie médicale schématisant les concepts clés de la pratique médicale; (2) une ontologie familiale illustrant les relations familiales ; (3) une ontologie documentaire représentant l'organisation logique d'un dossier électronique du patient ; (4) une ontologie de vie privée montrant les niveaux de confidentialité et les droits d'accès. Les deux premières ontologies ont été adoptées à partir de l'existant pendant que les deux dernières ont été créées par nous-mêmes. L'édition et l'utilisation de ces ontologies est fait grâce à l'éditeur d'ontologies PROTEGE.

La sémantification consiste à créer des annotations sémantiques (RDF) de parties du dossier électronique du patient par des concepts émanant de l'une des quatre ontologies précitées. La sémantification peut se faire selon deux approches : semi-automatique et automatique. Dans la première, les annotations sont créées par un expert humain à travers un outil logiciel (annotateur sémantique). Dans la seconde, les annotations sont générées automatiquement à partir du dossier électronique du patient à travers une feuille de style XSLT (générateur d'annotations sémantiques). Les annotations sémantiques résultantes sont ensuite interrogées grâce à SPARQL.

Le présent mémoire est structuré comme suit. Après cette introduction la première partie est dédiée au background. Elle est composée de quatre chapitres. Le chapitre 1 présente le dossier électronique de patient. Dans le deuxième chapitre, nous présentons en détails le web sémantique. Le troisième chapitre est dédié aux ontologies. Nous présentons le domaine des annotations dans le quatrième chapitre. La deuxième partie est dédiée à notre contribution. Enfin nous terminons par une conclusion, où nous rappelons notre principale contribution et nous traçons quelques perspectives.

Partie I:

Background

Chapitre I :

Dossier Electronique de Patient

1. Introduction



De nos jours, les personnes sont extrêmement mobiles ; elles changent de lieu de résidence, de médecin ou de caisse maladie. Pour les patients et les professionnels de la santé qui les soignent, il s'ensuit que les informations médicales actuelles et indispensables peuvent, à un instant crucial, faire défaut ou nécessiter de fastidieuses recherches.

Dans ce premier chapitre, nous allons présenter un aperçu de différentes sortes du dossier médical et de son évolution.

2. Historique du dossier Médical

Au début, les observations des soignants étaient une simple prise de notes sur papier. Au XIVème siècle a apparu la notion de « dossier de patient ». Comme un support écrit, il servait à la réunion et l'archivage des notes du médecin. Qui veut dire, il permettait au médecin de ne rien oublier sur son patient. De plus, il permet le partage des informations du malade avec d'autres médecins, équipes soignantes et/ou la famille. A la fin du XVIIème siècle avec l'arrivage de concept de l'informatisation des données a apparu « le dossier médical informatisé » pour chaque patient. Il était alors utilisé comme un support d'enregistrement. La complexité de la prise en charge, l'optimisation de la qualité des soins et l'évolution sociétale en termes de droits des personnes hospitalisées ont contribué à la valorisation de dossier médical. Depuis 1970, les dossiers de soins infirmiers font partie des dossiers médicaux. L'ensemble est devenu un outil de communication et de transmissions des données entre les professionnels de santé et ce quel que soit leur type d'exercice (secteurs hospitalier et libéral). Puis l'objectif du dossier médical a évolué en ajoutant de la sémantique hypertextuelle afin de relier entre les concepts du Dossier Electronique de Patient « DEP » en conclusion, le dossier médical comportait des notes prises par les médecins lors des consultations est aujourd'hui un document électronique de patient permet la facilité d'exploitation et de manipulation des connaissances. [Ben, 2011]

3. *Types de dossier Médical*

a. Dossier Santé Electronique

Un dossier santé électronique (DSE) est un outil qui permet aux spécialistes de la santé (médecin et professionnel) d'avoir accès à des données supposé essentiels pour intervenir rapidement et assurer un suivi de qualité auprès de leurs patients. [7]

b. Dossier Médical Electronique

Un dossier médical électronique (DME) englobe différents informations concernant un seul patient. Ces informations peuvent être des renseignements sociodémographiques, des antécédents médicaux, un profil pharmaceutique ou des diagnostics (résultats de laboratoire et d'imagerie) etc… Ce type de dossier est informatisé, qui est exploité par un clinicien, une organisation ou cabinet, peut être intégré à d'autres logiciels tels que la facturation et la gestion des rendez-vous etc… pour plus de performance. [7]

c. Dossier Electronique de Patient

Le Dossier Electronique de Patient (DEP) est une mémoire dans laquelle, on regroupe tout ce qui est connu d'un patient, toutes les informations nécessaires à sa prise en charge (ce qui s'est passé, ce qui a été dit, ce qui a été fait) et à la surveillance d'un patient. De plus il aide à la communication avec d'autres professionnels (médecin remplaçants, pharmaciens,…). Son utilisation est ainsi à la fois individuelle et collective. [Emi, 2010]

4. Bénéfice du dossier électronique de patient

La traçabilité du suivi permet une meilleure qualité de la prise en charge. C'est un moyen d'améliorer la qualité des soins, à fin d'avoir une meilleure gestion de la santé de population. De plus, l'accès du patient à son dossier médical permet de lui fournir une information claire de son état et de sa prise en charge. Par conséquent l'intérêt est double, d'un côté pour le professionnel et d'un autre pour le patient [Mor,2017]

Chapitre II :

Web Sémantique

1. Introduction

Dans le premier chapitre, nous avons exposé l'évolution de la technologie de l'informatisation du dossier médical.

L'objectif de ce chapitre est la description du web sémantique en vue de mieux comprendre la notion de la sémantification et ses objectifs.

2. Web sémantique

Dès la création du web par Tim Berners Lee, au début des années 1990, il a été prévu de partager des informations sous forme de pages HTML, affichables par un «navigateur web », et généralement destinées à être lues par un utilisateur humain.

Très rapidement, on s'est rendu compte que cette conception du web était bien trop limitée, et ne permettait pas un réel partage du savoir. Les limites du Web actuel sont nombreuses, nous mentionnons:

- L'hétérogénéité des formats d'information (Word, pdf …)
- la nature dynamique et évolutive de l'information
- L'absence de sémantique ou structure de l'information elle-même, etc

D'après son créateur, le web sémantique est défini comme suit : " Le Web sémantique est une extension du Web actuel dans lequel l'information est munie d'une signification bien définie permettant aux ordinateurs et aux personnes de mieux travailler en coopération. " [Sci,2001]

3. Historique du web

3.1. Web traditionnel

Le Web traditionnel reposant sur l'hypertexte, on a écrit le langage HTML (HyperText Markup Language). Il est conçu comme un langage spécifiant le contenu d'un document avec d'importantes extensions d'hypertexte sans en spécifier la forme à l'origine. Il n'est pas conçu pour être le langage de traitement de texte d'impression conforme à la visualisation

(WYSIWYG) telle que Word et WordPerfect. Ce choix a été fait parce qu'un même document HTML peut être affiché par de nombreux navigateurs de différentes capacités. Ainsi, par exemple, le HTML permet de marquer des parties de texte comme les titres ou les paragraphes, et puis laisse l'interprétation de ces éléments marqués au navigateur ce qui donne une interprétation différente suivant le navigateur utilisé. Mais tout le monde y a rajouté des choses bien spécifiques. [Rep, 2002]

3.2. Web structuré

Le web structuré reposant sur le "Standard Generalized Markup Language " a été très utilisé pour le traitement de la documentation technique dans l'industrie (aviation, automobile..) et relativement peu pour le traitement de la documentation institutionnelle ou commercial. [Rep, 2002]

3.3. Web sémantique

Le Web sémantique " le Web de données " permet aux machines de comprendre la sémantique, l'idée est de parvenir à un Web intelligent, où les informations ne seraient plus stockées mais comprises par les ordinateurs. Il étend le réseau des hyperliens entre des pages Web classiques par un réseau de lien entre données structurées permettant ainsi aux agents automatisés d'accéder plus intelligemment aux différentes sources de données contenues sur le Web. [Rep, 2002]

3.4. Différences entre le WEB traditionnel et le WEB sémantique

WEB TRADITIONNEL	WEB SEMANTIQUE
Web informel, contenus hétérogènes Le web ne respecte pas les critères de spécification du W3C	**Web formalisé, contenus structurés** Le web répond à des normes et Spécifications
Web opaque L'information est noyée dans le flux de Données	**Web transparent** L'information est visible, claire et accessible
Web quantitatif, contenus quelconques On table sur l'accumulation d'informations disponibles sur le web	**Web qualitatif, contenus de valeur** On exploite la valeur ajoutée des informations disponibles sur le web
Recherche de mots On ne s'intéresse qu'aux termes isolés	**Recherche de concepts** On s'intéresse au sens des mots, au contexte

4. *Objectif du web sémantique*

Le Web sémantique a pour objectif de permettre à des machines d'accéder aux sources d'information et aux services du Web. Cette évolution pourra permettre d'automatiser la découverte des services pertinents par rapport à une requête, et également de combiner automatiquement plusieurs services lorsque la demande de l'utilisateur ne peut être satisfaite par un seul service prédéfini mais correspond à un service complexe. En effet, une description sémantique des ressources du Web permettra de les identifier de façon plus précise. Par ailleurs, il sera possible d'exécuter automatiquement les services du Web. Les travaux que nous effectuons se situent dans le cadre du Web Sémantique.

Les deux objectifs principaux du web sémantique sont :

- Le Web doit devenir un moyen très puissant de coopération entre les humains.
- Les ordinateurs présents sur le réseau doivent coopérer : ils seront capables d'analyser toutes les données qui y circulent et de chercher les informations pertinentes pour l'utilisateur dans les bases de données accessibles aussi bien que dans les sites existants.

Ces données sont ensuite traitées et présentées correctement à l'utilisateur.

Il est important de noter que le terme "Sémantique" est employé dans un sens différent de celui utilisé dans le langage naturel. En effet, "Sémantique" signifie ici "interprétable par les machines". Les machines devront être capables d'utiliser des ressources provenant de diverses sources. C'est pourquoi ces ressources devront être disponibles, c'est-à-dire ouvertes et compréhensibles. [Bou, 2008]

5. Architecture du web sémantique

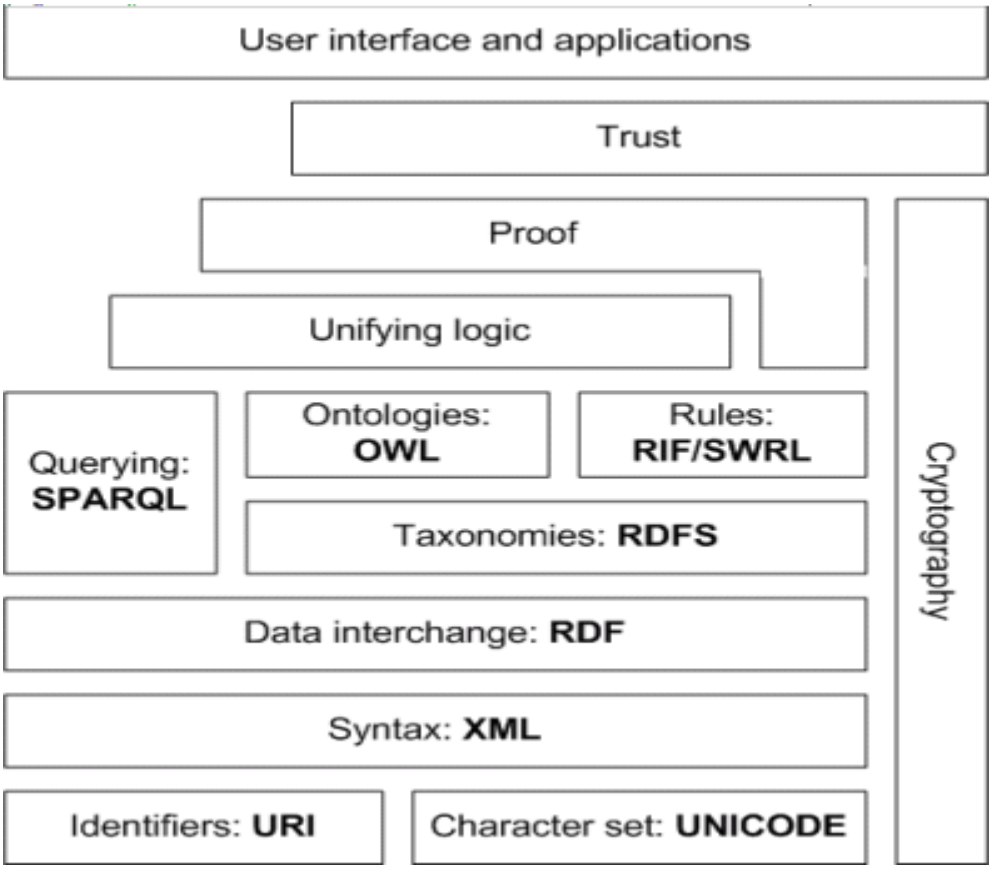


Figure 1: Architecture du web sémantique 'layer cake'[Geo, 2013].

5.1. La couche URI

L'URI chaîne de caractères respectant les normes d'Internet. Est un identifiant de façon unique une ressource ou un concept sur le Web et ne mène pas nécessairement à une ressource existant effectivement sur le réseau.

Exemple :

RFC 3986 ➔ http://www.gbiv.com/protocols/uri/rfc/rfc3986.html

5.2. La couche XML

Le langage **XML** est un format général de documents orienté texte. Il s'est imposé comme un standard incontournable de l'informatique. Il est aussi bien utilisé pour le stockage de documents que pour la transmission de données entre applications. Sa simplicité, sa flexibilité et ses possibilités d'extension ont permis de l'adapter à de multiples domaines allant des données géographiques au dessin vectoriel en passant par les échanges commerciaux. De

nombreuses technologies se sont développées autour de XML et enrichissent ainsi son environnement.

Le langage XML dérive de SGML et de HTML. Comme ces derniers, il s'agit d'un langage orienté texte et formé de balises qui permettent d'organiser les données de manière structurée.

NS Les espaces de nommage XML offrent une méthode simple pour qualifier les noms des éléments et des attributs utilisés dans des documents XML, en associant ceux-ci avec des espaces de nommage désignés par des références d'URI.

XML Schéma est un langage de définition de la structure et du typage des documents XML.

5.3. La couche RDF

RDF (Resource Description Framework) .En 1999, le W3C présente le RDF C'est un modèle simple qui permet de décrire de la donnée (metamodel data model). Le RDF permet de relier des entités en les présentant sous la forme de triplets (Sujet - Prédicat - Objet). Le résultat peut être représenté sous la forme d'un graphe orienté dans lequel les nœuds sont des entités et les arcs des relations. La Figure 2 présente un graphe généré à partir de la définition de deux triplets RDF ("Ahmed", "est_file_de", "mohamed"). [8]

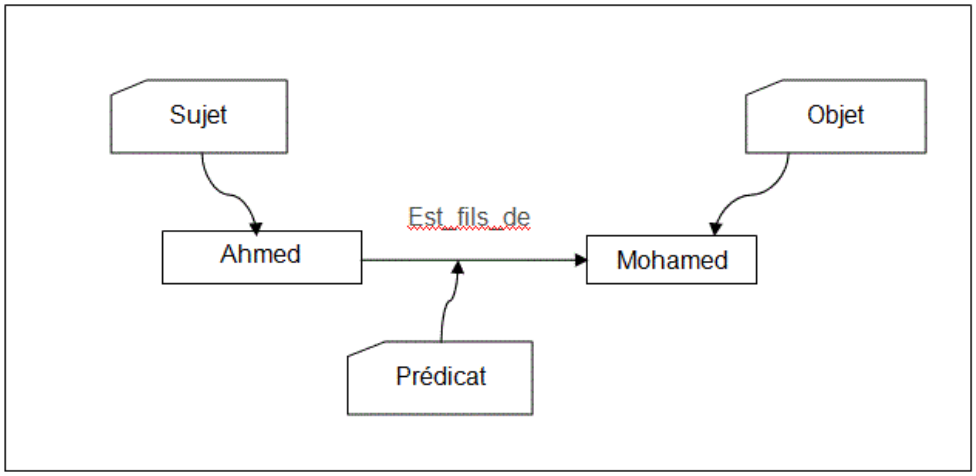


Figure 2: Exemple d'un model RDF

5.3.1. Type de RDF

❖ *RDF/XML*

La syntaxe RDF/XML, est basée sur la syntaxe XML. Elle est recommandée pour le RDF

Exemple :

```
<rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:cd="http://www.recshop.fake/cd#">
<rdf:Description
        rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
         <cd:artist>Bob Dylan</cd:artist>
         <cd:country>USA</cd:country>
        <cd:company>Columbia</cd:company>
        <cd:price>10.90</cd:price>
        <cd:year>1985</cd:year>
</rdf:Description>
</rdf:RDF>
```

❖ *N3*

La syntaxe N3 est une notation de RDF proposée par Tim Berners-Lee. Elle est plus lisible et facile pour l'humain. Sa syntaxe permet de repérer facilement les triplets et est donc plus intuitive pour imaginer le graphe correspondant.

Exemple :

```
@prefix ns0: <http://www.recshop.fake/cd#>.

<http://www.recshop.fake/cd/Empire Burlesque>

  ns0:artist "Bob Dylan" ;

  ns0:country "USA" ;

  ns0:company "Columbia" ;

  ns0:price "10.90" ;

  ns0:year "1985" .
```

❖ *Turtle*

« C'est un sous-ensemble de la syntaxe Notation3. La spécification de Turtle est actuellement un brouillon créé par le *RDF Working Group* du W3C le 9 août 2011 » [6]

Exemple :

```
@prefix ns0: <http://www.recshop.fake/cd#> .

<http://www.recshop.fake/cd/Empire Burlesque>

  ns0:artist "Bob Dylan" ;

  ns0:country "USA" ;

  ns0:company "Columbia" ;

  ns0:price "10.90" ;

  ns0:year "1985" .
```

❖ *N-Triplet*

C'est la syntaxe la plus élémentaire, structurée par une ponctuation et où chaque triplet est entièrement développé.

Exemple

```
<http://www.recshop.fake/cd/Empire Burlesque> <http://www.recshop.fake/cd#artist>
"Bob Dylan" .

<http://www.recshop.fake/cd/Empire Burlesque> <http://www.recshop.fake/cd#country>
"USA" .

<http://www.recshop.fake/cd/Empire Burlesque> <http://www.recshop.fake/cd#company>
"Columbia" .

<http://www.recshop.fake/cd/Empire Burlesque> <http://www.recshop.fake/cd#price> "
10.90" .

<http://www.recshop.fake/cd/Empire Burlesque> <http://www.recshop.fake/cd#year> "1
985" .
```

RDFS (RDF Schéma) : 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. [Fran, 2006]

5.4. La couche Ontologie

Une ontologie en informatique est un ensemble structuré de concepts permettant de donner un sens aux informations. son objectif premier est de modéliser un ensemble de connaissances dans un domaine donnée ; de plus les ontologies informatiques sont des outils qui permettent précisément de représenter un corpus de connaissances sous une forme utilisable par un ordinateur. [Inst,2013]

5.5. Les couches restantes

Logique, Preuve et Confiance

Une ontologie définit les concepts d'un domaine particulier, mais n'explique pas comment il faut les utiliser. Par exemple, considérons un agent électronique qui doit être capable de déduire que si une commande est de 10 objets, que 4 ont été commandés à un premier fournisseur et que 6 ont été commandés à un second fournisseur, alors la commande est passée. Ce type de connaissance ne fait pas partie de l'ontologie mais concerne les actions, processus et déductions logiques sur les objets de l'ontologie. Le raisonnement sur la connaissance est donc un point important du Web Sémantique.

Un des inconvénients du raisonnement automatique est que l'utilisateur prendra pour acquis des conclusions qui pourraient provenir de faits erronés. Il est nécessaire de savoir si l'on peut faire confiance aux informations disponibles. La solution proposée est d'offrir des outils permettant d'évaluer et de commenter les informations disponibles sur le Web, puis grâce aux signatures électroniques, il sera possible d'identifier formellement la personne qui aura annoté le document, et par conséquent de s'assurer de la pertinence des commentaires.

Chapitre III :

Les Ontologies

1. *Introduction*

Le terme « ontologie », est un mot composé de deux partie ontos qui veut dire existe et *logos* c'est-à-dire l'étude. L'informatique a emprunté à la philosophie au début des années 1990.

Une ontologie en informatique est un ensemble structuré de concepts permettant de donner un sens aux informations. Les relations qui peuvent être dans le graphique sont:

- Des relations sémantiques;
- Des relations de composition et d'héritage (au sens objet).

2. *Définition*

Les premiers qui ont défini l'ontologie sont *Neeches et ses collègues* [Nee, 1991]:
« 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 »

• En 1993 [Gru, 1993] propose la définition suivante :
« Spécification explicite d'une conceptualisation »

D'après GRUBER, cinq critères permettent de mettre en évidence des aspects importants d'une ontologie :

1. la clarté;
2. la cohérence;
3. l'extensibilité;
4. une déformation d'encodage minimale;
5. un engagement ontologique minimal.

3. *Composantes d'une ontologie*

Les ontologies sont très étendu l'utilisation dont l' intelligence artificielle, web sémantique, génie logiciel, médicaux et bio-informatique et l'architecture de l'information comme une forme de représentation des connaissances dans le monde entier ou une partie de ce monde. [Inst, 2013]

Les ontologies décrivent généralement :

- ✓ Individus : les objets de base.
- ✓ Classes : ensembles, collections, ou types d'objets.
- ✓ Attributs : propriétés, fonctionnalités, caractéristiques ou paramètres que les objets peuvent posséder et partager.
- ✓ relations : les liens que les objets peuvent avoir entre eux.
- ✓ Évènement : changements subis par des attributs ou des relations.

4. *Classification des ontologies*

La classification des ontologies peuvent être fonctionné par plusieurs dimensions. Parmi celle-ci, nous examinons deux, à savoir :

- ✓ L'objet de conceptualisation ;
- ✓ Le niveau de formalisation de la représentation des connaissances.

4.1 Typologie selon l'objet de conceptualisation

Gomez et ses collègues [Gom, 2004] proposent une classification selon le sujet de conceptualisation des ontologies come montre la figure suivante:

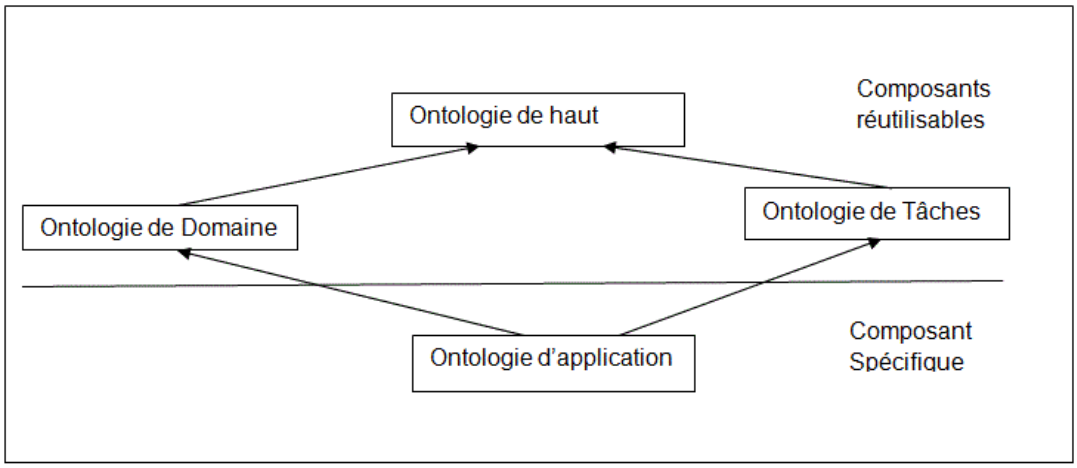


Figure 3: Classification des ontologies Selon l'objet de conceptualisation [Gom, 2004].

4.2 Typologie selon le niveau de formalisation utilisé

Selon le degré de formalise les ontologies peuvent être distinguées. Uschold et Grüninger [Usc, 1996] suggèrent une classification contenant les quatre catégories : Informelles ; Semi-informelles ; Semi-formelles et formelles.

- Informelle : l'ontologie est exprimée en langage naturelle (sémantique ouverte).
- Semi-informelle **:** l'ontologie est exprimée dans une forme restreinte et structurée de langage naturel.
- Semi-formelle: l'ontologie est exprimée dans un langage artificiel défini formellement.
- formelle : l'ontologie est exprimée dans un langage contenant une sémantique formelle, des théorèmes, et des preuves pour vérifier les propriétés telles que la validité et la complétude.

5 Langage

5.1 OWl

Le vocabulaire du langage OWL est un ensemble de primitives, décrites dans le langage RDF, qui étendent le vocabulaire du langage RDFS. Ces primitives sont identifiées par des URI construits chacun par concaténation d'un nom local et de l'espace de nommage http ://www.w3.org/2002/07/owl#.[Ben, 2011]

La définition d'une ontologie dans le langage OWL est constituée de définitions de classe, de définitions de propriété, une méta-description de l'ontologie et éventuellement des descriptions d'individus.

Le langage OWL offre trois sous langages d'expression conçus pour des communautés de développeurs et d'utilisateurs spécifiques.

- **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 certaines 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.

Il y a une stricte compatibilité ascendante de ces trois langages: [Deb, 2017]

- Toute ontologie OWL Lite valide est une ontologie OWL DL valide ;
- Toute ontologie OWL DL valide est une ontologie OWL Full valide ;
- Toute conclusion d'une ontologie OWL Lite est une conclusion valide de OWL DL ;
- Toute conclusion OWL DL est une conclusion valide de OWL Full

5.2 SPARQL

Le protocole SPARQL « *SPARQL Protocol and RDF Query Language* » est aussi un langage de requête qui permet de rechercher, d'ajouter, de modifier ou de supprimer des données RDF disponibles à travers Internet.

SPARQL est l'équivalent de SQL , comme en SQL on accède aux données d'une base de données *via* ce langage de requête alors qu'avec SPARQL, on accède aux données du Web des données. Il est doté :

- d'un langage de requêtes avec syntaxe basée sur des triplets
- d'un protocole d'accès comme un service Web (SOAP)
- d'un langage de présentation des résultats (XML)

Cela signifie qu'en théorie, on pourrait accéder à toutes les données du Web avec ce standard. Il cible donc l'interrogation de métadonnées RDF, structure de base du Web sémantique.

Il fonctionne en parfaite synergie avec les autres technologies Web sémantique du W3C :

- RDF pour la représentation des données ;
- RDFS ;
- OWL pour la création de vocabulaires.

La structure d'une requête SPARQL est très similaire à celle employée dans le langage SQL :

SELECT ?v1 ?v2 … ?vn

FROM WHERE {

(sujet1 | vi) (predicat1 | vj) (objet1 | vk).

. …

(sujetx | va) (predicaty | vb) (objetz | vc).

}

Exemple

Requête :

```
PREFIX dc:  <http://purl.org/dc/elements/1.1/> .
PREFIX :    <http://example.org/book/> .
PREFIX ns:  <http://example.org/ns#> .

SELECT ?book ?title

WHERE
 { ?book dc:title ?title .
   ?book ns:price ?price . FILTER ( ?price < 40 )
 }
```

Résultat :

book	title
:book2	“The Semantic Web”

6 *Editeurs d'ontologie*

Aujourd'hui, le nombre d'éditeurs d'ontologies a considérablement augmenté. Il existe différents éditeurs qui utilisent des formalismes variés et offrent différentes fonctionnalités. Parmi ces outils on trouve : OILed, OntoEdit, , Web ode, DOE, Protégé2000…. Voici la description de quelques-uns :

6.1 *OntoEdit*

OntoEdit est environnement de construction d'ontologies indépendant de tout formalisme. il contient des outils graphiques dédiés à la visualisation d'ontologies. OntoEdit intègre un serveur destiné à l'édition d'une ontologie par plusieurs utilisateurs. Un contrôle de la cohérence de l'ontologie est assuré à travers la gestion des ordres d'édition. [Sur,2002].

6.2 *Ontolingua*

Ontolingua de l'Université Stanford. Le serveur Ontolingua est le plus connu des environnements de construction d'ontologies en langage Ontolingua. Il consiste en un ensemble d'environnements et de services qui supportent la construction en coopération d'ontologies, entre des groupes séparés géographiquement. Il supporte plusieurs langages et dispose de traducteurs permettant de passer de l'un à l'autre. [Far,1996]

6.3 *Protégé2000*

Protégé2000 est une interface modulaire, développée au Stanford Medical Informatics de l'Université de Stanford7, permettant l'édition, la visualisation, le contrôle (vérification des contraintes) d'ontologies, l'extraction d'ontologies à partir de sources textuelles, et la fusion semi-automatique d'ontologies. Le modèle de connaissances de Protégé2000 est issu du modèle des frames et contient des classes (concepts), des slots (propriétés) et des facettes (valeurs des propriétés et contraintes), ainsi que des instances des classes et des propriétés. De nombreux plug-ins sont disponibles ou peuvent être ajoutés par l'utilisateur. [Noy,2001]

6.4 INA

L'éditeur développé à l'INA se compose d'au moins 2 fenêtres correspondant à une taxonomie de concepts et à une taxonomie de relations. Sur la partie gauche, 2 navigateurs permettent de consulter l'arborescence ontologique. Le premier affiche pour un concept, les listes de ses pères (jusqu'à la racine), de ses frères et de ses fils. Le second explorateur offre une visualisation plus classique. Dans la partie droite, l'utilisateur peut dénier les principes de différentiation propres à l'ontologie différentielle, mais aussi les aspects terminologiques

(langue, synonyme, terme préféré) ou les relations sémantiques associées (relation instance). Le multilinguisme génère autant de nouvelles fenêtres que de langues, toutes synchronisées.

L'application a été développée en JAVA et permet d'exporter les ontologies crées en XML, en HTML ou encore en LATEX. L'import des ontologies se fait, lui, en XML validé par une DTD.

Chapitre IV :

Annotations

1. Introduction

Pour donner une description plus précise à un contenu on lui associe des annotations. Les applications informatiques existantes qui ont proposé de manipuler des annotations se sont inspirées de la pratique sur le support papier. [Bri,2003]. Alors les annotations numériques sont juste une adaptation de cette notion aux documents électroniques.

I. Annotation classique

1. Définition

Le Petit Robert définit le terme annotation comme une « note critique ou explicative qui accompagne un texte – une note de lecture qu'on inscrit sur un livre .L'annotation est « l'action d'annoter, au résultat de cette action ».

- Dans [Des, 2002], l'auteur définit comme : « une annotation est une information graphique ou textuelle attachée à un document et le plus souvent placée dans ce document »

- Dans le contexte des recherches sur les IHM, [Bal. 2000] énonce que «une annotation est un commentaire sur un objet tel que le commentateur veut qu'il soit perceptiblement distinguable de l'objet lui-même et le lecteur l'interprète comme perceptiblement distinguable de l'objet lui-même ».

- de point de vue des psycholinguistes et cogniticiens tels que [Ver, 1997] cité par [Azo, 2006], une annotation est une « trace de l'état mental du lecteur et une trace de ses réactions vis-à-vis du document. ».

- Dans [Des. 2002] l'annotation est : « tout objet qu'une personne ajoute à un document avec un objectif spécifique. ».

2. Méta-donnée ou Annotation

La méta-donnée « est une donnée sur une donnée ». Son utilité est d'indexer les ressources du web, elle est différente de l'annotation et peut former une annotation particulière, qui peut être une méta-donnée portant sur [Pri, 2003]:

- ✓ le document annoté (nom auteur, date production…).
- ✓ l'annotation (nom de l'annotateur, date de production de l'annotation…).

Une méta-donnée a une pertinence à posteriori. Elle sera utile après la saisie ou au moment de la recherche» [Brin, 2003].

La figure illustre la relation entre une méta-donnée et une annotation

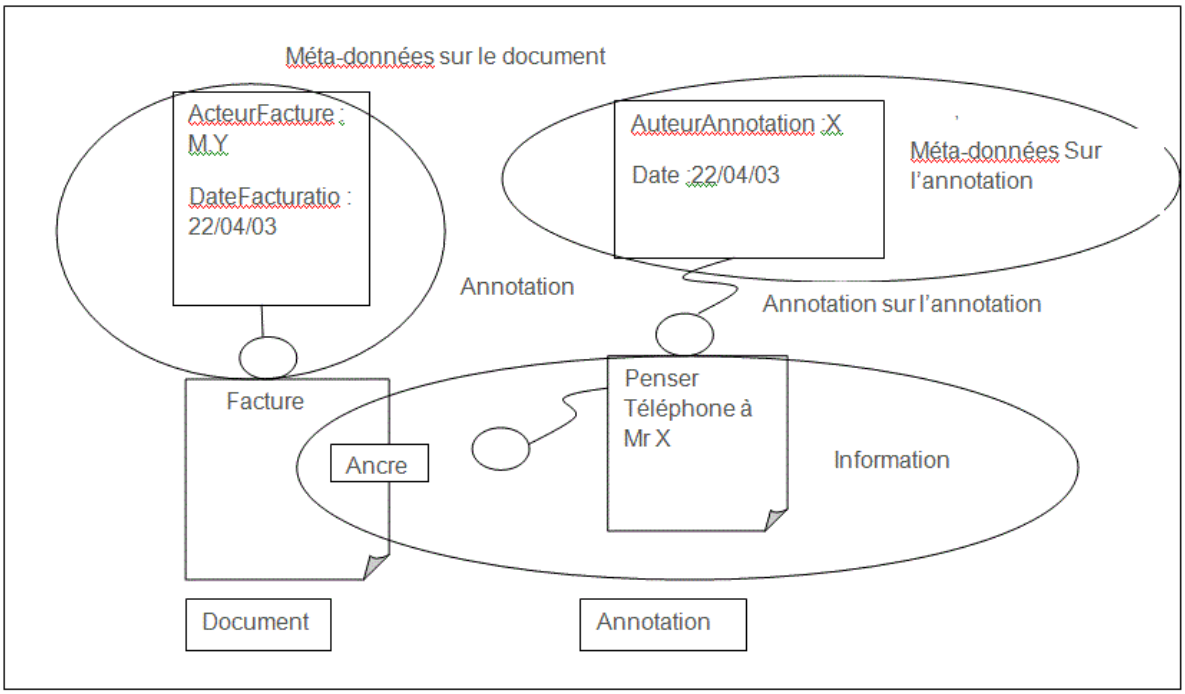


Figure 4: Relation entre l'annotation et la Méta-donnée [Brin, 2003].

3. Cycle de vie des annotations classiques

Ce cycle de vie intègre trois étapes [Brin, 2003]:

• La production de l'annotation (annotateur) : consiste à choisir la cible, y attacher une ancre, puis inscrire des traces sur le support de l'annotation pour produire une information.

• Le stockage et l'indexation de l'annotation (gestionnaire) : consiste à classer les annotations (base d'annotations) et associer à cette classification des informations avec la possibilité d'identifier, de localiser et d'accéder à ces annotations.

• La consultation d'une annotation (consultant) : consiste à localiser une cible sur laquelle porte l'annotation et associer une signification à son information.

4. Acteurs interagissant avec l'annotation classique

Les principaux acteurs qui peuvent traiter un document, à différentes étapes de son cycle de vie [Brin, 2003] :

- ✓ L'annotateur : auteur de l'annotation et peut être l'auteur du document annoté.
- ✓ Le consultant : celui qui perçoit l'annotation et qui peut être l'annotateur. Il peut être visé (connu par l'annotateur au moment de la rédaction de l'annotation).
- ✓ Le gestionnaire des annotations : en cas de documents numériques, c'est celui qui gère les annotations et les organise, pour pouvoir les retrouver ultérieurement avec les

objets annotés. Il peut être l'annotateur, le consultant ou les deux, comme il peut être humain ou un agent logiciel.

5. Quelques outils pour l'annotation classique

A cause de l'importance d'utilisation des annotations dans les documents numériques les chercheurs se focalisent sur le développement des outils d'annotation performant et facile à manipuler. Dans le but de développement des outils d'annotations des documents on peut trouver de nombreux logiciels commerciaux et des prototypes de recherche.

Selon le contexte de création de ces outils et des matériels sur lesquels ces applications fonctionnent, Les fonctionnalités et les interfaces utilisateur sont différentes et dépendent selon Bringay [Brin, 2003]. Il existe cinq catégories d'outils d'annotation :

- Outils destinés à l'annotation de page Web : Imarkup, Yawas,
- Outils destinées à l'annotation de documents pouvant être utilisés pour des lectures personnelles ou pour la construction de documents : Adobe Acrobat et Microsoft Office (Word),
- Outils destinés à l'annotation de contenus multimédias : Debora permet à des experts d'annoter des livres scannés,
- Outils destinés aux dispositifs mobiles, comme les PDA, les e-books, les tablettes PC : Adobe eBook Reader, Xlibris,
- Outils destinés à organiser un ensemble de ressources disponibles en ligne par la création de signets (marque-pages pouvant être vues comme des annotations) (exemple Nestor).

II. Annotation sémantique

1. Définition

L'annotation sémantique a pour but de donner une description formelle pour les documents électroniques sous la forme de méta-données attachées aux textes. Cette description s'exprime couramment en termes ontologiques quand il s'agit d'associer un type sémantique aux noms des entités mentionnées dans le ou de les associer à un concept ([Kir,2004]; [Ama,2005]). Ce type de méta-données fournit, à propos des entités du document annoté, des informations de classes et d'instances liées à une ontologie décrivant le contenu de la ressource annotée, ce qui est illustré dans la figure.

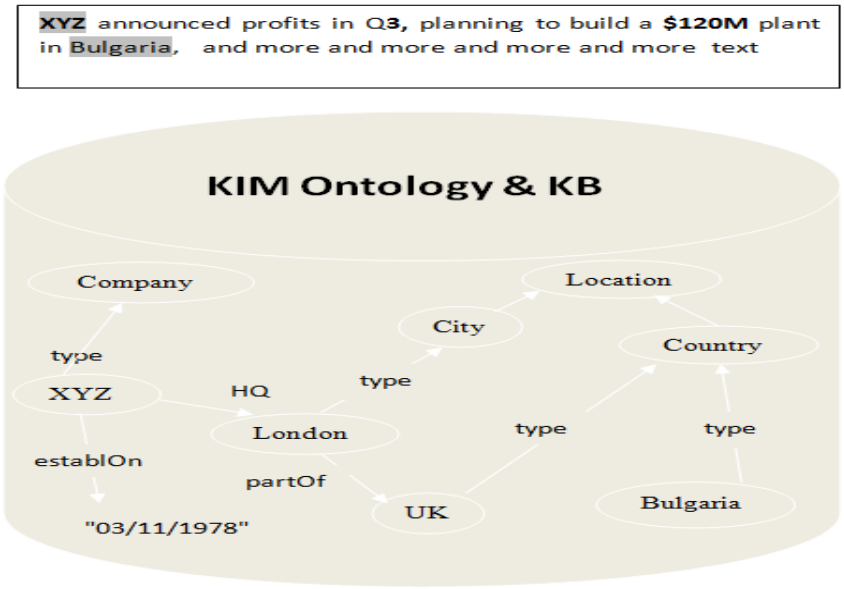


Figure 5: Annotation sémantique [Kir, 2017]

2. Annotation et méta-donnée sur le web sémantique

La communauté anglophone du web sémantique considère que les annotations de pages web deviennent des méta-données dès qu'elles sont stockées dans une base sur un serveur. D'un point de vue plus lié à la pratique des annotations et méta-données, une métadonnée sera plutôt attachée à une ressource identifiée en tant que telle sur le web et aura une pertinence à priori. Elle est saisie suivant un schéma, ce qui permet de mettre en place des inférences. Tandis qu'une annotation est plus située au sein de cette ressource et écrite au cours d'un processus d'annotation (elle peut être complètement libre). Evidemment, cette distinction n'a rien de définitive, il s'agit simplement de mettre l'accent sur le caractère plus situé au sein de la ressource de l'annotation, par rapport à une méta-donnée plus indépendante. Des pages Web sont annotées à partir de connaissances disponibles dans une ou plusieurs ontologies (qui ont pour objectif de normaliser la sémantique des annotations), et ces annotations, regroupées en entrepôts de méta-données deviennent utiles pour des agents de recherche d'information, faisant ou non appel à des moteurs d'inférence permettant de déduire de nouvelles connaissances formelles des annotations. [Ser, 2017]

3. Caractéristiques des annotations sémantiques

3.1. Processus des annotations sémantiques

Le processus général d'annotation sémantique de documents par des ontologies contient plusieurs étapes qui font référence à des annotations de natures un peu différentes. Ces étapes sont au nombre de trois :

- Repérer: peut-être un processus manuel [Han, 2002] ou automatique [Des, 2002] qui consiste à placer dans le document des références aux concepts de l'ontologie qu'il contient. Ces éléments sont considérés comme des méta-données.
- Instancier : processus manuel ou automatique permettant de valuer les attributs des concepts à l'aide des informations présentes dans le document (là encore, ce sont des méta-données).
- Enrichir : processus manuel visant à ajouter des informations par l'intermédiaire des attributs de concepts qui n'ont pas pu être valués à la phase précédente.

Les deux premières étapes sont des étapes d'insertion de méta-données car il n'y a pas ajout d'information mais plutôt localisation et caractérisation de l'information déjà présente.
Par contre, la dernière est plutôt une étape d'annotation plus classique, car il y a ajout d'information, le document est enrichi d'information qui n'est pas explicitement présente dans le document. Cette annotation est directement "formalisée" par des métadonnées.

3.2. Exploitation des annotations sémantiques

Les annotations sémantiques sont en effet exploitables pour améliorer la gestion des documents annotés. Elles peuvent être utilisées dans le cadre de la recherche et la classification de documents, le résumé automatique et généralement, pour favoriser l'interopérabilité. L'annotation sémantique peut être applicable pour toute sorte de texte : page web, documents réguliers, champs texte dans des bases de données, etc.
Les deux tâches pour lesquelles les annotations sémantiques sont utilisées sont la recherche d'informations et la composition de services et de documents. La construction de résumés automatiques de documents est une autre application qui prend en considération les parties annotés, Denoue [Den, 2002] a proposé une méthode de classification de documents basée sur les termes annotés.

3.3. Méthodes et éléments de base pour l'annotation sémantique

On définit les standards existants, et qui forment l'ossature technologique du web, pouvant servir dans le cadre de l'annotation sémantique et qui sont:

• Le protocole HTTP : pour la transmission.

• Les langages HTML/XML et les feuilles de style : pour la présentation des résultats.

• Le langage RDF (Ressource Description Framework) : pour l'expression et l'échange des méta-données, sous la forme de triplets, même si ceux-ci peuvent également être stockés dans des bases de données pour une gestion plus efficace. Le stockage des annotations au début était dans les documents eux-mêmes (balises ad hoc ou RDF), dans les URL (Uniform Ressource Locator) ou dans des bases sur des serveurs centralisés et les annotations étaient typées, plus ou moins partagées et souvent textuelles. Actuellement, il existe un ensemble d'éléments qui constituent, tous à la fois, un référentiel aussi bien technique que méthodologique sur lequel l'annotation sémantique de documents web peut s'appuyer .Il s'agit de :

• Méthodes de conception de schémas de méta-donnée (thésaurus, ontologies).

• Outils et principes d'utilisation de méta-données, d'annotation et de présentation et de présentation de celles-ci.

• Architectures de stockage, de requêtes et de diffusion d'annotation.

• Méthodes, outils et techniques d'extraction d'information des documents (fouille de données textuelles, méthodes statistiques, remplissage automatique de champs de méta-données, propositions aux utilisateurs, etc.). [Pri,2003]

4. Les outils d'annotation sémantique

Toute logiciel a pour objectif de gérer et d'insérer des annotations sémantiques liées au moins à une ressource documentaires données est considéré comme un outil d'annotation sémantique.

Il existe plusieurs d'outils d'annotations sémantiques leurs tâches sont la génération de méta-données pour les ressources Web.

Les outils d'annotations classiques existant permettent à l'utilisateur d'annoter seulement avec du texte. Les outils d'annotations sémantiques utilisent une ontologie qui formalise et structure les annotations produites en fonction des concepts et des contraintes définies dans cette ontologie. L'évolution de ces outils d'annotations suit les recommandations formulées par le W3C aux niveaux des langages utilisés pour l'annotation. Les premiers outils comme

SHOEKA, OntoAnnotate ou MnM utilisent des langages non standardisés, respectivement SHOE, HTML-A et OCML.

Partie II- Contribution :

Sémantification

Du Dossier Electronique Médical

1. Introduction

Le présent travail de master rentre dans le cadre du Web sémantique et son application au domaine médical. Il consiste à faire une sémantification du dossier électronique du patient (XML) à base d'ontologies (OWL). Quatre ontologies (OWL) ont été choisies : (1) une ontologie médicale schématisant les concepts clés de la pratique médicale; (2) une ontologie familiale illustrant les relations familiales ; (3) une ontologie documentaire représentant l'organisation logique d'un dossier électronique du patient ; (4) une ontologie de vie privée montrant les niveaux de confidentialité et les droits d'accès. Les deux premières ontologies ont été adoptées à partir de l'existant pendant que les deux dernières ont été créées par nous-mêmes. L'édition et l'utilisation de ces ontologies est fait grâce à l'éditeur d'ontologies PROTEGE.

La sémantification consiste à créer des annotations sémantiques (RDF) de parties du dossier électronique du patient par des concepts émanant de l'une des quatre ontologies précitées. La sémantification peut se faire selon deux approches : semi-automatique et automatique. Dans la première, les annotations sont créées par un expert humain à travers un outil logiciel (annotateur sémantique). Dans la seconde, les annotations sont générées automatiquement à partir du dossier électronique du patient à travers une feuille de style XSLT (générateur d'annotations sémantiques). Les annotations sémantiques résultantes sont ensuite interrogées grâce à SPARQL.

Nous commençons par faire le design du DEP comme un document XML en considérant son contenu comme un ensemble de données semi structurées. Ensuite nous passons en revue l'architecture détaillée de notre système.



Figure 6: Interface du DEP

2. Composition du DEP

Le dossier médical est un élément très important du domaine des soins de santé. Il permet aux professionnels de la santé de gérer et d'évaluer les soins à fournir à leurs patients. Le dossier médical constitue également un outil de communication efficace qui permet aux médecins de suivre avec précision la situation de leurs patients, peu importe le lieu où ils se trouvent.

Ainsi, un dossier médical est ouvert pour chaque patient dès sa première consultation dans un nouvel établissement de santé ou une nouvelle clinique privée. On y retrouve ses renseignements personnels ainsi que les informations qui concernent les soins et les services qui lui ont été dispensés jusqu'à ce moment.

a. Renseignements personnels

Tous les renseignements qui permettent d'identifier le patient sont considérés comme des renseignements personnels.

Dans le but de déterminer l'identité du patient, le médecin et les professionnels de la santé doivent donc inscrire plusieurs renseignements dans le dossier médical, notamment : (***nom; sexe; date de naissance; adresse; ses antécédent familiaux; numéro d'assurance).***

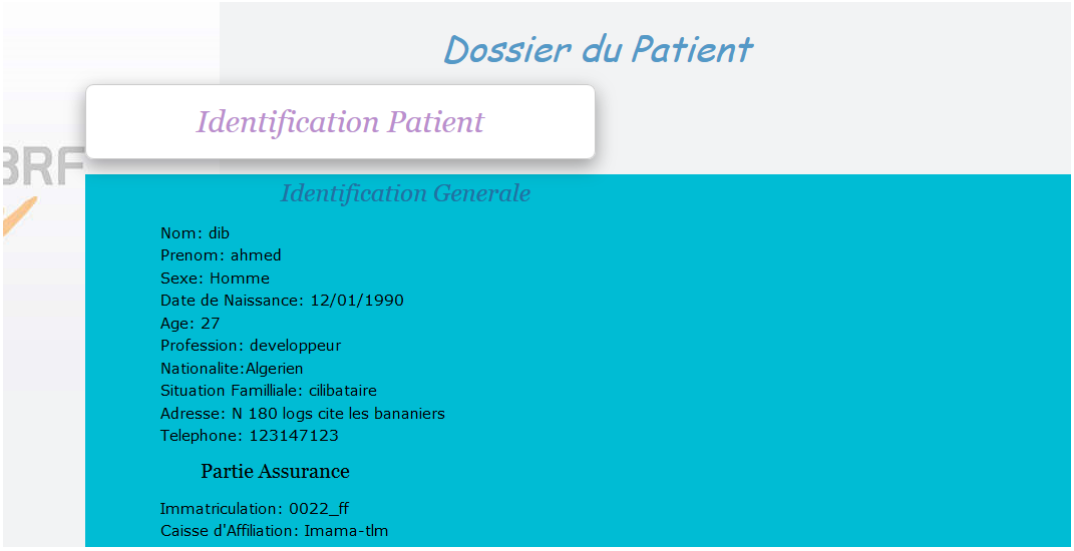


Figure 7: Partie renseignements personnels

b. Renseignements médicaux

Le médecin et les professionnels de la santé indiquent également au dossier médical du patient plusieurs renseignements médicaux qui le concernent, notamment :

La date de chaque consultation; toute information quant à un risque d'allergie; leurs observations médicales; leurs diagnostics; toute information pouvant influencer les soins à fournir.

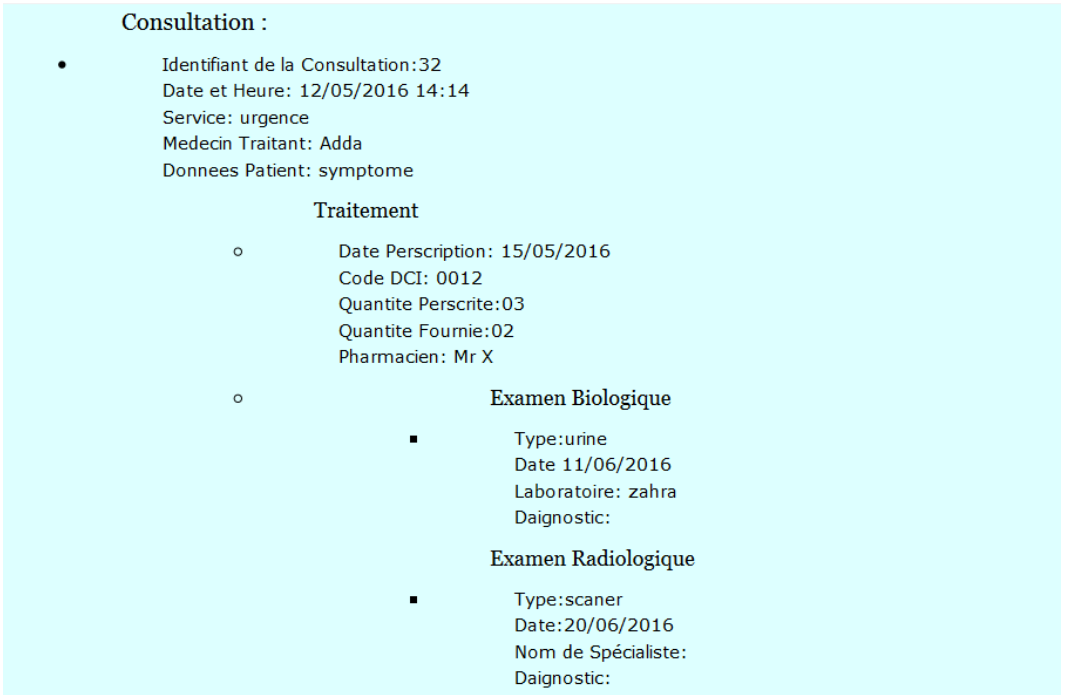


Figure 8: Extrait du renseignement médical

3. *Architecture du système*

L'architecture du système est illustrée dans la figure 21. Elle est constituée des composants suivants :

a. Module 1 : « Editeur d'ontologie »

Nous avons utilisé comme éditeur d'ontologie le logiciel « Protégé », qui permet la mise à jour des ontologies par un expert en ontologies. Dans le cadre de ce travail nous avons utilisé les Ontologies suivantes :

- ***Ontologie : « Structure du DEP »***

Cette ontologie illustre la structure fondamentale du fichier XML. Elle se compose d'une partie identification de patient et d'un autre de consultations faites par le patient.

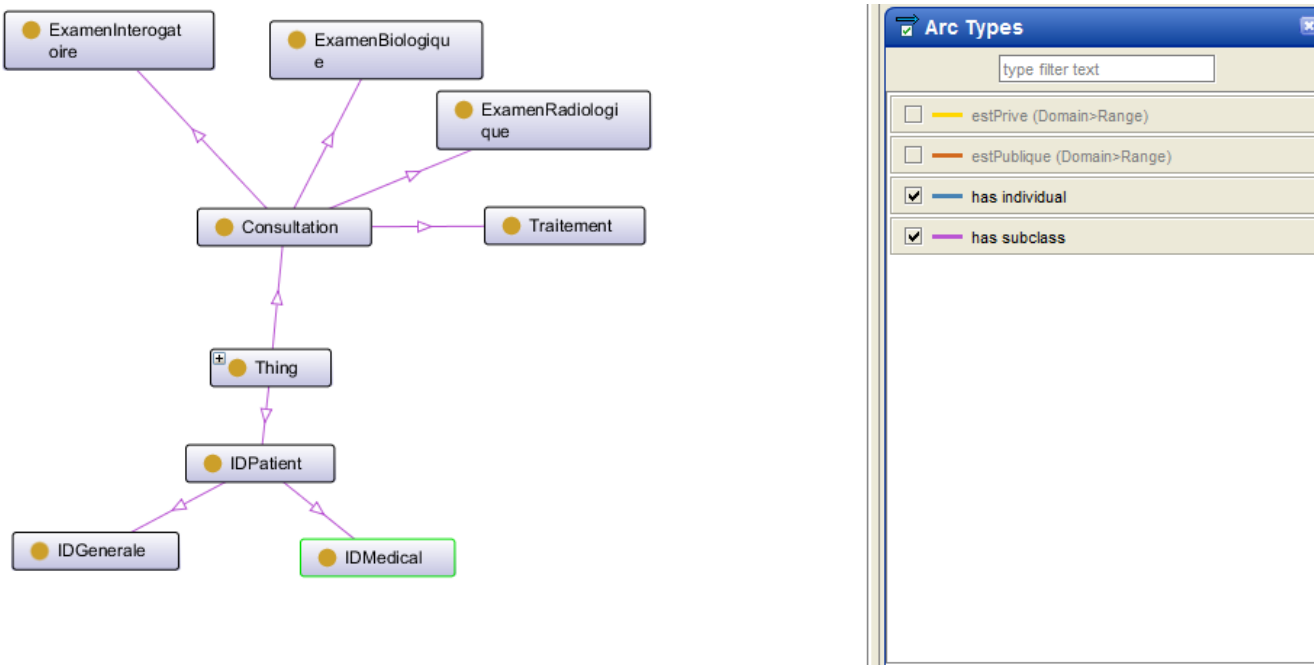


Figure 9: Ontologie: Structure du DEP

➢ *Ontologie : « Vie Privé »*

Dans cette ontologie nous décrivons les différents acteurs (acteur médical, patient) qui peuvent agir sur le DEP avec l'ensemble des droits d'accès possibles. Le patient est le seul responsable sur l'attribution des droit à l'ensemble des acteurs médicaux (médecin, pharmacien, infirmier ou bien un administrateur). Les droits sont compris entre privé et public c'est-à-dire il y'a des parties dans le DEP qui peuvent être visibles pour tout le monde et d'autres non.

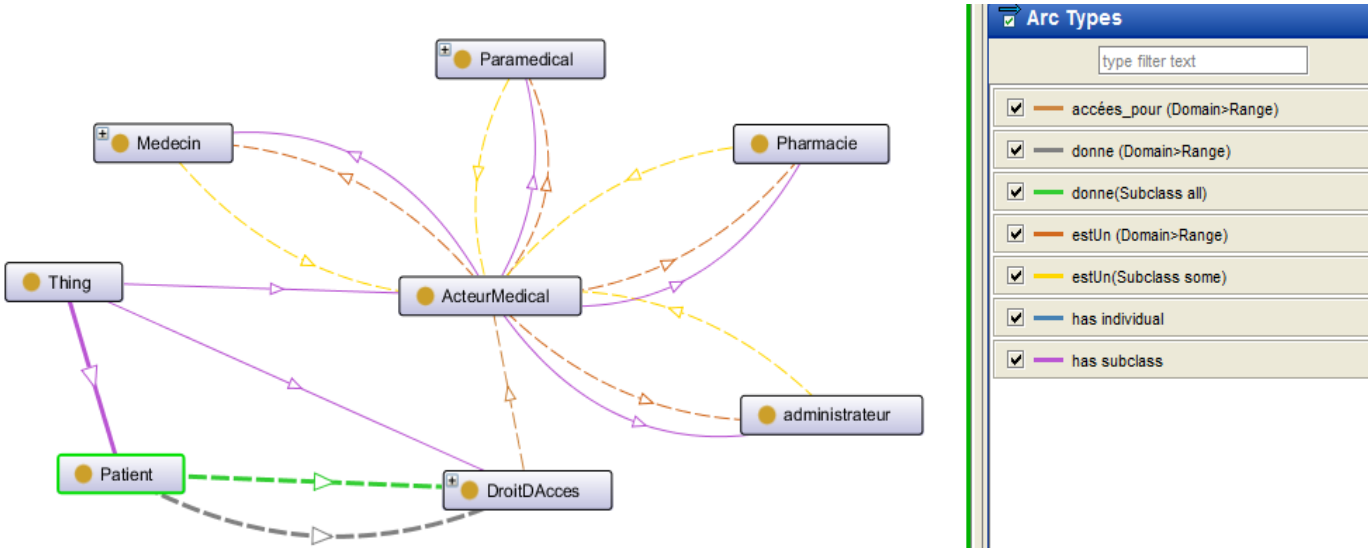


Figure 10: Ontologie de vie privée

➢ *Ontologie : Relations familiales*

Nous avons adopté cette ontologie qui décrit les relations entre membres de la famille.

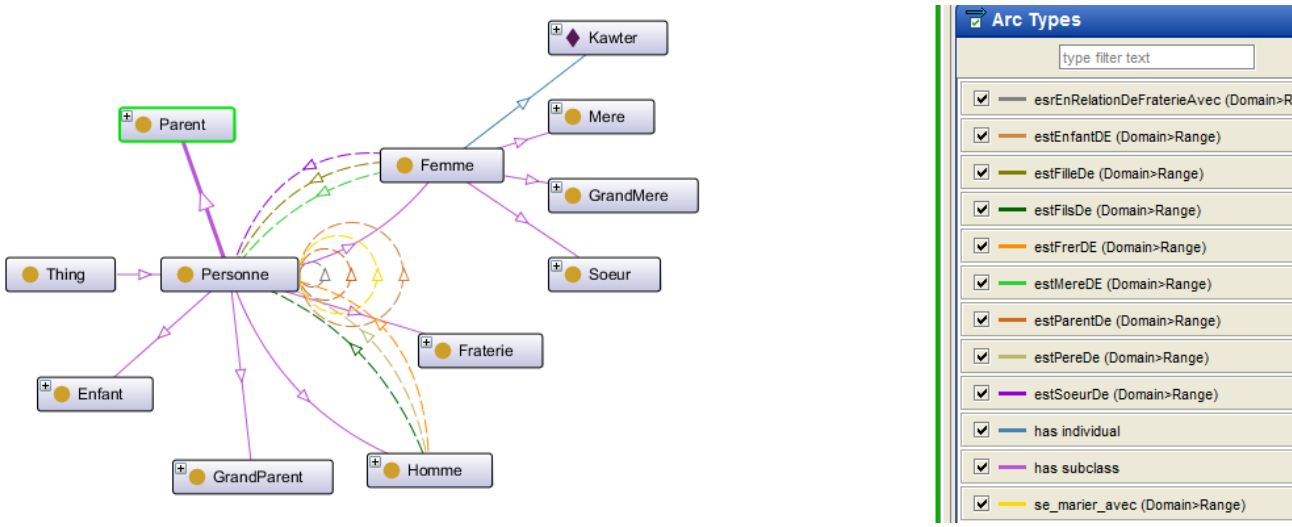


Figure 11: Ontologie de la Famille

➢ *Ontologies de maladies*

Nous avons adopté les ontologies médicales:

DMTO :

L'ontologie DMTO est une ontologie OWL 2 pour créer des plans de traitement personnalisés pour les patients diabétiques. L'ontologie est basée sur BFO et OGMS. En outre, il étend l'ontologie DDO en ajoutant les classes de traitement et les axiomes à la partie de diagnostic existante.

OGMS :

L'ontologie pour les sciences médicales générales (OGMS) est basée sur les documents portant sur un traitement ontologique des maladies et des diagnostics et sur les carcinomes et d'autres entités pathologiques. L'ontologie tente d'aborder certaines des questions soulevées lors de l'Atelier sur l'ontologie des maladies (Dallas, TX) et l'Atelier des signes, symptômes et constatations (Milan, Italie). L'OGMS était autrefois appelée ontologie du phénotype clinique. Les termes d'OGMS s'arrêtent à l'ontologie formelle de base.

BFO :

L'ontologie BFO se développe à partir d'une orientation philosophique qui chevauche avec celle de DOLCE et SUMO. Contrairement à ceux-ci, cependant, il est étroitement axé sur la tâche de fournir une véritable ontologie supérieure qui peut être utilisée à l'appui d'ontologies de domaine développées pour la recherche scientifique, comme par exemple en biomédecine dans le cadre de la Fonderie OBO. Ainsi, BFO ne contient pas de termes physiques, chimiques, biologiques ou autres qui relèvent correctement des domaines des sciences spéciales.

DDO :

Une ontologie pour le diagnostic du diabète contenant les complications liées au diabète, les symptômes, les médicaments, les tests de laboratoire, etc.

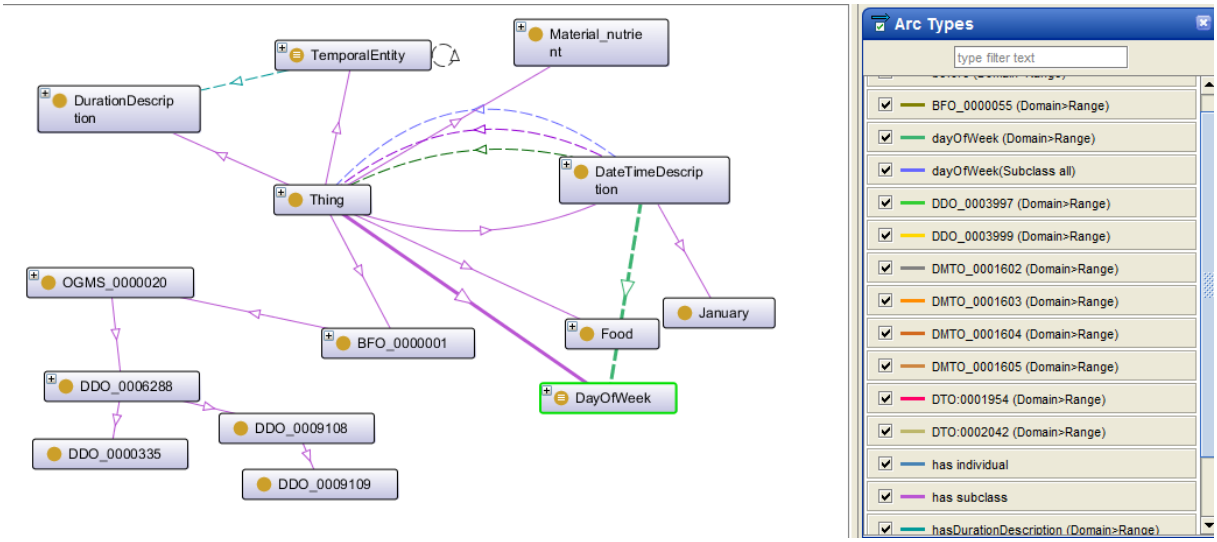


Figure 12: Ontologie Médicale "DMTO"

b. Module 2 : « Editeur DTD/XSD »

Nous avons choisi l'éditeur de schéma « Editix- Xmleditor». Il a comme input un fichier xml-schéma ou bien DTD et comme output ce fichier mis à jour. Il est utilisé par un expert en structure de DEP.

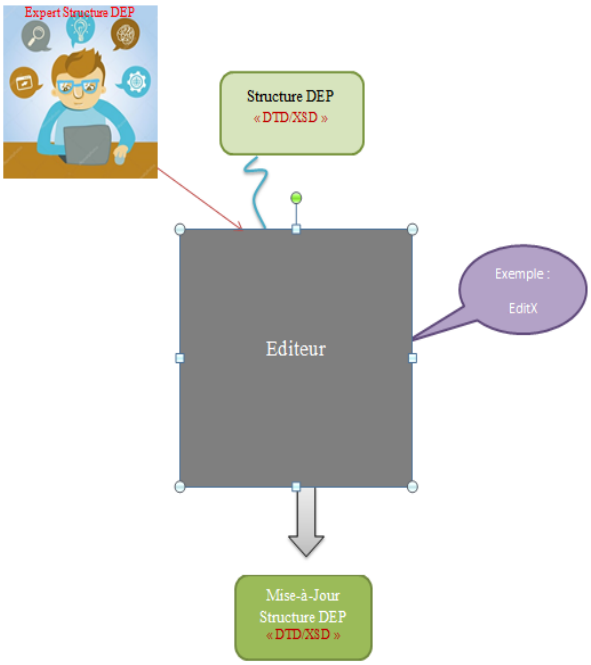


Figure 13: Module "Editeur DTD/XSD"

Un extrait du code DTD :

```
<!ELEMENT DPI (IdentificationPatient,Consultations,FACTURE)>

<!ELEMENT IdentificationPatient (IdentificationGeneral,IdentificationMedical)>
        <!ELEMENT IdentificationGeneral (nom,Prenom,sexe,DateNaissance,Age,Profession*,Nationalite,SituationFamilliale,Adresse,telephone*,IDAssure)>

                                <!ELEMENT nom (#PCDATA)>
                                <!ELEMENT Prenom (#PCDATA)>
                                <!ELEMENT sexe (#PCDATA)>
                                <!ELEMENT DateNaissance (#PCDATA)>
                                <!ELEMENT Age (#PCDATA) >
                                <!ELEMENT Profession (#PCDATA) >
                                <!ELEMENT Nationalite (#PCDATA)>
                                <!ELEMENT SituationFamilliale (#PCDATA)>
                                <!ELEMENT Adresse (#PCDATA)>
                                <!ELEMENT telephone (#PCDATA)>
                                <!ELEMENT IDAssure (Immatriculation,caisseDAffiliation)>
                                <!ELEMENT Immatriculation (#PCDATA)>
                                <!ELEMENT caisseDAffiliation (#PCDATA) >

        <!ELEMENT IdentificationMedical (Taille,Poids,GroupeSanguin,fil_le_s,Allergies*,maladies*)>

                                <!ELEMENT Taille (#PCDATA)>
                                <!ELEMENT Poids (#PCDATA)>
                                <!ELEMENT GroupeSanguin (#PCDATA)>
                                <!ELEMENT fil_le_s (pere,Mere)>
                                                        <!ELEMENT pere (#PCDATA)>
                                                        <!ELEMENT Mere EMPTY>
                                                        <!ATTLIST Mere   nom CDATA #REQUIRED
                                                                         prenom CDATA #REQUIRED>

                                <!ELEMENT Allergies (Allergie*)>
                                <!ELEMENT Allergie (#PCDATA)>
                                <!ELEMENT maladies (maladie*)>
                                                        <!ELEMENT maladie (#PCDATA)>

<!ELEMENT Consultations (Consultation*)>
```

Figure 14: Extrait du DTD du DEP

c. Module 3 : « Editeur DEP »

Ce module consiste à remplir les données de patient par un utilisateur tel que (l'agent de la réception de l'hôpital ou bien la secrétaire du médecin,…). Ce fichier est validé par rapport au schéma créé dans le module 2.

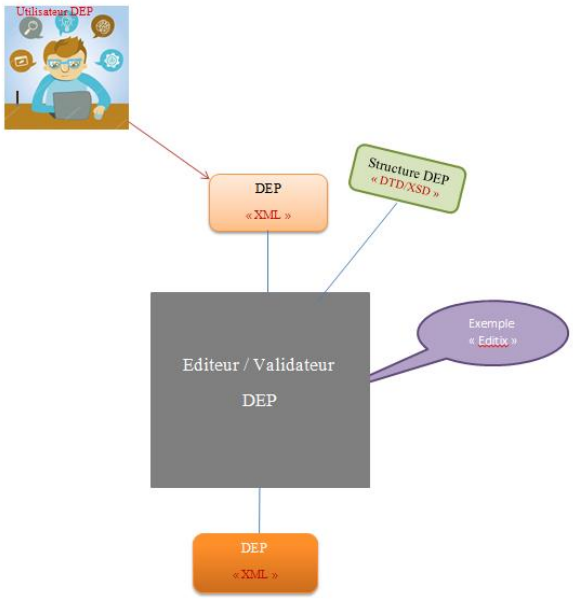


Figure 15: Module "Editeur DEP"

Un extrait du code XML :



Figure 16: Extrait du fichier XML DEP

La validation du code XML par DTD

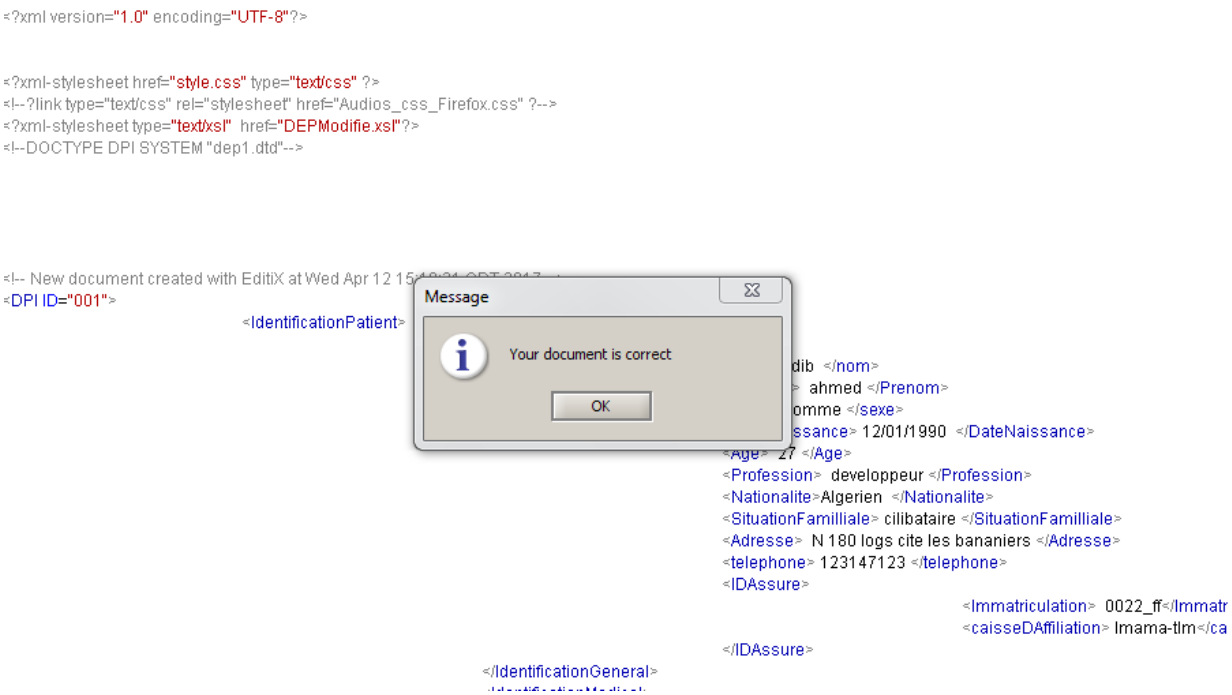


Figure 17: Validité du document xml

d. Module 4 : « Générateur d'Annotations Sémantiques »

Ce module adopte l'approche automatique, il génère des annotations sémantiques à base de l'ontologie sélectionnée, via la feuille de style XSLT correspondante que nous avons conçue.

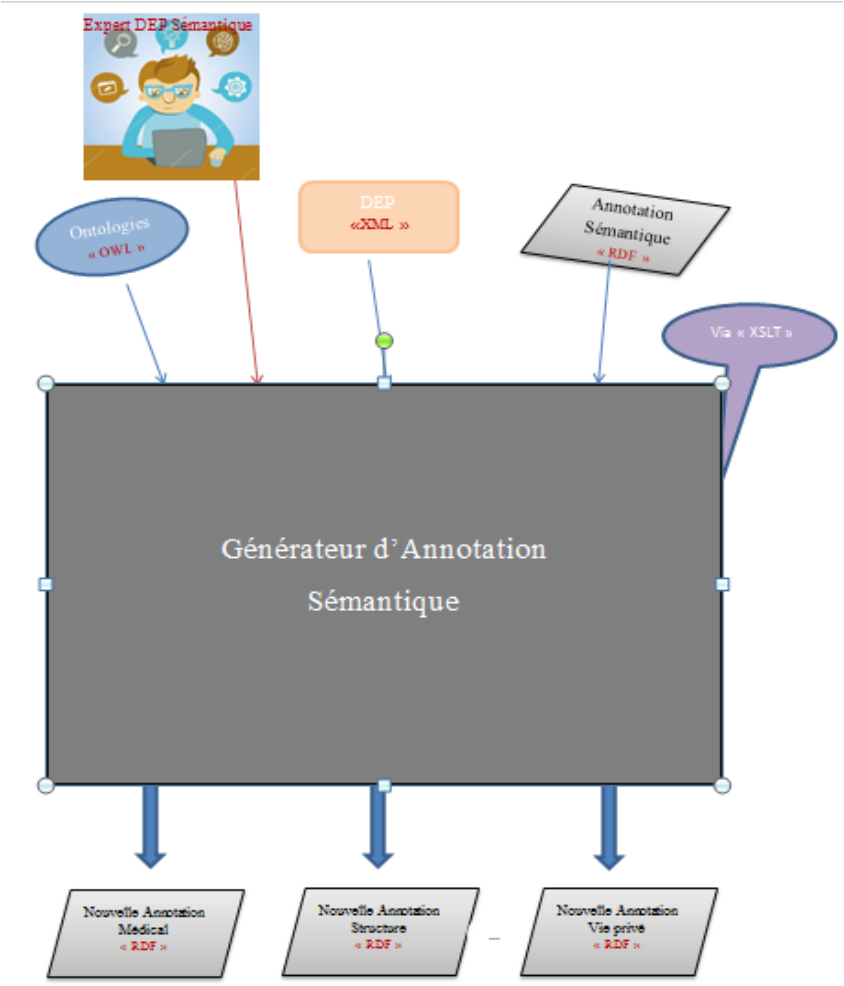


Figure 18: Module "Générateur d'annotation sémantique"

e. Module 5 : « Annotateur sémantique »

Ce module adopte l'approche semi-automatique. Il assiste un expert humain à créer des annotations sémantiques à base d'ontologies..

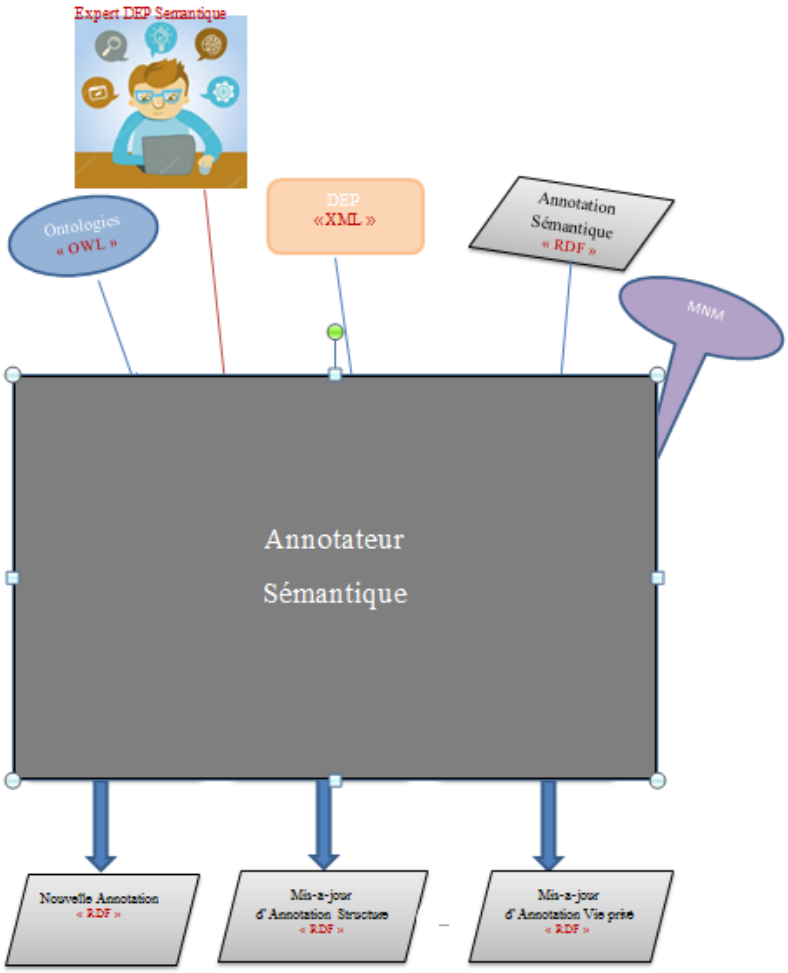


Figure 19: Module "Annotateur Sémantique"

f. Module 6 : « Recherche Sémantique »

Ce module est dédié à la recherche à base d'annotations sémantiques obtenues dans les deux modules précédents 4 et 5. Ce module est fondé sur le protocole SPARQL .

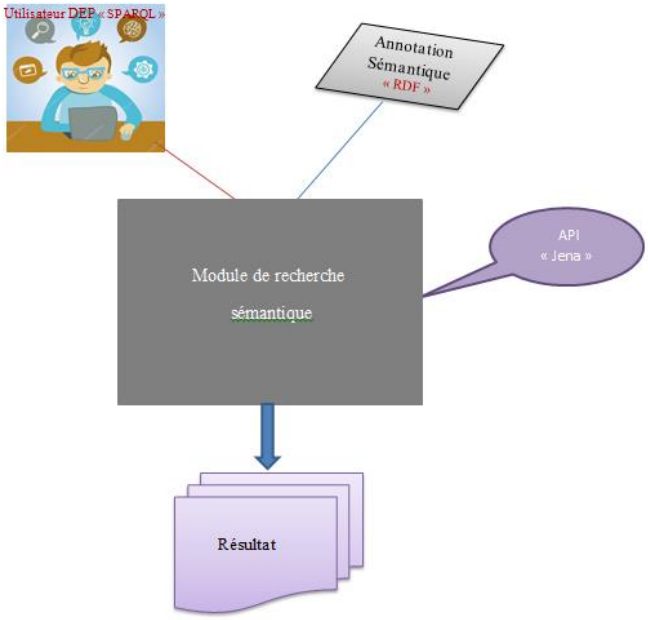


Figure 20: Module "Recherche sémantique"

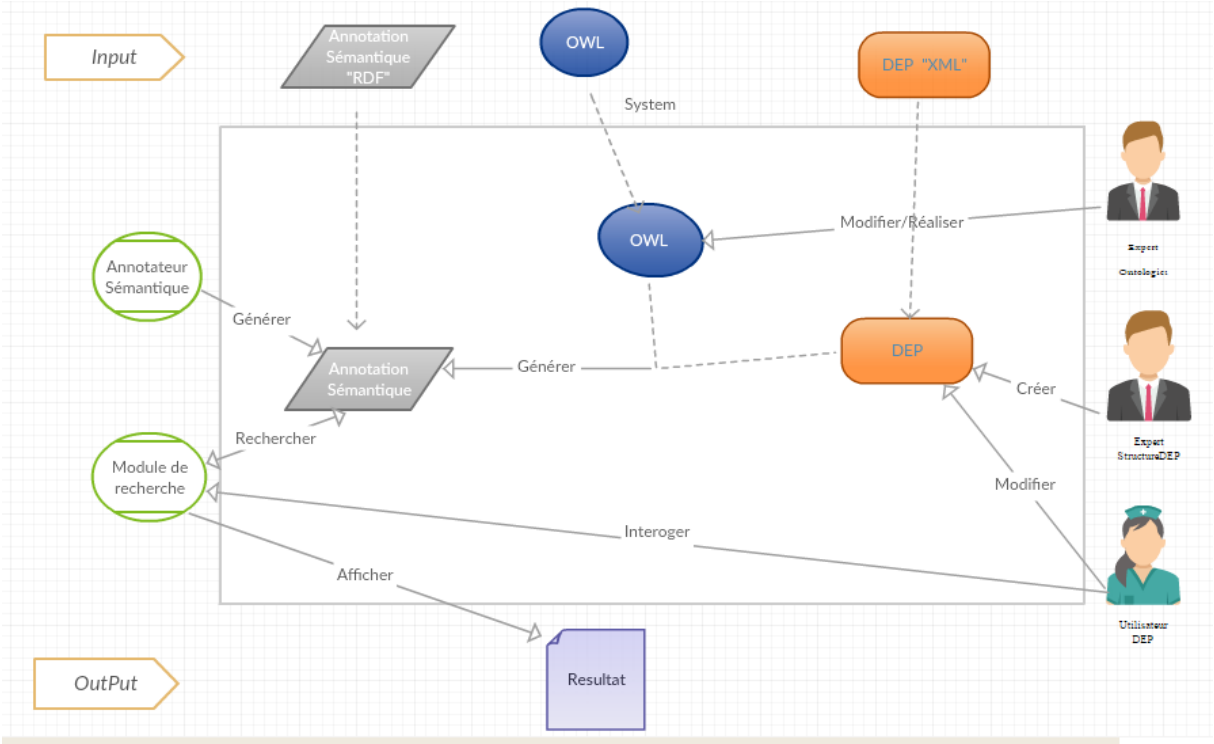


Figure 21: Architecture du système global

4. *Annotateur*

L'apparition du web sémantique a exigé le développement de nouvelles méthodes qui permettraient d'enrichir les annotateurs avec des étiqueteurs sémantiques. Ces méthodes se divisent en deux types :

- ✓ Les annotateurs sémantiques semi-automatiques
- ✓ Les annotateurs sémantiques automatiques

4.2 L'Annotateur Sémantique « semi-automatique »

L'annotateur humain doit spécifier pour chaque annotation le concept choisi dans une ontologie et la partie du DEP (Propriété) (Sujet de cette annotation). Cette méthode est précise mais prend un temps important lors de la réalisation plus le risque élevé des erreurs (présence de l'acteur humain).

Nous avons utilisé L'API JENA pour créer les annotations sémantiques de notre DEP

```
− <NamedIndividual rdf:about="http://www.semanticweb.org/nissa/ontologies/2017/5/5/untitled-ontology-85#Ahmed">
    <rdf:type rdf:resource="http://www.semanticweb.org/nissa/ontologies/2017/3/untitled-ontology-12#Fils"/>
    <rdf:type rdf:resource="http://www.semanticweb.org/nissa/ontologies/2017/3/untitled-ontology-12#Homme"/>
    <untitled-ontology-85:NomSpecialiste>Mr Y</untitled-ontology-85:NomSpecialiste>
    <untitled-ontology-85:Consultation>32</untitled-ontology-85:Consultation>
    <untitled-ontology-85:CaisseAffiliation>Imama-tlm</untitled-ontology-85:CaisseAffiliation>
    <untitled-ontology-85:Prenom>Ahmed</untitled-ontology-85:Prenom>
    <untitled-ontology-85:Taille>187</untitled-ontology-85:Taille>
    <untitled-ontology-85:Maladies>cardiaque</untitled-ontology-85:Maladies>
```

Figure 22: Extrait du code généré par l'API Jena

4.3 Les annotateurs sémantiques automatiques

Ce type d'annotateur laisse la main à la machine pour la tache de sélection et de position de l'annotation et même la création, l'enregistrement et l'affichage.

Nous avons utilisé XSLT pour créer une annotation sémantique à notre DEP.

```
<?xml version="1.0" encoding="UTF-8" ?>

<!-- New document created with EditiX at Wed Jun 07 02:25:18 GMT+01:00 2017 -->

<xsl:stylesheet version="1.0"
            xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
            xmlns:html="http://www.w3.org/1999/xhtml"
            xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
            xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
            xmlns:OnBCGO="C://Users/ILHAM/Desktop/mémo/mai/onto/BCGO"
            xmlns:OnVP="C://Users/ILHAM/Desktop/mémo/mai/onto/ViePrivé (1)"
            xmlns:OnP="C://Users/ILHAM/Desktop/mémo/mai/onto/personne"
            xmlns:OnStruct="C://Users/ILHAM/Desktop/mémo/mai/onto/Structure3105.owl"
 >

<xsl:template match="/">
    <rdf:RDF>
        <rdf:Description rdf:about="C://Users/ILHAM/Desktop/mémo/mai/xml/DP/DPI/IdentificationPatient/IdentificationGeneral">
        <OnStruct:nom><xsl:value-of select="//nom"/>
        </OnStruct:nom>
         <OnStruct:prenom><xsl:value-of select="//Prenom"/>
        </OnStruct:prenom>
        <OnStruct:age><xsl:value-of select="//Age"/>
        </OnStruct:age>
         <OnStruct:sexe><xsl:value-of select="//sexe"/>
        </OnStruct:sexe>
         <OnStruct:Profession><xsl:value-of select="//Profession"/>
        </OnStruct:Profession>
       <OnStruct:Nationalite><xsl:value-of select="//Nationalite"/>
        </OnStruct:Nationalite>
```

Figure 23: Extrait du code de transformation via XSLT

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<rdf:Description rdf:about="C://Users/ILHAM/Desktop/mémo/mai/xml/DP/DPI/IdentificationPatient/IdentificationGeneral'
<OnStruct:nom xmlns:OnStruct="C://Users/ILHAM/Desktop/mémo/ViePrivé (1)">    dib  </OnStruct:nom>
<OnStruct:prenom xmlns:OnStruct="C://Users/ILHAM/Desktop/mémo/ViePrivé (1)">  ahmed </OnStruct:prenom>
<OnStruct:age xmlns:OnStruct="C://Users/ILHAM/Desktop/mémo/ViePrivé (1)">  27 </OnStruct:age>
<OnStruct:sexe xmlns:OnStruct="C://Users/ILHAM/Desktop/mémo/ViePrivé (1)"> M </OnStruct:sexe>
<OnStruct:Profession xmlns:OnStruct="C://Users/ILHAM/Desktop/mémo/ViePrivé (1)">  developpeur </OnStruct:Profession>
<OnStruct:Nationalite xmlns:OnStruct="C://Users/ILHAM/Desktop/mémo/ViePrivé (1)">Algerien  </OnStruct:Nationalite>
```

Figure 24: Extrait du fichier annoté en RDF

Conclusion générale

L'évolution du web sémantique a apporté de la structuration, de l'indexation et l'interopérabilité des données.

Dans ce mémoire, nous avons décrit tout d'abord le dossier électronique du patient. Ensuite nous avons montré l'utilité du web sémantique pour la sémantification de ce dossier.

Ensuite nous avons présenté l'architecture globale de notre système, l'élément crucial de cette architecture est l'annotation du document DEP qui fait l'objet de notre travail. Elle constitue une relation sémantique entre les différentes ontologies utilisées et les éléments du contenu du DEP « XML ».

Les ontologies utilisées sont : (1) l'ontologie « ***structure du DEP »*** inspirée des dossiers électroniques existants des patients. Elle consiste en deux grandes parties : (a) l'identification générale du patient; (b) les actes subis par ce dernier ; (2) l'ontologie « ***vie privé »*** ajoute la sécurité et la performance a ce DEP puisque le patient est le premier responsable de l'affectation des droits d'accès aux différents acteurs médicaux ; (3) l'ontologie « Relation familiales » décrit quelques relation entre les membres d'une famille ; (4) l'ontologie ***médical***e qui décrit des plans de traitement personnalisés pour les patients diabétiques.

La création du DEP était inspiré des dossiers médicaux des CHUs de de Maghnia, Beni-Saf et Avicenne. Par ailleurs, nous avons construit une feuille de style XSLT pour produire un affichage web (HTML).

Enfin la création des annotations sémantiques nous a permis de lancer une recherche sémantique du contenu du DEP à l'aide du protocole « SPARQL ».

En perspectives, nous pensons que ce type d'architecture est très ouvert car les technologies du Web sémantique sont en perpétuelle évolution et les techniques de représentation des connaissances sont de plus en plus complexes et performantes. Plusieurs améliorations peuvent être apportées au système telles que :

- Rendre le DEP dynamique grâce à sa liaison à une base de données médicale.
- Etudier l'aspect inférence dans le DEP afin de détecter d'éventuelles contradictions ou de déduire de nouvelles connaissances.

Références bibliographiques

[Ama ,2005] AMARDEILH F., LAUBLET P. & MINEL J. L. (2005). Annotation documentaire et peuplement d'ontologie à partir d'extractions linguistiques. In Actes des 16emes journées francophones d'Ingénierie des Connaissances, p. 25–36

[Azo, 2006] Azouaou, F., Desmoulins, C. "Teachers' Document Annotating: Models for a Digital Memory Tool." Int. J. Cont. Engineering Education and Lifelong Learning (IJCEELL). 16 (1/2). pp.18-34. 2006.

[Bal, 2000] Baldonado. M, Cousins. S, Gwizdka. J. et Paepcke. A. '' Notable: At the Intersection of Annotations and Handheld Technologies''. Bristol. 2000, HUC, pp. 100-113.

[Bar, 2005] Catherine Barry,et al « Les annotations pour gérer les connaissances du dossier patient ». IC - 16_emes Journées francophones d'Ingénierie des Connaissances, May 2005, Nice, France. Presses universitaires de Grenoble, pp.73-84, 2005. <hal-01023761>

[Bat, 2002] Eric Bataille et al. , « Vers le concept de document sémantique », Document numérique 2002/1(Vol.6),p.99-114.

DOI 10.3166/dn.6.1-2.99-114

[Ben, 2011]Benhmidi. H, Mazouz. A « Construction et Manipulation D'une Ontologie » ,2011

[Bou, 2008] Boufaida Z et al, «Construction d'une Ontologie pour l'annotation des CVs/Offres d'emplois »,2008.

[Bri, 2003] S. Bringay, S. Barry, '' Information du dossier patient Tâches TV2 : Veille technologique : -Annotations''. Projet HTSC, 1er rapport de veille « Document numérique médicaux », Réf : RV/03/TV2/02, Septembre 2003.

[Brin, 2003] Bringay, S., Barry, C. et Charlet, J. "Les documents et les annotations du dossier patient hospitalier." Information - Interaction - Intelligence. 4 (1). pp.191-211. 2003.

[Cha, 2009] Jean Charlet et al« Utiliser et construire des ontologies en médecine : le primat de la terminologie »

[Deb, 2017] Deborah L.Mc Guinness, Frank Van Harmelen « vue d'ensemble du langage d'ontologie web OWl » visité le 17-03-2017

[Dec ,2012] Gunnar Declerck et al«Construire une ontologie médicale pour la recherche d'information : problématiques terminologique et de modélisation. »

[Den, 2000] Denoue, L. De la création à la capitalisation des annotations dans un espace personnel d'informations. Informatique. Annecy, Université de Savoie.p. 159. Sous la direction de L. Vignollet. 2000.

[Des, 2002] Desmontils. E, Jacquin. C. ‘‘Indexing a Web Site with a Termi nology Oriented Ontology’’. In The Emerging Semantic Web Amsterdam : IOS Press, 2002, pp. 181-197.

[Doc, 2012] Comité éditorial pédagogique de l'UVMaF, Support de Cours « Le dossier médical » 2011-2012

[Emi, 2010] « La Mise en place du Dossier Du Patient Informatisé en Soins De Suite et Réadaptation, vecteur d’interdisciplinarité » écrit par Emi Seino, Julie Lionti, Myriam Delsarte, Nathalie Desbief le vendredi 12 novembre 2010.

[Far, 1996] Farquhar, R. Fikes and J. Rice,“The Ontolingua Server: A Tool for Collaborative Ontology Construction”, In Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop, Banff, Alberta, Canada, 44.1-44.19, 1996, http://www.ksl.stanford.edu/software/ontolingua/.

[Fran, 2006] Francis «le langage d’ontologie Web OWL »,2006.

[Geo, 2013]George Abraham «Semantic Agents for Learning Style Prediction» 2013

[Gru, 1993] Gruber,T.Rm A translation approach to portable ontology specification. Knowledge Acquisition, 5 (2), 199–220. 1993.

[Gom, 2004] Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). Ontological Engineering (with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web).Springer. Cité dans [Hadjoui, 2012].

[Han, 2002] Handschuh. S, Staab. S et Ciravegna. F. ‘‘S-CREAM – Semiautomatic CREAtion of Metadata’’ .The 13th International Conference on Knowledge Engineering and Management (EKAW 2002), ed Gomez-Perez, A., Springer Verlag, 2002.

[Inf, 2017] Inforoute Santé du Canada, La santé numérique et vous ‘Aperçu du DES, du DMEet du DSP », 2017

[Inst, 2013] « les-ontologies »-51f7a7f9e921c,2013

[Kir, 2004]KIRYAKOV A., POPOV B., TERZIEV I., MANOV D. & OGNYANOFF D. (2004).Semantic annotation, indexing, and retrieval. J. Web Sem., 2(1), 49–79.

[kir, 2017] Kiryakov A « Open Source AI Platforms for knowledge Management and Sémantic web » visité le 10-05-2017

[Mor, 2017] Support de Cours, campus CERMIS.fr UE santé société humanité, *Professeur Patrice MORAND* visité le 20-03-2017

[Nee, 1991] Neeches, Finin T, Fikes R.E, & all. (1991). Enabling technology for knowledge sharing. Al Magazine, 12 (3), 36–56. [Khalfi, 2009].

[Noy, 2001] Natalya F. Noy and Deborah L. McGuinness, “Ontology Development 101: A Guide to Creating Your First Ontology”. Stanford Knowledge Systems Laboratory Technical

Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

[Pri, 2003] Pié.Y. ‘‘ Annotations et méta-données pour le web sémantique’’. Dans les actes des journées scientifiques « web sémantique et SHS » mai, 2003.

[Rep, 2002] du web traditionnel au web sémantique écrit par Elodie MUSSO et Marie Garreau, en cadreés par Fabien Gandon

[Sci, 2001] T. Berners-Lee, J. Hendler et O.Lassila, Scientific American 2001.

[Ser, 2017] Serge garlatti , Yannik Prié « Méta-données et Annotations dans le web Sémantique » visité le 10-04-2017

[Sur, 2002] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer and Wenke, D., “OntoEdit: Collaborative Ontology Engineering for the Semantic Web”. In Proceedings of the International Semantic Web Conference 2002 (ISWC 2002), Sardinia, Italia, June 2002.

[Usc, 1996] Uschold, M., & Grüninger, M. (1996). Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11 (2), 93–155. Cité dans [Hadjoui, 2012]

[Ver, 1997] Veron. M ‘‘ Modélisation de la composante annotative dans les documents éléctroniques, Master de recherche, IRIT, Toulouse, 1997.

ملخص

إن ملف المريض الإلكتروني يسمح بجمع كل المعلومات الطبية ذات الصلة بعلاج المريض. يندرج عمل الماستر هذا في إطار الويب الدلالي وتطبيقاته في المجال الطبي .هذا العمل يقوم على إنشاء الشروح الدلالية لأجزاء من سجل الالكتروني للمريض من المفاهيم المنبثقة من الأنطولوجيا .

لقد تم اختيار أربعة الأنطولوجيا: (1) الأنطولوجيا الطبية للمفاهيم الأساسية في مجال الطبي الممارسة الطبية. (2) أنطولوجيا الأسرة تظهر العلاقات الأسرية. (3) أنطولوجيا وثيقيه تمثل منظمة منطقية للسجل الإلكتروني للمريض؛ (4) انطولوجيا الحياة الخصوصية تظهر مستويات حقوق السرية والوصول. الانطولوجيا الأولى و الثانية استعملناهما مما هو موجود في الاعمال السابقة اما فيما يخص الباقي فقد انشاناهم بانفسنا.

ان عملية الشرح الدلالي يمكن القيام بها بطريقتين: نصف آلية وآلية. في الطريقة الأولى، يتم إنشاء شروح من قبل خبير البشري من خلال أداة برمجية في المجموعة الثانية، يتم إنشاء الشروح تلقائيا من الملف الالكتروني للمريض من خلال ورقة أنماط (XSLT).ثم في الاخير يتم الإستجواب عن طريق (SPARQL)

الكلمات المفتاحية : الأنطولوجيا، الأنطولوجيا الطبية، الويب الدلالي، الشرح الدلالي، ملف المريض الإلكتروني.

Résumé

Le Dossier Electronique du Patient (DEP) permettra de regrouper toutes les données médicales pertinentes pour le traitement d'un patient. Le présent travail de master rentre dans le cadre du Web sémantique et son application au domaine médical. Il consiste à faire une sémantification du dossier électronique du patient (XML) à base d'ontologies (OWL). Quatre ontologies (OWL) ont été choisies : (1) une ontologie médicale schématisant les concepts clés de la pratique médicale; (2) une ontologie familiale illustrant les relations familiales ; (3) une ontologie documentaire représentant l'organisation logique d'un dossier électronique du patient ; (4) une ontologie de vie privée montrant les niveaux de confidentialité et les droits d'accès. Les deux premières ontologies ont été adoptées à partir de l'existant pendant que les deux dernières ont été créées par nous-mêmes. La sémantification consiste à créer des annotations sémantiques (RDF) de parties du dossier électronique du patient par des concepts émanant de l'une des quatre ontologies précitées. La sémantification peut se faire selon deux approches : semi-automatique et automatique. Dans la première, les annotations sont créées par un expert humain à travers un outil logiciel (annotateur sémantique). Dans la seconde, les annotations sont générées automatiquement à partir du dossier électronique du patient à travers une feuille de style XSLT (générateur d'annotations sémantiques). Les annotations sémantiques résultantes sont ensuite interrogées grâce à SPARQL.

Mots clés : Dossier Electronique du Patient , Web sémantique, Ontologie, Annotation sémantique.

Abstract

The Electronic File of the Patient (EFP) will enable collecting all the medical data relevant to the treatment of a patient. The present Master work is part of the Semantic Web and its application to the medical field. It consists in making a semantization of the patient's electronic file (XML) based on ontologies (OWL). Four ontologies (OWL) have been chosen: (1) a medical ontology outlining the key concepts of medical practice; (2) a family ontology illustrating family relationships; (3) a documentary ontology representing the logical organization of an electronic patient's file; (4) a private life ontology showing the levels of privacy and access rights. The first two ontologies were adopted from the existing ones while the two last ones were created by us. Semantization consists in creating semantic annotations (RDF) of parts of the patient's electronic file by concepts emanating from one of the four ontologies mentioned above. Semantization can be done in two ways: semi-automatic and automatic. In the first way, annotations are created by a human expert through a software tool (semantic annotator). In the second, the annotations are generated automatically from the patient's electronic file through an XSLT style sheet (semantic annotation generator). The resulting semantic annotations are then queried using SPARQL.

Keywords: Electronic File of the Patient, Semantic web, Ontology, Semantic annotation.