

UNIVERSITE MOHAMMED V-AGDAL

Faculte des Sciences



Département d'Informatique

Filière Licence fondamentale en Science Mathématiques et
Informatique

PROJET DE FIN D'ETUDES

intitulé :

Web Sémantique

Application : **Forum du cancer**

Présenté par :

ELASRI CHAIMAE

NOUSSARI AYOUB

soutenu le 24 septembre 2020

Membre de jury :

Pr. Mohamed Benkhalifa	<i>Président</i>
Pr. Oussama Mohamed Reda	<i>Encadrant</i>
Dr. Yassine Touimi Benjelloune	<i>Examineur</i>

Année universitaire 2019-2020

Remerciements

Nous souhaitons avant tout remercier nos encadrants de Mémoire Oussama Mohamed REDA et Yassine Benjelloune Touimi pour le temps qu'ils ont consacré à nous apporter les outils méthodologiques indispensable à la conduite de cette recherche. leurs exigences nous a grandement stimulée. L'enseignement de qualité dispensé par la licence SMI à également su nourrir nos réflexions et a représenté une profonde satisfaction intellectuelle. Nous adressons aussi nos sincères remerciements à tous les professeurs, intervenants et toutes les personnes qui, par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé nos réflexions et ont accepté à nous rencontrer et répondre à nos questions durant nos recherches. Nous exprimons nos vifs remerciement aux membre du jury qui nous ont honoré par leur présence lors de la soutenance de notre travail et à tous les professeurs que nous avons eu lors de ces trois années de licence. Bien sur ce travail n'aurait pas eu lieu sans le soutien de nos collègues, nos amis, nos frères et bien évidemment notre famille et surtout nos parents, qui nous ont vivement encouragé dans cette expérience, que cela soit moralement ou financièrement. A tous ces intervenants, nous présentons nos sincères remerciement, notre respect et notre gratitude

Résumé

Le Web sémantique (plus techniquement appelé « le Web de données ») permet aux machines de comprendre la sémantique, la signification de l'information sur le Web.

Le concept du « modèle de réseau sémantique » a été inventé dans les années 1960 par le chercheur en sciences cognitives Allan M. Collins, le linguiste Ross Quillian (en) ainsi que la psychologue Elizabeth Loftus et exposé dans diverses publications comme manière de représenter des connaissances structurées. Appliqué au contexte de l'internet, ce modèle étend le réseau de liens hypertextes des pages Web lisibles humainement en insérant des métadonnées lisibles par la machine sur les pages. Ces métadonnées sont liées les unes aux autres, permettant à des agents d'accéder au Web de manière plus intelligente et effectuer des tâches pour le compte d'utilisateurs.

Le Web sémantique est considéré comme une passerelle pour accéder aux données entre différentes applications et systèmes. Ses applications sont nombreuses dans l'édition, les blogs, et dans plusieurs autres domaines.

Pour notre application on a créé un forum de discussion en PHP sur le cancer qui se base au début sur une ontologie qui range tous les informations nécessaires concernant le cancer, après sur chaque question posée dans le forum on applique LSA en python pour avoir les mots les plus pertinents puis on exécute une requête SPARQL sur l'ontologie pour obtenir la réponse de la question.

Mots clés : web sémantique, ontologie, LSA, SPARQL, PHP, PYTHON, forum.

Abstract

The Semantic Web (more technically called "the Web of data") allows machines to understand the semantics, the meaning of information on the Web.

The concept of the "semantic network model" was invented in the 1960s by cognitive scientist Allan M. Collins, linguist Ross Quillian and psychologist Elizabeth Loftus and expounded in various publications as a way of representing structured knowledge. Applied to the context of the internet, this model extends the hyperlink network of human readable web pages by inserting machine readable metadata on the pages. This metadata is linked to each other, allowing agents to access the web more intelligently and perform tasks on behalf of users.

The Semantic Web is seen as a gateway to access data between different applications and systems. Its applications are numerous in publishing, blogs, and in several other fields. For our application we created a discussion forum in PHP on cancer which is based at the beginning on an ontology which stores all the necessary information concerning cancer, after on each question asked in the forum we apply LSA in python to have the words most relevant then we run a SPARQL query on the ontology to get the answer to the question.

Keywords : semantic web, ontology, LSA, SPARQL, PHP, PYTHON, forum.

Table des matières

Remerciements	1
Résumé	2
Abstract	3
1 Web sémantique	7
1.1 Histoire	7
1.2 Définition	8
1.3 Objectif	9
1.4 Le Web Sémantique et le Web 3.0	9
2 Notion de l'Ontologie	10
2.1 Définition de l'ontologie	10
2.2 Ontologie de l'information	11
2.3 Ontologie du domaine	11
2.4 Ontologie Informatique	12
2.5 L'Ontologie, Outil principale du Web Sémantique	12
3 Installation du protégé, WampServer, PYTHON	13
3.1 Installation du protégé	13
3.2 Installation du WampServer	14
3.3 Installation du PYTHON	15
4 Les étapes engagées durant le projet	17
4.1 Etape1 : Créer l'Ontologie	17
4.2 Etape2 : Forum de Discussion	20
4.2.1 Définition du Forum de Discussion	20
4.2.2 Création du Forum de Discussion	20
4.3 Etape3 : Poser les Questions	26
4.4 Etape4 :LSA "Latent Semantic Analysis	28
4.4.1 Définition du LSA	28
4.4.2 code du LSA en PYTHON	29
4.5 Etape5 : SPARQL	30

4.5.1	Définition	30
4.5.2	Exécution des requêtes SPARQL	31
4.6	Etape6 : Réponse à la question	32
4.7	Automatisation	33
4.7.1	Web Scraping	33
4.7.1.1	Définition du web scraping	33
4.7.1.2	Code du web scraping en python	34
4.7.2	LSA	35
4.7.3	Perspective du projet dans l'avenir	36

Conclusion		37
-------------------	--	-----------

Chapitre 1

Web sémantique

1.1 Histoire

En 1994, lors de la première conférence WWW à Genève, plus précisément au CERN, a lieu l'annonce de la création du W3C. C'est d'ailleurs à cette période que Tim Berners-Lee dresse les objectifs du W3C et montre les besoins d'ajouter de la sémantique au Web futur. Il montre alors en quoi les liens hypertextes ou, plus précisément, la façon dont on met en relation les documents sur le Web est trop limitée pour permettre aux machines de relier automatiquement les données contenues sur le Web à la réalité. Compte tenu de l'ambition d'un tel projet, cette idée suscite quelques résistances et controverses qui sont classiquement rencontrées dès qu'on aborde des problématiques liées au domaine de l'intelligence artificielle.

Après cette conférence, mise à part la mise en place des recommandations nécessaires à la construction des documents, le W3C nouvellement créé entame les premières réflexions sur la mise en place du Web sémantique. Ces réflexions aboutissent à la publication d'un premier draft de recommandations sur le Web sémantique en octobre 1997 et d'une seconde en avril 1998. Cette même année, Tim Berners-Lee publie un document sur les toutes premières moutures de ce qui sera plus tard appelé le Web sémantique. Ces moutures consistent à mettre en place les différentes technologies du Web sémantique. Dans ce document, il présente le Web sémantique comme une sorte d'extension du Web des documents, qui constitue une base de données à l'échelle mondiale, afin que toutes les machines puissent mieux lier les données du Web. Cette feuille de route se matérialise par une représentation graphique, la figure1, qui montre l'agencement des différentes briques technologiques du Web sémantique.

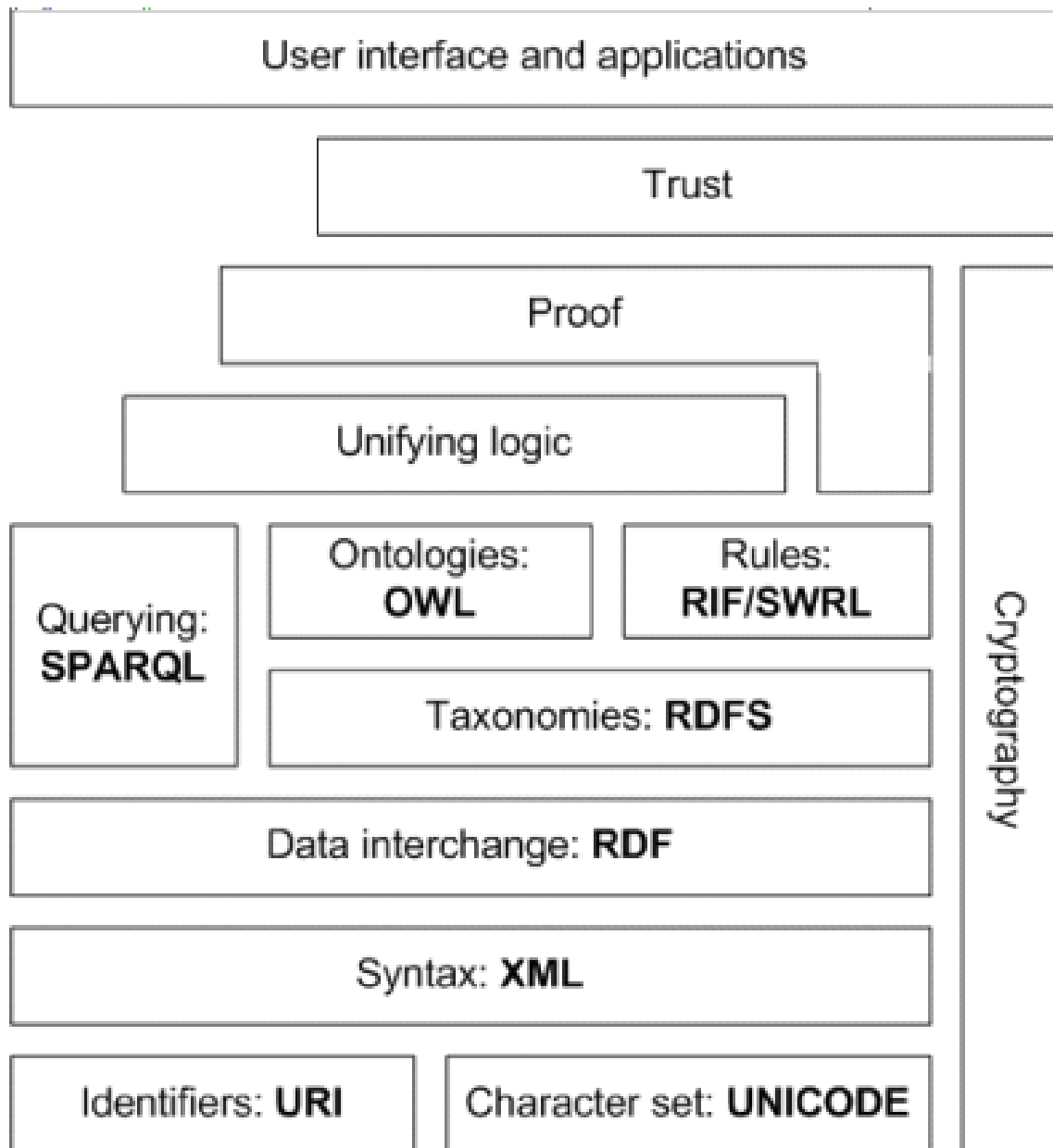


FIGURE 1.1 – Schéma descriptif

1.2 Définition

Le Web sémantique, ou toile sémantique, est une extension du Web standardisée par le World Wide Web Consortium (W3C). Ces standards encouragent l'utilisation de formats de données et de protocoles d'échange normés sur le Web, en s'appuyant sur le modèle Resource Description Framework (RDF).

Le Web sémantique est par certains qualifié de web 3.0 .

Selon le W3C, « le Web sémantique fournit un modèle qui permet aux données d'être partagées et réutilisées entre plusieurs applications, entreprises et groupes d'utilisateurs ». L'expression a été inventée par Tim Berners-Lee (inventeur du Web et directeur du W3C), qui supervise le développement des technologies communes du Web sémantique. Il le définit comme « une toile de données qui peuvent

être traitées directement et indirectement par des machines pour aider leurs utilisateurs à créer de nouvelles connaissances ». Pour y parvenir, le Web sémantique met en œuvre le Web des données qui consiste à lier et structurer l'information sur Internet pour accéder simplement à la connaissance qu'elle contient déjà.

Alors que ses détracteurs ont mis en doute sa faisabilité, ses promoteurs font valoir que les applications réalisées par les chercheurs dans l'industrie, la biologie et les sciences humaines ont déjà prouvé la validité de ce nouveau concept. L'article original de Tim Berners-Lee en 2001 dans le *Scientific American* a décrit une évolution attendue du Web existant vers un Web sémantique, mais cela n'a pas encore eu lieu. En 2006, Tim Berners-Lee et ses collègues ont déclaré : « Cette idée simple... reste largement inexploitée. »

1.3 Objectif

Un des principaux objectifs du Web sémantique est de permettre aux utilisateurs d'utiliser la totalité du potentiel du Web : ainsi, ils pourront trouver, partager et combiner des informations plus facilement. Aujourd'hui tout le monde est capable d'utiliser des forums, d'utiliser des réseaux sociaux, de chatter, de faire des recherches ou même d'acheter différents produits. Néanmoins, il serait mieux que la machine fasse tout ceci à la place de l'homme, car actuellement, les machines ont besoin de l'homme pour effectuer ces tâches. La raison principale est que les pages Web actuelles sont conçues pour être lisibles par des êtres humains et non par des machines. Le Web sémantique a donc comme principal objectif que ces mêmes machines puissent réaliser seules toutes les tâches fastidieuses comme la recherche ou l'association d'informations et d'agir sur le Web lui-même.

1.4 Le Web Sémantique et le Web 3.0

La communauté du Web dans son ensemble a tendance à dire que les deux termes « Web sémantique » et « Web 3.0 » représentent à peu près le même concept, si ce n'est pas totalement interchangeable. La définition continue de varier en fonction des gens avec qui vous parlez. L'avis général est que le Web 3.0 est très certainement la prochaine grande révolution, mais il se trouve que, pour le moment, ce ne sont que des spéculations quant à ce qu'il pourrait être. Il y aura encore de grosses améliorations, mais en gardant la plupart des propriétés du Web 2.0. Il y en a certains qui prétendent que le Web 3.0 sera plus applicatif et centrera ses efforts vers des environnements plus graphiques, d'autres qui prétendent qu'il sera plus axé sur la recherche d'informations géographiques basées sur la géolocalisation ou encore même utiliser les nombreux progrès en intelligence artificielle.

Chapitre 2

Notion de l'Ontologie

2.1 Définition de l'ontologie

Selon Gruber, “une ontologie est une spécification explicite d’une conceptualisation” et “une conceptualisation est une vision du monde abstraite et simplifiée qu’on veut représenter pour un but donné” (Gruber, 1993). La représentation explicite de l’information occupe une place importante dans le développement logiciel. Pour citer Linus Torvald, “les mauvais programmeurs s’inquiètent du code” alors que “les bons programmeurs s’inquiètent des structures de données et de leurs relations.” (Torvalds, 2006) Dans le cas du développement logiciel classique, on s’intéresse à des structures de données centrées sur les algorithmes, c’est-à-dire qui permettront une facilité de compréhension et de maintenance ainsi qu’une efficacité du point de vue de la mémoire et du temps d’exécution. Ce qui intéresse le développeur de l’ontologie est plutôt une représentation centrée sur une conceptualisation répondant aux buts de l’ontologie. Cette conceptualisation se base sur le monde réel, car elle consiste en une vision simplifiée de celui-ci. Pour illustrer ceci, nous pouvons donner un exemple de Grady Booch sur des visions simplifiées du chat (Booch, 2006). Une propriétaire de chat ne s’intéressera qu’à certaines caractéristiques du chat, c’est-à-dire au fait qu’il ronronne, qu’il a besoin d’eau et d’un certain type de nourriture, etc. La vétérinaire aura une vision différente du chat, car elle s’intéressera à d’autres propriétés comme sa structure osseuse, ses organes internes, sa prédisposition à divers types de maladies, etc. On voit ici que dans les deux cas, les visions de la propriétaire et de la vétérinaire ne représentent pas parfaitement le chat, mais seulement les caractéristiques qui sont importantes pour chacune. C’est ce qu’on entend par conceptualisation. Dans ce cas-ci, une représentation plus complète du chat exposerait chacune de ses cellules, son phénotype, etc. La représentation complète est rarement nécessaire, voire impossible, et c’est pourquoi on s’intéresse seulement à une vision simplifiée du monde réel permettant de répondre à des questions spécifiques. La représentation des connaissances par l’ontologie présente plusieurs avantages qui rendent cette technologie intéressante. Les attributs qu’on retient ici sont sa simplicité, sa flexibilité, la possibilité d’y appliquer du raisonnement et la possibilité

de l'interroger à divers niveaux d'abstraction.

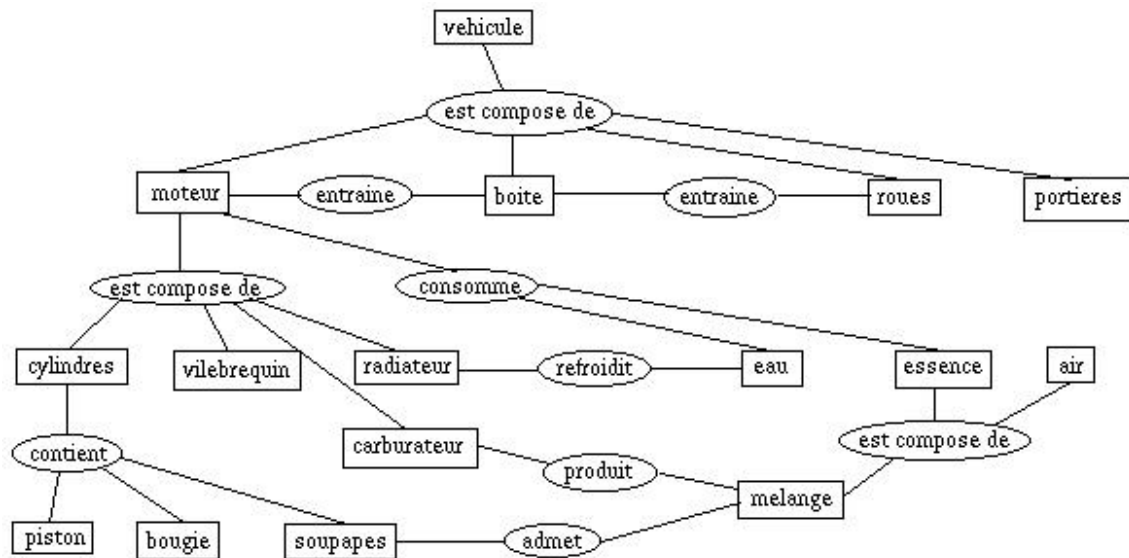


FIGURE 2.1 – Exemple d'une ontologie

2.2 Ontologie de l'information

L'ontologie de l'information contribue à organiser et clarifier les idées des collaborateurs sur un projet en exposant le schéma global du système avec tous ses liens et ses raisonnements. Ce type d'ontologie est plutôt un outil de Knowledge Management utilisé dans un projet dans le but de réduire les incompréhensions et quiproquos. C'est un schéma qui est fait à la main en groupe devant un tableau pour se mettre d'accord sur la définition des termes employés, avec les liens entre eux

2.3 Ontologie du domaine

L'ontologie du domaine est fonctionnelle et orientée objet. Elle est utilisée pour représenter un domaine (les composants informatiques, l'immobilier, le droit, la génétique, la vie de l'ornithorynque [U+0085]) sous forme de base de connaissances. Par exemple, faire l'ontologie d'un site web est intéressant pour comprendre sa structure. Elle peut être réalisée en amont de la création du site ou bien sur des sites de grande envergure en activité. Elle présente les concepts-clés, les attributs, les instances relatifs au domaine. Pour réaliser ce type d'ontologie, il existe des éditeurs de structure de base de connaissances. Le plus connu et le plus utilisé est Protégé. Open-source et gratuit, il est implémenté à partir d'une plate-forme Java J2EE. Développé par l'Université de Stanford, il intègre les standards du Web sémantique et notamment OWL (pour Web Ontology Language). Il offre de

nombreux composants optionnels : raisonneurs, interfaces graphiques. Ses ontologies peuvent être exportées dans les formats :RDF(S), OWL, et XML Schema

2.4 Ontologie Informatique

Les ontologies informatiques sont des outils qui permettent de représenter précisément un corpus de connaissances sous une forme utilisable par une machine. Elles représentent un ensemble structuré de concepts. Les concepts sont organisés dans un graphe dont les relations peuvent être des relations sémantiques et/ou des relations de composition et d'héritage (au sens objet). Par exemple, le langage UML (Unified Modeling Language) est un formalisme graphique qui est utilisé pour faire des ontologies informatiques.

2.5 L'Ontologie, Outil principale du Web Sémantique

Dans le cadre de ses travaux sur le Web sémantique, le W3C a mis en place en 2002 un groupe de travail dédié au développement de langages standards pour modéliser des ontologies utilisables et échangeables sur le Web. Ce groupe a publié en 2004 une recommandation définissant le langage OWL, fondé sur le standard RDF et en spécifiant une syntaxe XML. OWL est actuellement le standard le plus utilisé. Il est possible de transformer un modèle de données UML en ontologie pour le Web sémantique. Il existe néanmoins des différences entre UML et OWL qui ne les rendent pas équivalents et donc pas complètement traduisibles de l'un vers l'autre. Les taxonomies ne sont pas identiques. Ce problème est typique des langages de représentation de la connaissance.

Chapitre 3

Installation du protégé, WampServer, PYTHON

3.1 Installation du protégé

Pour commencer notre projet il fallait installer en premier PROTEGE : Protégé est un système auteur pour la création d'ontologies. Il a été créé à l'université Stanford et est très populaire dans le domaine du Web sémantique et au niveau de la recherche en informatique. Protégé est développé en Java. Il est gratuit et son code source est publié sous une licence libre (la Mozilla Public License). Protégé peut lire et sauvegarder des ontologies dans la plupart des formats d'ontologies : RDF, RDFS, OWL, etc. Il possède plusieurs concurrents tels que Hozo, OntoEdit et Swoop. Il est reconnu pour sa capacité à travailler sur des ontologies de grandes dimensions.



FIGURE 3.1 – logo du protégé

3.2 Installation du WampServer

L'installation su WampServer regroupe :

- MySQL, qui est un logiciel de gestion de bases de données qui permet d'enregistrer les données de manière organisée



FIGURE 3.2 – logo de MYSQL

- Apache, qu'on l'appelle un serveur web il s'agit du plus important de tous les programmes car c'est lui qui est chargé de délivrer les pages web aux visiteurs.



FIGURE 3.3 – logo d'Apache

- PHP, qui est un plug-in pour Apache qui le rend capable de traiter des pages web dynamiques en PHP, ce qui nous aidera dans notre écriture du script.

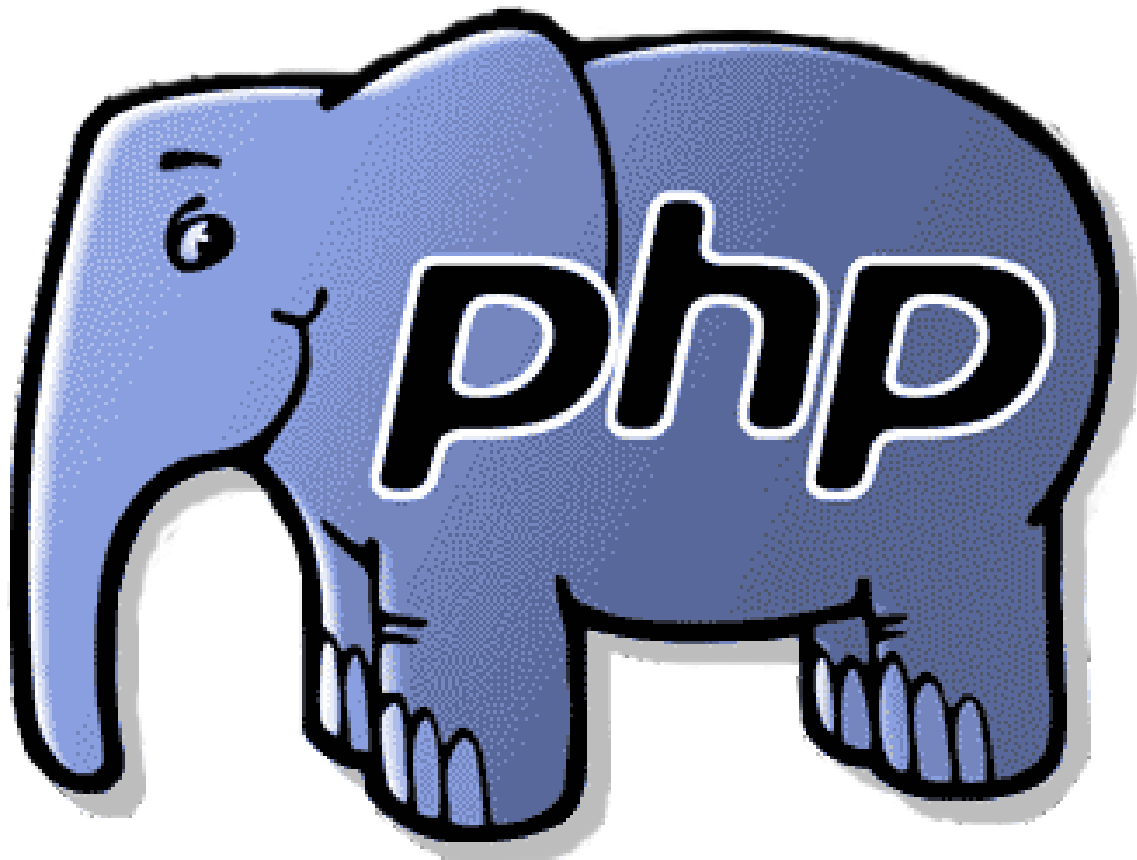


FIGURE 3.4 – logo de PHP

3.3 Installation du PYTHON

PYTHON est un puissant langage de programmation interprété et orienté objet. Il convient aux débutants en programmation, car il n'a pas besoin d'être compilé. Pour l'installer, il faut :

- Aller vers le site de python (www.python.org/downloads), le site détecte que nous sommes sous Windows et la version que nous utilisons, donc il affiche l'installation Windows appropriée
- Choisir la version à installer : dans notre cas la version est 3.7.8
- Exécuter l'installateur une fois le téléchargement terminé
- Cocher la case Add Python 3.5 To Path, on peut aussi l'exécuter directement depuis l'invite des commandes
- Cliquer sur Installer maintenant, pour qu'il s'installe avec tous ses paramètres par défaut.

Maintenant PYTHON est bien installé sur notre PC.



FIGURE 3.5 – logo de PYTHON

Chapitre 4

Les étapes engagées durant le projet

4.1 Etape1 : Créer l'Ontologie

Avant tout ,il faut d'abord ranger tous les informations nécessaires qu'on va stocker dans l'ontologie, Dans notre cas on a cherché tous ce qui a une relation avec le cancer (définition, types, causes...).

Tout d'abord on doit lancer Protégé puis définir l'ONTOLOGY IRI, pour notre ontologie on a défini l'URI : ontologies.

Dans le menu système FILE/SAVE , on indique le format de sauvegarde de l'ontologie, dans notre cas on a considéré RDF/XML

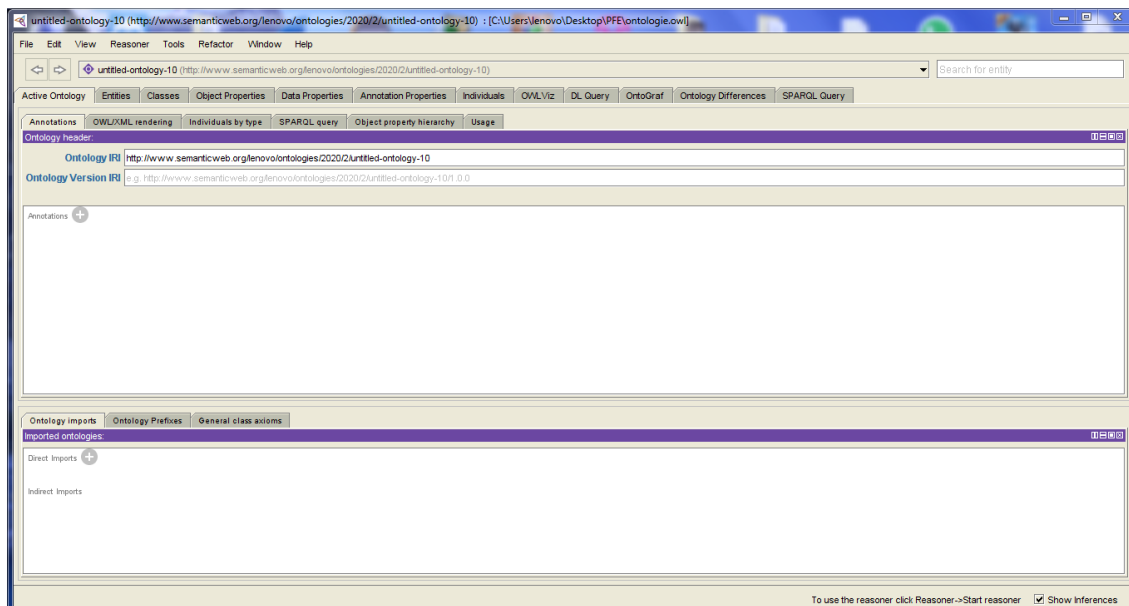


FIGURE 4.1 – Ecran d'affichge1

On choisit l'onglet classes, on clique sur le mot « Thing » dans la vue « class hierarchy » qui est la racine de toutes les classes qu'on a créé. En appuyant sur le premier bouton en haut à gauche de cette vue, un dialogue apparait qui permet de créer une classe qui sera un enfant de la classe sélectionnée. On peut ainsi définir des sous-classes. Le deuxième bouton permet de créer une classe soeur de

la classe sélectionnée. Le troisième bouton détruit la classe sélectionnée ainsi que ses sous-classes.

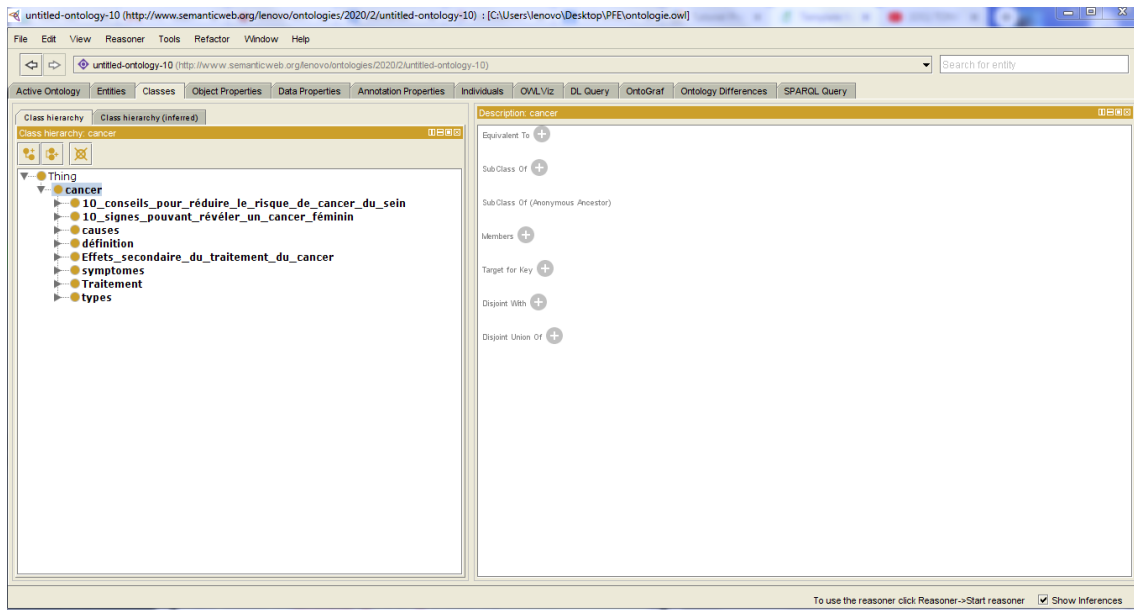


FIGURE 4.2 – ajout des classes et sous classes

Pour définir la relation entre chaque deux classes ou plus (les verbes) on appuie sur l'onglet « Object properties », on clique sur le mot « topObjectProperty » qui est la racine de tous les verbes qu'on a créé .

Dans la fenêtre « Description », on appuie sur + de l'item « Domains (intersection) » pour définir la classe concernée puis sur + de l'item « Ranges (Intersection) » pour définir la classe qui a une relation avec la première.

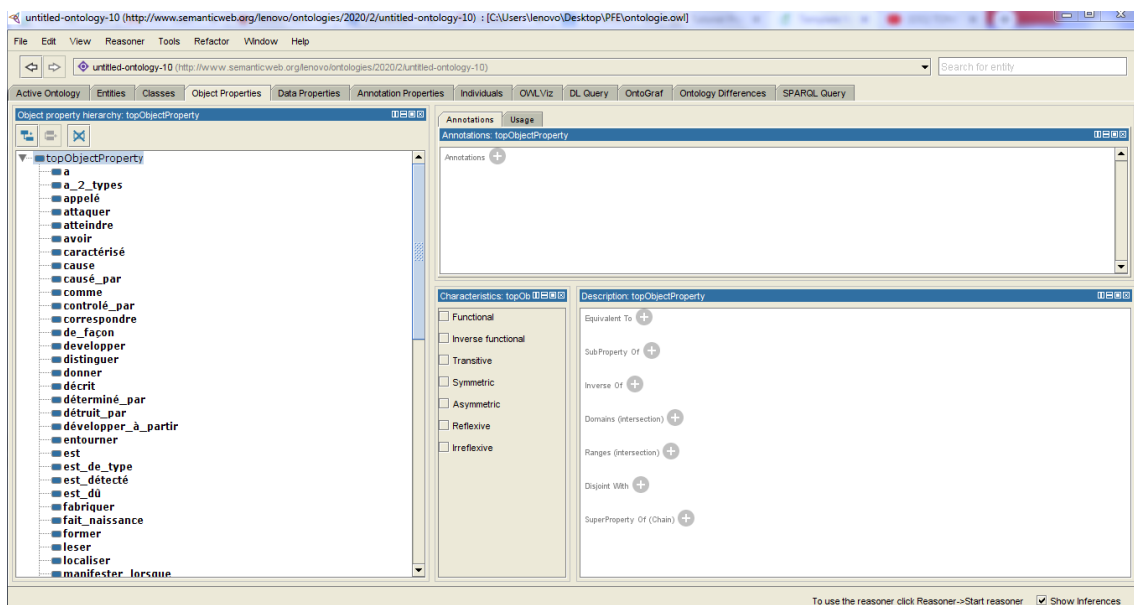


FIGURE 4.3 – définition des propriétés

Pour instancier nos classes, dans l'onglet « Individuals » on appuie le bouton à gauche de la vue « Individuals » pour ajouter les instances puis on revient aux

nos classes, et on sélectionne une.

Après on appuie sur + de l'item « Members » dans la vue « Description » de l'onglet classes, et on choisit les individus correspondants qu'on veut ajouter.

Dans notre cas on a instancier nos classes avec les différents diagnostics des types de cancer.

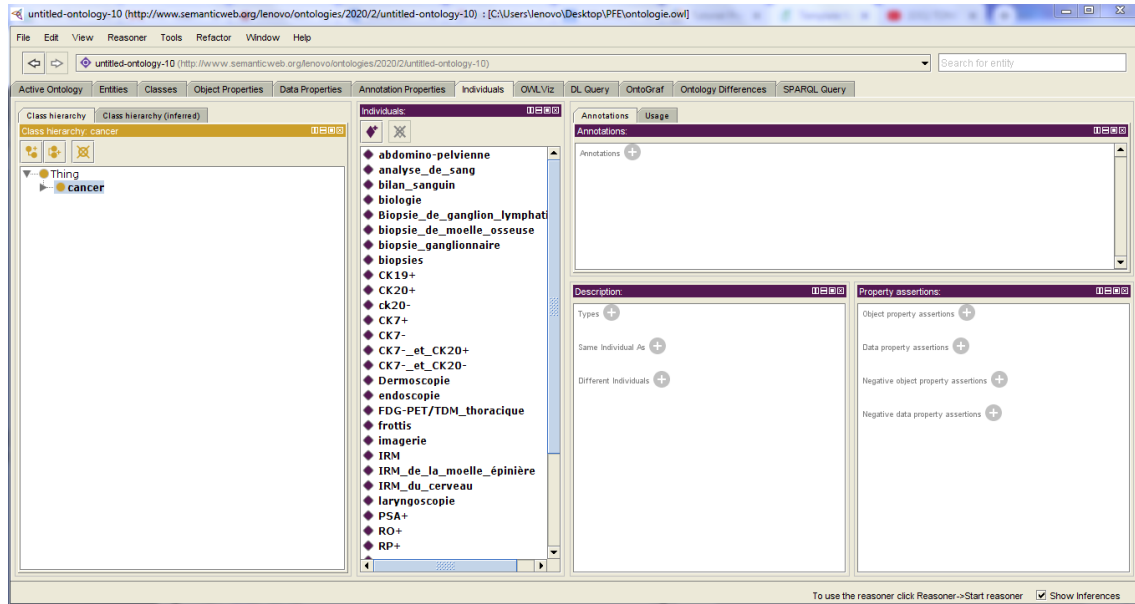


FIGURE 4.4 – ajout d'individues

Et par cette manière on a créé notre ontologie.

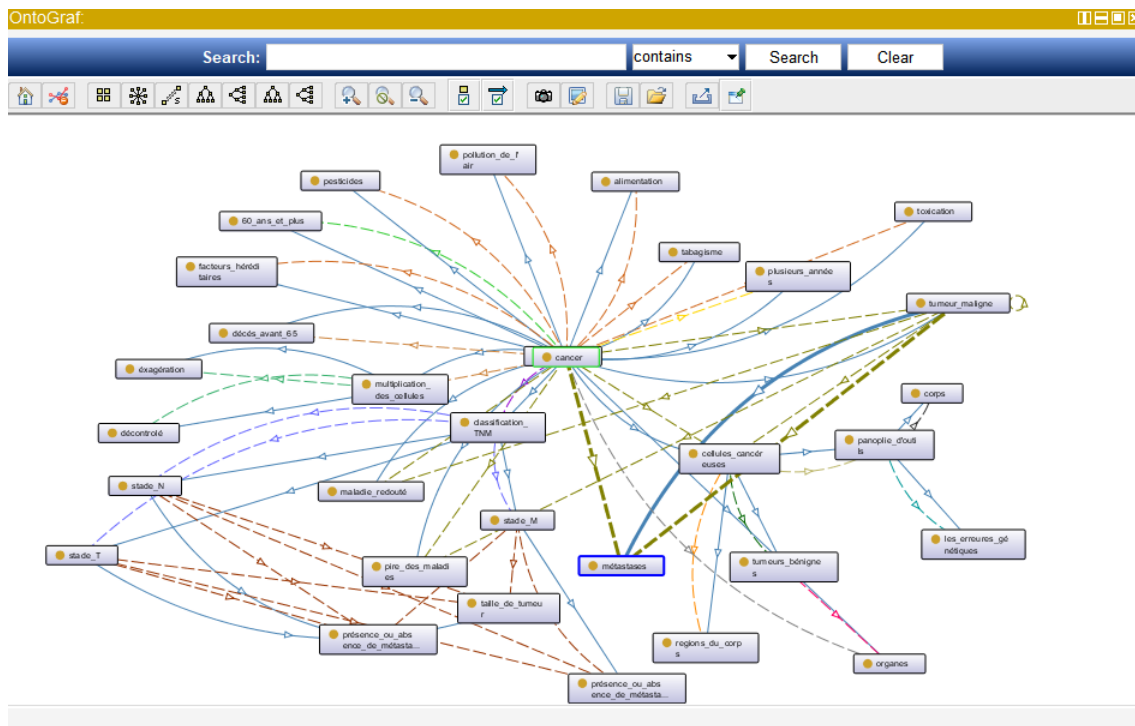


FIGURE 4.5 – ontologie

4.2 Etape2 : Forum de Discussion

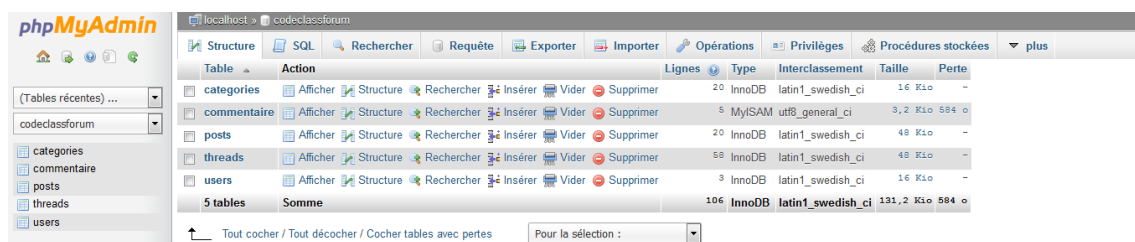
4.2.1 Définition du Forum de Discussion

En informatique, un forum est un espace de discussion publique (ou au moins ouvert à plusieurs participants). Les discussions y sont archivées ce qui permet une communication asynchrone (c'est ce qui différencie les forums de la messagerie instantanée). Il y a deux sortes de forum, en fonction du classement des messages : soit les « forums de discussion » dont les messages sont classés par date chronologique, soit les « forums de questions / réponses » dont les messages sont classés par votes. Forum est un terme d'origine latine (popularisé par l'anglais) désignant une place de la ville consacrée à la discussion et au commerce.

On regroupe maintenant sous ce thème Usenet qui existait déjà avant l'apparition d'Internet, et les forums Web qui ont accompagné le développement du Web dynamique. On peut aussi considérer les listes de diffusion – qui sont composées de courriels – comme étant des forums.

4.2.2 Création du Forum de Discussion

Tous d'abord on a créé une base de données avec phpMyAdmin pour que chaque détail ajouté au forum soit enregistré



The screenshot shows the phpMyAdmin interface for a database named 'codeclassforum'. The left sidebar lists the tables: categories, commentaire, posts, threads, and users. The main panel displays the structure of these tables with columns for Lignes, Type, Interclassement, Taille, and Perte.

Table	Action	Lignes	Type	Interclassement	Taille	Perte
categories	Afficher Structure Rechercher Insérer Vider Supprimer	20	InnoDB	latin1_swedish_ci	16 Kio	-
commentaire	Afficher Structure Rechercher Insérer Vider Supprimer	5	MyISAM	utf8_general_ci	3,2 Kio	584 o
posts	Afficher Structure Rechercher Insérer Vider Supprimer	20	InnoDB	latin1_swedish_ci	48 Kio	-
threads	Afficher Structure Rechercher Insérer Vider Supprimer	58	InnoDB	latin1_swedish_ci	48 Kio	-
users	Afficher Structure Rechercher Insérer Vider Supprimer	3	InnoDB	latin1_swedish_ci	16 Kio	-
5 tables	Somme	106	InnoDB	latin1_swedish_ci	131,2 Kio	584 o

FIGURE 4.6 – les tables de la base de donnée

-Categories : sont les grands titres qui précisent chaque partie du forum pour qu'il soit bien organisé

cat_id	cat_name	cat_description
4	CANCER	
7	cancer de sein	
8	Conseils pour reduire le risques du cancer de sein	
9	Cancer de cerveau	
10	cancer de l'endometre	
11	Cancer de l'estomac	
12	Cancer de l'oesophage	
13	Cancer de la gorge	
14	cancer de la peau	
15	Cancer de la thyroide	
16	Cancer de la vessie	
17	Cancer de testicules	
18	Cancer du col de l'uterus	
19	Cancer du colon (cancer colorectal)	
20	Cancer du foie	
21	Cancer du pancreas	
22	Cancer du poudmon	
23	Cancer du prostate	
24	Cancer du systeme lymphotique	
25	Signes pouvant reveler un cancer feminin	

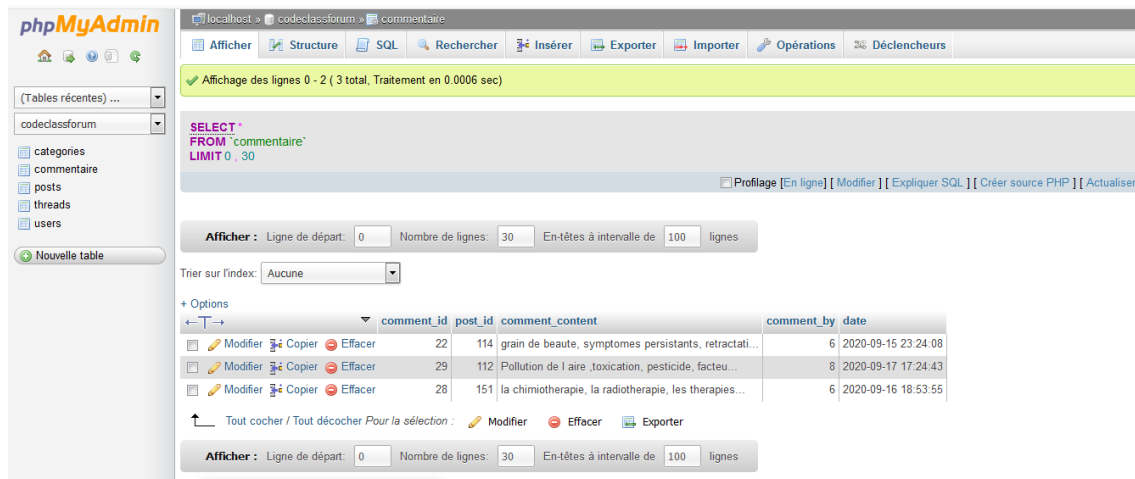
FIGURE 4.7 – categories

-Exemple : cancer et cancer de sein sont des catégories

Category	Sub-category	Posts
CANCER	Definition du cancer	2
	Un des pires des maladies	2
	Causes	2
	le cancer a plusieurs causes	2
	Symptômes	2
cancer de sein	Traitement	1
	Effets secondaires du traitement du cancer	1
	Causes	1
	Symptômes	1
		1

FIGURE 4.8 – exemple1

-Commentaire : sont les réponses de chaque question posé sur le forum



phpMyAdmin

localhost » codeclassforum » commentaire

Afficher Structure SQL Rechercher Insérer Exporter Importer Opérations Déclencheurs

✓ Affichage des lignes 0 - 2 (3 total, Traitement en 0.0006 sec)

SELECT *
FROM `commentaire`
LIMIT 0 , 30

Profilage [En ligne] [Modifier] [Expliquer SQL] [Créer source PHP] [Actualiser]

Afficher : Ligne de départ: 0 Nombre de lignes: 30 En-têtes à intervalle de 100 lignes

Trier sur l'index: Aucune

+ Options

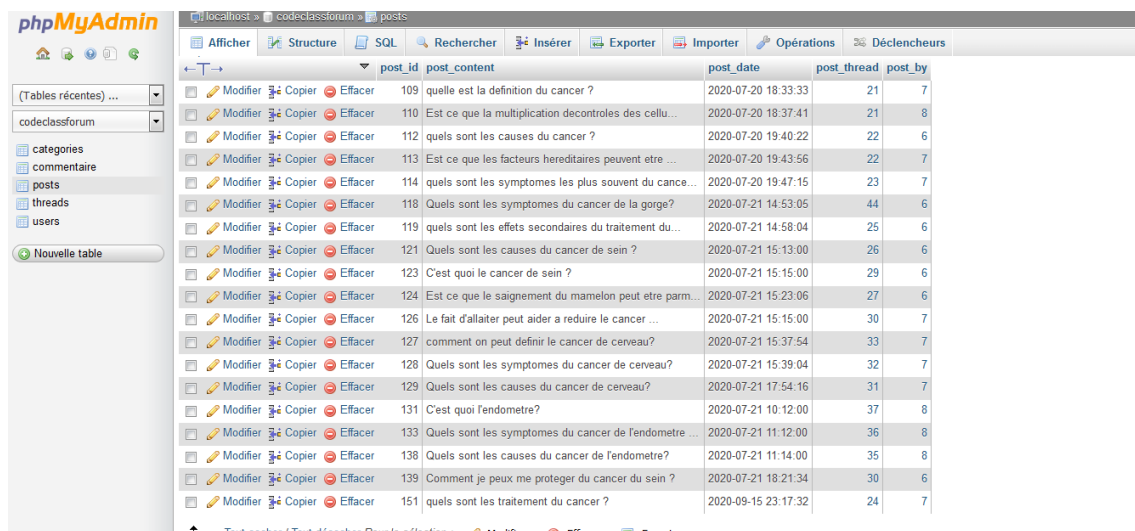
	comment_id	post_id	comment_content	comment_by	date
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	22	114	grain de beauté, symptômes persistants, retractati...	6	2020-09-15 23:24:08
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	29	112	Pollution de l'aire ,toxication, pesticide, facteu...	8	2020-09-17 17:24:43
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	28	151	la chimiothérapie, la radiothérapie, les therapies...	6	2020-09-16 18:53:55

Tout cocher / Tout décocher Pour la sélection : ☐ Modifier ☐ Effacer ☐ Exporter

Afficher : Ligne de départ: 0 Nombre de lignes: 30 En-têtes à intervalle de 100 lignes

FIGURE 4.9 – commentaire

-Posts : sont les postes de chaque utilisateurs, dans notre cas se sont les questions posée concernant le cancer. Chacune est stocké dans la base de donnée.



phpMyAdmin

localhost » codeclassforum » posts

Afficher Structure SQL Rechercher Insérer Exporter Importer Opérations Déclencheurs

commentaire

posts

Nouvelle table

	post_id	post_content	post_date	post_thread	post_by
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	109	quelle est la définition du cancer ?	2020-07-20 18:33:33	21	7
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	110	Est ce que la multiplication decontroles des cellu...	2020-07-20 18:37:41	21	8
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	112	quels sont les causes du cancer ?	2020-07-20 19:40:22	22	6
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	113	Est ce que les facteurs hereditaires peuvent etre ...	2020-07-20 19:43:56	22	7
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	114	quels sont les symptômes les plus souvent du cance...	2020-07-20 19:47:15	23	7
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	118	Quels sont les symptômes du cancer de la gorge?	2020-07-21 14:53:05	44	6
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	119	quels sont les effets secondaires du traitement du...	2020-07-21 14:58:04	25	6
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	121	Quels sont les causes du cancer de sein ?	2020-07-21 15:13:00	26	6
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	123	C'est quoi le cancer de sein ?	2020-07-21 15:15:00	29	6
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	124	Est ce que le saignement du mamelon peut etre parm...	2020-07-21 15:23:06	27	6
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	126	Le fait d'allaiter peut aider a reduire le cancer ...	2020-07-21 15:15:00	30	7
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	127	comment on peut definir le cancer de cerveau?	2020-07-21 15:37:54	33	7
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	128	Quels sont les symptômes du cancer de cerveau?	2020-07-21 15:39:04	32	7
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	129	Quels sont les causes du cancer de cerveau?	2020-07-21 17:54:16	31	7
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	131	C'est quoi l'endometre?	2020-07-21 10:12:00	37	8
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	133	Quels sont les symptômes du cancer de l'endometre ...	2020-07-21 11:12:00	36	8
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	138	Quels sont les causes du cancer de l'endometre?	2020-07-21 11:14:00	35	8
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	139	Comment je peux me proteger du cancer du sein ?	2020-07-21 18:21:34	30	6
<input type="checkbox"/> Modifier <input type="checkbox"/> Copier <input type="checkbox"/> Effacer	151	quels sont les traitement du cancer ?	2020-09-15 23:17:32	24	7

Tout cocher / Tout décocher Pour la sélection : ☐ Modifier ☐ Effacer ☐ Exporter

FIGURE 4.10 – posts

-Threads :sont les sous-titres de chaque catégorie, pour bien préciser le besoin de la question d'un utilisateur

thread_id	thread_name	thread_date	thread_desc	original_post	cat_id	created_by
21	Definition du cancer	2020-07-08 19:58:03	Un des pires des maladies		4	6
22	Causes	2020-07-20 14:56:59	le cancer a plusieurs causes		4	6
23	Symptômes	2020-07-20 15:04:01			4	6
24	Traitement	2020-07-20 15:05:04			4	6
25	Effets secondaires du traitement du cancer	2020-07-20 15:06:07			4	6
26	Causes	2020-07-20 17:28:56			7	6
27	Symptômes	2020-07-20 17:29:55			7	6
29	Definition	2020-07-20 17:32:57			7	6
30	Conseils	2020-07-20 17:33:35			8	6
31	Causes	2020-07-20 17:34:07			9	6
32	Symptomes	2020-07-20 17:34:21			9	6
33	Definition	2020-07-20 17:34:39			9	6
34	Causes	2020-07-20 17:37:53			11	6
35	Causes	2020-07-20 17:38:12			10	6
36	Symptômes	2020-07-20 17:38:28			10	6
37	Definition	2020-07-20 17:38:36			10	6
38	Symptômes	2020-07-20 17:41:25			11	6
39	Definition	2020-07-20 17:41:34			11	6
40	Causes	2020-07-20 18:09:54			12	6

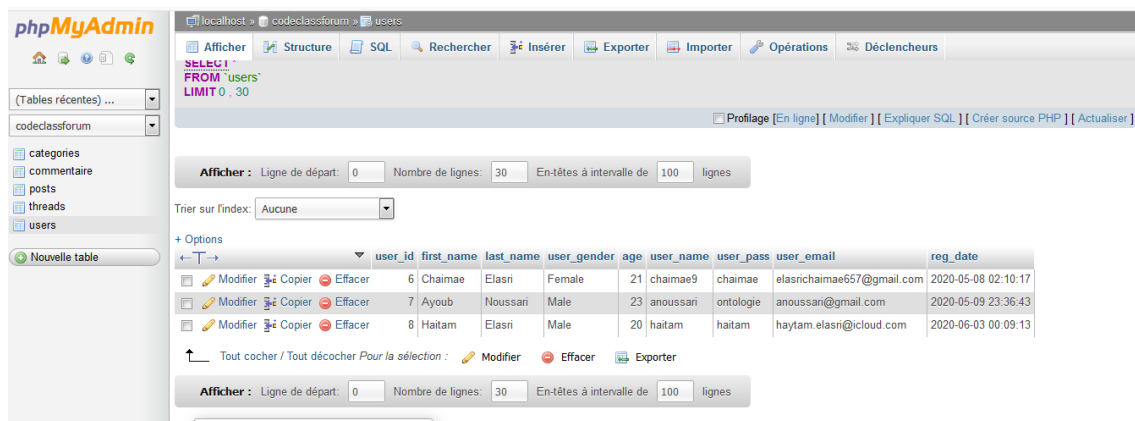
FIGURE 4.11 – Threads

-Exemple : « causes , symptômes et Definition » sont des threads de la catégorie « Cancer de l'endomètre » ainsi que le nombre de postes dans chacune

cancer de l'endomètre	
Causes	1 Posts
Symptômes	1 Posts
Definition	1 Posts

FIGURE 4.12 – exemple2

-Users : sont les utilisateurs enregistré et qui ont le droit de poster dans le forum



user_id	first_name	last_name	user_gender	age	user_name	user_pass	user_email	reg_date
6	Chaimae	Elasri	Female	21	chaimae9	chaimae	elasrichaimae657@gmail.com	2020-05-08 02:10:17
7	Ayoub	Noussari	Male	23	anoussari	ontologie	anoussari@gmail.com	2020-05-09 23:36:43
8	Haitam	Elasri	Male	20	haitam	haitam	haytam.elasri@icloud.com	2020-06-03 00:09:13

FIGURE 4.13 – users

-la page d'accueil de notre forum est comme si dessous :

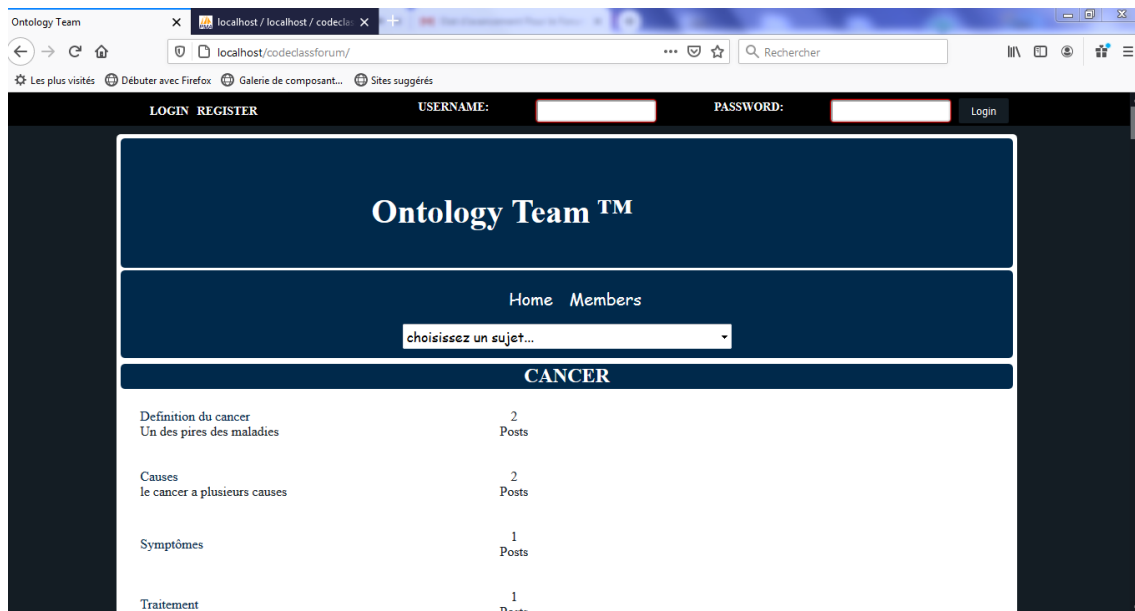


FIGURE 4.14 – Home

-Home est la page d'accueil qui apparait dans la figure 4.14, affiche tous les catégories et leurs threads ainsi que le nombre de poste de chaque sous-titre

-Members : redirige vers les membres du forum et il y'a aussi le nombre de postes de chaque utilisateur ainsi que la date d'enregistrement , comme on peut chercher un utilisateur dans la liste

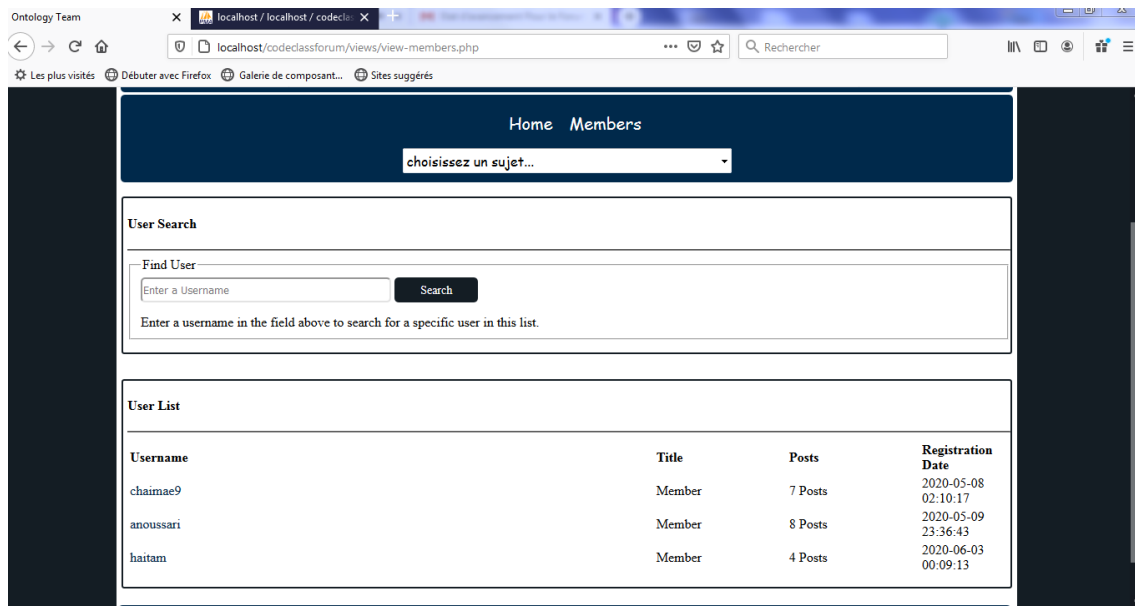


FIGURE 4.15 – Members

-Liste déroulante : avec le titre « choisissez un sujet... » où choisir la catégorie correspondante pour poser la question, dans notre cas se sont les types de cancer, càd chaque catégorie est un type de cancer. Après choisir le sujet, la page va glisser vers le choix sélectionné. Comme ça l'utilisateur va savoir que la page d'accueil contient plusieurs catégories

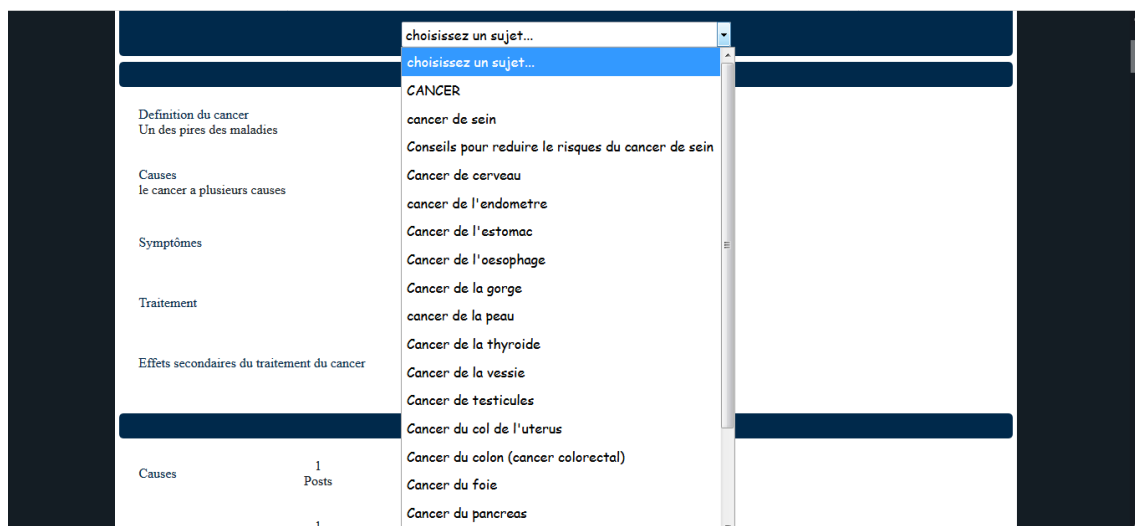
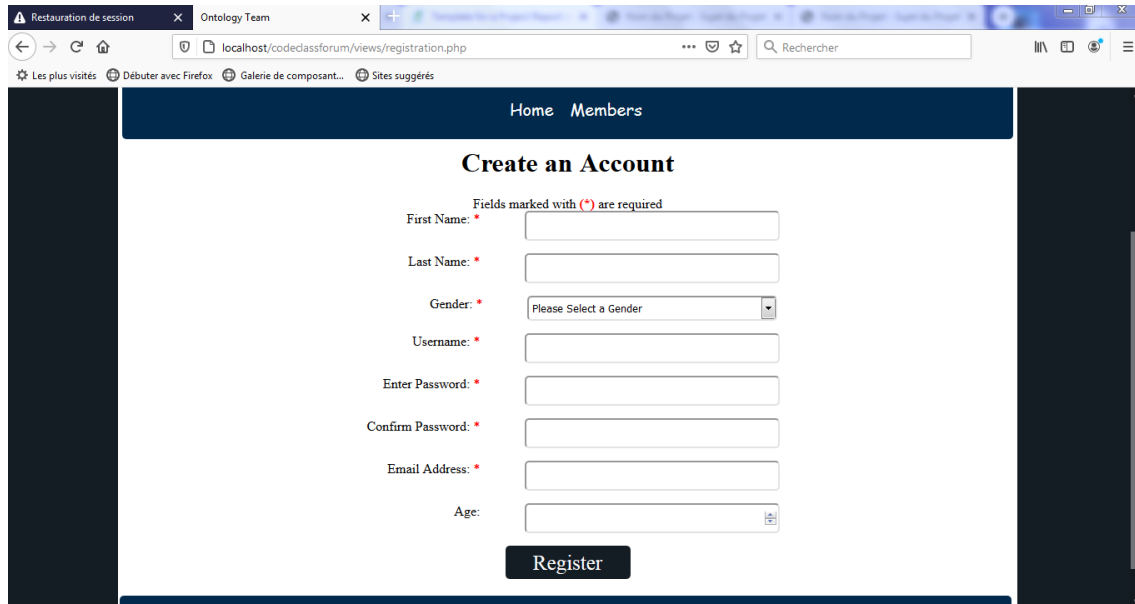


FIGURE 4.16 – Liste déroulante

On a utilisé le CSS pour décorer notre forum et aussi le java script pour les messages d'erreurs.

4.3 Etape3 : Poser les Questions

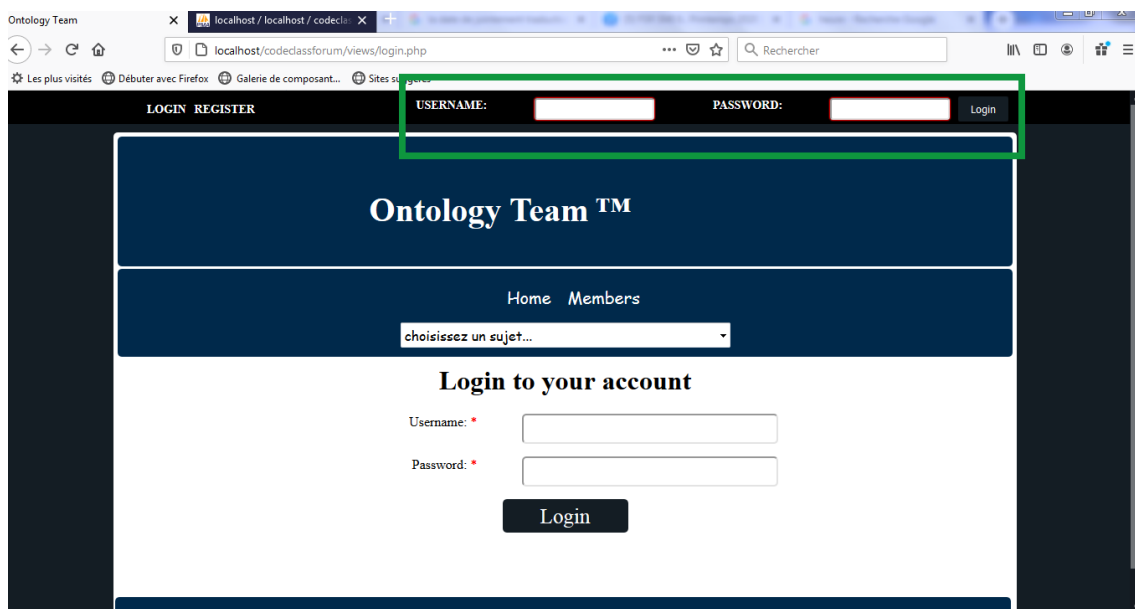
-Pour poser une question sur notre forum tous d'abord l'utilisateur doit créer un compte dans la partie « REGISTER » qui se trouve dans la partie supérieur à gauche qui va lui rediriger vers une page où se trouve le formulaire de registration.



The screenshot shows a web browser window with the URL `localhost/codeclassforum/views/registration.php`. The page has a dark blue header with 'Home' and 'Members' links. Below the header, the title 'Create an Account' is centered. A note states 'Fields marked with (*) are required'. The form includes the following fields: 'First Name:', 'Last Name:', 'Gender:' (a dropdown menu with 'Please Select a Gender'), 'Username:', 'Enter Password:', 'Confirm Password:', 'Email Address:', and 'Age:'. A 'Register' button is located at the bottom of the form.

FIGURE 4.17 – Register

-Après cliquer sur « Register » l'utilisateur crée son compte, il peut se connecter soit directement dans la partie supérieur à droite et écrire dans les champs « USERNAME » et « PASSWORD » ses informations, soit dans « LOGIN » qui se trouve dans la partie supérieur à droite qui va rediriger vers une page et remplir le formulaire avec les informations d'utilisateurs.



The screenshot shows a web browser window with the URL `localhost/codeclassforum/views/login.php`. The page has a dark blue header with 'LOGIN' and 'REGISTER' links. Below the header, the title 'Login to your account' is centered. A green box highlights the 'USERNAME:' and 'PASSWORD:' fields, which are input boxes. A 'Login' button is located at the bottom of the form. The page also features the 'Ontology Team' logo and a 'Home' link.

FIGURE 4.18 – Login

-Maintenant l'utilisateur est connecté , on voit en haut à droite une petite image , «WELCOME » username et « SIGN OUT » pour déconnecter, donc il peut poster sur le forum.

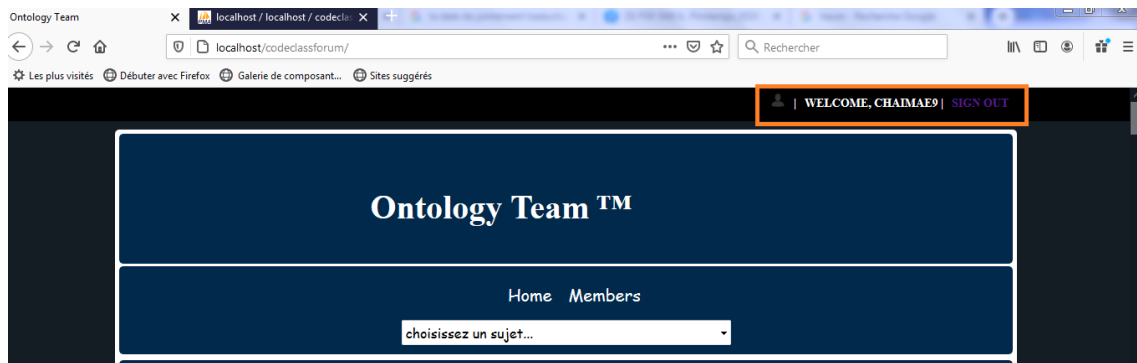


FIGURE 4.19 – Welcome

-d'après le besoin de la question, l'utilisateur se redirige vers la page correspondante où se trouve une zone de texte « Post a Reply » où écrire et le bouton « Post » pour poster.

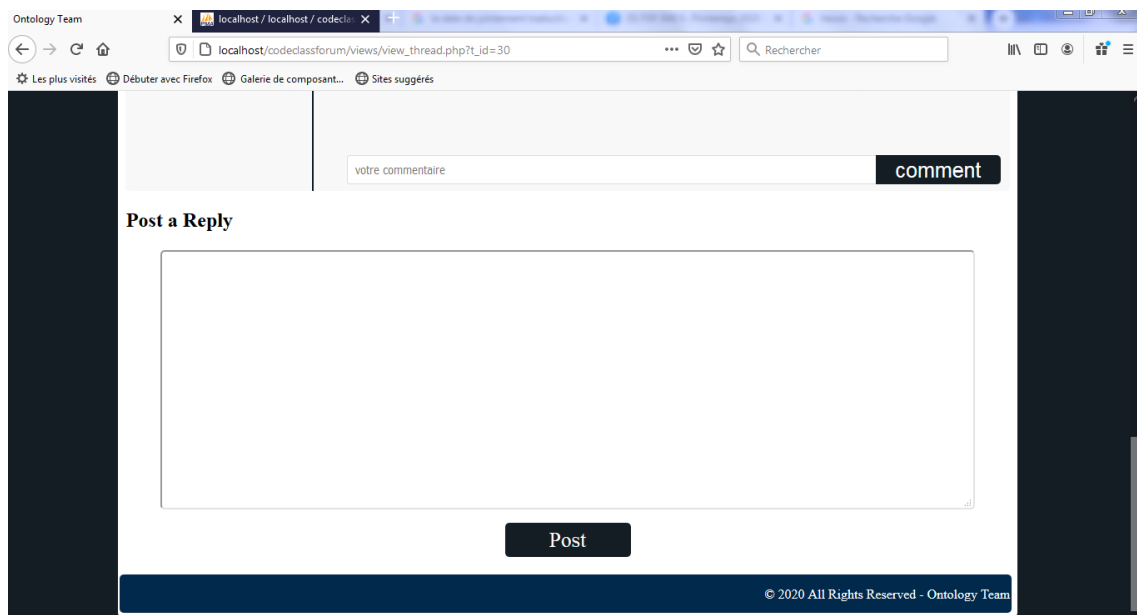


FIGURE 4.20 – poster

-Exemple d'une question posée, on trouve des informations en haut du poste (l'utilisateur qui a posté, la date et l'heure) et à gauche on trouve une image profile et au dessous il y'a username de l'utilisateur, sa position, nombre de poste dans le forum ainsi que la date et l'heure de jointement.



FIGURE 4.21 – question

-il existe deux images profiles : une pour un utilisateur male et l'autre pour femme.

4.4 Etape4 :LSA "Latent Semantic Analysis

4.4.1 Définition du LSA

Analyse sémantique latente (LSA) est une technique dans le traitement du langage naturel , en particulier la sémantique de répartition , de l' analyse des relations entre un ensemble de documents et les termes qu'ils contiennent en produisant un ensemble de concepts en rapport avec les documents et les termes. LSA suppose que les mots qui sont proches de sens se produiront dans des morceaux de texte similaire (l' hypothèse distributive). Une matrice contenant mot compte par le paragraphe (lignes représentent des mots uniques et les colonnes représentent chaque paragraphe) est construit à partir d' un grand morceau de texte et une technique mathématique appelée décomposition en valeurs singulières (SVD) est utilisé pour réduire le nombre de lignes tout en préservant la structure de similitude entre les colonnes. Les mots sont alors comparées en prenant le cosinus de l'angle entre les deux vecteurs (ou le produit scalaire entre les normalisations des deux vecteurs) formées par deux rangées. Les valeurs proches de 1 représentent des mots très similaires alors que les valeurs proches de 0 représentent des mots très dissemblables.

Application :

-LSA pourrait être utilisé pour extraire des résumés textuels à partir de docu-

ments texte ou même des descriptions de produits . Il peut s'agir de résumer les descriptions de produits, de rapports médicaux non structurés ou même de résumer un résumé.

-Les modèles de sujets sont construits autour de l'idée que la sémantique de notre document est en fait régie par des sujets cachés ou qui façonnent le sens de notre document et de notre corpus. LSA avec SVD peut aider à la modélisation de sujets sur un corpus de texte.

-LSA et SVD sont utilisés comme précurseurs pour trouver des similitudes entre différents mots, différents documents ou des comparaisons sur des requêtes sur des documents qui se font en appliquant la similarité cosinus . Cela pourrait être exploité dans les systèmes de référencement et de recommandation.

4.4.2 code du LSA en PYTHON

-Gensim est une bibliothèque Python pour la modélisation de sujets , l'indexation de documents et la recherche de similitudes avec de grands corpus. Le public cible est la communauté du traitement du langage naturel (PNL) et de la recherche d'informations (IR).

-NLTK est un package Python puissant qui fournit un ensemble d'algorithmes de langages naturels divers. C'est gratuit, open source, facile à utiliser, large communauté et bien documenté. NLTK comprend les algorithmes les plus courants tels que la création de jetons, le marquage d'une partie du discours, la recherche de racines, l'analyse des sentiments, la segmentation de sujets et la reconnaissance d'entités nommées. NLTK aide l'ordinateur à analyser, prétraiter et comprendre le texte écrit.

```
1 from gensim import corpora, models, similarities
2 from nltk.corpus import stopwords
```

FIGURE 4.22 – gensim

- la variable document prend deux chaînes de caractères, la première est notre bagage concernant le cancer , et la deuxième c'est la question qu'on a besoin traiter.

```
2 documents = [" cancer causes definition traitement symptomes types cerveau sein l'endomètre l'estomac l'oésophage gorgi",
3 "quels sont les symptomes du cancer ?",
4 ]
```

FIGURE 4.23 – document

-Éliminer les stopwords, rendre tous les mots en minuscule « lower() », et aussi séparer chaque mot de l'autre « split() »

```
stoplist=set(stopwords.words('french'))
texts = [[word for word in document.lower().split() if word not in stoplist] for document in documents]
print("remove stopwords")
print(texts)
```

FIGURE 4.24 – élimination des stop words

-Afficher les mots répétés dans les deux chaînes de caractères, et le nombre de chaque mot ainsi que son identifiant.

```
18 texts = [[word for word in text if word not in tokens_once]
19           for text in texts]
20 print("repeated word")
21 print(texts)
22 dictionary = corpora.Dictionary(texts)
23 print("nombre de mots")
24 print(dictionary)
25 # save as binary file at the dictionary at local directory
26 dictionary.save('deerwester.dict')
27 # save as text file at the local directory
28 dictionary.save_as_text('deerwester_text.dict')
29 # show pairs of "word : word-ID number"
30 print("mot avec son identifiant")
31 print(dictionary.token2id)
```

FIGURE 4.25 – les mots répétés

-Les mots trouvés dans la variable "dictionary" par LSA sont les mots les plus pertinents, c'est à dire les mots à retenir d'après la question posée sur le Forum

```
Dictionary(2 unique tokens: ['cancer', 'symptomes'])
```

FIGURE 4.26 – exemple de résultat

4.5 Etape5 : SPARQL

4.5.1 Définition

SPARQL est un langage de requête et un protocole qui permet de rechercher, d'ajouter, de modifier ou de supprimer des données RDF disponibles à travers Internet. Son nom est un acronyme récursif qui signifie SPARQL Protocol and RDF Query Language.

SPARQL est considéré dès 2007 comme l'une des technologies clés du Web sémantique par Tim Berners-Lee l'inventeur du Web sémantique qui explique « Tenter d'utiliser le Web sémantique sans SPARQL revient à exploiter une base de données relationnelle sans SQL ».

Aujourd'hui, le Web des données (ou Linked Open Data) est constitué de centaines de services SPARQL qui mettent à disposition de plus en plus de données

au travers d'Internet. L'ambition du W3C est d'offrir une interopérabilité non pas seulement au niveau des services, comme avec les services Web, mais également au niveau des données.

4.5.2 Exécution des requêtes SPARQL

Après avoir les mots les plus pertinents à partir su LSA, on est donc en courant du besoin de la question,alors on doit cherché la réponse.

Comme on a cité au début, l'ontologie est une graphe ou on a stockés tous les informations nécessaires .

SPARQL, le langage des graphe c'est notre outil de recherche, et qui est disponible sur PROTEGE dans l'onglet "SPARQL query", et on écrit les requêtes correspondante à la question.

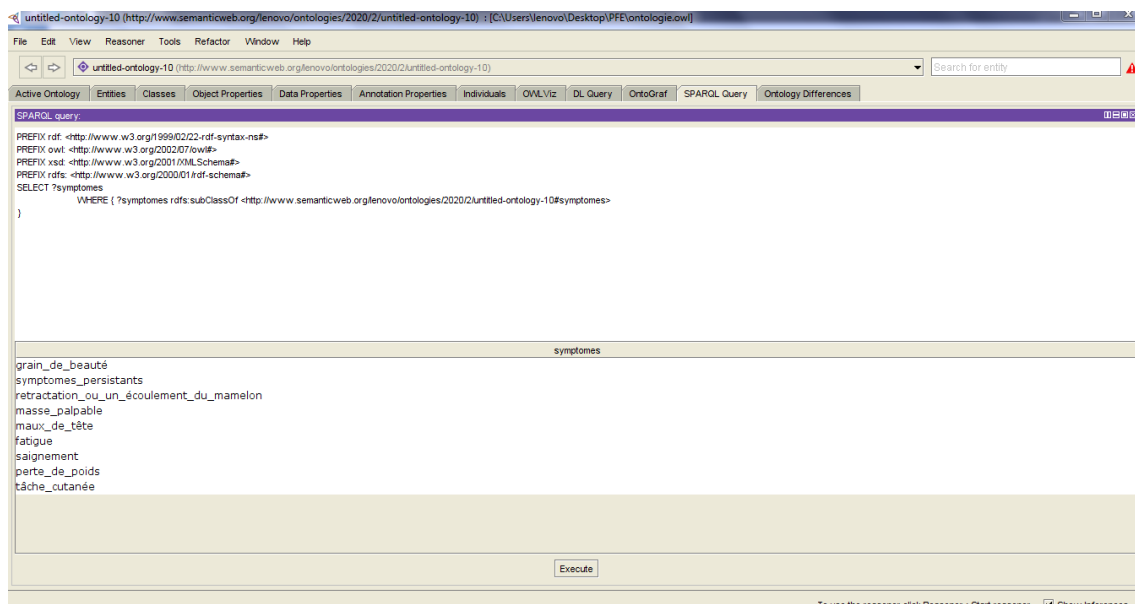


FIGURE 4.27 – SPARQL

4.6 Etape6 : Réponse à la question

-Le résultat de l'exécution de la requête SPARQL est exactement la réponse de la question.

-SPARQL fait sa recherche dans l'ontologie et extraire la relation entre chaque classe.

-Pour répondre sur le forum on doit écrire la réponse dans la partie dédiée, sous chaque question on trouve une zone de texte avec le titre "Votre commentaire" où écrire la réponse et un bouton "commenter" pour poster la réponse .

-On peut ajouter plusieurs commentaires, et à côté on trouve username d'utilisateur , la date et l'heure du commentaire.

-Si on a plusieurs commentaires sous le même poste on a ajouté une barre de défilement pour faire glisser les commentaires et toujours le dernier commentaire qui va être visible,càd la barre de défilement est toujours positionnée en bas.

-L'utilisateur doit être connecté pour ajouter des commentaires si ce n'est pas le cas, zone du texte et bouton "commenter" ne seront pas afficher à l'utilisateur.



FIGURE 4.28 – user connecté

- le cas où l'utilisateur n'est pas connecté ou n'a pas de compte.

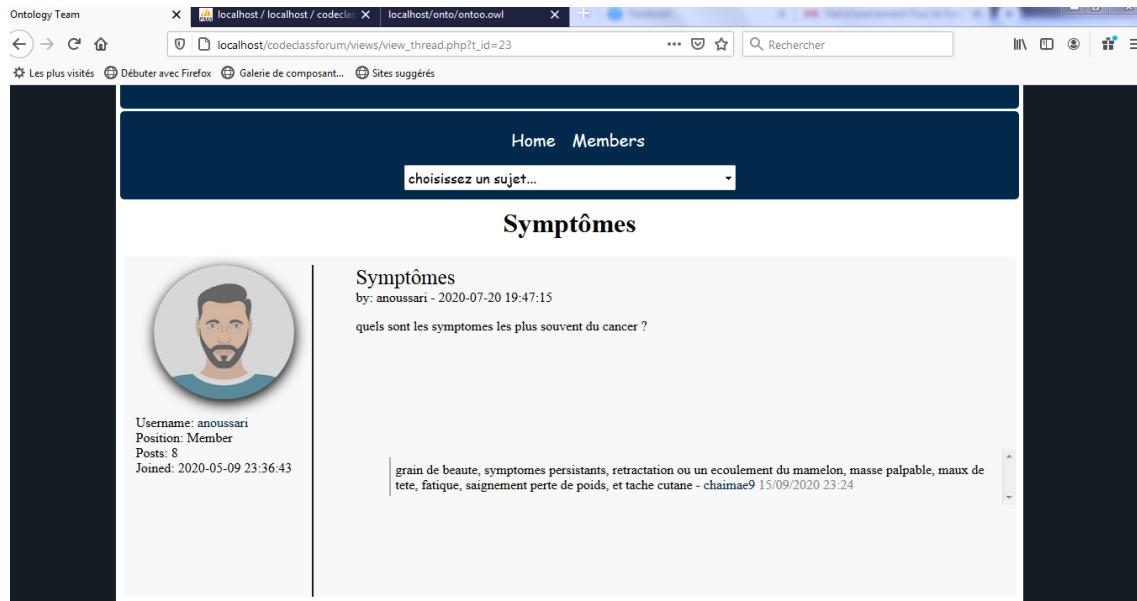


FIGURE 4.29 – user non connecté

4.7 Automatisation

Tous les étapes qu'on a cité ont été traité manuellement, Maintenant on veut réaliser le traitement automatiquement c'est-à-dire chercher la réponse et la poster sur le forum d'une façon automatique.

4.7.1 Web Scraping

4.7.1.1 Définition du web scraping

Recueillir des données sur le web est parfois compliqué et quand cela est possible, il est difficile de pouvoir les télécharger ou d'effectuer un copier-coller. Le web scraping est une technique permettant l'extraction des données d'un site via un programme, un logiciel automatique ou un autre site. L'objectif est donc d'extraire le contenu d'une page d'un site de façon structurée. Le scraping permet ainsi de pouvoir réutiliser ces données.

Dans quel cas utiliser le web scraping ?

L'intérêt principal du web scraping est de pouvoir récolter du contenu sur un site web, qui ne peut être copié collé sans dénaturer la structure même du document. Ainsi cette technique est souvent utilisée dans le cadre d'une veille concurrentielle, notamment sur des sites e-commerce.

-Dans notre cas on veut extraire les questions posés sur le forum pour les traiter.

4.7.1.2 Code du web scraping en python

-Le requests module vous permet d'envoyer des requêtes HTTP en utilisant Python.

La requête HTTP renvoie un objet de réponse avec toutes les données de réponse (contenu, codage, état, etc.).

-Beautiful Soup (littéralement « Belle Soupe ») est une bibliothèque Python d'analyse syntaxique de documents HTML et XML créée par Leonard Richardson.

Elle produit un arbre syntaxique qui peut être utilisé pour chercher des éléments ou les modifier. Lorsque le document HTML ou XML est mal formé (par exemple s'il manque des balises fermantes), Beautiful Soup propose une approche à base d'heuristiques afin de reconstituer l'arbre syntaxique sans générer d'erreurs. Cette approche est aussi utilisée par les navigateurs web modernes.

Donc elle permet d'extraire des informations d'un site web, ou encore d'un document XML



```
webS.py > ...
1  import requests
2  from bs4 import BeautifulSoup
3  for id in range(21,80):
4      url = 'http://localhost/codeclassforum/views/view_thread.php?t_id='+ str(id)
5      r = requests.get(url)
6      if r.ok:
7          s = BeautifulSoup(r.text, 'lxml')
8          ps = s.findAll('p')
9          for p in ps:
10             t = str(p.text)+'\n'
11             with open(r'C:\Users\lenovo\Desktop\PFE\web.txt','a') as filout:
12                 filout.write(t)
13
```

FIGURE 4.30 – code en python

-On importe les bibliothèques nécessaire.

-url est variable qui prend le lien de la page ou se trouve les questions, chaque page a son propre id.

-Les questions sont écrit dans la balise html <p>.

-le retour du code est stocké dans un fichier texte "web.txt".

```

C: > Users > lenovo > Desktop > PFE > web.txt
1  quelle est la definition du cancer ?
2
3  Est ce que la multiplication decontroles des cellules peut reveler un cancer ?
4
5  quels sont les causes du cancer ?
6
7  Est ce que les facteurs hereditaires peuvent etre les causes du cancer ?
8
9  quels sont les symptomes les plus souvent du cancer ?
10
11  quels sont les traitement du cancer ?
12
13  quels sont les effets secondaires du traitement du cancer ?
14
15  Quels sont les causes du cancer de sein ?
16
17  Est ce que le saignement du mamelon peut etre parmi les symptomes du cancer du sein ?
18
19  C'est quoi le cancer de sein ?
20
21  Le fait d'allaiter peut aider a reduire le cancer du sein ?
22
23  Comment je peux me proteger du cancer du sein ?
24
25  Quels sont les causes du cancer de cerveau ?
26

```

FIGURE 4.31 – web.txt

4.7.2 LSA

le code LSA traite les questions.

on doit ouvrir le fichier "web.txt" en mode lecture pour lire les questions et les traiter

```

Users > lenovo > Desktop > PFE > LSI.py > ...
from gensim import corpora, models, similarities
from nltk.corpus import stopwords
file = open(r'C:\Users\lenovo\Desktop\PFE\web.txt', "r")
for line in file.readlines():
    documents = [" cancer causes definition traitement symptomes types cerveau sein l'endometre
line,
]

```

FIGURE 4.32 – lsa1

-line est la variable qui stocke chaque question.

-Aussi le retour de LSA doit être stocké dans un autre fichier "lsa.txt" .

-la variable dictionnaire a les mots les plus pertinents donc c'est la variable qui va être stocké dans le fichier.

```

24 dictionary = corpora.Dictionary(texts)
25 print("nombre de mots")
26 print(dictionary)
27 with open(r'C:\Users\lenovo\Desktop\PFE\lsa.txt','a') as filout:
28     filout.write(str(dictionary) + '\n')

```

FIGURE 4.33 – lsa2

-le fichier "lsa.txt" après l'exécution du code LSA.

```

C: > Users > lenovo > Desktop > PFE > lsa.txt
1 Dictionary(2 unique tokens: ['cancer', 'definition'])
2 Dictionary(1 unique tokens: ['cancer'])
3 Dictionary(2 unique tokens: ['cancer', 'causes'])
4 Dictionary(2 unique tokens: ['cancer', 'causes'])
5 Dictionary(2 unique tokens: ['cancer', 'symptomes'])
6 Dictionary(2 unique tokens: ['cancer', 'traitement'])
7 Dictionary(4 unique tokens: ['cancer', 'effets', 'secondaires', 'traitement'])
8 Dictionary(3 unique tokens: ['cancer', 'causes', 'sein'])
9 Dictionary(3 unique tokens: ['cancer', 'sein', 'symptomes'])
10 Dictionary(2 unique tokens: ['cancer', 'sein'])
11 Dictionary(3 unique tokens: ['cancer', 'reduire', 'sein'])
12 Dictionary(2 unique tokens: ['cancer', 'sein'])
13 Dictionary(3 unique tokens: ['cancer', 'causes', 'cerveau'])
14 Dictionary(3 unique tokens: ['cancer', 'cerveau', 'symptomes'])
15 Dictionary(2 unique tokens: ['cancer', 'cerveau'])
16 Dictionary(3 unique tokens: ['cancer', 'causes', "l'endometre"])
17 Dictionary(3 unique tokens: ['cancer', "l'endometre", 'symptomes'])
18 Dictionary(1 unique tokens: ["l'endometre"])
19 Dictionary(3 unique tokens: ['cancer', 'gorge', 'symptomes'])

```

FIGURE 4.34 – lsa.txt

4.7.3 Perspective du projet dans l'avenir

SPARQL

Cette partie permet de manipuler le SPARQL automatiquement, c'est-à-dire exécuter des requêtes SPARQL d'une façon automatique, via un script qui prend les mots d'après le fichier texte où LSA a stocké les mots les plus pertinents : « lsa.txt ». Les requêtes cherchent dans l'ontologie et retournent la réponse, le script stocke le résultat dans un autre fichier texte pour poster la réponse à question correspondante via ce fichier.

Problème trouvé : liaison de l'ontologie avec le script,

Conclusion

Comme vous avez pu le constater durant la lecture du rapport, nous avons fait face à quelques difficultés tout au long de notre travail , mais nous avons réussi à y faire face et à les surmonter . Nous avons fait en sorte de donner le meilleur de nous pour arriver au résultat escompté. Ce projet a été très enrichissant notamment pour mettre en pratique nos connaissances déjà acquises durant les années antérieures mais aussi dans l'acquisition de nouvelles informations. Cette étude nous a offert l'opportunité de travailler avec le web sémantique qui est très à la pointe aujourd'hui . Cela nous a ouvert de nouveaux horizons et donné une idée plus concrète sur l'orientation que nous pourrions prendre, que cela soit pour nos études futures ou pour intégrer le milieu professionnel.

ANNEXE

-Exemple d'une sous-classe qui contient aussi d'autres sous-classes/

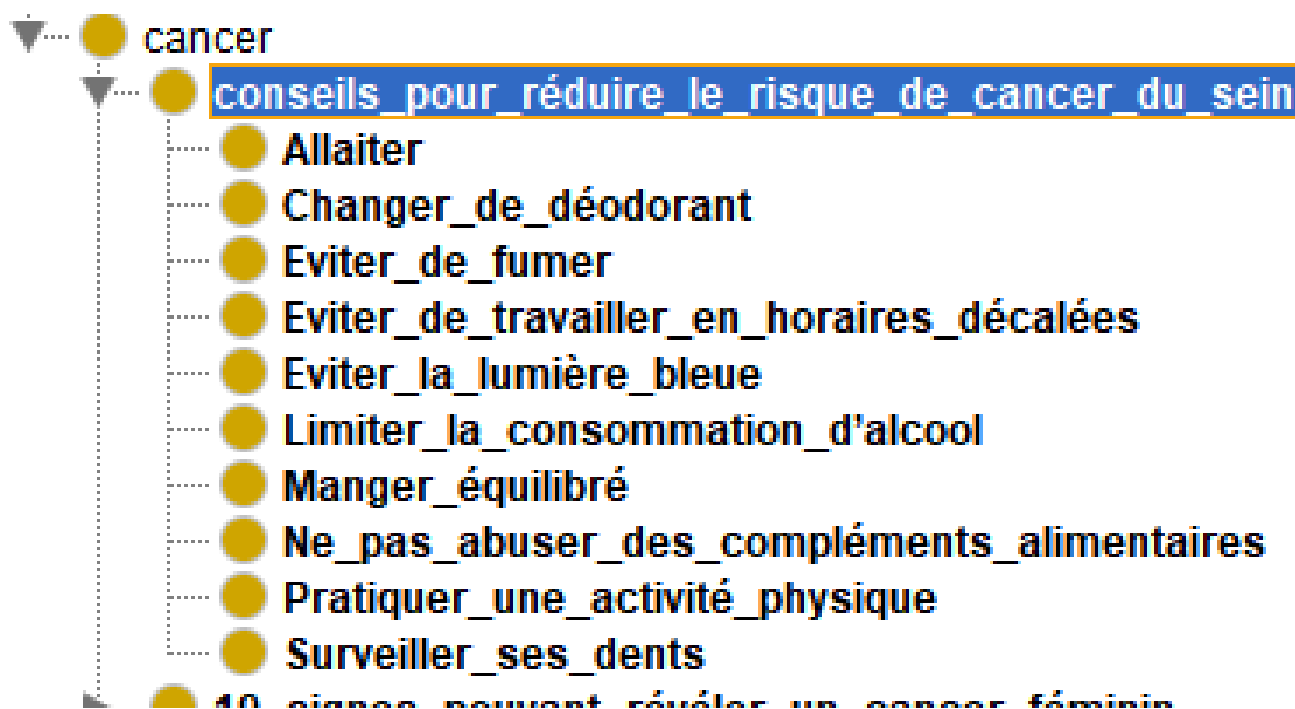


FIGURE 4.35 – la classe conseils pour réduire le risque du cancer du sein

-Graphe qui détermine la classe conseil :

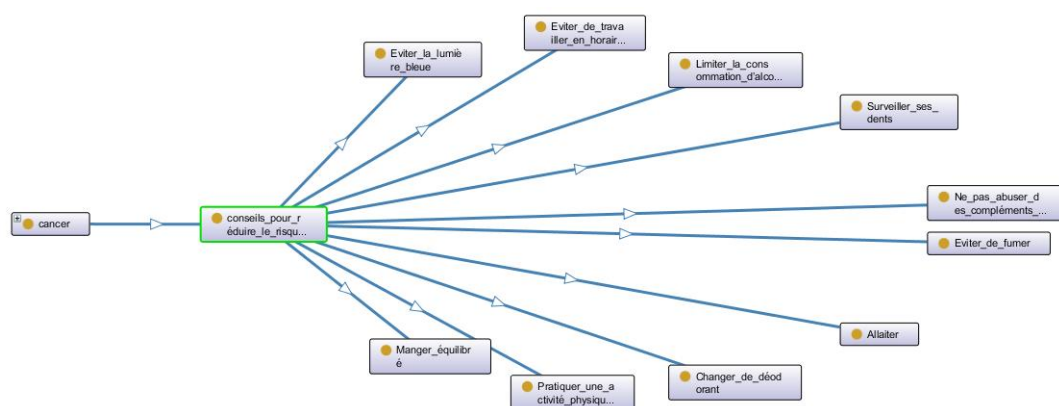
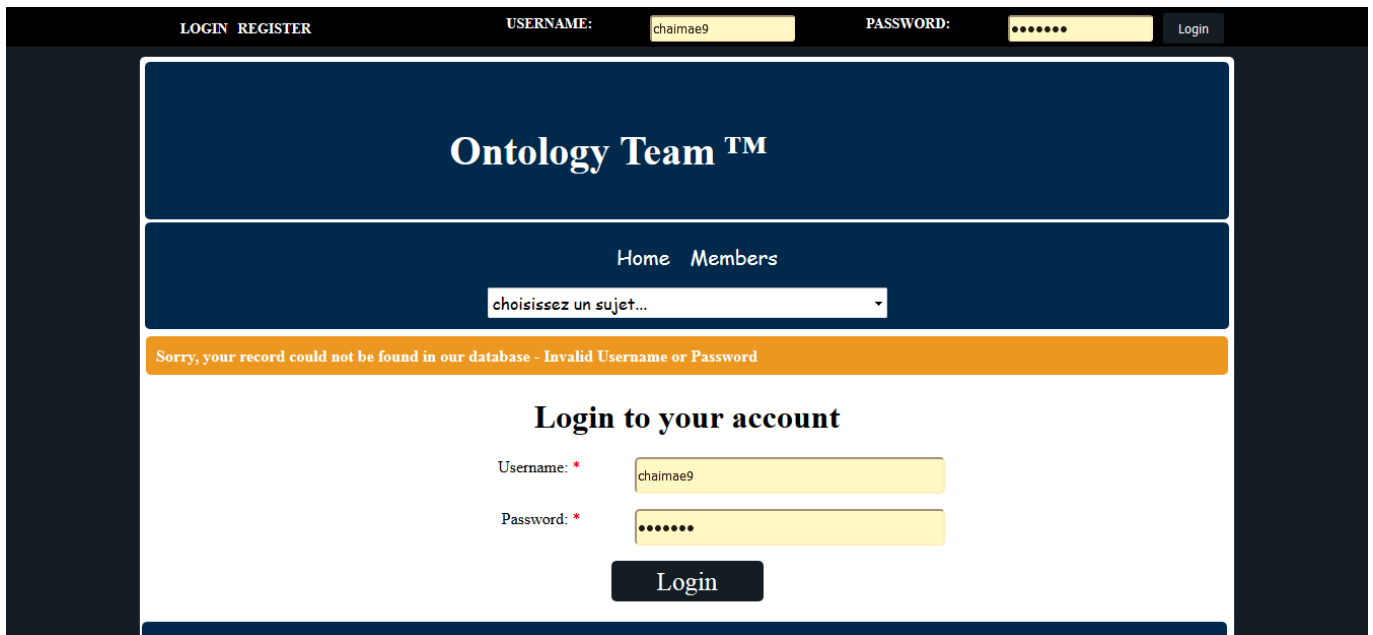


FIGURE 4.36 – graphe de la classe conseils

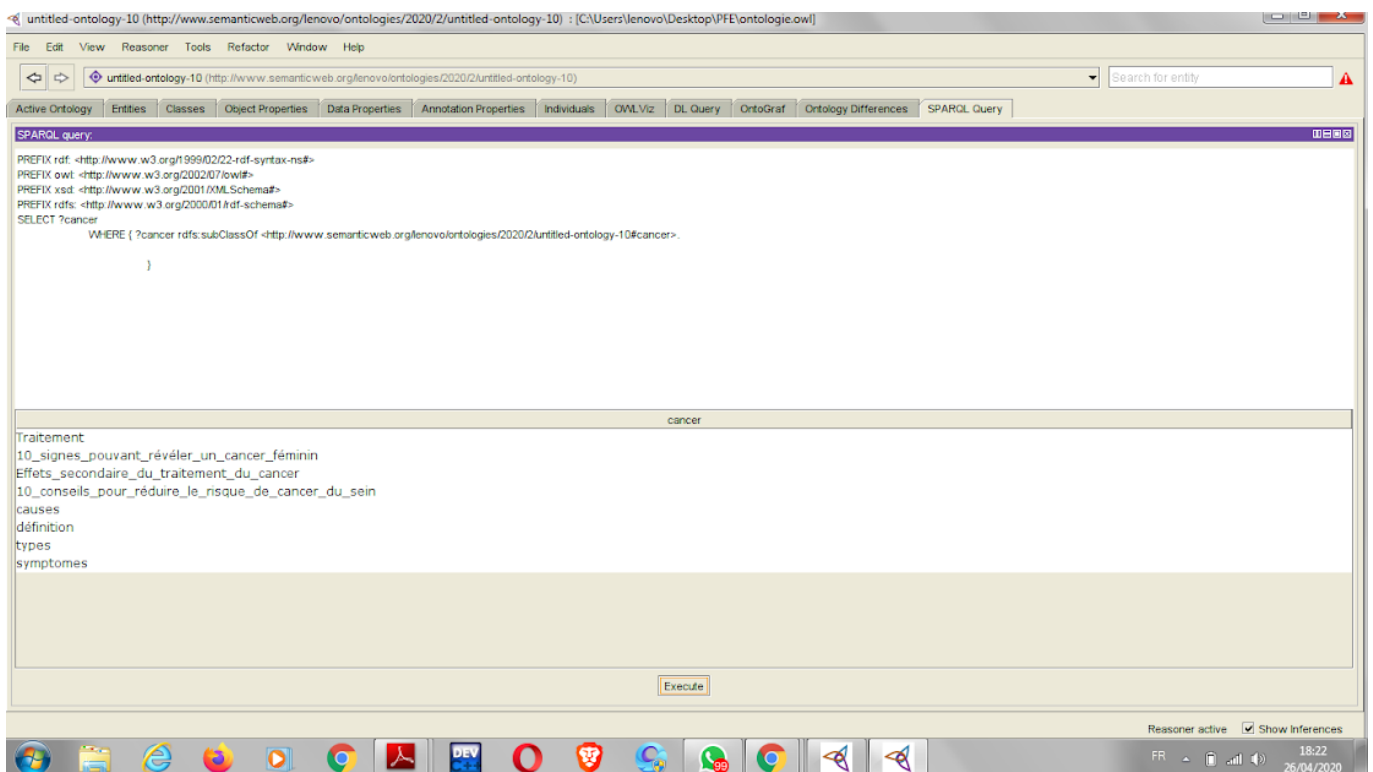
-Le message affiché si l'utilisateur met un erreur pour se connecter :



The screenshot shows the login page of the Ontology Team. At the top, there are links for 'LOGIN' and 'REGISTER'. Below them, the 'USERNAME:' field contains 'chaimae9' and the 'PASSWORD:' field contains a masked password. A 'Login' button is to the right. The main content area has a dark blue header with 'Ontology Team™'. Below it, a navigation bar has 'Home' and 'Members' links. A dropdown menu shows 'choisissez un sujet...'. An orange error message banner reads: 'Sorry, your record could not be found in our database - Invalid Username or Password'. Below this, a section titled 'Login to your account' contains 'Username: *' and 'Password: *' fields, both with the same values as the top section, and a 'Login' button.

FIGURE 4.37 – message d'erreur

-requête SPARQL qui cherche tous les sous-classes de la classe mère "cancer" :



The screenshot shows a SPARQL query interface. The query is:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?cancer
WHERE {
  ?cancer rdfs:subClassOf <http://www.semanticweb.org/lenovo/ontologies/2020/2/untitled-ontology-10#cancer>.
}
```

 The results show a table with one row for 'cancer' and its subclasses: 'Traitement', '10_signes_pouvant_révéler_un_cancer_féminin', 'Effets_secondaire_du_traitement_du_cancer', '10_conseils_pour_réduire_le_risque_de_cancer_du_sein', 'causes', 'définition', 'types', and 'symptomes'. An 'Execute' button is at the bottom.

FIGURE 4.38 – requête SPARQL

-Ajouter les instances au cancer du sein :

The screenshot shows the OntoGraf web interface. The top navigation bar includes tabs for 'Active ontology', 'Entities', 'Classes', 'DL Query', 'OWLax', and 'OntoGraf'. Below this, there are tabs for 'Annotation properties', 'Datatypes', 'Individuals', 'Classes', 'Object properties', and 'Data properties'. The main content area is divided into two panels. The left panel, titled 'Class hierarchy: cancer du sein', shows a tree structure starting from 'owl:Thing' and branching into 'cancer'. Under 'cancer', there are several sub-classes, including 'types', which lists various cancer types like 'cancer du colon(cancer colorectal)', 'cancer d_cerveau', 'cancer de l'endomètre', 'cancer de l'estomac', 'cancer de l'oesophage', 'cancer de la gorge', 'cancer de la peau', 'cancer de la thyroïde', 'cancer de la vessie', 'cancer de testicules', 'cancer du col de l'utérus', 'cancer du foie', 'cancer du pancréas', 'cancer du poumon', 'cancer du prostate', 'cancer du sein', and 'cancer du système lymphatique'. The right panel, titled 'Annotations: cancer du sein', shows the 'Description: cancer du sein' section. It includes a 'General class axioms' section with a '+' icon, a 'SubClass Of (Anonymous Ancestor)' section, an 'Instances' section with a '+' icon, and a 'Target for Key' section with a '+' icon. The 'Instances' section lists four instances: 'ck+', 'ck20-', 'RO+', and 'RP+', each preceded by a diamond icon. The 'Target for Key' section is empty. The 'Disjoint With' and 'Disjoint Union Of' sections are also empty.

FIGURE 4.39 – instances

Bibliographie

-Pour la notion du web sémantique :

<https://jplu.developpez.com/tutoriels/web-semantic/introduction/>

-Pour la notion de l'ontologie :

www.journaldunet.com/developpeur/tutoriel/theo/070403-ontologie.shtml

-Pour installer Protege :

<http://protege.stanford.edu/download/protege/4.3/installanywhere.exe>
Installers/

-les informations du cancer :

<https://www.passeportsante.net/>

-Pour créer l'ontologie :

<http://www.iro.umontreal.ca/~lapalme/ift6282/OWL/EtapesCreationOntologie.html>
sIg-nzkRrHnPndXO1DjsX47Zzklo7zH-IRA3gaiHemUO8k21FlsZ

-Pour visualiser l'ontologie :

<http://owlgred.lumii.lv/online-visualization/0kmf>

-Pour le codage :

Visual Studio Code

-Pour apprendre SPARQL :

<https://web-semantique.developpez.com/tutoriels/jena/arq/intro-sparql/>

-LSA :

[https://fr.wikipedia.org/wiki/Analyse_s%C3%A9mantique-latente](https://fr.wikipedia.org/wiki/Analyse_s%C3%A9mantique_latente)

-Pour le code de LSA :

<http://github.com/iam-mhaseeb/Python-Implementation-of-LSA>