

**République Algérienne Démocratique et Populaire**

**Ministère de l'enseignement supérieur et de la recherche scientifique**

**Université de 8 Mai 1945 – Guelma -**

**Faculté des Mathématiques, d'Informatique et des Sciences de la matière**

**Département d'Informatique**



**Mémoire de Fin d'études Master**

**Filière : Informatique**

**Option : SIQ**

**Thème :**

---

---

**Proposition d'une annotation sémantique des  
données issues d'IoT**

---

---

**Encadré Par :  
DJAKDJAKHA LYNDA**

**Présenté Par :  
Hesna Mrabti**

**Juin 2022**

# Remerciements

**En premier lieu, je tiens à exprimer ma profonde gratitude à mon encadreur :**

**Dr. Djakhdjakha Lynda, je la remercie pour sa disponibilité, ses encouragements, et surtout pour ses sacrifices.**

**Mille mercis à Hamadoun Cissé, le guide de mon travail.**

**Je suis très reconnaissante à Pr. Mohammed Nadjib Kouahla, de m'avoir fait l'honneur de présider le jury,**

**Merci à Dr. Zedadra Amina d'avoir accepté d'examiner le contenu du présent travail.**

# Dédicace

*A celle qui m'a arrosé de tendresse et d'espairs, à la source d'amour...*

*A mes parents,*

*A mes chers frères : Chemseddine et Zakaria,*

*A ma chère sœur Wafa et son mari Ibrahim et son fils Younes,*

*A ma très chère amie Wissem,*

*A Saad.*

# Résumé

*Les datasets regroupent un ensemble de données cohérents qui peuvent se présenter sous différents formats (textes, chiffres, images, vidéos, ...). Les données collectées dans des datasets constituent une source d'informations importante pour les applications IoT dans les différents domaines. C'est dans ce contexte que se situe ce travail de master qui s'intéresse à la proposition d'une annotation sémantique pour l'interprétation du contenu des datasets et en particulier les datasets image afin de réduire la difficulté liée à la sélection de ces datasets par les analystes de données.*

**Mots-clés :** *IoT, capteur IoT, imageur, dataset, Ontologie, annotation sémantique,*

## Abstract

*Datasets bring together a set of coherent data that can be presented in different formats (texts, figures, images, videos, etc.). data collected in datasets represent an important source of information for IoT applications in different domains. It is in this context that this master's work is located, which focuses on the proposition of a semantic annotation for the interpretation of the content of datasets and in particular image datasets in order to reduce the difficulty linked the the selection step of these datasets by data analysts.*

**Keywords:** *IoT, IoT sensor, image sensor, dataset, Ontology, semantic annotation.*

# TABLE DES MATIERES

Remerciements .....	i
Dédicace.....	ii
Résumé .....	iii
TABLE DES MATIERES .....	iv
Introduction générale .....	1
Chapitre 01 : Concepts de base .....	3
<b>1.1. Introduction</b> .....	3
<b>1.2. Les capteurs IoT et les données IoT</b> .....	3
<b>1.2.1. Un capteur IoT</b> .....	3
<b>1.2.2. Un capteur d'images</b> .....	4
<b>1.2.3. Les caractéristiques d'un capteur d'image :</b> .....	4
<b>1.2.4. Une image numérique</b> .....	4
<b>1.2.5. Les caractéristiques d'une image numériques</b> .....	5
<b>1.2.6. Type d'image numérique</b> .....	5
<b>1.3. Ontologie</b> .....	8
<b>1.3.1. Les composants d'une ontologie :</b> .....	8
<b>1.3.2. Pourquoi développer une ontologie ?</b> .....	9
<b>1.3.3. Langages de spécification d'ontologie</b> .....	9
<b>1.3.4. Méthodes et Méthodologies de construction des ontologies</b> .....	10
<b>1.4. Annotation sémantique</b> .....	12
<b>1.4.1. Définition d'annotation</b> .....	12
<b>1.5. Travaux connexes</b> .....	14
<b>1.6. Conclusion</b> .....	15
Chapitre 02 : Conception .....	16
<b>2.1. Introduction</b> .....	16
<b>2.2. Démarche globale de la solution</b> .....	16
<b>2.3. Construction de l'ontologie</b> .....	18
<b>2.3.1. Spécification</b> .....	18
<b>2.3.2. Conceptualisation</b> .....	19
<b>2.3.3. La formalisation</b> .....	23

<b>2.4. Un système pour l'annotation sémantique .....</b>	<b>23</b>
<b>2.5. Conclusion .....</b>	<b>27</b>
Chapitre 03 : Implémentation .....	28
<b>3.2. Outils d'implémentation .....</b>	<b>28</b>
<b>3.3. Implémentation de l'ontologie .....</b>	<b>29</b>
<b>3.4. Implémentation du processus d'annotation .....</b>	<b>32</b>
<b>3.6. Conclusion .....</b>	<b>41</b>
Conclusion générale .....	42
Références .....	43
Bibliographie.....	43

# LISTE DES FIGURES

Figure 1. Définition d'une image [6].....	4
Figure 2. Une image panchromatique [9] .....	5
Figure 3. Exemple d'une image multispectrale à 8 bandes .....	6
Figure 4. Exemple d'une image hyperspectrale [11] .....	7
Figure 5. Exemple d'une image thermique [13] .....	7
Figure 6. Une image satellite à 3 bandes (RGB).....	8
Figure 7. Démarche globale de la solution .....	17
Figure 8. Document de spécification de l'ontologie .....	19
Figure 9. Diagramme de cas d'utilisation.....	24
Figure 10. Diagramme d'activité du système d'annotation .....	25
Figure 11. Diagramme de séquence du système .....	26
Figure 12. Création de l'ontologie avec Owlready de Python.....	30
Figure 13. Une partie du fichier OWL. ....	30
Figure 14. La hiérarchie des concepts de l'ontologie avec Protégé.....	31
Figure 15. Visualisation de la liste de propriétés de données (data properties) avec Protégé ..	31
Figure 16. Le code d'extraction du format des images du dataset .....	32
Figure 17. Code pour l'Annotation .....	33
Figure 18. Instanciation d'un nouveau dataset.....	34
Figure 19. Visualisation du contenu d'un dataset .....	35
Figure 20. Page d'accueil du système d'annotation .....	35
Figure 21. Message d'erreur en cas de répertoire existant. ....	36
Figure 22. Affiche des caractéristiques du dataset .....	36
Figure 23. Détails des caractéristiques du dataset.....	37
Figure 24. Détails des caractéristiques de l'image .....	37
Figure 25. Bouton annoter et changer image .....	38
Figure 26. Changement de l'image du dataset .....	38
Figure 27. Etape 1 de l'annotation d'une image .....	39
Figure 28. Etape 2 de l'annotation d'une image .....	39
Figure 29. Etape 3 de l'annotation d'une image .....	40
Figure 30. Etape 4 de l'annotation d'une image .....	40
Figure 31. Passage de la souris sur une zone annotée. ....	41
Figure 32. Annotation du type de photographie (Image) .....	41

## **LISTE DES TABLEAUX**

Tableau 1. Avantages et inconvénients de méthodologies de développement [19].....	12
Tableau 2. Question de compétences de l'ontologie .....	19
Tableau 3. Glossaire de quelques termes utilisés par l'ontologie .....	20
Tableau 4. Liste de quelques concepts / sous concepts utilisés .....	21
Tableau 5. Liste de quelques relations utilisées .....	21
Tableau 6. Liste des attributs ajoutés. ....	22

## LISTE D'ABREVIATIONS

Abréviation	Définition
<b>IoT</b>	Internet of Thing
<b>OWL</b>	Web Ontology Language
<b>Px</b>	Pixel
<b>RDF</b>	Resource Description Framework
<b>RDFS</b>	RDF Schema
<b>UML</b>	Langage de Modélisation Unifié
<b>XML</b>	eXtensible Markup Language

# Introduction générale

L'Internet se transforme jour après jour en un hyper réseau appelé Internet des objets ou en anglais Internet of Things (IoT) connectant des milliards d'êtres humains mais aussi d'objets. L'IoT est considéré comme l'une des technologies les plus importantes du 21<sup>ème</sup> siècle. Avec le développement rapide des technologies IoT, les données capteurs jouent un rôle de plus en plus important dans différents domaines (médecine, agriculture, urbain) et dans différentes applications telles que la cartographie, la surveillance des ressources et des cultures, et la prévention des catastrophes. La plupart de ces applications sont basées sur des images qui sont générées par des capteurs d'image et stockées dans des datasets disponibles sur le web. Ces datasets contiennent des images de même ou de types variés, qui ont les mêmes ou différentes caractéristiques et générées par les mêmes ou bien par des capteurs différents.

Les développeurs qui utilisent les données IoT doivent faire une recherche pour sélectionner le dataset approprié, où le manque d'information sur le contenu des dataset représente un grand problème qui entraîne un parcours de tout le contenu pour choisir un dataset.

Dans ce contexte, et dont le but est de faciliter la recherche, l'accès et l'utilisation des datasets, l'objectif de ce travail est de proposer une annotation sémantique pour l'interprétation des éléments des datasets de type image. L'annotation sémantique des ressources Web permet une meilleure interprétation de leur contenu [1]. C'est l'utilisation des informations descriptives permettant de donner un sens à une ressource. Elle a pour objectif de formaliser l'interprétation qui peut être faite sous forme de métadonnées attachées aux textes ou à certain de leurs segments [2]. L'annotation repose sur des ontologies enrichies de connaissances. Les ontologies constituent l'une des technologies les plus importantes du web sémantique. Elles permettent d'enrichir l'annotation en lui associant une sémantique. Une ontologie est une spécification formelle, explicite et consensuelle de la conceptualisation d'un domaine [3]. Elle est constituée d'un ensemble de concepts organisés et liés par des rôles, comportée des axiomes et ainsi comportées des instances de concepts et des instances de rôles.

Ce mémoire est structuré en trois chapitres :

- Le premier chapitre présente un état de l'art sur les concepts de base de notre travail à savoir les ontologies, ses langages de spécification et ses outils de développement. Il décrit la définition d'un dataset, ses éléments et en particulier la notion d'image capteur ainsi les différents types d'images. Il donne aussi une synthèse de quelques travaux liés à notre proposition.
- Le deuxième chapitre présente la démarche de construction de l'ontologie proposée et le processus d'annotation sémantique.
- Le troisième chapitre présente les étapes de l'implémentation.
- Et la fin une conclusion pour conclure et donner quelques perspectives du travail présenté.

# Chapitre 01 : Concepts de base

## 1.1. Introduction

L'objectif principal de ce travail est de proposer une annotation sémantique pour l'interprétation des données IoT de type image. De ce fait, nous allons dans un premier temps faire un état de l'art pour comprendre le contexte du travail.

Dans ce chapitre, nous présentons les concepts de dataset et capteur IoT, nous présentons aussi les concepts représentant le contenu d'un dataset image tel que : le concept d'image numérique, ses types et ses caractéristiques. Ensuite, en vue de proposer une annotation sémantique (ontologique), nous allons cerner le domaine des ontologies, nous décrivons les concepts liés aux ontologies. Nous établissons un aperçu sur les langages et les méthodologies de développement des ontologies. Enfin, nous présentons un ensemble de travaux qui ont traité la problématique de l'annotation sémantique pour mieux positionner notre proposition.

## 1.2. Les capteurs IoT et les données IoT

L'IoT est défini comme un système d'appareils, de capteurs et de machines numériques interconnectés qui peuvent communiquer et partager des données sur un réseau sans nécessiter d'interaction humaine [4].

Il offre un potentiel unique pour collecter des données afin de les utiliser dans différentes applications. Le développement de ce domaine dépend des appareils ou capteurs IoT qui sont considérés comme son système nerveux. Les capteurs IoT jouent un rôle primordial dans le fonctionnement de l'IoT. Ils détectent les changements d'un environnement, mesurent, collectent et transforment ses données pour pouvoir ensuite être exploitées par les analystes des données.

### 1.2.1. Un capteur IoT

Un capteur IoT ou un capteur d'un objet connecté permet de déceler les formes d'énergie (lumière, force, etc.) et de les transformer en informations numérique. On peut distinguer deux types de capteurs IoT, des capteurs non-imageur et des capteurs d'images :

### 1.2.2. Un capteur d'images

Un capteur d'image ou un imageur est un dispositif qui permet de convertir de la lumière en une grandeur électrique pour former une image visualisable, transformable et mémorisable [5].

### 1.2.3. Les caractéristiques d'un capteur d'image :

Les principales caractéristiques d'un imageur sont :

- Le champ d'observation total : de quelques dizaines à quelques millions de km.
- La résolution spatiale : qui exprime son aptitude à distinguer les détails dans le champ total d'observation
- Les bandes spectrales : domaines de sensibilité spectrale des détecteurs et de leurs accessoires (filtres...)
- La résolution radiométrique : qui caractérise l'aptitude des détecteurs à distinguer des signaux porteurs de plus ou moins d'énergie

### 1.2.4. Une image numérique

Une image numérique est un signal fini échantillonné à valeurs quantifiées dans un certain espace de couleurs. Dans [6], une image est définie par le nombre de points (pixels) qui la constituent. La définition est obtenue par la multiplication du nombre de pixels formant la largeur de l'image par celui de la hauteur [6].

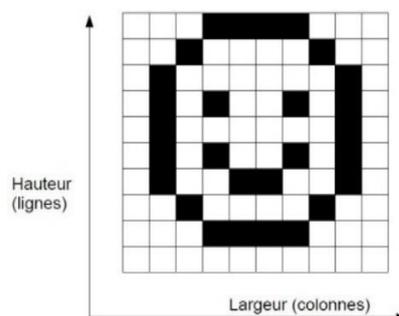


Figure 1. Définition d'une image [6]

Les pixels sont la plus petite composante d'une image numérique. Le nom "pixel" est abrégé en px et est dérivé de l'expression anglaise *picture element*, qui signifie "élément d'image" ou

"point élémentaire". Les pixels sont généralement rectangulaires ou presque carrés, avec des tailles allant de 0,18 mm à 0,66 mm de côté [7].

Les images composées de pixels sont appelées images "raster", par opposition aux images "vectorielles", et sont composées d'une géométrie disposée géographiquement (points, lignes ou surfaces) avec des attributs associés [8].

### 1.2.5. Les caractéristiques d'une image numériques

Les caractéristiques principales des images numériques sont :

- *La résolution spectrale* : est la capacité d'un capteur à distinguer des rayonnements électromagnétiques de fréquences différentes.
- *La résolution spatiale* : cette propriété est donnée en fonction de nombre de pixels par millimètres.

### 1.2.6. Type d'image numérique

#### *L'image panchromatique*

Le concept panchromatique enregistre seulement l'intensité totale du rayonnement sur chaque pixel. Une image panchromatique offre une résolution spatiale importante et obtenue à partir de l'enregistrement du rayonnement dans un unique intervalle de longueur d'onde [0.4..0.7]. Dans ce type d'images, les données sont acquises dans un seul canal. De ce fait les images obtenues sont en noir et blanc. La figure suivante montre une image panchromatique ou le canal vert est visualisé en intensité de gri.



Figure 2. Une image panchromatique [9]

### ***L'image multispectrale***

Une image multispectrale est une image qui capture des données à des fréquences spécifiques à travers le spectre électromagnétique [10]. Les images multispectrales sont obtenues par des capteurs qui ont la capacité d'enregistrer l'information dans des enregistrements simultanés dans un petit nombre de bandes spectrales qui n'ont pas nécessairement contiguës. Le nombre de bandes est en générale dans l'intervalle [3. 8].

Pour obtenir des images multispectrales en couleur, les bandes d'information numérique sont représentées par combinaison de trois couleurs primaires (rouge, vert et bleu).



*Figure 3. Exemple d'une image multispectrale à 8 bandes*

### ***L'image hyperspectrale***

Ces images sont obtenues par des enregistrements d'information dans une multitude de bandes spectrales (Elles pourraient avoir des centaines de milliers de bandes) beaucoup plus étroites (10-80 nm) et souvent contiguës.



Figure 4. Exemple d'une image hyperspectrale [11]

### ***L'image thermique***

L'imagerie thermique est une technologie capable de détecter des personnes et des objets dans l'obscurité totale et dans diverses conditions météorologiques [12].



Figure 5. Exemple d'une image thermique [13]

### ***L'imagerie satellitale***

L'imagerie satellitale appelée aussi 'imagerie spatiale' fait référence à la capture d'images depuis l'espace par des capteurs situés sur le satellite. Visuellement, l'imagerie satellite ressemble beaucoup à une photographie, mais elle contient beaucoup plus d'informations [8].



Figure 6. Une image satellite à 3 bandes (RGB)

### 1.3. Ontologie

Dans la littérature, il existe plusieurs définitions pour le concept d'ontologie. Une ontologie permet la structuration et la formalisation des concepts dans un domaine. Elle est définie comme un réseau sémantique qui regroupe un ensemble de concepts décrivant complètement un domaine. Ces concepts sont liés les uns aux autres par des relations taxonomiques d'une part, et sémantique d'autre part.

Gruber [3] a défini l'ontologie comme *la spécification d'une conceptualisation d'un domaine de connaissance*. Pour [14], une ontologie est *une description formelle d'entités et de leurs propriétés, relations, contraintes, comportement*, et elle est simplifiée dans [15] comme *un ensemble de définitions de concepts et leurs relations* [16].

#### 1.3.1. Les composants d'une ontologie :

Selon [3] les connaissances intégrées à l'ontologie sont formalisées en utilisant cinq types de composants : concepts, relations, fonctions, axiomes et instances.

- **Concepts** : un concept représente un ensemble d'objets (des objets matériels, des évènements, des concepts profonds, des idées, etc.) et leurs propriétés communes. Un

concept a (i) une intention qui désigne la sémantique et (ii) une extension : qui est l'ensemble des objets qu'englobe un concept [17].

- **Relations** : les relations dans une ontologie sont formellement définies comme un sous-ensemble du produit de n ensembles. On peut distinguer deux types de relations :
  - o **Relation de subsumption** : cette relation est dite aussi taxonomique, La relation de subsumption « est-un » (is-a), lie un élément supérieur (le concept plus général) et un élément inférieur (le concept spécifique) [17].
  - o **Relation associative** : ces relations sont des relations d'interaction entre deux concepts qui ne sont pas la relation de subsumption [18].
- **Fonctions** : un héritage de relation qui définit le nième élément de la relation en fonction de l'élément n-1 précédent [17].
- **Axiome** : les axiomes sont des contraintes ou des règles sur des concepts ou des instances ontologiques qui sont considérées toujours comme vraies [17].
- **Instances (ou individus)** : Ils forment une définition étendue de l'ontologie. Ils représentent un élément unique qui véhicule des connaissances sur le domaine [18].

### 1.3.2. Pourquoi développer une ontologie ?

Une ontologie est développée pour les raisons suivantes [19] :

- Partager une compréhension commune de la structure de l'information entre les personnes, ou créateur de logiciels.
- La réutilisation du savoir sur un domaine.
- Analyser le savoir.
- Expliciter ce qui est considéré comme implicite sur un domaine.

### 1.3.3. Langages de spécification d'ontologie

Une ontologie repose sur un langage de spécification. Il existe plusieurs langages qui sont basés sur la logique de prédicat et d'autre sur les logiques de description. Nous citons quelques

langages :

- **XML** (eXtensible Markup Language) est un langage de balises qui permet de décrire la structure des documents. Il dérive de SGML (Standard Generalized Markup Language) et HTML (Hyper Text Markup Language).
- **RDF** (Resource Description Framework) est un langage développé par le W3C pour représenter des informations sur le World Wide Web, comme des pages Web. RDF apporte une manière standard de coder des informations afin que celles-ci puissent être échangées de façon totalement automatisée entre applications [20].
- **RDF Schema (RDFS)** est un langage de définition pour la description de propriétés et de classes représentées par des ressources RDF. Il définit des graphes de triplets RDF, avec une sémantique de généralisation / hiérarchisation de ces propriétés et de ces classes.
- **OWL** (Web Ontology Language) : Le langage d'ontologie Web OWL a été conçu pour les applications qui doivent traiter le contenu de l'information plutôt que de simplement présenter des informations aux humains. OWL permet une meilleure interprétabilité machine du contenu Web que celle prise en charge par XML, RDF et RDF Schéma (RDFS) en fournissant un vocabulaire et une sémantique formelle supplémentaires. OWL a trois sous-langages de plus en plus expressifs. OWL Lite, OWL DL et OWL Full [21]. Plus expressif que son prédécesseur, RDFS, OWL s'est rapidement imposé dans le monde des ontologies et est aujourd'hui pratiquement le standard le plus utilisé [22].
- **DAML+OIL** est un langage d'ontologie spécialement conçu pour une utilisation sur le web. Il Utilise les normes Web (XML et RDF) et ajoute des primitives d'ontologie orientées objet [23]

#### **1.3.4. Méthodes et Méthodologies de construction des ontologies**

Dans la littérature, il n'existe pas encore un consensus sur les bonnes pratiques à adopter lors du processus de création, ni même sur les normes techniques régissant le processus de développement d'ontologies [24]. Mais, on peut utiliser plusieurs façons pour la création d'une ontologie, construction à partir de zéro, la réutilisation d'autres ontologies ou la fusion d'ontologies, et il existe plusieurs méthodologies pour le développement des ontologies telles que : Celui d'**Uchuld** et **King** [25], **Ontologie 101** de Noy et McGuinness [26], la

méthodologie **On-To-Knowledge** [27] et **METHONTOLOGY** [28] :

- **La méthodologie d'Uchuld et King** [25]: Cette approche divise le développement de l'ontologie en un ensemble d'étapes : Identifier les Objectifs , Construire l'ontologie : Cette étape est divisée en trois sous-étapes :
  - Capture : Les concepts et les relations sont définis textuellement, aussi même les termes qui se réfèrent à ces concepts et relations sont identifiés et définis.
  - Coder : la représentation de l'ontologie de l'étape de capture est transformée en un langage formel « d'ontologie » par exemple OWL, à cette étape on peut utiliser un environnement de développement d'ontologies comme Protégé.
  - Intégrer : Comme l'ontologie qui est en cours de développement couvre en partie la portée des ontologies déjà existante, on doit toujours rechercher quelles ontologies peuvent être réutilisées.
  
- **METHONTOLOGY** [29] Cette méthode a été développée par le groupe d'ontologie de l'Institut polytechnique d'intelligence artificielle de Madrid. Elle vise à construire une ontologie au « niveau connaissance » et s'appuie sur : un processus de développement d'ontologies comprenant des activités de gestion de projet (planification, assurance qualité), des activités orientées vers le développement (spécifications, conceptualisation, formalisation) et des activités de support (documentation). Le cycle de vie d'ontologie basé sur des prototypes évolutifs [30].
  
- **La Méthode 101** [26] tente de construire une ontologie formelle en adoptant et en s'adaptant à l'ontologie existante et propose les étapes suivantes :
  - Déterminer le domaine et la portée de l'ontologie.
  - Envisager la réutilisation de l'ontologie existante.
  - Quels sont les termes les plus importants de l'ontologie ?
  - Définir les classes et les hiérarchies de classes.
  - Définir les propriétés de la classe.
  - Définir les facettes d'attribut.
  - Créer une instance.

Dans cette méthode, les outils Protégé 2000 [31] et Ontolingua [32] sert de support [33].

- **La méthodologie On-To-Knowledge** [27] recommande un processus de

développement itératif, qui comporte quatre phases principales : une phase de spécification d'état, une phase d'amélioration, une phase d'évaluation et une phase d'application et de développement. OnToKnowledge propose l'acquisition de connaissances par une spécialisation générale en ontologie. Par conséquent, l'ontologie développée avec cette méthodologie dépend fortement de l'application. OnToKnowledge recommande la suite d'outils OntoStudio pour prendre en charge le développement d'ontologies [24].

Le tableau résume les avantages et les inconvénients des méthodologies citées :

<b>Méthodologies</b>	<b>Avantages</b>	<b>Inconvénients</b>
<b>Ushuld et King</b>	Simple et directe	La documentation n'est pas appliquée
<b>Ontologie 101</b>	Simple et compréhensible Faible effort	Rarement utilisée La documentation n'est pas appliquée
<b>METHONTOLOGY</b>	Largement utilisé La documentation est appliquée	

*Tableau 1. Avantages et inconvénients de méthodologies de développement [34]*

## **1.4. Annotation sémantique**

### **1.4.1. Définition d'annotation**

Une annotation est une note particulière attachée à une partie de document (un paragraphe, un groupe de mot, une image, une partie d'image, etc.), un document, une collection de documents, ou à une autre annotation [35].

### **1.4.2. Types d'annotation**

Plusieurs classifications sont proposées dans la littérature :

- La classification de [36]

- **Annotations informelles** : ces annotations sont exprimées dans un langage informel.
  - **Annotations formelles** : ces annotations sont exprimées dans un langage formel.
  - **Annotations ontologiques** : se composent des termes ontologiques acceptés et compris dans un domaine donné.
- La classification de [37]
- **Annotation textuelle** : consiste à ajouter des notes ou des commentaires à un objet.
  - **Annotation de lien** : consiste à relier un objet à un contenu d'annotation
  - **Annotation sémantique** : consiste à utiliser des informations lisibles à la fois par l'homme et par la machine.

### 1.4.3. Annotation sémantique

L'annotation sémantiques est l'association de sens à l'information contenue dans les documents par des annotations ou des métadonnées, qui sont enregistrées dans le document lui-même ou bien dans des supports externes [38]. L'objectif d'annotation sémantique est d'associer une interprétation stabilisée aux données brutes [39]

### 1.4.4. Types d'annotation sémantique

Selon [2], on distingue cinq types d'annotation sémantique, cette classification est basée sur la nature de l'élément ontologique auquel les annotations se rattachent :

- Les termes peuvent rattachés aux instances de concepts : Le processus d'annotation consiste à *créer des instances* de concepts et à leur rattacher les annotations appelées dans ce cas 'entités nommées'.
- Les termes peuvent rattachés aux concepts : le processus d'annotation consiste à créer de nouveaux concepts considérés pertinents.
- Les termes peuvent rattachés aux rôles : le processus d'annotation consiste à créer des nouveaux rôles entre concepts
- Des textes peuvent renvoyer à des relations entre instances de concepts : le processus consiste à créer des nouvelles instances entre instances de classes.
- Des textes peuvent exprimer des axiomes ontologiques : Le processus d'annotation consiste à créer des nouveaux axiomes.

## 1.5. Travaux connexes

Cette section, est consacré à présentation des travaux de recherche ayant traité la problématique de ce mémoire. L'objectif étant d'approfondir davantage de cette dernière, d'analyser et de mettre en avant les avantages et les limites des travaux existants et aussi à mieux positionner notre proposition qui est une annotation sémantique pour l'interprétation des éléments des datasets image. Nous citons des travaux récents :

En 2017, les auteurs de [40] ont présenté une bonne comparaison des méthodes d'annotation sémantique proposées entre 2006 et 2015. Plusieurs méthodes ont proposé des méthodes d'annotation dans le domaine d'IoT [41], [42], [43], [41], [44], [45]. [41] et [43] proposent des méthodes d'annotation pour les réseaux de capteurs, [42] propose une méthode pour annoter les données IoT, [41] propose une méthode pour annoter les entités d'information IoT, [44] pour les documents et [45] pour annoter les fichiers WSDL (Web Service Description Language) de services Web.

Dans le même travail [40], ses auteurs ont proposé une méthode automatique d'annotation sémantique des données IoT. Le processus proposé utilise une ontologie de capteurs OWL, et est basé sur un ensemble d'étapes : extraction d'information, classification de texte, division des propriétés d'information, la sélection sémantique des labels et l'intégration d'information.

Les auteurs de [46] ont proposé une méthode d'annotation sémantique pour annoter les données des capteurs IoT dynamique. Cette méthode construit une ontologie initiale basée sur l'ontologie du réseau de capteurs sémantiques (SSN). Ils ont utilisé la méthode de clustering K-means, pour extraire des nouvelles connaissances, qui sont utilisées pour mettre à jour l'ontologie initiale. Le résultat est une nouvelle ontologie qui forme alors la base de l'annotation sémantique.

Les auteurs de [47] ont présenté un modèle de gestion de flux de données IoT pour les systèmes de surveillance en temps réel. Le modèle permet d'intégrer les données IoT hétérogènes avec des annotations sémantiques. Les auteurs ont utilisé un ensemble de standards pour représenter les annotations et les données IoT tels que : SWE (Sensor Web Enablement).

[48] a proposé un service appelé SiSEG, pour l'annotation des données et résoudre le problème d'intégration de données hétérogènes. SiSEG met en place une procédure automatisée qui reçoit des données brutes hétérogènes du secteur de l'énergie ou d'autre secteurs en format JSON ou XML et les ontologies requises en tant que connaissances de domaine, et ajoute des métadonnées à ces données pour générer des données annotées et normalisées selon sa base de

connaissances en JSON-LD, RDF et OWL.

[49] a présenté un framework d'annotation sémantique d'images appelé HCASAI. Les auteurs ont combiné différentes fonctionnalités de bas et de haut niveau, entourant le texte et des techniques de cartographie ontologique. HCASAI propose plusieurs niveaux d'annotation basés sur cinq caractéristiques de bas niveau de l'image pour l'identification de son contenu (CLD 'color layout descriptor', SCD 'scalable color descriptor', EHD 'edge histogram descriptor', CSD 'Color Structure descriptor' et DCD 'dominant color descriptor'). Il classe l'image à l'aide de ANN (artificial neural network). Les annotations sont codées en RDF.

Dans [50], les auteurs ont proposé une technologie automatique pour l'annotation sémantique des nouvelles image IoT. Elle est basée sur la fusion de poids et des caractéristiques d'image telles que les caractéristiques de couleur, de texture et de forme pour déterminer leur stabilité en utilisant l'écart-type de ces caractéristiques. Puis elle calcule la pondération des coefficients pour annoter l'image.

Dans [51], les auteurs ont proposé une annotation sémantique pour description des datasets de data mining en utilisant une ontologie. L'annotation combine trois type d'information : de provenance, des informations spécifiques au data mining et des informations spécifiques au domaine étudié qui est dans ce travail ' l'observation de la terre'.

Les travaux récents sur l'annotation sémantique ont montré que l'annotation sémantique représente une solution efficace pour résoudre le problème d'hétérogénéité et l'intégration de des données IoT, quelques soit le type de données : image, texte, etc.

D'après nos recherches, il n'existe que le travail [51] qui traite le problème de l'interprétation des éléments de datasets IoT. Cependant, ce dernier considère une image IoT comme un objet observé spécifique à un domaine donné sans prendre en considération les caractéristiques des images observées.

Dans notre travail, nous allons profiter des avantages de travaux cités pour proposer une interprétation des éléments des datasets IoT.

## **1.6. Conclusion**

Nous avons dans ce chapitre évoqué l'état de l'art afin de mieux cerner les concepts de base, les technologies existantes dans notre contexte et de mieux comprendre notre problématique. Le chapitre suivant présentera la conception.

# Chapitre 02 : Conception

## 2.1. Introduction

Afin de donner une annotation sémantique pour l'interprétation des IoT datasets de type image, il est important de conceptualiser les connaissances relatives aux datasets. Ces connaissances appartenant à des domaines variés (médical, agriculture, ...).

Dans ce chapitre, nous présentons la démarche de développement de l'ontologie qui constitue la contribution principale de ce travail. De ce fait, nous basons sur METHONTOLOGIE une méthodologie d'élaboration d'ontologies qui se réfère à un ensemble de tâches à accomplir pour la construction de l'ontologie. Ensuite, nous modélisons le développement du système d'annotation en utilisant les diagrammes d'UML.

## 2.2. Démarche globale de la solution

La démarche globale de la solution est divisée en deux parties principales : (i) la construction de l'ontologie qui sera le support de l'annotation sémantique et (ii) le processus d'annotation sémantique qui est basé sur plusieurs types d'annotation. La figure suivante présente une vue globale de cette démarche.

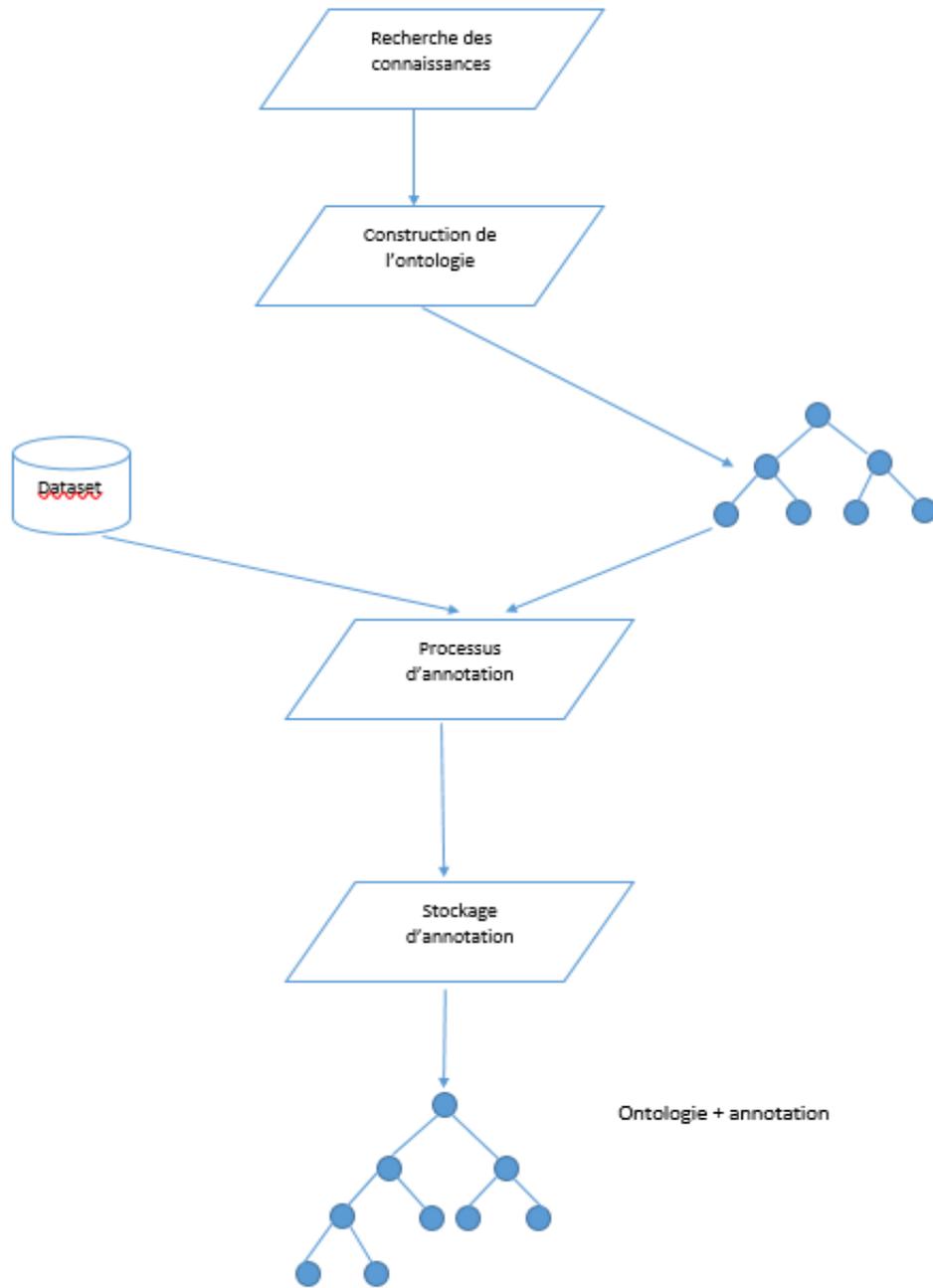


Figure 7. Démarche globale de la solution

## 2.3. Construction de l'ontologie

Le développement d'une ontologie consiste à construire l'ontologie en travaillant à sa spécification, à sa conception, à sa formalisation, à son implémentation et à sa maintenance. [52]. Dans ce travail, nous avons appliqué la méthodologie METHONTOLOGIE pour construire une ontologie qui permet de représenter les éléments des datasets de type image. Ce choix est motivé par le fait qu'elle est considérée comme mature et complète, et la plus appropriée pour développer une ontologie. Elle fournit ainsi suffisamment de détails sur ses activités [28], [53], [54], [55], [56].

Avec METHONDOLOGIE, la construction de l'ontologie est réalisée selon un cycle constitué de quatre étapes [57]:

- Spécification : pour identifier le but et les utilisateurs visés de l'ontologie,
- Conceptualisation : pour la structuration du domaine au niveau des connaissances,
- Formalisation : permet la traduction automatique du modèle conceptuel en utilisant des traducteurs) et
- Implémentation : permet l'expression du modèle formel à l'aide d'un langage d'implémentation.

### 2.3.1. Spécification

**Document de spécification.** La phase de spécification est une phase primordiale pour la construction d'une ontologie. Elle s'agit d'établir un *document de spécification* qui permet de spécifier : Domaine de connaissances, objectifs, utilisateurs et sources d'informations. La figure suivante représente les points de spécification de l'ontologie :

**Domaine de connaissances :**

**Objectifs :** L'objectif de cette ontologie est de capitaliser et représenter les connaissances des IoT datasets de type image. Elle est aussi utilisée comme une structure de stockage pour faciliter l'identification de type de contenu de ces dataset.

**Utilisateurs :** Les programmeurs dans les domaines d'apprentissage automatique, les utilisateurs des applications IoT, Analyste de données, ...

**Sources d'informations :** [www.Kaggle.com](http://www.Kaggle.com), ...

Figure 8. Document de spécification de l'ontologie

**Les questions de compétences.** Elles constituent l'élément-clé qui permet de caractériser de façon rigoureuse les connaissances que doit inclure une ontologie [52].

Pour cette ontologie nous avons dégagé un certain nombre de questions de compétences que nous résumons comme suit :

N°	Questions de compétences
QC1	Pour un dataset donné, quelles sont ses images ?
QC2	Pour chaque image appartient à un dataset, quelles sont ses caractéristiques ?
QC3	Pour chaque image, quel est son type ?
QC4	Pour une image donnée, quel est sa source (capteur IoT) ?

Tableau 2. Question de compétences de l'ontologie

### 2.3.2. Conceptualisation

Les réponses aux questions de compétences nous ont permis d'identifier l'ensemble de connaissances qui doivent être retenues pour représenter les éléments liés à un dataset de type image. Le tableau suivant montre un glossaire des termes utilisés.

Terme	Description
Dataset image	Un ensemble d'images
Image IoT	Est une donnée générée par un capteur IoT, et stockée dans un dataset image
Capteur IoT	Ou capteur : est un appareil qui peut générer des données images
Caractéristique	Est une caractéristique d'une image

Caractéristique spatiale	L'ensemble de caractéristique spatiale caractérisant une image
Caractéristique spectrale	L'ensemble de caractéristique spectrale caractérisant une image
Image hyperspectrale	Est un type d'image, obtenue par des enregistrements d'information dans une multitude de bandes spectrale
Image multispectrale	Est un type d'image
Image thermique	Est un type d'image
Image satellitale	Est un type d'image, elle fait référence à la capture d'images depuis l'espace par des capteurs situés sur le satellite
Image camera	Est un type d'image

*Tableau 3. Glossaire de quelques termes utilisés par l'ontologie*

**- La liste de concepts et sous concepts :**

Le tableau suivant présente quelques concepts et sous concepts utilisés dans notre ontologie :

<b>Nom du concept/ sous concept</b>	<b>Attributs</b>	<b>Type</b>
Dataset	<ul style="list-style-type: none"> <li>- date_creation_dataset</li> <li>- date_modification_dataset</li> <li>- url_dataset</li> <li>- taille_dataset</li> </ul>	Concept
Image	<ul style="list-style-type: none"> <li>- date_creation_image</li> </ul>	Concept
Capteur		Concept
Caractéristique		Concept
Caractéristique_spatiale		Sous concept
Caractéristique_spectrale		Sous concept

Image_Camera		Sous concept
Image_Hyperspectrale		Sous concept
Image_Multispectrale		Sous concept
Image_satellitaire		Sous concept
Image_Thermique		Sous concept
Bande		Sous concept

*Tableau 4. Liste de quelques concepts / sous concepts utilisés*

**- La liste des relations :**

Le tableau suivant présente quelques relations utilisées dans notre ontologie.

<b>Nom de relation</b>	<b>Concept source</b>	<b>Concept cible</b>	<b>Relation inverse</b>	<b>Cardinalité</b>
Has_a	Dataset	Image	/	1..n
has	Image	Caractéristique	describe	1..n
generated	Capteur	Image	Generated_by	1..n

*Tableau 5. Liste de quelques relations utilisées*

**- Liste des attributs :**

La liste des attributs suivante est ajoutée pour faire l'annotation :

<b>Nom du concept/ sous concept</b>	<b>Attributs</b>

Dataset	-
Image	- - format_image - taille_image - type_vue
Capteur	- marque_capteur
Caractéristique	
Caractéristique_spatiale	- resolution
Caractéristique_spectrale	-
Image_Camera	-
Image_Hyperspectrale	-
Image_Multispectrale	-
Image_satellitaire	-
Image_Thermique	-
Bande	- numero_bande - nom_bande - nombre_bande - longueur_bande_max - longueur_bande_min - valeur_max_pixel - valeur_min_pixel

*Tableau 6. Liste des attributs ajoutés.*

### 2.3.3. La formalisation

Pour la formalisation de l'ontologie, nous utilisons les logiques de description. Ces logiques sont considérées comme des fragments de la logique du premier ordre où les formules ont une variable libre pour décrire le concept et deux variables pour décrire les relations [58] Nous donnons dans la suite une partie des axiomes terminologique du modèle conceptuel proposé :

- $Capteur \subseteq Thing$
- $Caractéristique \subseteq Thing$
- $Dataset \subseteq Thing$
- $Image \subseteq Thing$
- $Image\_Camera \subseteq Image \subseteq Thing$
- $Image\_Hyperspectrale \subseteq Image \subseteq Thing$
- $Image\_Multispectrale \subseteq Image \subseteq Thing$
- $Image\_Satellitaire \subseteq Image \subseteq Thing$
- $Image\_Thermique \subseteq Image \subseteq Thing$
- $Caractéristique\_spatiale \subseteq Caractéristique \subseteq Thing$
- $Caractéristique\_spectrale \subseteq Caractéristique \subseteq Thing$
- $Bande \subseteq Caractéristique\_spectrale \subseteq Caractéristique \subseteq Thing$
- 
- ...

L'étape d'implémentation de l'ontologie fait l'objet du chapitre 03.

## 2.4. Un système pour l'annotation sémantique

Après la construction de l'ontologie qui présente les concepts pertinents, les relations entre concepts et les propriétés des datasets image, nous utilisons dans cette section quelques diagrammes UML pour modéliser le développement du système d'annotation :

- **Les fonctionnalités offertes par le système :**

Notre application permet d'offrir une liste de fonctionnalités à l'utilisateur qui peut être un développeur, un analyste de données, un utilisateur d'application IoT, etc. Ces fonctionnalités sont représentées comme des besoins fonctionnels sous forme de cas d'utilisation (figure suivante) :

**L'acteur :** Utilisateur

### Les cas d'utilisation :

- Télécharger un dataset
- Annoter un dataset
- Enregistrer annotation
- Afficher une annotation

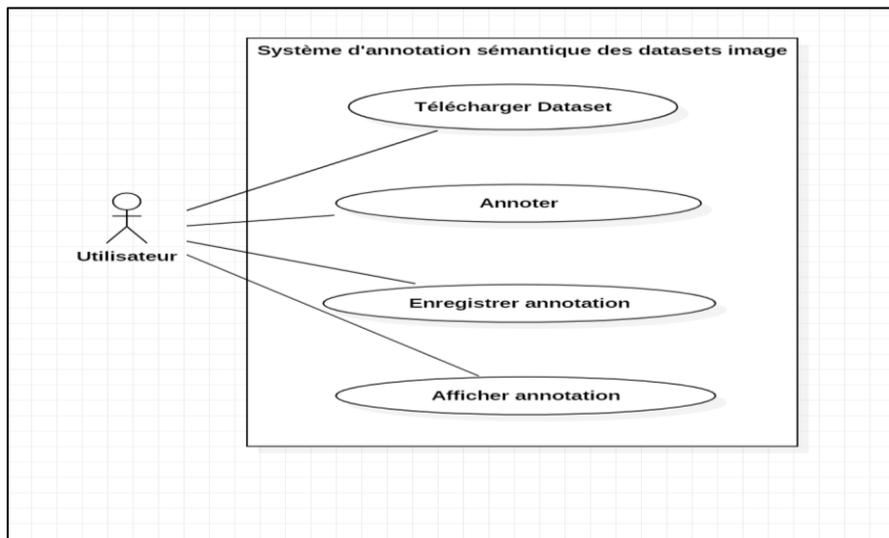


Figure 9. Diagramme de cas d'utilisation

### - Processus d'annotation

Nous modélisons la dynamique du système proposé par un diagramme d'activité d'UML. Un diagramme d'activité permet modéliser les processus en exposant l'enchaînement d'activités séquentielles et/ou parallèles à l'aide de sa notation riche.

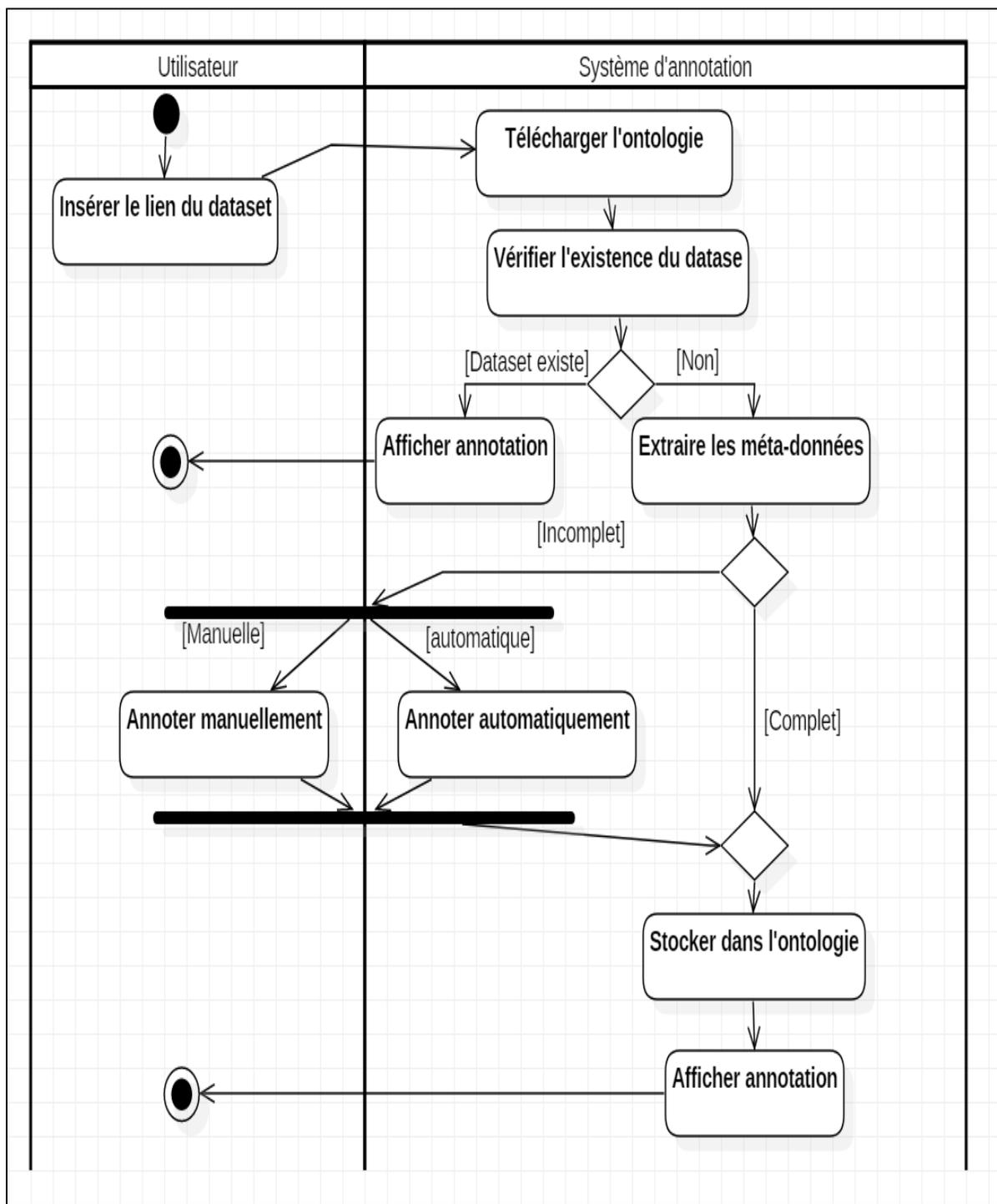


Figure 10. Diagramme d'activité du système d'annotation

Et pour donner plus de détails sur l'enchaînement et les objets qui interagissent pour la réalisation des cas d'utilisation nous utilisons le diagramme de séquence. Le diagramme de séquence donne une représentation graphique chronologique des interactions entre les acteurs

et les objets du système :

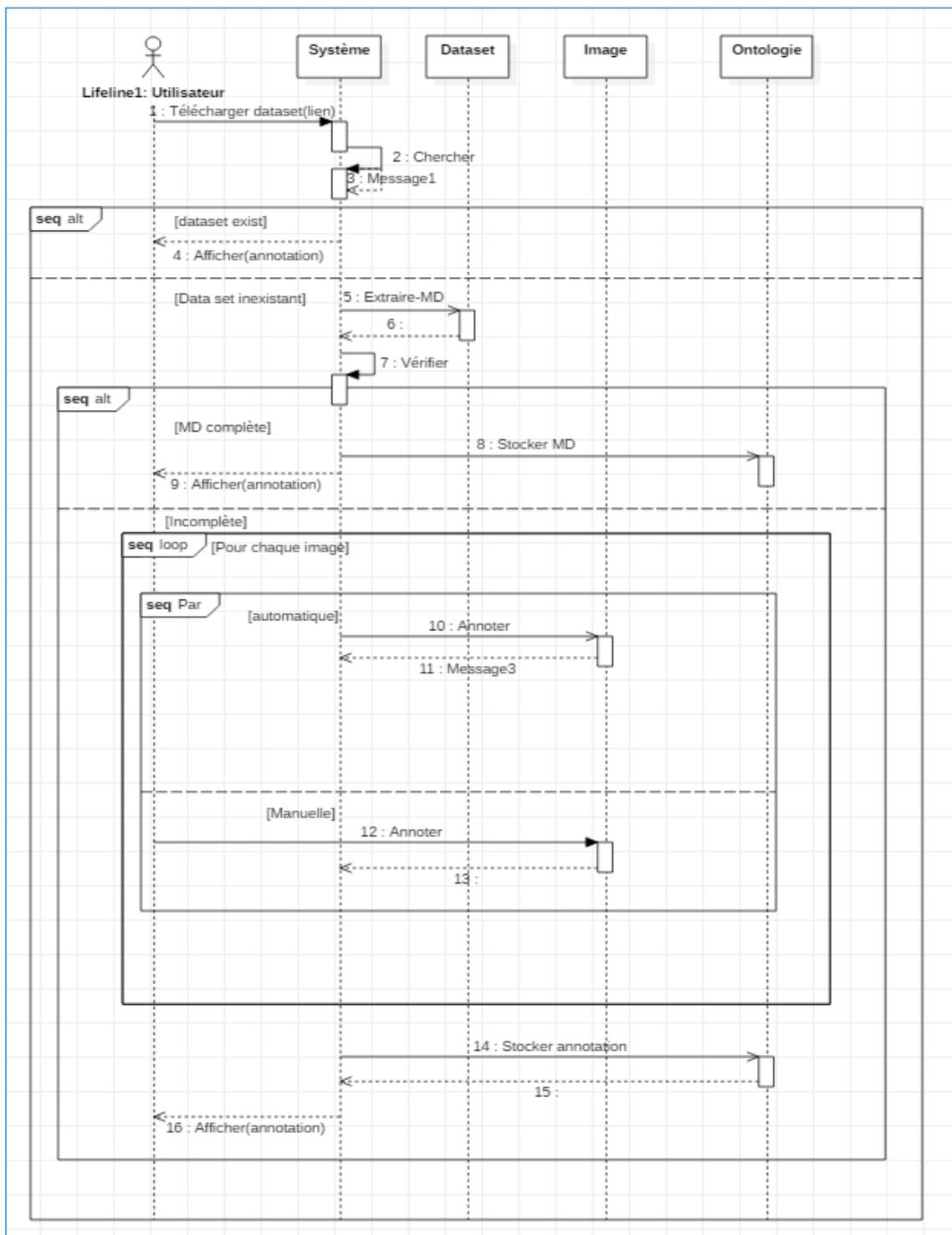


Figure 11. Diagramme de séquence du système

## **2.5. Conclusion**

Dans ce chapitre, nous avons décrit les étapes de construction de l'ontologie qui permet de représenter les concepts structurels d'un dataset, à savoir : dataset image, image, imageur, etc. Nous avons modélisé le développement du système d'annotation avec des diagrammes UML. Le diagramme des cas d'utilisation, d'activité et de séquences permettent de d'identifier les fonctionnalités, l'enchaînement et les différents objets du système.

L'implémentation de l'ontologie pour qu'elle soit manipulable ainsi du système d'annotation font l'objet du chapitre 03.

# Chapitre 03 : Implémentation

## 3.1. Introduction

Dans ce chapitre, nous allons présenter les étapes d'implémentation. Nous décrivons premièrement les outils d'implémentation. Nous décrivons ainsi l'implémentation de l'ontologie proposée à travers Owlready de Python. Ensuite, nous détaillons l'implémentation des étapes du processus d'annotation sémantique. Enfin, nous illustrons la vérification de l'ontologie et la validation du processus proposés à travers l'exécution d'un scénario.

## 3.2. Outils d'implémentation

Avant de commencer l'implémentation de notre système de stockage et d'annotation, nous allons spécifier les différents outils utilisés :

### **Python :**

Python est un langage de programmation interprété. C'est le langage de programmation open source le plus employé par les informaticiens [59].

Grâce à des bibliothèques spécialisées, Python s'utilise pour de nombreuses situations comme le développement logiciel, l'analyse de données, la gestion d'infrastructures, ou le traitement d'images. C'est ce qui en fait le langage de choix pour le développement de notre projet.

### **Flask :**

C'est est un Framework python permettant la réalisation d'un site web ou d'une api web [60]. Son principal avantage est d'être simple à utiliser mais sans perdre de fonctionnalités, de plus il peut quasiment tout faire grâce à de nombreuses extensions. On l'utilise pour construire notre API qui nous permet d'interagir avec notre ontologie à partir de requête HTTP (Hypertext Transfer Protocol).

### **Os :**

Le module Os de Python permet d'effectuer des opérations courantes liées au système

d'exploitation [61]. Il est indépendant par rapport au système d'exploitation de la machine.

Le module `os` en Python permet d'interagir avec les fonctionnalités du système d'exploitation et d'accéder aux informations comme la date de création des fichiers.

#### **Pillow :**

C'est une bibliothèque python permettant la manipulation des images [62]. Elle nous permet d'extraire quelques informations sur l'image telles que la résolution, l'espace couleur, ... Il est utilisé pour le traitement de nos images légers issues du dataset.

#### **Rasterio :**

Rasterio est un module très utile pour le traitement des données images géographiques [63]. Il nous permet de manipuler ces données et d'en extraire des informations pertinentes comme le nombre de bandes, ... On l'utilise pour le traitement de nos images lourds issues des dataset.

#### **ImgNotes :**

C'est une bibliothèque JQuery permettant d'ajouter des notes à une image [64]. Nous avons apporté quelques modifications à la bibliothèque pour l'adapter à nos objectifs. La bibliothèque est open source.

### **3.3. Implémentation de l'ontologie**

Après les étapes de spécification, conceptualisation et formalisation de l'ontologie, nous arrivons à son implémentation. En utilisant `Owlready2` de Python pour la manipulation des ontologies OWL. Une ontologie OWL permet de réaliser des inférences logiques et relier les connaissances entre elles [65]. Ainsi Python permet de rendre facile l'intégration des ontologies.

`Owlready` est module sous licence LGPG (Licence publique générale limitée) qui permet la programmation orientée ontologies en Python. Avec `Owlready` les entités d'une ontologie formelle sont manipulées comme s'il s'agissait d'objets Python classique.

La figure suivante montre une partie de la création de l'ontologie avec `Owlready` de Python :

```

Entrée [1]: from owlready2 import *
import numpy as np
import pandas as pd
import os
import time
from exif import Image

onto = get_ontology("http://test.org/onto.owl#")

* Owlready2 * Warning: optimized Cython parser module 'owlready2_optimized' is not available, defaulting to slower Python implementation

Entrée [2]: with onto:
class Image(Thing): pass
class date_creation_image (Image >> str, FunctionalProperty):pass
class type_vue (Image >> str, FunctionalProperty):pass
class taille_image (Image >> str, FunctionalProperty):pass
class nombre_band (Image >> int, FunctionalProperty):pass
class format_image (Image >> str, FunctionalProperty):pass
class Dataset(Thing): pass
class date_creation_dataset (Dataset >> str, FunctionalProperty):pass
class url_dataset (Dataset >> int, FunctionalProperty):pass
class date_modification_dataset (Dataset >> str, FunctionalProperty):pass
class taille_dataset (Dataset >> int, FunctionalProperty):pass
class Capteur(Thing): pass
class marque_capteur (Capteur >> str, FunctionalProperty):pass
class Caractéristique(Thing): pass
AllDisjoint([Image, Capteur, Caractéristique, Dataset])

```

Figure 12. Création de l'ontologie avec Owlready de Python

La figure suivante montre une partie du fichier OWL de l'ontologie créée :

```

146 <http://test.org/onto.owl#longueur_bande_min> <http://www.w3.org/2000/01/rdf-schema#domain> <http://test.org/onto.owl#Bande> .
147 <http://test.org/onto.owl#longueur_bande_min> <http://www.w3.org/2000/01/rdf-schema#range> <http://www.w3.org/2001/XMLSchema#integer> .
148 <http://test.org/onto.owl#generated> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#ObjectProperty> .
149 <http://test.org/onto.owl#generated> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#FunctionalProperty> .
150 <http://test.org/onto.owl#generated> <http://www.w3.org/2000/01/rdf-schema#domain> <http://test.org/onto.owl#Capteur> .
151 <http://test.org/onto.owl#generated> <http://www.w3.org/2000/01/rdf-schema#range> <http://test.org/onto.owl#Image> .
152 <http://test.org/onto.owl#generated_by> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#ObjectProperty> .
153 <http://test.org/onto.owl#generated_by> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#FunctionalProperty> .
154 <http://test.org/onto.owl#generated_by> <http://www.w3.org/2000/01/rdf-schema#domain> <http://test.org/onto.owl#Image> .
155 <http://test.org/onto.owl#generated_by> <http://www.w3.org/2000/01/rdf-schema#range> <http://test.org/onto.owl#Capteur> .
156 <http://test.org/onto.owl#has> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#ObjectProperty> .
157 <http://test.org/onto.owl#has> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#FunctionalProperty> .
158 <http://test.org/onto.owl#has> <http://www.w3.org/2000/01/rdf-schema#domain> <http://test.org/onto.owl#Image> .
159 <http://test.org/onto.owl#has> <http://www.w3.org/2000/01/rdf-schema#range> <http://test.org/onto.owl#Caractéristique> .
160 <http://test.org/onto.owl#describe> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#ObjectProperty> .
161 <http://test.org/onto.owl#describe> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#FunctionalProperty> .
162 <http://test.org/onto.owl#describe> <http://www.w3.org/2000/01/rdf-schema#domain> <http://test.org/onto.owl#Caractéristique> .
163 <http://test.org/onto.owl#describe> <http://www.w3.org/2000/01/rdf-schema#range> <http://test.org/onto.owl#Image> .
164 <http://test.org/onto.owl#has_a> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#ObjectProperty> .
165 <http://test.org/onto.owl#has_a> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#FunctionalProperty> .
166 <http://test.org/onto.owl#has_a> <http://www.w3.org/2000/01/rdf-schema#domain> <http://test.org/onto.owl#Dataset> .
167 <http://test.org/onto.owl#has_a> <http://www.w3.org/2000/01/rdf-schema#range> <http://test.org/onto.owl#Image> .

```

Figure 13. Une partie du fichier OWL.

On peut visualiser la hiérarchie des concepts en utilisant Protégé. Les figures suivantes montrent la hiérarchisation des concepts et l'ensemble des attributs en utilisant protégé :

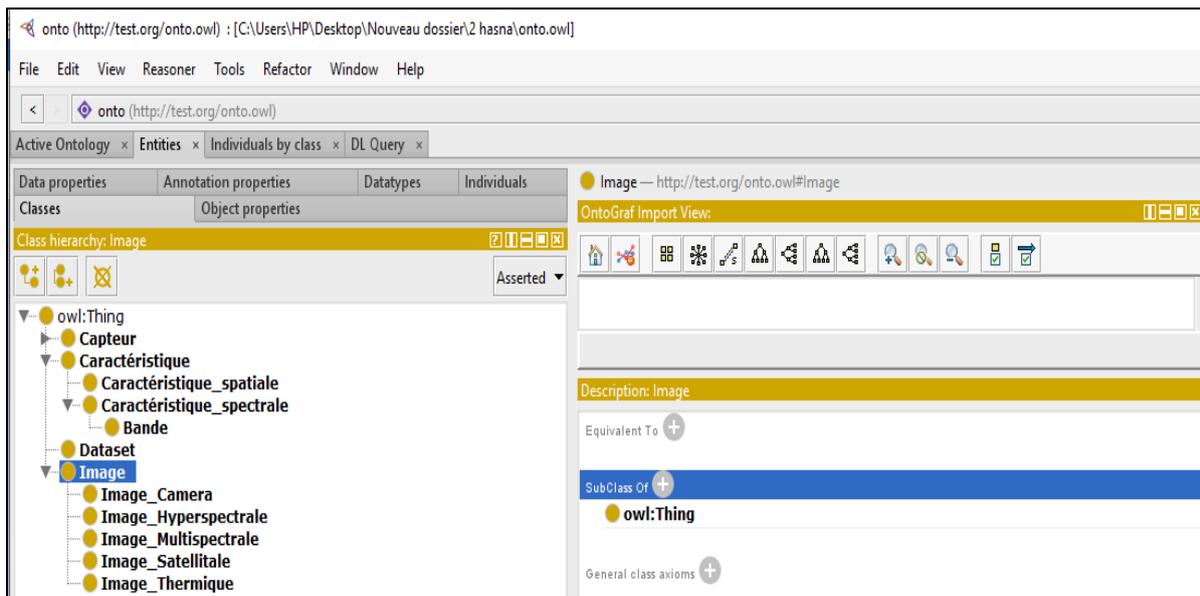


Figure 14. La hiérarchie des concepts de l'ontologie avec Protégé

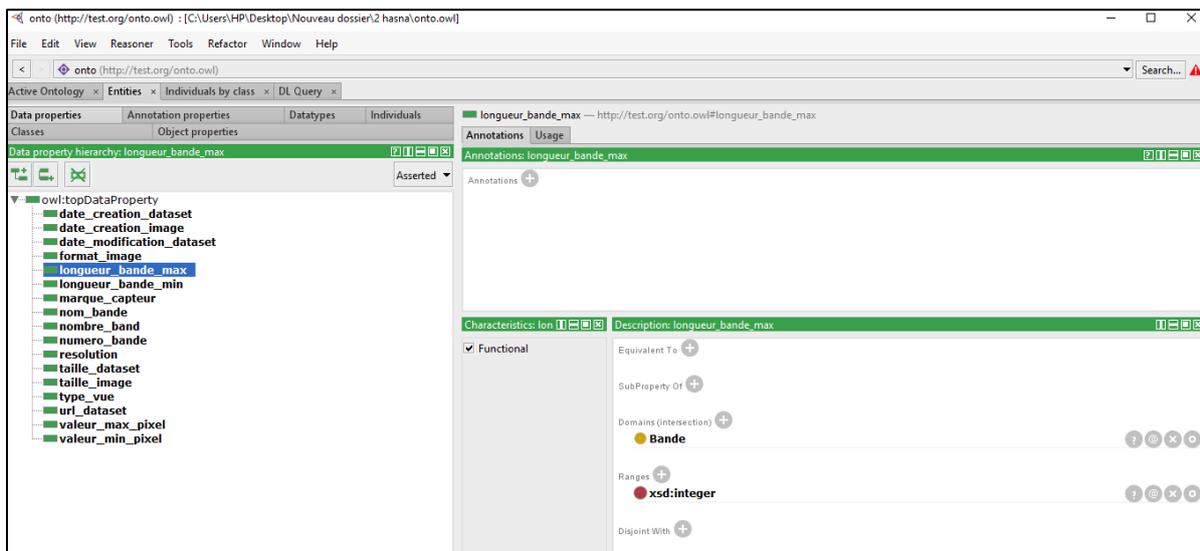


Figure 15. Visualisation de la liste de propriétés de données (data properties) avec Protégé

### 3.4. Implémentation du processus d'annotation

Dans cette section, nous montrons l'implémentation des quelques étapes du processus de l'annotation sémantique. Dans la figure suivante, on présente le code pour l'extraction du format des images du dataset. Elle prend en entrée le dataset et retourne la liste des formats des images.

```
Entrée [16]: #format de donnée dans Le dataste
# list pour enregistrer Les fichiers
res = []
# boucle
for file in os.listdir(dataset):
    # retourne Le nom du fichier
    split_tup = os.path.splitext(file)
    if(split_tup[1] not in res):
        #retourne l'extension
        res.append(split_tup[1])
print(res)

['.jpg', '.jfif', '.jpeg', '.db']
```

Figure 16. Le code d'extraction du format des images du dataset

La figure suivante permet de décrire le code pour l'annotation d'une image. Elle permet d'afficher le menu d'annotation de l'image, qui après remplissage par l'utilisateur, transfère l'annotation à l'ontologie pour instantiation de l'annotation.

```

221     maj(parseInt(position.value), parseInt(total.value));
222     $(window).load(function() {
223
224         cust_onEdit = function(ev, elem) {
225             var $elem = $(elem);
226             $('#noteDialog').remove();
227             return $('<div id="NoteDialog"></div>').dialog({
228                 title: "Editeur D'annotation",
229                 resizable: false,
230                 modal: true,
231                 height: "300",
232                 width: "450",
233                 position: {
234                     my: "left bottom",
235                     at: "right top",
236                     of: elem
237                 },
238                 buttons: {
239                     "Save": function() {
240                         var txt = $('textarea', this).val();
241                         $elem.data("note").note = txt;
242                         $(this).dialog("close");
243                     },
244                     "Delete": function() {
245                         $elem.trigger("remove");
246                         $(this).dialog("close");
247                     },
248                     "Cancel": function() {
249                         $(this).dialog("close");
250                     }
251                 },
252                 open: function() {
253                     $(this).css("overflow", "hidden");
254                     var textarea = $('<textarea id="txt" style="height:100%; width:100%;>');
255                     $(this).html(textarea);
256                     textarea.val($elem.data("note").note);
257                 }
258             });
259         };
260

```

Figure 17. Code pour l'Annotation

La figure suivante décrit l'instanciation d'un nouveau dataset dans notre ontologie. La fonction *Get* vérifie d'abord si le dataset n'est pas déjà instancié. Si c'est le cas une notification est transférer à l'utilisateur. Dans le cas contraire on fait une instanciation de l'ontologie avec l'ensemble des images qu'elle contient après extraction de caractéristiques.

```

class Data(Resource):
    def get(self, lien):
        if dataset_exist(lien):
            base={}
            ligne = {'lien': 1}
            base[0]=ligne
            i=1
            for file in os.listdir(lien):
                # this will return a tuple of root and extension
                ligne = {'lien': file}
                base[i]=ligne
                i=i+1
            return base
        else:
            base={}
            ligne = {'lien': 0}
            base[0]=ligne
            i=1
            for file in os.listdir(lien):
                # this will return a tuple of root and extension
                ligne = {'lien': file}
                base[i]=ligne
                i=i+1
            #instanciation du dataset
            dataset=Dataset(lien);
            addInfo(dataset,base)
            return base

```

Figure 18. Instanciation d'un nouveau dataset

### 3.5. Un scénario d'exécution

#### Un scénario d'exécution

Nous allons montrer l'exécution du système à travers un scénario d'exécution : Le processus d'annotation commence par une interaction entre l'utilisateur et le système. L'utilisateur comme nous avons indiqué peut-être un analyste de données qui veut sélectionner un dataset image parmi de nombreux dataset disponible sur le Web. L'utilisateur peut visualiser le contenu du dataset qu'il contient 51 fichiers images plus un fichier caché. La figure suivante montre le contenu du dataset test\_crop\_image.

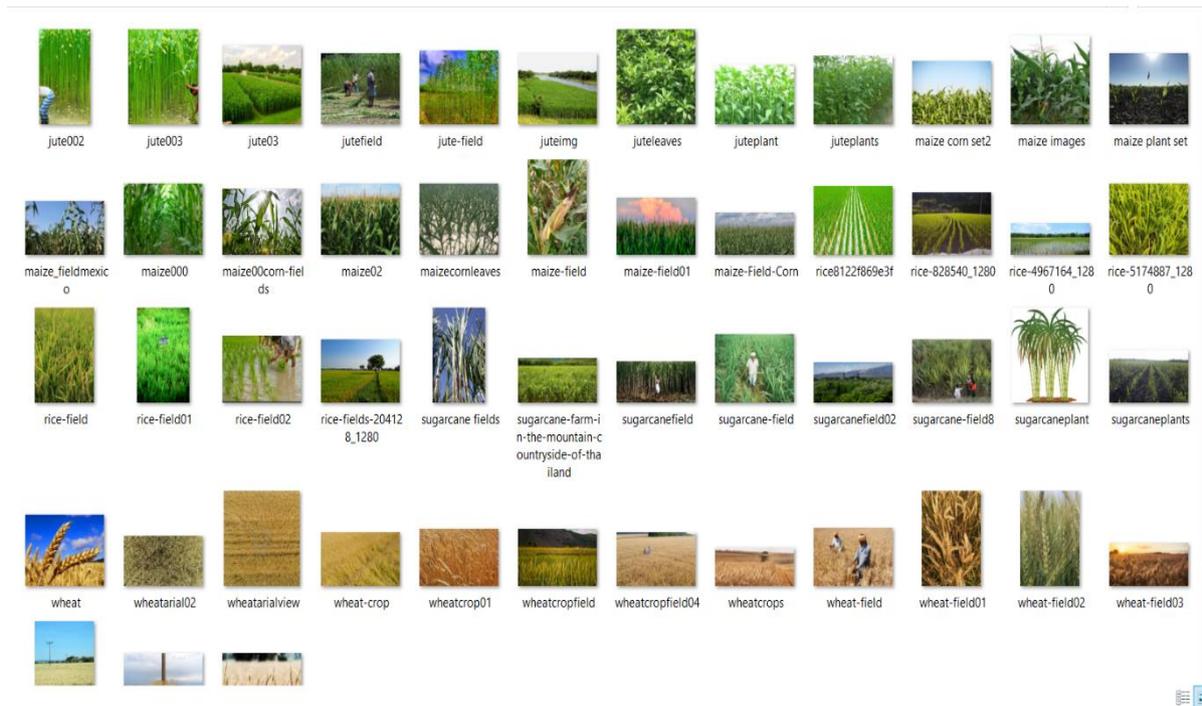


Figure 19. Visualisation du contenu d'un dataset

Le parcours de ces fichiers ne peut pas donner suffisamment d'informations sur le contenu du dataset pour permettre à un analyste de prendre la décision que ce dataset est approprié pour son application (domaine) ou non. L'utilisateur peut alors télécharger le dataset via notre système. La figure suivante montre la fenêtre qui permet de charger le dataset. L'utilisateur saisit le nom du répertoire du dataset et après appuie sur le bouton charger.

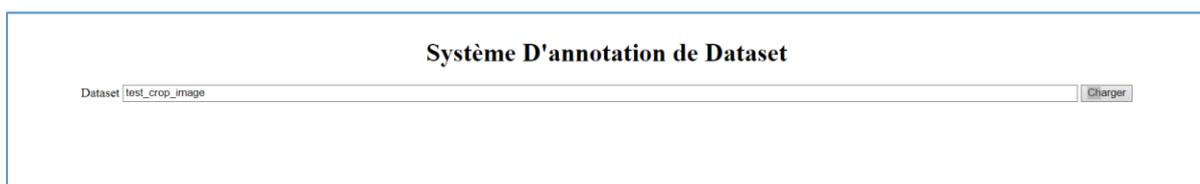


Figure 20. Page d'accueil du système d'annotation

La figure suivante indique le message d'erreur au cas où le répertoire du dataset est inexistant.

## Système D'annotation de Dataset

Dataset

**Ce repertoire n'existe pas. Veuillez saisir s'il vous un qui est valide.**

Figure 21. Message d'erreur en cas de répertoire existant.

La figure suivante permet d'afficher les informations du dataset après extraction des caractéristiques. On distingue les informations du dataset en haut et celles de l'image en bas à droite.

The screenshot shows the 'Système D'annotation de Dataset' interface. At the top, the dataset name is 'test\_crop\_image' and there is a 'Charger' button. Below this, there are two columns of metadata. The left column contains: 'Nom Fichier' (test\_crop\_image), 'Date Creation' (Wed Jun 8 10:37:28 2022), 'Taille' (9949622), and 'Auteur' (empty). The right column contains: 'Nombre Fichier' (52), 'Date Modification' (Wed Jun 8 10:39:58 2022), 'URL' (empty), and 'Format de Fichier' (['jpg', 'jif', 'jpeg', 'tiff']). In the center, there is a large image of a green field under a cloudy sky. To the right of the image, there is a section titled 'Caractéristique de l'image' with the following fields: 'Nom du Fichier' (jute-field), 'Format de l'image' (jpg), 'Taille image' (83630), 'Resolution image' (466 720), 'Nombre de band' (3), 'Valeur Min-Max Pixel' ((0, 255) (0, 255) (0, 255)), 'Espace couleur' (empty), 'Appareil D'acquisition' (empty), and 'Date de creation' (Wed Jun 8 10:37:28 2022).

Figure 22. Affiche des caractéristiques du dataset

La figure suivante indique de façon plus visible les informations du dataset. On remarque que le champ auteur et url sont vides. Le système n'a pas pu les extraire car ils sont inexistantes. L'utilisateur peut les remplir manuellement s'il a l'information.

### Système D'annotation de Dataset

Dataset

<p>Nom Fichier <input type="text" value="test_crop_image"/></p> <p>Date Creation <input type="text" value="Wed Jun 8 10:37:26 2022"/></p> <p>Taille <input type="text" value="9949622"/></p> <p>Auteur <input type="text"/></p>	<p>Nombre Fichier <input type="text" value="52"/></p> <p>Date Modification <input type="text" value="Wed Jun 8 10:39:56 2022"/></p> <p>URL <input type="text"/></p> <p>Format de Fichier <input type="text" value="['jpg', 'jif', 'jpeg', 'dib']"/></p>
---	---

*Figure 23. Détails des caractéristiques du dataset*

La figure suivante en détails une image du dataset et ses caractéristiques à droite. On aperçoit aussi que le champ appareil d'acquisition n'a pas pu obtenu car il est inexistant dans les métadonnées du dataset. L'utilisateur peut le saisir ou encore spécifier le type d'image de photographie qui permettra de faciliter la classification de l'image.



#### Caractéristique de l'image

Nom du Fichier	<input type="text" value="jute-field"/>
Format de l'image	<input type="text" value="jpg"/>
Taille image	<input type="text" value="83630"/>
Resolution image	<input type="text" value="466 720"/>
Nombre de band	<input type="text" value="3"/>
Valeur Min-Max Pixel	<input type="text" value="(0, 255) (0, 255) (0, 255)"/>
Espace couleur	<input type="text" value="RGB"/>
Appareil D'acquisition	<input type="text"/>
Date de creation	<input type="text" value="Wed Mar 10 10:37:26 2021"/>
Type Photographie	<input type="text" value="Selectionner le type d'image"/>

*Figure 24. Détails des caractéristiques de l'image*

La figure suivante indique les boutons pour annoter une image ou encore pour passer à l'image suivante du dataset.



1 sur 4

Annoter Change Image

Figure 25. Bouton annoter et changer image

La figure suivante présente le changement d'image après avoir utilisé le bouton image. On remarque le change du 1 sur 4 précédant par 2 sur 4. On constate également ce changement dans les caractéristiques de l'image à droite où les informations de l'image ont changé.



[ '.jpg', '.tif', '.png', '.dib' ]

### Caractéristique de l'image

Nom du Fichier	jule-001
Format de l'image	.jpg
Taille image	65610
Resolution image	466 720
Nombre de band	3
Valeur Min-Max Pixel	(0, 255) (0, 255) (0, 255)
Espace couleur	RGB
Appareil D'acquisition	
Date de creation	Sat Mar 10 10:37:26 2022
Type Photographie	Selectionner le type d'image

2 sur 4

Enregistrer Change Image

Figure 26. Changement de l'image du dataset

La figure suivante indique le clic sur le bouton annoter qui est la première étape pour annoter l'image.



Figure 27. Etape 1 de l'annotation d'une image

La figure suivante indique la phase suivante de l'annotation où l'utilisateur choisit une zone de l'image à annoter. Ensuite s'affiche le menu où l'utilisateur saisit sa note et ensuite il l'enregistre.

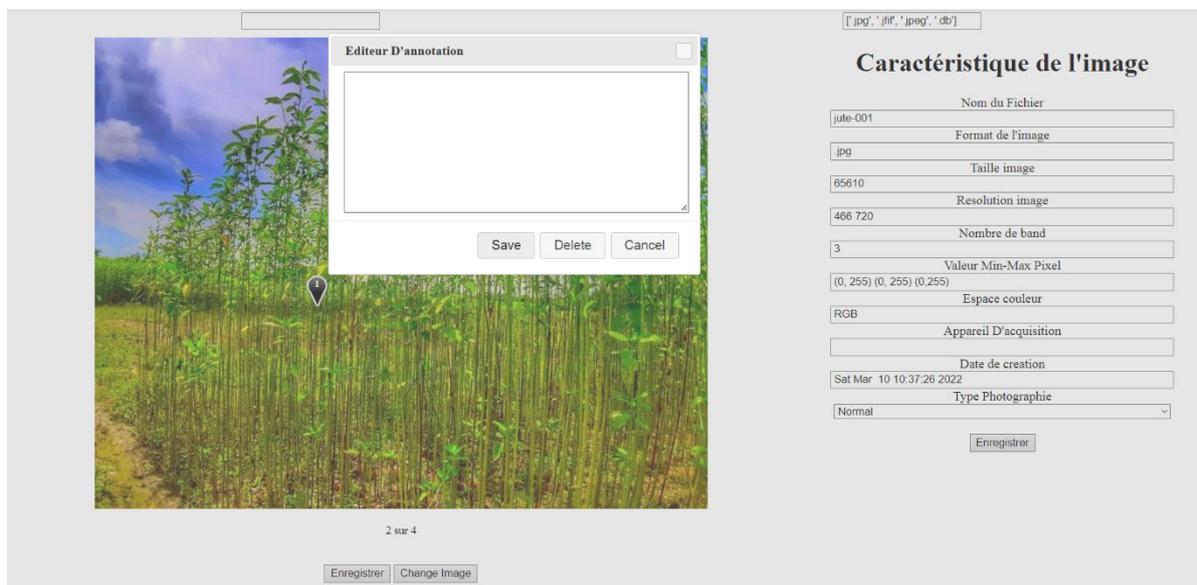


Figure 28. Etape 2 de l'annotation d'une image

La figure suivante indique l'annotation du point choisie comme plante. L'utilisateur en y cliquant dessus peut supprimer ou modifier l'annotation.

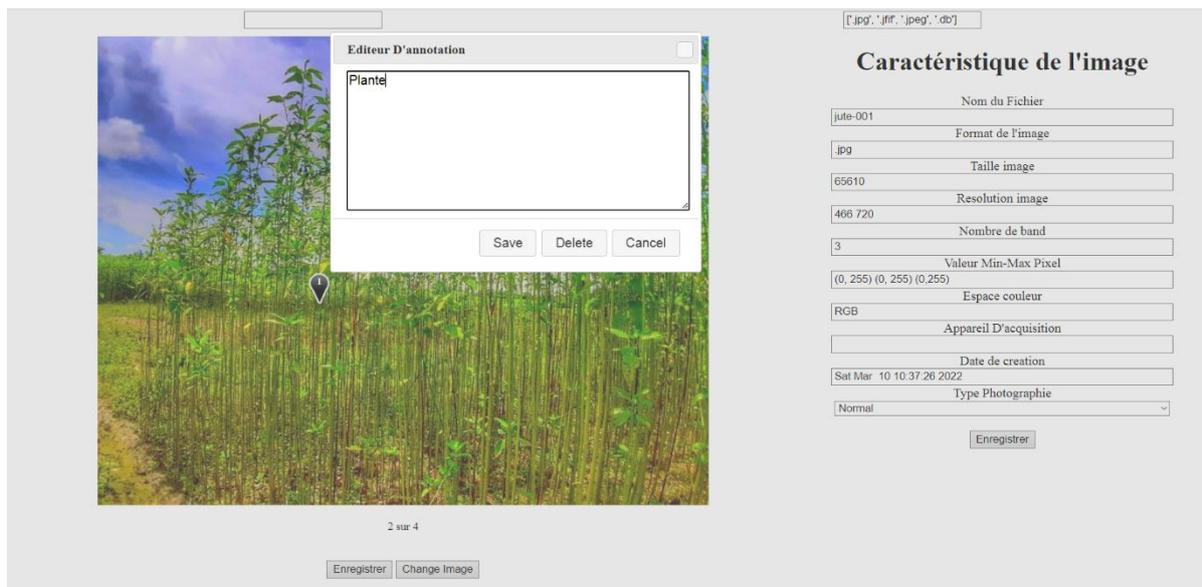


Figure 29. Etape 3 de l'annotation d'une image

La figure suivante présente le point annoté par l'utilisateur. Il porte le numéro 1 vue que c'est la première annotation. L'utilisateur clique sur enregistrer pour sauvegarder l'annotation de l'image.

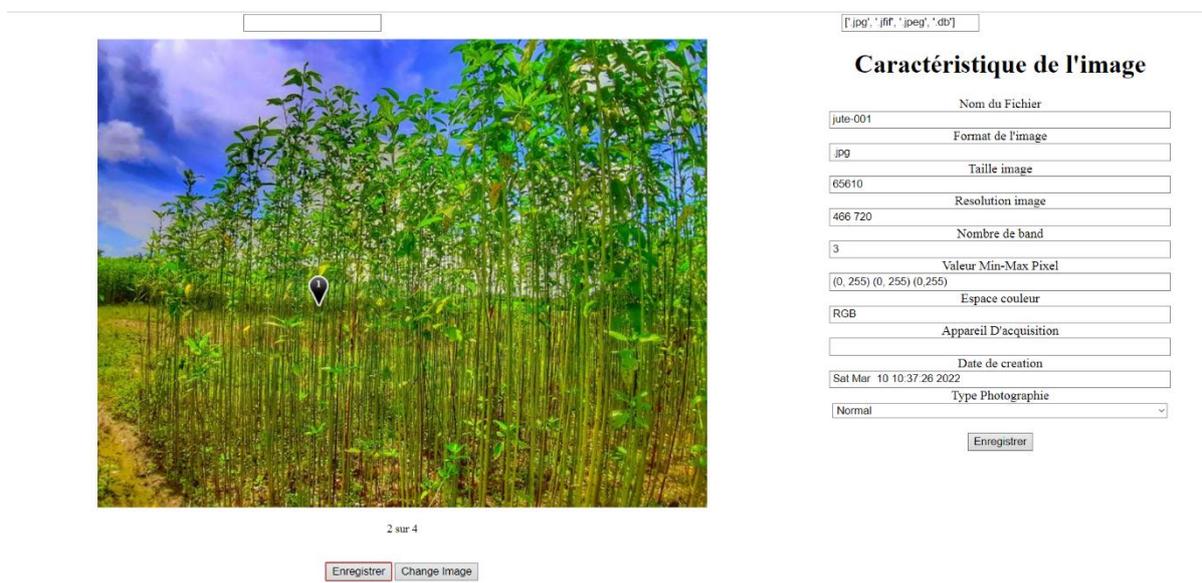


Figure 30. Etape 4 de l'annotation d'une image

La figure suivante indique le passage de la souris sur une zone annotée. On y aperçoit l'annotation de la zone (Plante).

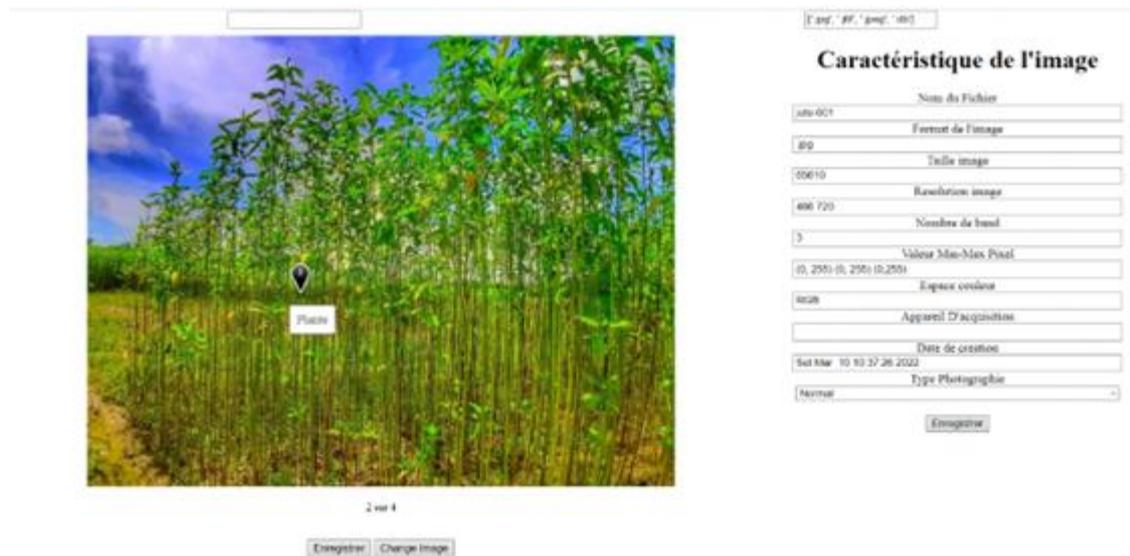


Figure 31. Passage de la souris sur une zone annotée.

La figure suivante indique un autre type d'annotation de l'image ou l'utilisateur spécifie le type de la photographie (Image) comme normal ou comme vue du ciel. Elle permet d'apporter un impact considérable lors de la classification de l'image. A la fin l'utilisateur clique sur enregistrer pour sauvegarder les changements dans les caractéristiques de l'image.

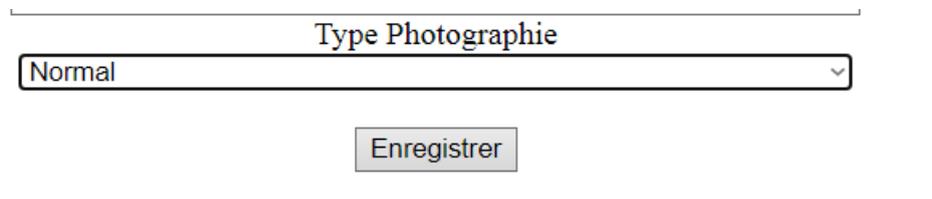


Figure 32. Annotation du type de photographie (Image)

### 3.6. Conclusion

Nous avons dans ce chapitre décrit les outils et les étapes de l'implémentation de la solution proposée. L'exécution est faite à travers un scénario.

# Conclusion générale

Dans ce mémoire de master, nous avons présenté la théorie autour (i) des datasets et les données IoT de type image (ii). Les ontologie et l'annotation sémantique. Nous avons constaté la difficulté de sélectionner un dataset avec le manque d'information sur son contenu. Afin de simplifier la recherche et la sélection des datasets pour des utilisations ultérieures. Nous avons proposé un simple système pour annoter le contenu ces datasets. Les connaissances de dataset sont décrites dans une ontologie développée s'appuyant sur la méthodologie METHONTOLOGIE. La solution que nous avons proposée est indépendante du domaine. Elle

- L'annotation des datasets de structures et de données hétérogènes et générées à partir des sources hétérogènes.
- Le stockage des métadonnées et l'instanciation de l'ontologie au fur et à mesure de l'annotation des datasets.
- L'instanciation permet d'évoluer l'ontologie.
- L'instanciation et le stockage de l'ontologie permet d'annoter à chaque fois donc on peut arriver à une bonne structure de stockage qui permet de limiter la recherche les recherches des datasets sur le Web.

Comme perspectives pour ce travail, nous raffinerons ce système basé sur d'autres caractéristiques des différents types d'image et les règle SWRL.

# Références

## Bibliographie

- [1] N. Aussenac-gilles, “Annotation Sémantique de Documents es pour la Recherche d’Information TH ESE EN CO-TUTELLE,” 2010.
- [2] Y. Ma, L. Audibert, and A. Nazarenko, “Ontologies étendues pour l’annotation sémantique,” 2009.
- [3] T. R. Gruber, “A translation approach to portable ontology specifications,” *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, 1993, doi: 10.1006/knac.1993.1008.
- [4] I. Saleh, “Les enjeux et les défis de l’ Internet des Objets ( IdO ) Issues and Challenges of the Internet of Things ( IoT ),” 2017.
- [5] J. Le hir, “Conception mixte d’un capteur d’images intelligent intégré à traitements locaux massivement parallèles,” 2019.
- [6] B. D. MONTASSAR, “IMAGE NUMÉRIQUE.” <https://imagenumeriqueblog.wordpress.com/les-caracteristiques-dune-image-bitmap/definition-dune-image/> (accessed Jun. 01, 2022).
- [7] “Les Pixels : Définition Et Caractéristiques.” <https://www.gralon.net/articles/photo-et-video/photo-et-video/article-les-pixels---definition-et-caracteristiques-2014.htm> (accessed Jun. 01, 2022).
- [8] “Comprendre une image satellitaire.” <https://cms.geobretagne.fr/content/comprendre-une-image-satellitaire> (accessed Jun. 01, 2022).
- [9] “Les images de télédétection.” <https://eo.belspo.be/fr/les-images-de-teledetection> (accessed Jun. 01, 2022).
- [10] Kaggle, “Kaggle Competition Review-Dstl Satellite Imagery Feature Detection,” 2017.
- [11] “IMAGERIE HYPERSPECTRALE.” <https://www.ino.ca/fr/solutions/hyperspectrale> (accessed Jun. 01, 2022).
- [12] “L’imagerie thermique: jusqu’où permet-elle de voir ?” [https://www.flirmedia.com/MMC/CSV/Tech\\_Notes/TN\\_0002\\_FR.pdf](https://www.flirmedia.com/MMC/CSV/Tech_Notes/TN_0002_FR.pdf) (accessed Jun.

- 01, 2022).
- [13] “L’imagerie thermique.” <https://parlonssciences.ca/ressources-pedagogiques/documents-dinformation/limagerie-thermique> (accessed Jun. 01, 2022).
  - [14] M. Gruninger and M. S. Fox, “Methodology for the Design and Evaluation of Ontologies,” 1995.
  - [15] Y. Hayashi *et al.*, “A Multiple View Authoring Tool for Modeling Training Materials,” 1999.
  - [16] “Quels sont les composants principaux d’une ontologie?” <https://123dok.net/article/quels-sont-les-composants-principaux-d-une-ontologie.q76dgkry> (accessed Jun. 01, 2022).
  - [17] “Constituants d’une ontologie.” <https://123dok.net/article/constituants-ontologie-ontologie-composants.z1d916p8> (accessed Jun. 01, 2022).
  - [18] “Composantes de l’ontologie.” <https://123dok.net/article/composantes-l-ontologie-qu-l-ontologie-ingenierie-connaissances.zggwxj7z> (accessed Jun. 01, 2022).
  - [19] N. F. Noy, “ontologie Développement d ’ une ontologie 101 : Guide pour la création de votre première ontologie,” no. June, 2014.
  - [20] “RDF.” <https://developer.mozilla.org/fr/docs/Glossary/RDF> (accessed Jun. 01, 2022).
  - [21] “OWL Web Ontology Language.” <https://www.w3.org/TR/owl-features/> (accessed Jun. 01, 2022).
  - [22] “Ontologie (informatique) - Définition et Explications.” <https://www.techno-science.net/definition/324.html> (accessed Jun. 01, 2022).
  - [23] I. Horrocks, “DAML + OIL : a Description Logic for the Semantic Web,” pp. 1–7, 2002.
  - [24] “Méthodes et méthodologies de construction d’ontologies.” <https://123dok.net/article/methodes-et-methodologies-de-construction-d-ontologies.yee2r4ry> (accessed Jun. 01, 2022).
  - [25] M. Uschold and K. Martin, “Towards a Methodology for Building Ontologies,” 1995.

- [26] N. F. Noy and D. L. McGuinness, “Ontology Development 101: A Guide to Creating Your First Ontology,” 2001. doi: 10.1016/j.artmed.2004.01.014.
- [27] Y. Sure, C. Tempich, and D. Vrandečić, “Ontology Engineering Methodologies,” in *Semantic Web Technologies*, Chichester, UK: John Wiley & Sons, Ltd, 2006, pp. 171–190.
- [28] M. Fernández, A. Gómez-Pérez, and N. Juristo, “MÉTHONTOLOGIE : de l’art ontologique vers l’ingénierie ontologique. Dans Actes de l’AAAI97 Symposium de printemps,” pp. 33–40, 1997.
- [29] A. Farquhar, R. Fikes, and J. Rice, “The ontolingua server: A tool for collaborative ontology construction,” *Int. J. Hum. Comput. Stud.*, vol. 46, no. 6, pp. 707–727, Jun. 1997, doi: 10.1006/ijhc.1996.0121.
- [30] “Méthodologies de construction d’ontologies.” [https://123dok.net/article/méthodologies-de-construction-d-ontologies.z126wovy](https://123dok.net/article/methodologies-de-construction-d-ontologies.z126wovy) (accessed Jun. 01, 2022).
- [31] M.-A. Storey *et al.*, “Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protégé,” 2001.
- [32] A. FARQUHAR, R. FIKES, and J. RICE, “The Ontolingua Server: a tool for collaborative ontology construction,” *Int. J. Hum. Comput. Stud.*, vol. 46, no. 6, pp. 707–727, Jun. 1997, doi: 10.1006/ijhc.1996.0121.
- [33] “Méthodes et méthodologie de construction d’ontologie.” <https://www.rapport-gratuit.com/methodes-et-methodologie-de-construction-dontologie/> (accessed Jun. 01, 2022).
- [34] S. P., “‘A Weather Ontology for Predictive Control in Smart Homes,’ MSc dissertation, Dept. of Software Engineering and Internet Computing,” vol. 2013, p. 183, 2013.
- [35] L. Denoue, P. Chiu, and T. Fuse, “Shared freeform input for note taking across devices,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2003, no. January, pp. 710–711, doi: 10.1145/765891.765943.
- [36] E. Oren, K. H. Möller, S. Scerri, S. Handschuh, and M. Sintek, “What Are Semantic Annotations,” Galway, Ireland, 2006.

- [37] S. Bechhofer, L. Carr, C. Goble, S. Kampa, and T. Miles-Board, “The Semantics of Semantic Annotation,” in *the 1st International Conference on Ontologies, Databases and Applications of SEMantics (ODBASE)*, 2002, pp. 1152–1167.
- [38] A. Ben Abacha and P. Zweigenbaum, “Annotation et interrogation sémantiques de textes médicaux,” no. i, 2009.
- [39] B. Habert, “Portrait de linguiste(s) à l’instrument.,” pp. 124–132, 2006.
- [40] F. Liu, P. Li, and D. Deng, “Device-Oriented Automatic Semantic Annotation in IoT,” *J. Sensors*, vol. 2017, Jun. 2017, doi: 10.1155/2017/9589064.
- [41] P. Barnaghi, W. Wang, L. Dong, and C. Wang, “A linked-data model for semantic sensor streams,” *Proc. - 2013 IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput. GreenCom-iThings-CPSCom 2013*, pp. 468–475, 2013, doi: 10.1109/GreenCom-iThings-CPSCom.2013.95.
- [42] S. Kolozali, M. Bermudez-Edo, D. Puschmann, F. Ganz, and P. Barnaghi, “A knowledge-based approach for real-time IoT data stream annotation and processing,” in *Proceedings - 2014 IEEE International Conference on Internet of Things, iThings 2014, 2014 IEEE International Conference on Green Computing and Communications, GreenCom 2014 and 2014 IEEE International Conference on Cyber-Physical-Social Computing, CPS 20*, 2014, no. iThings, pp. 215–222, doi: 10.1109/iThings.2014.39.
- [43] L. Gao, “Semantic annotation and reasoning for sensor data streams,” pp. 66–76, 2015.
- [44] J. Bing, “Research on semantic-based service architecture and key algorithms for the internet of things,” Jilin University, Changchun, China, 2013.
- [45] Z. Ming, “Research on several key issues in internet of things applications,” Beijing University of Posts and Telecommunications, Beijing, China, 2014.
- [46] S. Lin, J. Li, and C. Yu, “Dynamic Data Driven-based Automatic Clustering and Semantic Annotation for Internet of Things Sensor Data,” vol. 31, no. 6, pp. 1789–1801, 2019.
- [47] B. Sejdiu, F. Ismaili, and L. Ahmedi, “A management model of real-time integrated semantic annotations to the sensor stream data for the IoT,” *WEBIST 2020 - Proc. 16th Int. Conf. Web Inf. Syst. Technol.*, no. Webist, pp. 59–66, 2020, doi:

10.5220/0010111500590066.

- [48] M. Haghgoo, A. Nazary Aghche Mazary, and A. Monti, “SiSEG-Auto Semantic Annotation Service to Integrate Smart Energy Data,” *Energies*, vol. 15, no. 4, 2022, doi: 10.3390/en15041428.
- [49] T. Helmy, “A Generic Framework for Semantic Annotation of Images,” *Int. J. Image Graph.*, vol. 18, no. 3, pp. 1–22, 2018, doi: 10.1142/S0219467818500134.
- [50] Y. Wang, “Automatic Semantic Annotation of News Images in Mobile Internet of Things and Construction of Semantic Internet of Things System,” pp. 0–11, 2022.
- [51] S. Matwin and R. Goebel, *The series of articles in artificial intelligent*. 2019.
- [52] A. Gómez-Pérez, “Développement récents en matière de conception, de maintenance et d’utilisation des ontologies,” *Terminol. Nouv.*, vol. 19, pp. 9–20, 1999.
- [53] Z. Zakaria, S. Kasim, M. Husna Nur Hasbullah, and A. A. Azadin, “The Development of Personality Ontology Based on the Methontology Approach,” vol. 7, pp. 73–76, 2018.
- [54] M. Fernández-López and A. Gómez-Pérez, “Overview and analysis of methodologies for building ontologies,” *Knowledge Engineering Review*, vol. 17, no. 2. pp. 129–156, 2002, doi: 10.1017/S0269888902000462.
- [55] R. Iqbal, M. A. A. Murad, A. Mustapha, and N. M. Sharef, “An Analysis of Ontology Engineering Methodologies: A Literature Review,” *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, no. 16, pp. 2993–3000, Sep. 2013, doi: 10.19026/rjaset.6.3684.
- [56] “Méthodologies de création d’une ontologie.” Méthodologies de création d’une ontologie (accessed Jun. 01, 2022).
- [57] J. Mercantini, E. Chouraqui, and N. Tourigny, “Élaboration d ’ ontologies à partir de corpus en utilisant la méthode d ’ ingénierie des connaissances KOD,” 2007.
- [58] F. Amourache, Z. Boufaïda, and L. Yahiaoui, “Construction d’une ontologie basée compétence pour l’annotation des CVs/Offres d’emploi.,” in *Conference on Software Engineering and Artificial Intelligence (MCSEAI)*, 2008, pp. 1–7.
- [59] La redaction, “Python : définition et utilisation de ce langage informatique,” 2020.

- <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/> (accessed Jun. 11, 2022).
- [60] P. Chopinet, “Python : Comment faire une api web avec Flask,” 2021. <https://blog.jaaj.dev/2021/04/20/Comment-faire-une-api-web-en-python.html> (accessed Jun. 11, 2022).
- [61] Pythonforge, “Le module os en Python – système d’exploitation,” 2021. <https://pythonforge.com/module-os-systeme-dexploitation/#:~:text=Le module os de Python,importe quelle système d%27exploitation.> (accessed Jun. 11, 2022).
- [62] Q. Vaucher, “Python Imaging Library (PIL),” 2017. <https://he-arc.github.io/livre-python/pillow/index.html> (accessed Jun. 11, 2022).
- [63] M. Kaushik, “Reading and Visualizing GeoTiff | Satellite Images with Python.” <https://towardsdatascience.com/reading-and-visualizing-geotiff-images-with-python-8dcca7a74510> (accessed Jun. 11, 2022).
- [64] Waynegm, “ImgNotes,” 2017. <https://github.com/waynegm/imgNotes> (accessed Jun. 11, 2022).
- [65] J. Lamy and H. Berthelot, “Ontopy : programmation orientée ontologie en Python To cite this version : HAL Id : hal-01168286 Ontopy : programmation orient ´ ee ontologie en Python,” 2015.