

République algérienne démocratique et populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique Université

Kasdi Merbah – Ouargla



Faculté des Nouvelles Technologies de l'Information et de la Communication

Département des Sciences Informatiques et des Technologies de l'Information

Mémoire de Master Académique

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Informatique Industrielle

Présenté par : Lemlem Rihab et Kiouer Abir

Thème

Construction d'une ontologie pour la classification
des commentaires (cas des réseaux sociaux algériens)

Soutenu le : 23/06/2024

Membre du jury : Dr. Hamrouni et Dr. Bouhyaoui

Dr. Abderrahim Mohammed El Amine Encadreur Université Ouargla

Année académique : 2023/2024

Dédicace

“

Je remercie Allah, Le Tout-Puissant, pour Sa grâce et Son soutien dans l'achèvement de cette recherche.,

À celui qui m'a donné tout ce qu'il possède pour que je réalise ses espoirs, à celui qui m'encourageait toujours à avancer pour atteindre mes objectifs, à l'homme qui possède une humanité inébranlable, à celui qui a veillé sur mon éducation avec d'énormes sacrifices traduits par son respect pour la connaissance, à ma première école de vie, mon cher père, Mohammed Lemlem, que Dieu prolonge sa vie ; à celle qui a donné tout son être, tout son amour et sa tendresse, à celle qui a tout enduré, qui m'a protégé et a été mon soutien dans les moments difficiles, et dont les prières pour moi ont été un soutien constant, me guidant pas à pas dans mon travail, à celle envers qui je trouve un réconfort dès que je me souviens de son sourire dans mon visage, ma mère Hathat .F, l'être le plus cher à mon cœur et à mes yeux, qu'Allah la récompense de la meilleure récompense dans les deux mondes ; À eux deux, je dédie ce humble travail, à mon mari qui m'a soutenu dans son achèvement, à mes sœurs Asmaa, Ritadj, Israa, et à mon petit frère Adam, qui ont partagé avec moi le fardeau de la vie ; Je dédie également le fruit de mes efforts à mon cher professeur, le Professeur Abderrahim Mohammed Al-Amine, vers qui je me suis tournée chaque fois que le chemin était sombre devant moi, et qui l'a éclairé pour moi, semant l'espoir en moi quand le désespoir s'installait

dans mon âme, me donnant connaissance chaque fois que je demandais, et consacrant une partie précieuse de son temps malgré ses multiples responsabilités ; à tous les enseignants du département d'informatique et à tous ceux qui croient que les graines du succès du changement sont en nous-mêmes avant d'être dans autre chose.,

Merci.

”

- RIHAB

“

Par-dessus tout, le mérite revient à Dieu Tout-Puissant, merci et louange à Lui... Premièrement, à celui qui m'a donné la lumière dans cette vie et ils ont toujours été ma première école, à celui qui a travaillé, s'est efforcé et a été patient juste pour vivre dans la générosité de ma chère mère Chadau A et à celui qui m'a donné tout préparé pour nous, mon père Kiouer Ali, à mes sœurs Sajida, Basma et au petit de la maison Ahmed Anis. Je me souviens aussi particulièrement de ma mère, qui m'a élevé et qui a toujours été mon refuge contre tout. Elle était la tendresse Slimani O . Je tiens à adresser mes remerciements aux personnes qui m'ont aidé dans la réalisation de ce mémoire, Mes collègues, mes amis, à tous ceux qui ont une place dans mon cœur, et je tiens particulièrement à mentionner mon honorable professeur Abderrahim Mohammed Al-Amine il a été patient et gentil et n'a pas lésiné sur nous avec ses vastes connaissances,

Merci et respect.

”

- ABIR

Remerciements

Après louange et gratitude à Dieu pour Son soutien et Sa bienveillance dans l'achèvement de cette recherche, nous tenons à exprimer notre profonde gratitude et nos remerciements à tous ceux qui ont contribué au succès de ce travail.

Tout d'abord, nous adressons nos sincères remerciements et notre profonde appréciation à nos chers parents, qui ont été les premiers soutiens et les moteurs de notre parcours académique et de nos études. Grâce à leur soutien indéfectible et à leurs encouragements constants, nous avons pu réaliser cette réussite.

Deuxièmement, nous exprimons notre gratitude et notre reconnaissance spéciale au professeur éminent, le Professeur Abderrahim Mohammed Al-Amine, qui a supervisé avec dévouement et attention cette recherche. Ses conseils précieux et son soutien constant ont été l'une des raisons de notre succès dans l'achèvement de ce travail avec succès.

Troisièmement, nous tenons à exprimer nos sincères remerciements à tous les enseignants du département d'informatique, qui nous ont apporté leur soutien et leur assistance tout au long de la période de préparation de la recherche, et ont contribué à enrichir les discussions scientifiques et l'orientation académique.

Enfin, nous tenons à exprimer notre gratitude et notre reconnaissance à tous ceux qui ont contribué de quelque manière que ce soit au succès de ce travail, qu'ils soient proches ou éloignés. La collaboration et la contribution sincère sont ce qui a contribué au succès de cette recherche, et nous espérons que notre contribution sera utile et contribuera au domaine de la science et de la recherche.

Résumé

Les réseaux sociaux sont devenus une partie intégrante de la vie des gens à l'ère moderne, utilisés pour communiquer et interagir avec d'autres en ligne, y compris la publication de contenu, le partage de photos et de vidéos, et l'interaction avec le contenu des autres via des likes et des commentaires.

L'analyse des sentiments sur les réseaux sociaux fait référence à l'utilisation de techniques informatiques pour comprendre les sentiments exprimés par les utilisateurs dans leurs publications et commentaires, et peut être utilisée pour comprendre les sentiments du public à l'égard de produits spécifiques ou pour suivre les tendances et les sentiments généraux sur des questions particulières. En ce qui concerne l'ontologie, il s'agit de l'étude de la manière dont les individus interagissent avec les concepts et les connaissances dans un domaine particulier, et peut être utilisée dans les réseaux sociaux pour comprendre l'échange de connaissances et son évolution entre les utilisateurs. L'analyse des sentiments et l'ontologie peuvent se chevaucher dans le contexte des réseaux sociaux, où l'analyse linguistique et algorithmique peut être utilisée pour comprendre comment les individus interagissent avec le contenu et construisent leurs concepts et connaissances.

Mots clés : Analyse des sentiments, Ontologie, WordNet, SentiWordNet.

Abstract

Social media has become an integral part of people's lives in the modern era, used for communication and interaction with others online, including content publishing, sharing photos and videos, and engaging with others' content through likes and comments.

Sentiment analysis on social media refers to the use of computer techniques to understand the sentiments expressed by users in their posts and comments, and can be used to understand public sentiment towards specific products or to track trends and general sentiments on particular issues. As for ontology, it is the study of how individuals interact with concepts and knowledge in a particular domain, and can be used in social media to understand the exchange of knowledge and its evolution among users. Sentiment analysis and ontology can overlap in the context of social media, where linguistic and algorithmic analysis can be used to understand how individuals interact with content and build their concepts and knowledge.

Keywords : Sentiment Analysis, Ontology, WordNet, SentiWordNet.

ملخص

أصبحت مواقع التواصل الاجتماعي جزءاً لا يتجزأ من حياة الناس في العصر الحديث، حيث تستخدم للتواصل والتفاعل مع الآخرين عبر الإنترنت، بما في ذلك نشر المحتوى، ومشاركة الصور والفيديوهات، والتفاعل مع محتوى الآخرين من خلال الإعجابات والتعليقات. تشير تحليلات المشاعر على وسائل التواصل الاجتماعي إلى استخدام التقنيات الحاسوبية لفهم المشاعر التي يعبر عنها المستخدمون في منشوراتهم وتعليقاتهم، ويمكن استخدامها لفهم مشاعر الجمهور تجاه منتجات معينة أو لتتبع الاتجاهات والمشاعر العامة حول قضايا محددة. فيما يتعلق بالأنطولوجيا، فهي دراسة كيفية تفاعل الأفراد مع المفاهيم والمعرفة في مجال معين، ويمكن استخدامها في مواقع التواصل لفهم تبادل المعرفة وتطورها بين المستخدمين. يمكن أن تتداخل تحليلات المشاعر والأنطولوجيا في سياق مواقع التواصل، حيث يمكن استخدام التحليل اللغوي والخوارزمي لفهم كيفية تفاعل الأفراد مع المحتوى وبناء مفاهيمهم ومعارفهم.

كلمات مفتاحية :

SentiWordNet WordNet. تحليل المشاعر، أنطولوجيا،

Table des matières

Dédicace	I
Remerciements	III
Résumé	IV
Abstract	V
VI	ملخص
Introduction générale	1
1 Concepts de base : Les ontologies et les réseaux sociaux	2
1.1 Introduction	3
1.2 Historique et définition d'ontologie	3
1.2.1 Définition d'ontologie	3
1.2.2 Historique de l'ontologie (informatique)	4
1.3 L'objectif et Les types d'ontologie	5
1.3.1 L'objectif d'ontologie	5
1.3.2 Les types d'ontologies	6
1.4 Les outils d' ontologies	7
1.5 Web Sémantique	8
1.6 Réseaux sociaux	9
1.6.1 Définition des Réseaux sociaux :	9
1.6.2 Quelques réseaux sociaux	10
1.7 Conclusion	11
2 Analyse de sentiment et travaux connexes	12
2.1 Introduction	13
2.2 Analyse de sentiment	13
2.2.1 Définition	13
2.2.2 Importance de l'analyse des sentiments	14
2.2.3 Processus d'analyse des sentiments	15
2.2.4 Types d'analyse des sentiments	19
2.2.5 Applications de l'analyse des sentiments	21
2.2.6 Approches pour l'analyse des sentiments	21
2.2.7 Défis liés à l'analyse des sentiments	22
2.3 Solutions d'analyse des sentiments	22

2.4	Travaux connexes	23
2.5	Conclusion	24
3	Conception et Réalisation	26
3.1	Introduction	27
3.2	WordNet	27
3.2.1	Structuration des Relations Lexicales dans WordNet	28
3.2.2	Ordre hiérarchique dans WordNet	28
3.2.3	Relations Lexico-Sémantique dans WordNet (Relation entre les mots)	29
3.2.4	Relations sémantiques : (Relation entre les synsets)	30
3.3	WordNet Arabe (Arabic WordNet : AWN)	30
3.3.1	Description de AWN	30
3.3.2	Création et structuration de AWN	31
3.3.3	Concepts de base Arabes (Arabic Base Concepts)	31
3.4	Senti WordNet	32
3.5	Conception et Réalisation d'un Senti WordNet (DZSentiWordNet)	32
3.5.1	Caractéristiques du dialecte algérien :	33
3.5.2	Algorithmes de construction du DZSentiWordNet	33
3.5.3	Vérification et validation du DZSentiWordNet	35
3.6	Conclusion	35
	Conclusion	36
	Bibliographie	38
	Annexes	40
	A Définitions	41

Table des figures

- 1.1 Exemple (Hiérarchie de concepts d'une ontologie pour les recettes de cuisine) [3]. 4
- 1.2 Tim Berners-Lee [4]. 4
- 1.3 Différents types d'ontologies. 7
- 1.4 Exemples de réseaux sociaux [9]. 11

- 2.1 Différents types de sentiment [11]. 14
- 2.2 Etapes de processus de l'analyse des sentiments .[10] 15
- 2.3 Les niveaux d'analyse des sentiments. 19

- 3.1 Dictionnaire vs WordNet [22] 27
- 3.2 Hiérarchie du plus général au plus spécifique 29
- 3.3 Les relations dans WordNet. [22] 31

Liste des tableaux

1.2	Les choses que l'ontologie décrit	5
2.2	Exemple des avis des clients.	20

Introduction générale

Avec le début du troisième millénaire, les sites web et l'internet sont devenus un sujet d'intérêt pour les personnes de tous âges en raison de leur ouverture sur le monde. Il est devenu facile de communiquer entre individus grâce aux plateformes de médias sociaux, qui sont des plateformes numériques en ligne. Parmi ces plateformes figurent des sites populaires tels que Facebook, Twitter, Instagram, LinkedIn, Snapchat, et d'autres encore. Les gens utilisent ces plateformes pour partager des idées, des images, des vidéos, ainsi que pour exprimer leurs opinions et réactions au contenu qui y est publié.

Les commentaires des utilisateurs sur les publications sont l'un des éléments clés de l'interaction des utilisateurs avec le contenu publié sur ces plateformes. Les commentaires des utilisateurs varient considérablement, pouvant inclure des réponses au contenu, l'expression d'opinions, des éloges, des critiques, des blagues, voire parfois des discussions approfondies.

À travers les commentaires, les utilisateurs peuvent interagir les uns avec les autres, échanger des points de vue, et construire des interactions sociales. Les commentaires peuvent également être un lieu de dialogue et de discussion fructueuse ou un espace où les utilisateurs expriment librement leurs opinions. Cependant, les commentaires peuvent également poser un défi pour les utilisateurs et les publications avec lesquelles ils interagissent, car ils peuvent parfois contenir des contenus négatifs ou offensants, ce qui nécessite une gestion par la plateforme pour maintenir un environnement interactif positif et sûr pour tous. En général, les commentaires des utilisateurs sont une partie essentielle de l'expérience des médias sociaux, contribuant à enrichir le contenu, à élargir les discussions, et à renforcer la communication entre les individus. L'objectif de notre travail est de créer et de développer une ontologie qui peut être utilisée pour faire l'analyse des sentiments des individus en se basant sur les commentaires des publications sur les réseaux sociaux, afin de déterminer si le commentaire est positif, négatif ou neutre. Ce mémoire est composé de trois chapitres. Le premier chapitre décrit les concepts de base, il introduit la notion d'ontologie et des réseaux sociaux. Le deuxième chapitre aborde l'analyse des sentiments. Ensuite le troisième chapitre présente notre travail à savoir la conception et la réalisation d'une ontologie pour l'analyse des sentiments dans le cadre des réseaux sociaux algériens.

Chapitre 1

Concepts de base : Les ontologies et les réseaux sociaux

1.1 Introduction

L'objectif de ce chapitre est d'introduire les concepts liés aux différents aspects abordés tout au long de notre recherche. Nous commencerons par explorer l'historique et la définition de l'ontologie, ainsi que son objectif. Nous examinerons également les différents types d'ontologies et les outils disponibles pour travailler avec celles-ci. Ensuite, nous aborderons le domaine du Web sémantique, qui constitue un contexte important pour notre étude. Enfin, nous fournirons une définition des réseaux sociaux et discuterons de leur pertinence dans le cadre de notre recherche. L'analyse de chaque concept nous permettra de former une idée générale de l'approche à adopter pour mener à bien notre travail.

1.2 Historique et définition d'ontologie

1.2.1 Définition d'ontologie

L'ontologie est une branche de la philosophie qui explore l'essence de l'existence. Le terme vient du grec ancien *ôn*, *onton*, signifiant "être", et *logos*, désignant le discours ou l'étude. Dans le domaine de la métaphysique, l'ontologie se concentre sur la nature fondamentale de la réalité et cherche à comprendre les principes et les causes premières, ou encore ce qu'on appelle parfois "l'être en tant qu'être", c'est-à-dire l'essence ou le sens de l'existence [1]. Alors que la métaphysique se penche sur le "pourquoi" ou le sens de l'existence, l'ontologie examine le "comment" de l'existence, en étudiant la manière dont les êtres sont organisés dans le monde, leur étant. De manière générale, elle cherche à décrire, dans une approche analytique, ce qui constitue l'ensemble des entités existantes dans un domaine donné. En d'autres termes, elle vise à explorer ce qui "est" dans un cadre donné, en identifiant les éléments et en examinant leurs relations et leurs structures [1].

Dans le domaine de l'informatique et de la science de l'information, le concept d'ontologie est également utilisé pour désigner un ensemble organisé de termes et de concepts représentant le sens d'un domaine d'informations (voir Figure 1.1). Il s'agit d'un modèle de données représentatif des concepts dans un domaine donné, ainsi que des relations entre ces concepts [2]. Les ontologies sont employées pour raisonner sur les objets du domaine concerné, organisés dans un graphe avec des relations sémantiques et de subsomption.

L'idée d'ontologie en informatique et l'exemple de la programmation orientée objet partagent une similitude notable car ils ont une origine commune : les systèmes d'intelligence artificielle symbolique [2]. La branche de l'intelligence artificielle que nous abordons ici est qualifiée de symbolique car elle repose sur des représentations formelles des connaissances sous forme de symboles que le système peut stocker et manipuler. Contrairement à d'autres approches, ces représentations sont à la fois compréhensibles par les humains et manipulables par les systèmes grâce à l'application de règles de manipulation définies sur les symboles de ces représentations, permettant ainsi de simuler, par exemple, un raisonnement [2].

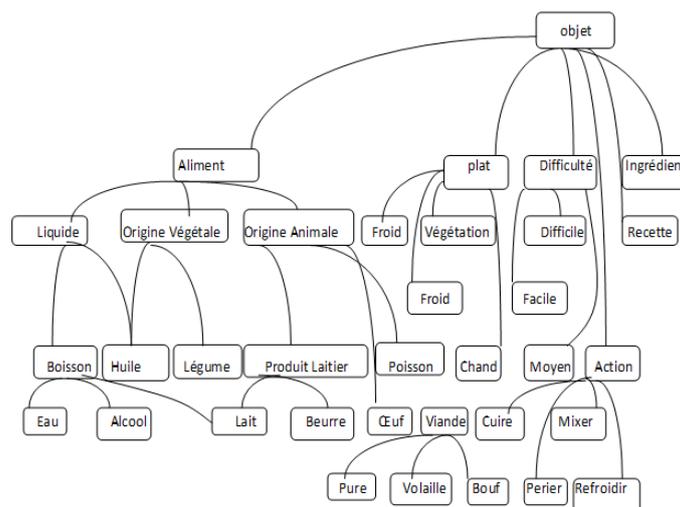


FIG. 1.1 : Exemple (Hiérarchie de concepts d'une ontologie pour les recettes de cuisine) [3].

1.2.2 Historique de l'ontologie (informatique)

L'ontologie est un terme informatique emprunté à la philosophie dans les années 90. À ses débuts, le World Wide Web, créé en 1990 par Tim Berners-Lee (Figure 1.2), était principalement destiné à partager des informations sous forme de pages HTML, accessibles via un navigateur web et destinées à être lues par des utilisateurs humains. Cependant, l'introduction de XML en 1998 a apporté un cadre pour structurer les connaissances, ouvrant la voie à la création de nouveaux langages web conçus non seulement pour la présentation graphique mais aussi pour le partage et la manipulation des connaissances [2].

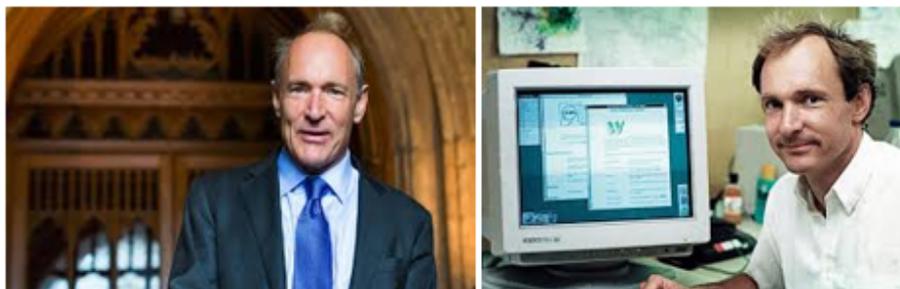


FIG. 1.2 : Tim Berners-Lee [4].

En 2004, le W3C publie RDF et OWL, deux technologies clés du Web sémantique. RDF (Resource Description Framework) et OWL (Web Ontology Language) offrent un cadre de travail pour la gestion des ressources, l'intégration, le partage et la réutilisation des données sur le Web. Ces technologies ont révolutionné la manière dont les informations sont représentées et traitées sur le Web, permettant aux machines de comprendre et de manipuler les données de manière plus sophistiquée, et ouvrant la voie à des applications plus intelligentes et plus interactives [2].

Dans le domaine de l'intelligence artificielle, l'utilisation des ontologies s'est également développée parallèlement à l'évolution du Web sémantique. Les ontologies sont devenues essentielles pour modéliser et représenter les connaissances dans des domaines spécifiques, permettant aux systèmes d'IA de raisonner sur les informations de manière plus précise et

de prendre des décisions plus éclairées. Ainsi, l'histoire de l'ontologie dans le contexte de l'informatique est étroitement liée à l'évolution du World Wide Web et des technologies associées, jouant un rôle crucial dans la transformation de la façon dont nous partageons, manipulons et comprenons les données et les informations.

1.3 L'objectif et Les types d'ontologie

1.3.1 L'objectif d'ontologie

Le but principal d'une ontologie est de modéliser un ensemble de connaissances dans un domaine spécifique, qu'il soit réel ou imaginaire. Les ontologies sont utilisées dans des domaines tels que l'intelligence artificielle, le Web sémantique, le génie logiciel, l'informatique biomédicale et l'architecture de l'information, où elles servent de forme de représentation formelle. Les ontologies sont utilisées pour réaliser une compréhension et une interaction intelligentes avec les données et les informations, en aidant à l'analyse des données, à la compréhension des concepts, à la découverte des relations entre les entités, à la perception du contexte et à l'autonomisation des systèmes intelligents pour la prise de décisions. En résumé, les ontologies jouent un rôle important dans l'habilitation de l'intelligence artificielle à comprendre le monde qui l'entoure et à interagir avec lui de manière intelligente et efficace.

TAB. 1.2 : Les choses que l'ontologie décrit

Les Concepts	La Description
Individus	Les Objets De Base
Classes	Ensembles, Collections, Ou Types D'Objets
Attributs	Propriétés, Fonctionnalités, Caractéristiques Ou Paramètres Que Les Objets Peuvent Posséder Et Partager
Relations	Les Liens Que Les Objets Peuvent Avoir Entre Eux
Événements	Changements Subis Par Des Attributs Ou Des Relations

En général, les objectifs des "ontologies" seraient [2] :

1. **Intelligence artificielle** : permettre aux machines de comprendre les humains.
2. **Interopérabilité** : permettre aux machines de comprendre d'autres machines.
3. **Rendre du sens pour l'utilisateur** : permettre aux humains de comprendre d'autres humains via des machines.

1.3.2 Les types d'ontologies

Il existe plusieurs types d'ontologie. Nous allons citer les catégories suivantes (voir Figure 1.3) :

- **Les ontologies dans le domaine** : Elles portent ce nom car elles expriment des conceptualisations spécifiques à un domaine particulier. Elles représentent le vocabulaire spécifique à ce domaine à travers des concepts et des relations qui modélisent les activités principales, les théories et les principes fondamentaux de ce domaine. La plupart des ontologies existantes sont des ontologies du domaine, elles sont réutilisables pour plusieurs applications dans le domaine pour lequel elles ont été créées car elles ont été conçues de manière aussi autonome que possible par rapport au type de manipulations qui seront opérées sur ces connaissances [5].
- **Les ontologies applicatives (ou ontologies d'application)** : Ce sont les ontologies les plus spécifiques, elles contiennent les connaissances nécessaires pour une application particulière et ne sont pas réutilisables. Elles peuvent également inclure une ontologie de domaine [5].
- **Les ontologies génériques ou de haut niveau (upperontology)** : Elles expriment des conceptualisations valables dans différents domaines avec une valeur relativement générale, comme les notions d'objets, de propriété, de valeur, d'état, ainsi que des concepts de temps, d'espace et d'événements. Elles sont conçues pour être utilisées dans des situations diverses et pour servir une large communauté d'utilisateurs.
- **Les ontologies de représentation** : Ce type d'ontologies regroupe les concepts utilisés pour formaliser les connaissances. Parmi les ontologies de représentation, on trouve celles qui décrivent les notions utilisées dans toutes les ontologies pour spécifier les connaissances, telles que les substances, les concepts, les relations, etc. Par exemple, la "Frame-Ontology" est une ontologie de représentation qui définit de manière formelle les concepts utilisés principalement dans les langages basés sur les frames : classes, sous-classes, attributs, valeurs, relations et axiomes. Les ontologies de représentation sont indépendantes des différents domaines de connaissances, car elles décrivent des primitives cognitives communes à divers domaines [5].

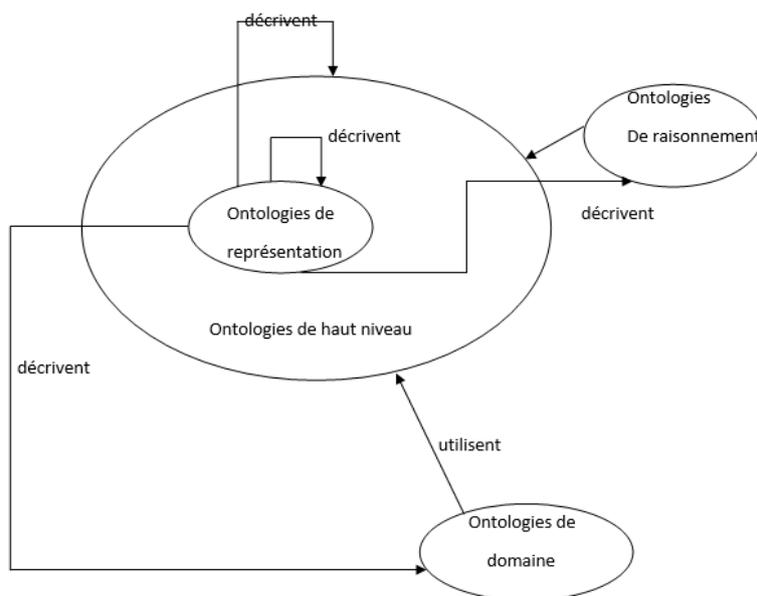


FIG. 1.3 : Différents types d'ontologies.

1.4 Les outils d'ontologies

Il existe plusieurs outils permettant la manipulation des ontologies (de la création à la mise à jour). Les éditeurs d'ontologie suivants sont disponibles gratuitement en téléchargement :

1. **Protégé** : Il s'agit du logiciel le plus renommé et le plus utilisé pour l'édition d'ontologies. Développé en open-source par l'université Stanford, Protégé a évolué depuis ses premières versions (Protégé-2000) pour intégrer les standards du Web sémantique, notamment OWL. Il propose une multitude de composants optionnels tels que des raisonneurs et des interfaces graphiques [6].
2. **SWOOP** : SWOOP est un éditeur d'ontologie développé par l'Université du Maryland dans le cadre du projet MINDSWAP. Contrairement à Protégé, il a été nativement conçu pour les standards RDF et OWL, et prend en charge leurs différentes syntaxes (pas seulement XML). Bien qu'il soit plus léger que Protégé et moins avancé en termes d'interface, il intègre également des outils de raisonnement [6].
3. **KMgen** : Il s'agit d'un éditeur d'ontologie pour le langage KM (KM : The Knowledge Machine) [6]. Avec l'avènement du marché des technologies du Web sémantique, plusieurs outils logiciels ont été proposés par des éditeurs commerciaux depuis 2005, notamment :
 - **SemanticWorks** : Il fait partie de la suite d'outils XML développée par Altova, et prend en charge le langage OWL via sa syntaxe XML.
 - **TopBraid Composer** : Développé par TopQuadrant, son interface et ses fonctionnalités sont très similaires à celles de Protégé.

- **Tedi** : C'est un éditeur d'ontotermologies développé par l'équipe Condillac "Terminologie & Ontologie" de l'Université Savoie Mont-Blanc. Les ontotermologies sont disponibles aux formats HTML, CSV, OWL et JSON.
- **KAD-Office** : Développé par Iknova Consulting, il est spécialisé dans la représentation graphique des connaissances industrielles d'ingénierie.
- **Suite d'outils TRC** : Développée par The ReuseCompany. En outre, il existe des outils informatiques permettant de construire une ontologie à partir d'un corpus de textes. Ces outils analysent le texte à la recherche de termes récurrents ou définis par l'utilisateur, puis examinent la manière dont ces termes sont mis en relation dans le texte. Le projet WordNet en est un exemple majeur.

1.5 Web Sémantique

Le Web sémantique est un concept élaboré par le Consortium World Wide Web (W3C), visant à rendre plus facile la compréhension et le traitement, par les machines, du sens des informations sur le Web. Il repose sur l'idée d'utiliser des formats de données normalisés et des métadonnées lisibles par machine pour décrire le contenu et les relations entre les ressources Web. Cela permet aux ordinateurs d'interpréter et d'utiliser plus aisément les informations disponibles en ligne [7].

Le Web sémantique s'appuie sur la structure existante du World Wide Web en utilisant des technologies telles que RDF (Resource Description Framework) et OWL (Web OntologyLanguage) pour ajouter une couche de signification au contenu Web. En annotant les ressources Web avec des métadonnées lisibles par machine selon ces normes, il devient possible aux ordinateurs de comprendre les relations et les significations des ressources en ligne, ce qui leur permet d'effectuer des tâches plus intelligentes telles que l'inférence et la réponse à des questions.

L'objectif du Web sémantique est de permettre aux machines de comprendre et de traiter de manière plus significative la grande quantité d'informations disponibles sur le Web, ainsi que de faciliter la recherche et l'utilisation des informations nécessaires. Il transforme notre façon d'utiliser le Web et d'interagir avec les ordinateurs, permettant aux machines de nous assister plus efficacement dans des activités telles que la recherche d'informations, la prise de décisions et la résolution de problèmes [7].

Pour atteindre ces objectifs, le Web sémantique repose sur plusieurs technologies, notamment :

1. **RDF (Resource Description Framework)** : est un standard utilisé pour décrire la signification des données sur le Web. Il propose un modèle de données simple et flexible qui permet de représenter les informations de manière structurée, en décrivant les ressources, leurs attributs et leurs relations [7].
2. **OWL (Web OntologyLanguage)** : est une norme qui permet de définir la signification des concepts et des relations dans un domaine spécifique. Il s'appuie sur un langage formel basé sur la logique, ce qui permet de décrire de manière précise les

termes et les relations entre eux, facilitant ainsi la compréhension et l'interopérabilité des données [7].

3. **SPARQL** : est un langage de requête standard utilisé pour interroger les données RDF. Il permet d'exprimer des requêtes complexes pour récupérer des informations spécifiques à partir de graphes RDF, en fournissant une syntaxe claire et puissante pour manipuler les données sémantiques sur le Web.

Ces technologies permettent aux développeurs web d'enrichir leurs données avec une signification supplémentaire, en utilisant un ensemble commun de normes et de langages. Cela donne la capacité aux ordinateurs de comprendre la sémantique des données, de les traiter et de les manipuler de manière plus intelligente.

Un aspect fondamental du Web sémantique est l'introduction de données liées, où les données sur le Web sont connectées entre elles grâce à l'utilisation de RDF et d'URL. Cette approche permet aux données provenant de différentes sources d'être liées de manière significative, ouvrant la voie à une vision plus intégrée et interconnectée du Web.

Dans son ensemble, le Web sémantique vise à transformer le Web en une ressource plus intelligente et plus utile, aussi bien pour les êtres humains que pour les machines, en rendant la signification des informations sur le Web plus explicite et plus accessible.

1.6 Réseaux sociaux

1.6.1 Définition des Réseaux sociaux :

Plusieurs définitions du terme « réseau social » sont présentées dans la littérature. Selon Halimi (2016) [8], il s'agit d'une « structure sociale composée d'individus ou d'organisations, dont les identités sociales sont interconnectées par différentes relations sociales telles que l'amitié, l'intérêt ou la connaissance ». Une autre définition, proposée par Boyd et Ellison (2007), décrit les « sites de réseaux sociaux comme des services en ligne permettant aux individus de créer un profil public ou semi-public dans un système délimité, d'énumérer d'autres utilisateurs avec lesquels ils ont une connexion, et de visualiser et parcourir leurs propres connexions ainsi que celles établies par d'autres utilisateurs dans le système. La nature et la terminologie de ces connexions peuvent varier d'un site à l'autre » [8].

Zhu et ses collègues (2016) définissent un réseau social comme une « plateforme pratique permettant aux utilisateurs de stocker et de partager des messages dans divers types de médias, ainsi que les comportements sociaux variés des utilisateurs, ce qui fait que les données sont liées socialement plutôt que d'exister de manière indépendante ».

En dépit de la grande variété de fonctionnalités techniques proposées par les réseaux sociaux, leur structure fondamentale repose sur des profils visibles affichant une liste détaillée d'amis, qui sont également des utilisateurs du système. Ces profils constituent des pages uniques où les individus peuvent se définir en ligne.

De ce fait, les réseaux sociaux sont souvent conçus pour être largement accessibles, attirant initialement des populations homogènes. Il est donc fréquent de trouver des groupes utilisant ces plateformes pour se regrouper selon des critères tels que la nationalité, l'âge, le niveau d'éducation, ou d'autres facteurs qui segmentent habituellement la société, même si ce n'était pas l'intention initiale des concepteurs.

1.6.2 Quelques réseaux sociaux

Voici une présentation de quelques sites de réseaux sociaux (voir Figure 1.4) :

1. **Facebook** : Lancé en 2004 par Mark Zuckerberg, Facebook est l'un des réseaux sociaux les plus populaires au monde. Il permet aux utilisateurs de créer des profils, de partager des publications, des photos et des vidéos, de se connecter avec des amis et de rejoindre des groupes d'intérêt commun.
2. **Twitter** : Fondé en 2006, Twitter est une plateforme de microblogging où les utilisateurs peuvent publier des messages courts appelés « tweets ». Les tweets peuvent contenir jusqu'à 280 caractères et sont souvent utilisés pour partager des pensées, des actualités et des liens vers d'autres contenus en ligne.
3. **Instagram** : Lancé en 2010, Instagram est une application de partage de photos et de vidéos appartenant à Facebook. Les utilisateurs peuvent publier des photos et des vidéos, les modifier avec des filtres et des outils de retouche, et les partager avec leurs abonnés. Instagram est également connu pour ses fonctionnalités telles que les stories éphémères et les IGTV.
4. **LinkedIn** : Lancé en 2003, LinkedIn est un réseau social professionnel conçu pour les professionnels et les entreprises. Il permet aux utilisateurs de créer des profils professionnels, de se connecter avec d'autres professionnels, de rechercher des opportunités d'emploi, de partager des articles et des actualités liés à leur secteur d'activité.
5. **Snapchat** : Lancé en 2011, Snapchat est une application de messagerie multimédia où les utilisateurs peuvent envoyer des photos et des vidéos éphémères à leurs amis. Il est également connu pour ses filtres faciaux et ses lentes AR (réalité augmentée), ainsi que pour sa section Discover où les utilisateurs peuvent consulter du contenu produit par des médias et des marques.
6. **TikTok** : Lancé en 2016, TikTok est une plateforme de partage de vidéos courtes qui est devenue extrêmement populaire auprès des jeunes. Les utilisateurs peuvent créer, éditer et partager des vidéos de 15 à 60 secondes, souvent basées sur des défis, des danses, des sketches comiques ou des tutoriels. Ces plateformes représentent une petite sélection des nombreux réseaux sociaux disponibles en ligne, chacun ayant ses propres caractéristiques, communautés et cultures.



FIG. 1.4 : Exemples de réseaux sociaux [9].

1.7 Conclusion

Au cours de ce chapitre, nous avons passé en revue le concept d'ontologie, en examinant son historique, sa définition, ses objectifs, ses types, ainsi que les outils disponibles pour travailler avec les ontologies, en plus d'aborder le domaine du Web sémantique. Aussi, nous avons fourni une description des réseaux sociaux. Cette revue nous a permis de mieux comprendre et d'appréhender les différentes notions essentielles pour notre travail à venir.

Chapitre 2

Analyse de sentiment et travaux connexes

2.1 Introduction

Ce chapitre explore les fondements de l'analyse des sentiments, un domaine vital dans le traitement du langage naturel. Il commence par une définition précise de l'analyse des sentiments et met en lumière son importance dans divers domaines. Ensuite, le chapitre explore le processus d'analyse des sentiments, de la collecte des données textuelles à l'interprétation des résultats. Le chapitre examine également les différents types d'analyse des sentiments, ainsi que les approches et les techniques pour les appliquer dans des contextes spécifiques. Ensuite, il aborde les principaux défis auxquels sont confrontés les chercheurs et les praticiens dans ce domaine.

Enfin, le chapitre présente un ensemble d'œuvres liées à l'ontologie et à l'analyse des sentiments, en mettant également en lumière les tendances émergentes dans ce domaine. Cette analyse détaillée vise à fournir une compréhension approfondie des défis auxquels sont confrontés les chercheurs et les spécialistes, ainsi qu'à identifier les opportunités disponibles dans le domaine de l'analyse des sentiments et de l'ontologie dans le contexte des évolutions technologiques et sociales actuelles.

2.2 Analyse de sentiment

2.2.1 Définition

Dans la littérature, l'analyse des sentiments (SA) reçoit différentes dénominations ou termes communs, y compris l'extraction d'opinions (OM), l'analyse de la subjectivité, l'analyse des émotions, et l'extraction d'évaluations, entre autres. Parmi les plus couramment utilisés dans la littérature, on trouve l'analyse des sentiments et l'exploration des opinions (MO). Ces deux concepts sont similaires et désignent le même domaine d'étude, qui peut lui-même être considéré comme un sous-domaine de l'analyse de la subjectivité. L'analyse des sentiments est un domaine de recherche dans le domaine de l'exploration de texte, défini comme le traitement informatique des opinions, des sentiments et de la subjectivité du texte. Il n'est pas surprenant qu'il y ait une certaine confusion parmi les chercheurs quant à la différence entre les sentiments et les opinions, ce qui alimente le débat sur l'appellation du domaine, soit l'analyse des sentiments, soit l'exploration des opinions [10].

Dans le dictionnaire collégial de Merriam-Webster, les sentiments sont définis comme des attitudes, des pensées ou des jugements suscités par le sentiment (voir Figure 2.1), tandis que les opinions sont définies comme des points de vue, des jugements ou des évaluations formés dans l'esprit sur une question particulière. La différence entre les deux est assez subtile, et chacun contient certains éléments de l'autre. Les opinions sont généralement des points de vue concrets sur quelque chose, tandis que les sentiments sont des émotions. Par exemple, l'expression "Je suis préoccupé par la situation politique actuelle" exprime un sentiment, tandis que "Je pense que la politique ne va pas bien" exprime une opinion. Les significations sous-jacentes des deux phrases sont étroitement liées, car le sentiment décrit dans la première phrase est susceptible d'être causé par l'opinion exprimée dans la deuxième phrase. Cependant, certaines opinions ne sont pas nécessairement

associées à des sentiments, comme "Je pense qu'il gagnera les prochaines élections présidentielles [10].

En général, l'analyse des sentiments est un processus visant à suivre les opinions des clients ou des auteurs sur un objet ou un sujet spécifique. Cela implique la création d'un cadre pour recueillir et examiner les opinions sur l'élément en question, telles qu'exprimées dans les entrées de blog, les commentaires, les critiques, ou les tweets.

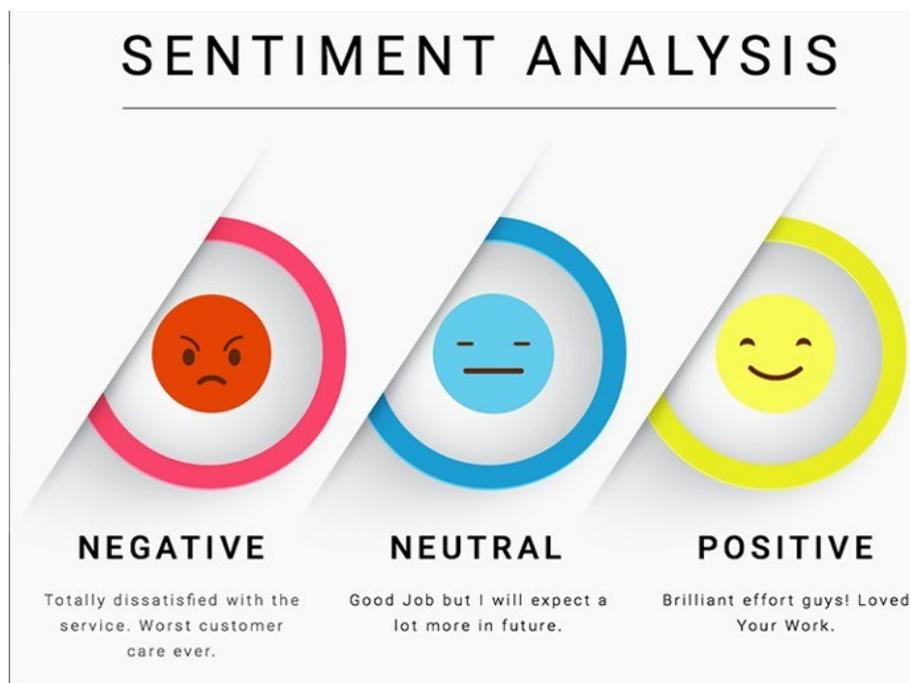


FIG. 2.1 : Différents types de sentiment [11].

2.2.2 Importance de l'analyse des sentiments

L'analyse des sentiments, également connue sous le nom d'exploration d'opinion, est un outil crucial en intelligence décisionnelle qui aide les entreprises à améliorer leurs produits et leurs services. Voici quelques avantages de l'analyse des sentiments :

Fournir des informations objectives : Les entreprises peuvent éviter les biais personnels associés aux évaluateurs humains en utilisant des outils d'analyse des sentiments basés sur l'intelligence artificielle (IA). Ainsi, elles obtiennent des résultats cohérents et objectifs lors de l'analyse des opinions des clients .

Prenons par exemple la phrase suivante : "Je suis étonné par la rapidité du processeur, mais déçu qu'il chauffe rapidement." Les spécialistes du marketing pourraient ignorer la partie décevante de l'évaluation et être partiaux en faveur des performances du processeur. Cependant, des outils d'analyse des sentiments précis trient et classent le texte pour capturer les émotions de manière objective.

Développer de meilleurs produits et services : Un système d'analyse des sentiments aide les entreprises à améliorer leurs produits et services en se basant sur les commentaires authentiques et spécifiques des clients. Les technologies d'IA identifient

les objets ou les situations réalistes (appelés entités) que les clients associent à un sentiment négatif. Par exemple, les ingénieurs produits se concentreront sur l'amélioration de la capacité de gestion thermique du processeur, car le logiciel d'analyse de texte a associé "déçu" (négatif) à "processeur" (entité) et "chauffe" (entité) [12].

Analyser à grande échelle : Les entreprises tirent constamment des informations d'une grande quantité de données non structurées telles que les e-mails, les transcriptions de chatbots, les enquêtes, les dossiers de gestion de la relation client et les commentaires sur les produits. Les outils d'analyse des sentiments basés sur le cloud permettent aux entreprises de mettre à l'échelle le processus de découverte des émotions des clients dans les données textuelles à un coût abordable.

Résultats en temps réel : Face aux crises potentielles ou aux tendances du marché en rapide évolution, les entreprises doivent réagir rapidement. Les spécialistes du marketing s'appuient sur un logiciel d'analyse des sentiments pour connaître en temps réel les opinions des clients sur la marque, les produits et les services de l'entreprise, et prendre des mesures immédiates en fonction de leurs conclusions. Ils peuvent configurer le logiciel pour envoyer des alertes lorsque des sentiments négatifs sont détectés pour des mots clés spécifiques.

2.2.3 Processus d'analyse des sentiments

Les étapes du processus de l'analyse des sentiments sont illustrées ci-dessus (voir Figure 2.2) :

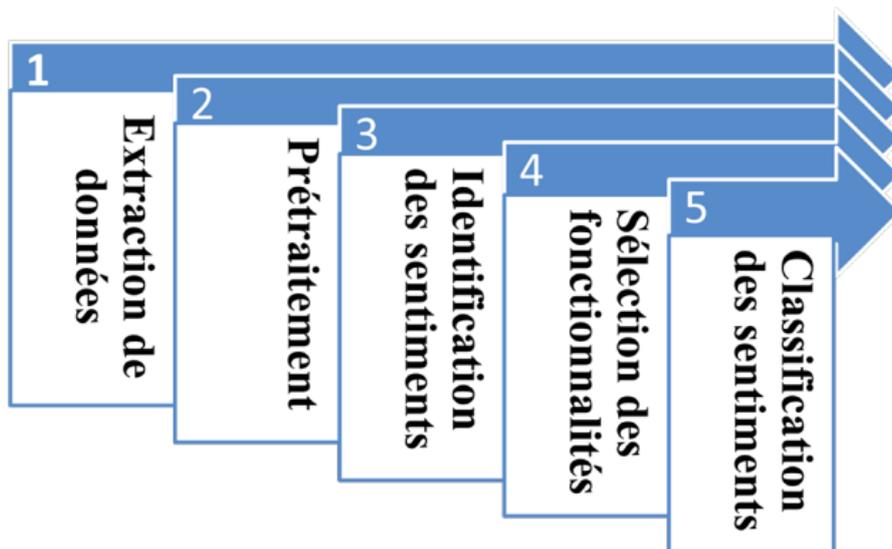


FIG. 2.2 : Etapes de processus de l'analyse des sentiments .[10]

1. Collecte des données

La collecte des données commence par l'agrégation de discussions provenant d'utilisateurs actifs sur divers forums publics, y compris des blogs, des forums de discussion et des plateformes d'évaluation de produits. De plus, des données sont également

collectées à partir de journaux privés accessibles via les réseaux sociaux tels que Twitter et Facebook. Ces sources de données contiennent souvent une quantité importante d'informations, et elles peuvent être désorganisées et réparties sur plusieurs plateformes de médias sociaux et sites web. Les données peuvent inclure une variété de formats tels que des publications textuelles, des commentaires, des critiques de produits, des tweets, etc .

Une fois que les données ont été collectées à partir de ces sources, elles sont généralement stockées dans une base de données centrale pour être traitées et analysées. Cette étape de collecte des données est cruciale car elle fournit le matériau brut sur lequel reposera toute analyse ultérieure.

2. Prétraitement

Le prétraitement des données est une étape cruciale dans le processus d'analyse des sentiments. Il vise à nettoyer les données extraites afin de préparer le texte pour la classification. Les textes en ligne sont souvent entachés de bruit et d'éléments indésirables tels que des balises HTML, des scripts, etc. Le prétraitement des données vise à réduire ce bruit, ce qui contribue à améliorer les performances du classificateur. De plus, il permet d'accélérer le processus de classification, ce qui est particulièrement bénéfique pour une analyse en temps réel .

La préparation du texte implique plusieurs étapes de nettoyage des données. Tout d'abord, il s'agit d'identifier et de supprimer tout contenu non textuel de l'ensemble de données, comme les balises HTML, les scripts et autres éléments indésirables. Ensuite, il est nécessaire d'éliminer tout contenu jugé non pertinent pour le domaine d'étude, tel que les mots vides ou les mots qui n'apportent pas de valeur ajoutée à l'analyse. Par exemple, les mots courants comme "et", "ou", "le", etc., peuvent être supprimés car ils n'influencent pas significativement le sentiment exprimé dans le texte.

Pour un système d'analyse des sentiments basé sur des flux de données en temps réel, la stratégie de prétraitement doit être soigneusement élaborée pour garantir des résultats précis et efficaces. Cela peut impliquer l'utilisation d'algorithmes de nettoyage des données avancés ainsi que des techniques de traitement du langage naturel pour identifier et supprimer le bruit de manière efficace tout en préservant les informations pertinentes pour l'analyse des sentiments.

- (a) **Élimination des mots vides :** Lorsque l'on effectue une analyse de texte, certains mots très courants, tels que "et", "ou", "le", etc., peuvent ne pas apporter de contribution significative à la compréhension du contenu. Pour cette raison, il est souvent judicieux de les éliminer afin de réduire le bruit et de concentrer l'analyse sur les mots qui portent une charge sémantique plus importante . Cette élimination se fait généralement en utilisant une liste prédéfinie de mots vides, également appelés "stop words", qui sont retirés du texte avant toute analyse ultérieure. En éliminant ces mots, on obtient une version plus concise et concentrée du texte, ce qui facilite l'identification des informations pertinentes.
- (b) **Structuration des phrases :** La structure d'une phrase, c'est-à-dire la manière dont les mots sont organisés et interconnectés, joue un rôle crucial dans la

transmission du sens et de l'émotion. En analysant la structure d'une phrase, on peut mieux comprendre la sémantique de chaque mot ainsi que les relations entre eux. Certaines techniques d'analyse des sentiments exploitent cette structure pour identifier l'opinion exprimée dans le texte. Par exemple, l'identification de phrases telles que "Je suis heureux que" ou "Je suis déçu de" peut aider à déterminer le sentiment associé à un certain sujet. En structurant les phrases de manière appropriée, on peut donc améliorer la précision de l'analyse des sentiments et obtenir des résultats plus fiables .

- (c) **Élimination des localisateurs de ressources uniformes (URL), hashtags, références, caractères spéciaux** : Lors du nettoyage des données, il est essentiel de supprimer les éléments qui n'apportent pas de valeur sémantique à l'analyse, tels que les URL, les hashtags, les références et les caractères spéciaux. Ces éléments peuvent introduire du bruit et perturber l'analyse des sentiments en ajoutant des informations non pertinentes. En éliminant ces éléments, on améliore la qualité des données et on facilite leur traitement ultérieur .
- (d) **Traduction des mots argotiques** : Cette étape consiste à remplacer les termes argotiques par leur équivalent significatif afin de garantir une compréhension précise du texte. En utilisant un dictionnaire d'argot Internet, les mots ou expressions argotiques sont identifiés et traduits en langage courant, ce qui permet une interprétation plus claire et cohérente du contenu .
- (e) **Suppression des lettres répétées dans les mots** : Les mots comportant des lettres répétées de manière excessive peuvent introduire de la redondance et de la confusion dans le texte. Pour éviter cela, on supprime les lettres répétées plus de deux fois dans un mot, en le réduisant à sa forme de base. Par exemple, le mot "Happyyyyy" est réduit à "Happy", ce qui simplifie le texte et facilite son analyse.
- (f) **Racinisation** : La racinisation est le processus de réduction d'un mot à sa forme racine ou canonique, ce qui permet de regrouper les variantes d'un même mot sous une seule forme. Cela permet de simplifier le texte et de réduire la complexité lors de l'analyse. Par exemple, les mots "waiting", "waits" et "waited" sont racinisés en "wait", ce qui facilite la comparaison et l'interprétation des données.

3. Détection des sentiments

La détection des sentiments constitue une étape fondamentale dans le processus d'analyse des données textuelles. Elle implique l'évaluation et l'extraction des opinions et des émotions exprimées dans le texte. Cette phase est essentielle pour comprendre le ressenti des utilisateurs à l'égard d'un produit, d'un service ou d'un sujet donné. Pour commencer, chaque phrase est examinée individuellement pour déterminer si elle contient une opinion ou une émotion. Les phrases qui expriment

des opinions sont conservées, tandis que celles qui ne véhiculent que des faits ou une communication objective sont écartées, car elles ne contribuent pas à l'analyse des sentiments .

La détection des sentiments peut être réalisée à différents niveaux, depuis l'analyse d'un seul terme jusqu'à l'évaluation de phrases complètes voire de documents entiers. Pour ce faire, différentes techniques d'analyse des sentiments sont utilisées, telles que l'analyse lexicale, l'apprentissage automatique et le traitement du langage naturel.

En résumé, la détection des sentiments permet de comprendre les opinions, les émotions et les attitudes des utilisateurs à partir des données textuelles, ce qui est essentiel pour prendre des décisions éclairées en matière de produits, de services et de stratégies commerciales.

4. Sélection des fonctionnalités

Dans cette étape, la sélection des fonctionnalités est cruciale pour l'analyse des sentiments. L'objectif principal est de réduire la dimensionnalité de l'espace des fonctionnalités, ce qui permet de minimiser le coût de calcul associé au traitement des données. Une dimensionnalité réduite contribue également à éviter la sur-adaptation du modèle aux données d'apprentissage, assurant ainsi sa capacité à généraliser à de nouveaux exemples .

Il est important de trouver un équilibre entre la richesse des fonctionnalités et les contraintes de calcul. Cela signifie sélectionner les fonctionnalités les plus informatives tout en évitant le surapprentissage et en garantissant des performances optimales de classification. Une sélection judicieuse des fonctionnalités permet de construire des modèles d'analyse des sentiments plus efficaces et plus généralisables, capables de traiter efficacement de nouveaux textes avec précision.

5. Classification des sentiments

La cinquième étape consiste à classer chaque phrase subjective de l'ensemble de données textuelles en groupes de polarité. Ces groupes sont généralement représentés sur un continuum à deux extrémités (positif, négatif, bon, mauvais, etc.).

Les techniques de classification des sentiments (SC) peuvent être divisées en deux grandes catégories : l'approche d'apprentissage automatique (Machine Learning) et l'approche basée sur le lexique. Les approches basées sur l'apprentissage automatique impliquent la formation d'un algorithme de classification sur un ensemble de fonctionnalités sélectionnées spécifiquement pour la tâche, puis leur test sur un autre ensemble de données pour évaluer leur capacité à détecter les caractéristiques pertinentes et à produire des classifications précises. Ces approches utilisent des algorithmes d'apprentissage automatique qui exploitent des attributs linguistiques pour la classification .

D'autre part, l'approche basée sur le lexique s'appuie sur un ensemble prédéfini de termes appelé "lexique des sentiments". Cette méthode attribue des polarités à chaque terme en fonction de son impact émotionnel, puis classe les textes en fonction de la présence et de la polarité des termes du lexique. La construction d'un tel lexique fait l'objet du présent mémoire de fin d'étude (Voir chapitre 3).

2.2.4 Types d'analyse des sentiments

Il y a divers types d'analyse des sentiments, variant selon le degré de détail et la complexité de l'analyse. On peut identifier trois types principaux : l'analyse des sentiments au niveau du document, au niveau de la phrase et au niveau de l'aspect (voir Figure 3.1).

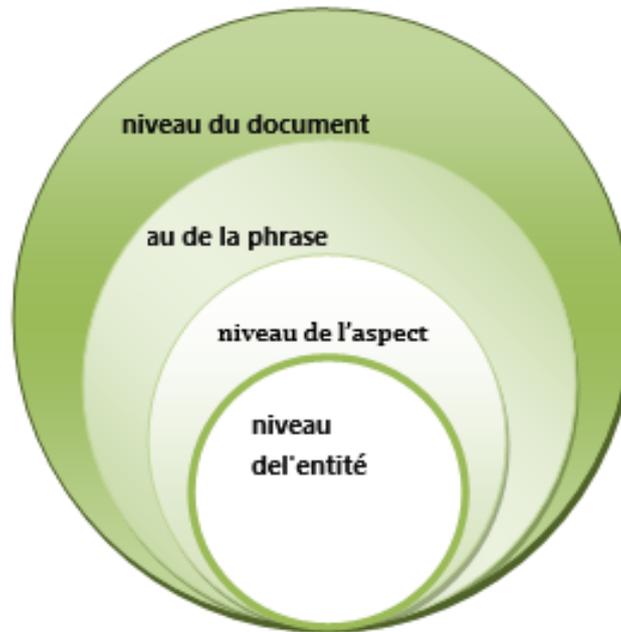


FIG. 2.3 : Les niveaux d'analyse des sentiments.

1. **L'analyse des sentiments au niveau du document** : consiste à évaluer la polarité globale d'un texte complet, qu'il s'agisse d'un avis, d'un article, d'un tweet, ou autre. Son but est de déterminer si le texte exprime une opinion positive, négative ou neutre sur le sujet en question [13]. Voici quelques exemples pour illustrer chaque polarité :
 - Négatif : "Il y a deux semaines, j'ai acheté ce produit et je suis extrêmement déçu. Il ne fonctionne pas correctement, il est bruyant et il consomme trop d'énergie. Je ne le recommande absolument pas."
 - Positif : "Ce produit est vraiment incroyable ! Il est facile à utiliser, silencieux et économe en énergie. Il répond parfaitement à mes besoins. Je le recommande vivement."
 - Neutre : "Ce produit est acceptable. Il fait ce qu'on attend de lui, sans plus. Il n'est ni bruyant ni silencieux, ni économe ni gourmand en énergie. Je n'ai pas grand-chose à dire à son sujet." En analysant l'ensemble du texte, on peut déterminer la tonalité générale ainsi que l'orientation des opinions exprimées, ce qui permet de mieux saisir la perception globale du sujet abordé.
2. **L'analyse des sentiments au niveau de la phrase** : implique de classer la polarité de chaque phrase d'un texte, en considérant que certaines phrases peuvent

ne pas exprimer d'opinion. Le but est de déterminer si chaque phrase est positive, négative, neutre ou objective [13]. Voici des exemples :

- Négatif : "J'ai acheté ce produit il y a deux semaines et je suis très déçu."
- Positif : "Ce produit est génial!"
- Neutre : "Il fait ce qu'on lui demande, sans plus."
- Objectif : "Il n'est ni bruyant ni silencieux, ni économe ni gourmand en énergie." En analysant chaque phrase individuellement, on peut déterminer la polarité de chaque opinion exprimée, ce qui permet une compréhension plus fine des nuances et des variations dans le ton du texte [13].

3. **L'analyse des sentiments au niveau de l'aspect** : vise à identifier et à classifier la polarité des opinions exprimées sur les différentes caractéristiques ou aspects d'un sujet [13]. Son objectif est de déterminer les points forts et les points faibles d'un produit, d'un service, d'une marque, etc., en fonction des avis des clients. Par exemple :

TAB. 2.2 : Exemple des avis des clients.

Aspect	Polarité
Fonctionnement	Négatif
Bruit	Négatif
Consommation d'énergie	Négatif
Facilité d'utilisation	Positif
Silence	Positif
Economie d'énergie	Positif
Satisfaction des besoins	Positif
Recommandation	Positif
Qualite générale	Neutre

L'analyse des sentiments au niveau de l'aspect offre une compréhension plus approfondie et détaillée des opinions des clients en identifiant les caractéristiques les plus louées ou les plus critiquées. Elle permet également de comparer les opinions sur différents aspects entre plusieurs produits, services ou marques, offrant ainsi des insights précieux pour comprendre les préférences des consommateurs et les domaines nécessitant une amélioration ou une mise en valeur.

4. **L'analyse des sentiments au niveau del'entité** : Au niveau de l'entité, l'objectif est de découper le texte pour identifier les différents aspects et les entités mentionnés, par exemple, dans le cadre du tourisme, une principale entité peut être "ville". Ensuite, il s'agit d'identifier les aspects associés à cette entité, tels que la qualité générale de la visite, la gestion des routes et la qualité de la nourriture. Enfin, pour chaque aspect identifié, nous évaluons le sentiment exprimé. Dans ce cas, le sentiment peut être positif pour la visite globale de ville, mais négatif pour la gestion des routes et la qualité de la nourriture. Cette approche permet d'avoir une vision plus détaillée des opinions exprimées dans le texte, en analysant les sentiments associés

à chaque aspect spécifique de l'entité principale [14].

2.2.5 Applications de l'analyse des sentiments

La gestion de contenu axée sur les opinions joue un rôle crucial dans le domaine de l'intelligence artificielle. Elle offre diverses applications essentielles, telles que l'évaluation des critiques concernant un produit spécifique par le biais de la classification des avis de produits en ligne. De même, elle permet la surveillance de l'évolution des opinions du public à l'égard d'un parti politique en extrayant des données à partir de sites d'actualités en ligne ou de contenus de blogs . Malgré la large utilisation des applications basées sur les opinions ou les retours, le champ du traitement du langage naturel se tourne actuellement vers les analyses de sentiments et les systèmes d'exploration des opinions pour une compréhension plus approfondie des attitudes et des émotions exprimées dans le langage humain [10]. Les principales applications des analyses de sentiments et de l'extraction d'opinions sont données ci-dessous :

- Achat de produits ou de services.
- Amélioration de la qualité du produit ou du service Haut du formulaire.
- Recherche marketing.
- Systèmes de recommandation.
- Détection de flamme.
- Détection d'opinion Spam.
- Élaboration de politiques.
- Prise de décision.

2.2.6 Approches pour l'analyse des sentiments

Il existe trois approches principales pour l'analyse des sentiments [20] :

1. **Basée sur des règles** : Cette méthode utilise des algorithmes manuellement codés pour évaluer le langage. Ces règles impliquent souvent la conversion en tokens, la lemmatisation, la racinisation et l'étiquetage morpho-syntaxique. Elles peuvent également utiliser des lexiques de mots. Cette approche analyse des mots spécifiques dans les phrases pour évaluer leur polarité et leur subjectivité afin de déterminer le sentiment global [20].
2. **Machine Learning** : Une version automatisée de l'analyse des sentiments basée sur des règles, elle utilise des capacités de Machine Learning. Ce modèle apprend à associer des mots à des polarités en étant entraîné sur des données d'entraînement. Une fois entraîné, il extrait et classe les caractéristiques à l'aide d'algorithmes de classification tels que le Deep Learning, Naïve Bayes, etc.

3. **Hybride** : Cette approche combine des méthodes basées sur des règles et des techniques de Machine Learning.

Chacune de ces approches a ses avantages et ses inconvénients, et le choix dépend souvent des besoins spécifiques de l'utilisateur ou de l'entreprise.

2.2.7 Défis liés à l'analyse des sentiments

Malgré les progrès rapides dans le domaine de l'analyse des sentiments et des technologies qui le sous-tendent, il reste encore relativement nouveau. Selon le livre "Analyse des sentiments" de Liu Bing (2020), le terme lui-même n'est devenu largement utilisé qu'à partir de 2003. Les outils nécessitent encore des améliorations et il reste beaucoup à apprendre de ces méthodes. Cependant, plusieurs inconvénients et défis sont couramment rencontrés dans ce domaine en pleine évolution [15].

- **Manque de contexte** : Le contexte joue un rôle fondamental dans l'interprétation de l'émotion véhiculée par un bloc de texte [15].
- **Utilisation de l'ironie et du sarcasme** : L'identification de l'ironie et du sarcasme dans un texte reste un défi pour les logiciels d'analyse des sentiments en raison de la nature subtile de ces formes de communication, qui reposent davantage sur des indices non verbaux que sur le choix des mots eux-mêmes [15].
- **Négation** : La notion de négation intervient lorsque l'utilisation d'un mot à connotation négative modifie le sens général d'une phrase pour exprimer l'opposé de sa signification littérale. Prenons par exemple la phrase : « Je ne dirais pas que les chaussures étaient bon marché ». Bien que cette phrase semble suggérer que les chaussures étaient probablement coûteuses ou du moins pas bon marché, elle n'exprime pas explicitement cette idée. Cependant, un outil d'analyse des sentiments pourrait avoir du mal à saisir cette subtilité et interpréter simplement le mot « bon marché » comme une indication positive [15].
- **Tournures idiomatiques** : Les expressions idiomatiques, telles que les proverbes ou les phrases figées, posent souvent problème aux outils d'analyse des sentiments et aux algorithmes d'apprentissage automatique sur lesquels ils reposent. Lorsqu'elles sont utilisées dans des contextes numériques tels que les réseaux sociaux ou les commentaires sur les produits, ces expressions peuvent être mal interprétées ou totalement ignorées par les systèmes de traitement automatique du langage naturel [15].

2.3 Solutions d'analyse des sentiments

Aujourd'hui, les solutions d'analyse des sentiments les plus efficaces pour le support client s'appuient sur l'intelligence artificielle (IA) et l'apprentissage automatique (ML) pour enrichir l'expérience client. Parmi ces solutions, IBM Watson Assistant se distingue

comme une plateforme de premier plan dans le domaine de l'intelligence artificielle conversationnelle. Grâce à l'utilisation de vastes modèles de langage (LLM), cette plateforme permet aux entreprises de développer des agents vocaux et des chatbots d'IA. Ces outils fournissent une assistance en libre-service automatisée de haute qualité à leurs clients, via une interface simple et conviviale [15].

2.4 Travaux connexes

L'article [2] traite de l'utilisation des moteurs de règles en informatique, en soulignant leur puissance et leur facilité d'utilisation. Les systèmes de règles sont considérés comme une extension ou une alternative à OWL, un langage formel pour la représentation des connaissances et le raisonnement. Les principaux avantages des moteurs de règles sont leur puissance et leur existence dans diverses applications. L'article [2] aborde l'utilisation d'ontologies dans divers domaines tels que l'intelligence artificielle, le Web sémantique, le génie logiciel, l'informatique biomédicale et l'architecture de l'information. Les ontologies sont utilisées comme une forme de représentation des connaissances sur un monde ou une partie spécifique de celui-ci, décrivant des individus, des classes, des attributs et des relations. Bien que l'article donne un aperçu de l'utilisation des ontologies pour guider l'accès à l'information, il reconnaît également que les techniques existantes du Web sémantique peuvent ne pas répondre pleinement aux besoins d'accès contextuel des utilisateurs. Les utilisateurs ont souvent leurs propres représentations des domaines, qui peuvent différer des ontologies établies, ce qui a un impact sur l'accès à l'information.

L'article [16] utilise une méthode de découverte de corpus ontologique pour améliorer l'exploration des sentiments et des opinions, en passant d'une approche statistique à une approche ontologique. Cette méthode précède la méthode symbolique, permettant une vue d'ensemble rapide des textes en extrayant des concepts de domaine et en fournissant une ontologie pour faciliter le développement de la grammaire de la méthode symbolique. Une méthode hybride est utilisée pour réduire les coûts de configuration en combinant l'annotation de texte avec des algorithmes d'apprentissage automatique pour les jugements initiaux au niveau du texte.

L'article [16] décrit une méthode symbolique avec une grammaire spécifique à un domaine et une méthode de distribution statistique pour la création d'ontologies.

Le document de recherche [17] a utilisé une méthodologie d'étude de cas pour valider la recherche et développer des ontologies raffinées pour l'analyse des sentiments. Les avis publiés sur le site www.amazon.com ont été sélectionnés pour trois modèles différents de téléphones mobiles intelligents, deux modèles différents de téléphones portables simples et deux modèles différents de téléphones portables humides et sales. Les caractéristiques et les attributs des téléphones mobiles ont été extraits manuellement pour créer des ontologies.

Le calcul de la polarité des opinions (OGC) à partir des commentaires des critiques a été effectué en suivant des étapes spécifiques. Les caractéristiques et les attributs ont été extraits de revues utilisées pour évaluer des ontologies spécialisées pour différents types de téléphones mobiles. Des ontologies raffinées pour les téléphones portables simples, intelligents et mouillés/sales ont été vérifiées après avoir analysé les sentiments de sites

Web tels que www.ebay.com. La méthodologie impliquait un examen et une évaluation approfondis des critiques et des opinions exprimées par les évaluateurs afin d'atteindre les objectifs de recherche.

Le document de recherche [18] utilise diverses méthodes d'analyse des sentiments, en se concentrant sur des modèles d'apprentissage automatique pour l'évaluation des sentiments.

Trois approches différentes sont présentées pour la classification des sentiments : les approches linguistiques, statistiques et hybrides. Des expérimentations sur des commentaires de films publiés sur des blogs d'opinion montrent que les méthodes statistiques surpassent les approches linguistiques. Les méthodes statistiques, en particulier l'apprentissage supervisé, sont mises en avant pour l'analyse des sentiments. Les commentaires sont représentés sous la forme d'un ensemble de variables, et un modèle est créé à l'aide d'exemples de texte étiqueté pour classer les nouveaux commentaires non étiquetés.

Plusieurs recherches montrent que les techniques d'apprentissage automatique donnent de meilleurs résultats que les méthodes linguistiques, même si les dictionnaires d'opinion utilisés ne sont peut-être pas optimaux. Ces méthodes montrent l'utilisation de l'apprentissage automatique, des approches statistiques et de l'apprentissage supervisé pour l'analyse des sentiments dans le document de recherche.

L'article [19] utilise des techniques de classification automatique des données de texte pour concevoir une application permettant d'analyser les commentaires et de fournir des recommandations. Il identifie les commentaires exprimant des opinions ou représentant du bruit, évalue les opinions réelles sur une échelle de 1 à 5 et calcule les notes moyennes pour générer une recommandation finale. Des études expérimentales ont été menées sur différents algorithmes de classification afin de sélectionner celui qui convient le mieux à la solution, avec des résultats expérimentaux satisfaisants obtenus. L'article [19] traite de la fréquence inverse des documents (IDF) dans le calcul de la fréquence des mots, en mettant en évidence sa comparaison entre des documents connus contenant le mot spécifique. Des méthodes d'apprentissage supervisé telles que les machines à vecteurs de support (SVM) et les classificateurs K-Nearest Neighbors (KNN) sont mentionnées pour les tâches de classification de texte, soulignant leur efficacité et leur rapidité. Une méthode basée sur l'identification de mots caractéristiques dans les avis en ligne est présentée comme un moyen de déterminer les faiblesses des produits. L'article [19] aborde également les méthodes de recommandation qui impliquent de faire des prédictions automatiques en fonction des intérêts des utilisateurs, recueillis à partir de nombreux avis d'utilisateurs.

L'article [14] utilise des approches basées sur le comptage, en utilisant des documents comme contexte pour saisir la similitude sémantique entre eux.

2.5 Conclusion

Dans ce chapitre, nous avons abordé la notion d'analyse de sentiment, mettant en lumière son rôle essentiel dans de nombreux domaines. Nous avons examiné le processus d'analyse des sentiments, explorant les différentes approches disponibles ainsi que les applications concrètes de cette discipline. De plus, nous avons passé en revue les principaux

défis auxquels nous sommes confrontés dans ce domaine. Nous avons procédé à l'examen et à l'étude des diverses recherches antérieures qui ont abordé un sujet similaire à celui de l'étude actuelle. Cette analyse et cette étude ont contribué à élargir notre compréhension du sujet, en nous permettant de découvrir les recherches antérieures et les résultats obtenus, tout en aidant à identifier les lacunes dans les connaissances et à déterminer les orientations de recherche futures.

Chapitre 3

Conception et Réalisation

3.1 Introduction

Dans ce chapitre nous allons décrire les ressources et la méthode utilisés dans la construction d'une ontologie pour l'analyse des sentiments dans le cadre des réseaux sociaux des algériens. Dans un premier temps nous allons décrire WordNet et SentiWordNet ensuite on aborde la méthode de conception et de vérification de notre ontologie (DZSentiWordNet) à partir de SentiWordNet.

3.2 WordNet

WordNet est un système de référence lexicale en ligne, inspiré par les théories psycholinguistiques actuelles sur la mémoire lexicale humaine. Les noms, verbes et adjectifs anglais sont organisés en ensembles de synonymes, chacun représentant un concept lexical sous-jacent. Différentes relations relient ces ensembles de synonymes.

WordNet constitue une ressource lexicale exhaustive couvrant le lexique général. Il a été largement utilisé comme référence dans de nombreuses applications de recherche d'informations. La compréhension des mots ne se limite pas à leurs significations, mais s'étend également au contexte dans lequel ils sont utilisés. L'association des mots avec les sens appropriés fournit la compréhension conceptuelle nécessaire. Les termes ayant des significations similaires sont organisés autour d'une idée appelée "Synset". Ces groupes de "Synsets" sont liés entre eux par des relations lexicales préétablies. En outre, les catégories de haut niveau de WordNet imposent certaines restrictions sur le nombre de sens des mots tout en maintenant l'espace de recherche limité dans les processus de généralisation. Les concepts sont les unités organisationnelles de WordNet et englobent plus d'un mot, y compris les composés, les collocations, les expressions idiomatiques et les verbes à particule. Les composés, les collocations, les expressions idiomatiques élargissent la notion de stockage des mots dans le lexique pour inclure les informations conceptuelles qui peuvent ne pas être représentées lexicalement par un seul mot [21].

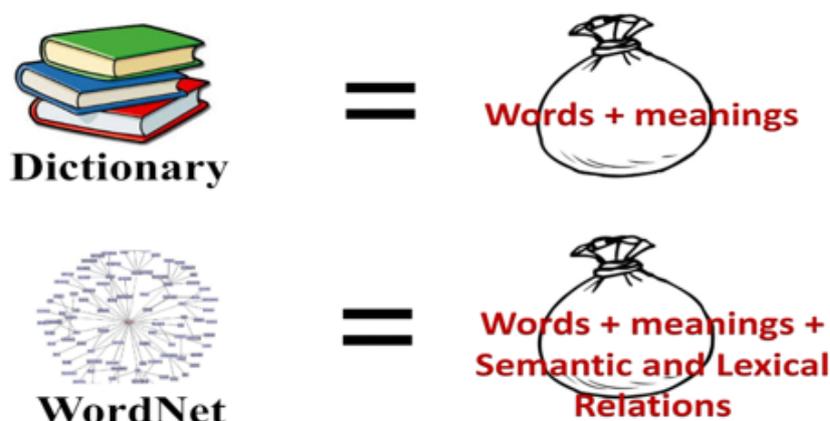


FIG. 3.1 : Dictionnaire vs WordNet [22]

3.2.1 Structuration des Relations Lexicales dans WordNet

La structure de WordNet est représentée par les concepts suivants [22] :

- **Le Mot** : Représente un seul mot dans la langue, tel que "voiture" ou "livre" en français.
- **Synset (Ensemble de Synonymes)** : Un ensemble de mots qui partagent le même sens, comme "voiture", "voiture de sport", "voiture berline", et "voiture tout-terrain". Ces ensembles sont organisés pour représenter les relations lexicales entre les mots.
- **Relations Lexicales** : Décrivent les relations entre les mots dans WordNet, telles que la meronymie, l'hyperonymie, l'hyponymie, la synonymie, etc. Ces relations montrent comment les mots sont liés dans différents ensembles de mots.
- **Structure Hiérarchique** : Représente la relation hiérarchique entre les Synset, où il y a un Synset principal (Hypernym) représentant un concept plus général partagé avec un Synset secondaire (Hyponym) qui représente un concept plus spécifique.
- **Définitions et Exemples** : Fournissent des détails supplémentaires sur les significations des mots et leurs utilisations courantes dans la langue.

3.2.2 Ordre hiérarchique dans WordNet

Le concept de hiérarchie dans WordNet est appelé "hiérarchie hypernymique". C'est un système de classification des mots basé sur leurs relations sémantiques. Dans WordNet, chaque mot est associé à un "synset", qui est un ensemble de synonymes. Les synsets sont organisés en une structure hiérarchique où les termes généraux sont placés au sommet et les termes plus spécifiques sont situés en bas.

Exemple :

Niveau le plus général (Voir Figure 3.2) :

- **"Véhicule (véhicule)"** : c'est le concept le plus large. Il englobe tous les types de véhicules tels que les voitures, les camions et les motos.

Niveau intermédiaire (plus spécifique) :

- **"Voiture (voiture)"** : ce concept est un type de "véhicule". Il inclut tous les types de voitures.
- **"Moto (moto)"** : ce concept est également un type de "véhicule", mais il se réfère spécifiquement aux motos.

Niveau le plus spécifique :

- **"Voiture de sport (voiture de sport)"** : ce concept est un type spécifique de "voiture". Il inclut uniquement les voitures de sport.

Relations dans la hiérarchie :

- **Hypernymie (relation de généralisation) :**
 "Voiture (voiture)" est l'hyperonyme de "voiture de sport (voiture de sport)".
 "Véhicule (véhicule)" est l'hyperonyme (terme général) de "voiture (voiture)".
- **Hyponymie (relation de spécification) :**
 "Voiture" est l'hyponyme (terme spécifique) de "véhicule (véhicule)".
 "Voiture de sport (voiture de sport)" est l'hyponyme de "voiture (voiture)".

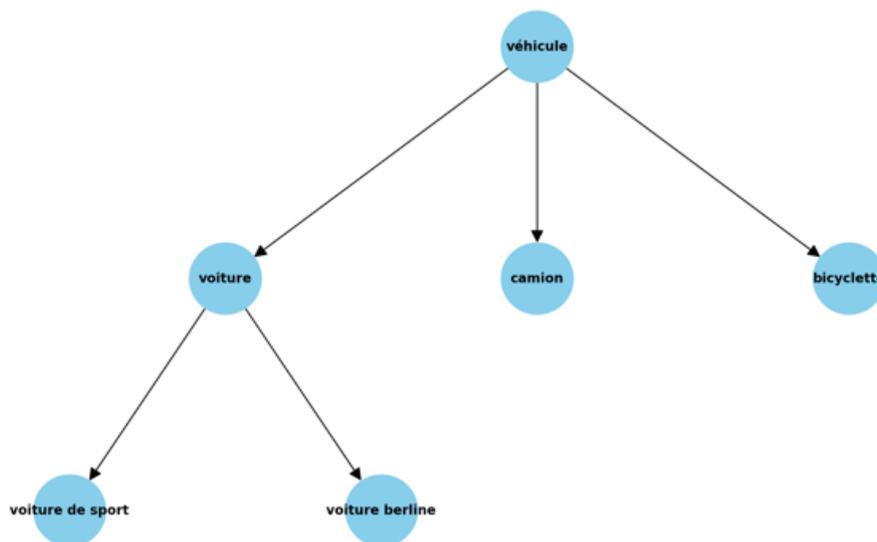


FIG. 3.2 : Hiérarchie du plus général au plus spécifique

3.2.3 Relations Lexico-Sémantique dans WordNet (Relation entre les mots)

- **Synonymie :** Il s'agit de la relation entre des mots ayant un sens identique ou similaire. Dans WordNet, les mots synonymes sont regroupés dans un ensemble appelé "Synset", représentant ainsi le même concept.

Exemple : Dans un Synset, on peut trouver les mots "plante" et "flore" (plant, Flora). Ces mots sont reliés par la synonymie car ils désignent le même concept.

- **Antonymie :** C'est la relation entre des mots ayant des significations opposées ou contradictoires.

Exemple : "Jour" et "nuit" (Day, night) sont des exemples de mots antonymes. Chacun est l'antonyme de l'autre.

- Gradation : Il s'agit de la relation entre des mots représentant une séquence temporelle, graduelle ou spatiale.

Exemple : "Matin", "après-midi", "soir" (matin, après-midi,soir)sont des mots représentant différentes périodes de la journée, et ils sont liés par la gradation. Ces relations aident à analyser et comprendre les liens linguistiques entre les mots, renforçant ainsi la compréhension de l'organisation des concepts et des significations dans la langue.

3.2.4 Relations sémantiques : (Relation entre les synsets)

- Hyperonymie / Hyponymie : Relation de type "is-a-kind-of" (est-un-genre-de).

Exemple :

"fruit" est un hyperonyme de "mangue" et "mangue" est un hyponyme de "fruit". "Fruit" est le terme général (hyperonyme) pour "mangue". "Mangue" est un terme spécifique (hyponyme) pour "fruit".

- Meronymie / Holonomie : Relation de type "part-whole" (partie-tout).

Exemple :

"main" est un meronymie de "corps" et "corps" est un holonyme de "main". "Main" est une partie (meronymie) de "corps". "Corps" est le tout (holonyme) qui inclut "main".

- Entraînement : Une action ou un état implique nécessairement une autre action ou état.

Exemple :

"ronfler" implique "dormir". Si quelqu'un "ronfle", cela signifie qu'il "dort".

- Attribut : Relation entre les synsets de noms et d'adjectifs. **Exemple :**

"chaud" est une valeur ou un attribut de "température". "Chaud" est une caractéristique de l'état de "température".

- Nominalisation (Nominalisation) : Relation entre les synsets de noms et de verbes.

Exemple :

"service" nominalise le verbe "servir". "Service" est le nom dérivé du verbe "servir".

3.3 WordNet Arabe (Arabic WordNet : AWN)

3.3.1 Description de AWN

Le WordNet Arabe comprend environ 10 000 synsets et est une traduction littérale d'un petit sous-ensemble de WordNet. Ce sous-ensemble a été créé en identifiant d'abord

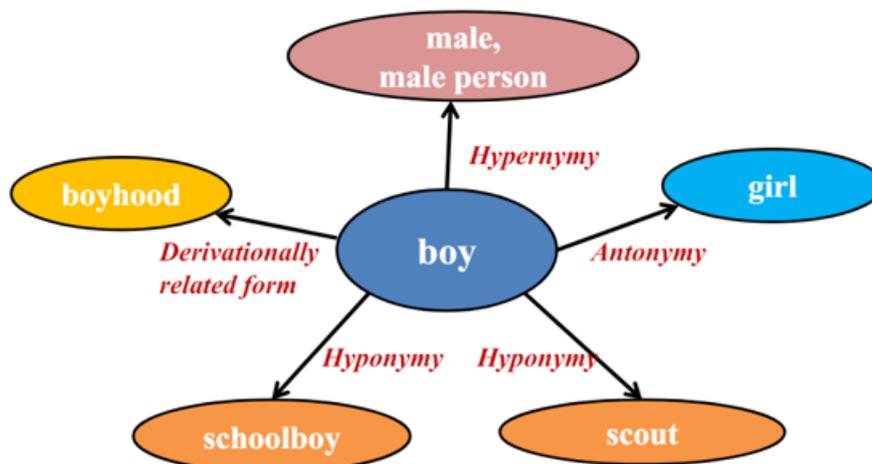


FIG. 3.3 : Les relations dans WordNet. [22]

un ensemble de synsets dans WordNet, appelés Concepts de Base Communs, qui ont ensuite été traduits manuellement en arabe. Ces concepts de base existent dans 12 langues (dans EuroWordNet et BalkaNet, voir Annex A), et on suppose donc qu'ils existent également en arabe. Ces concepts ont été principalement étendus vers le bas avec des concepts plus spécifiques et vers le haut avec des concepts plus généraux afin d'améliorer la connectivité maximale de ces concepts de base. Ainsi, le WordNet Arabe est un sous-ensemble de WordNet. Les relations sémantiques entre les synsets traduits ont été conservées telles quelles à partir de WordNet. Comme WordNet a été mappé à l'ontologie SUMO (Niles & Pease, 2001), le WordNet Arabe hérite de toutes les préoccupations logiques et ontologiques de WordNet, puisqu'il en est une traduction littérale et un sous-ensemble [23].

3.3.2 Création et structuration de AWN

Les concepts de base arabes sont identifiés et étendus via des relations hyponymiques pour créer un réseau lexical. Les concepts de base communs (CBCs, voir Annexe A) sont encodés en synsets à partir de plusieurs langues, et des concepts spécifiques sont ajoutés et traduits manuellement en arabe. Les synsets anglais liés à SUMO (voir Annexe A) sont ensuite traduits et alignés avec les concepts arabes. Cette méthode permet de créer un lexique cognitif interconnecté pour l'arabe, relié aux concepts des autres langues et ontologies.

3.3.3 Concepts de base Arabes (Arabic Base Concepts)

(a) Définition et extension via les relations hyponymiques

- Définition : Identification des concepts de base arabes.
- Extension : Utilisation des relations hyponymiques pour étendre ces concepts (ex : "animal" à "chien", "chat").
- Réseau Lexical de Base (Core WordNet) : Création d'un réseau lexical de base hiérarchique à partir des relations hyponymiques.

(b) Concepts de base communs (Common Base Concepts - CBCs)

- Langues Participantes : Concepts de base communs (CBCs) rassemblés de 12 langues dans les projets EuroWordNet et BalkaNet.
- Encodage en Synsets : Encodage des CBCs en synsets (ensembles synonymiques).

(c) Ajout et traduction des concepts spécifiques à la langue

- Ajout de Concepts Spécifiques : Ajout de concepts spécifiques à chaque langue.
- Traduction Manuelle : Traduction manuelle des concepts en synsets arabes.

(d) Correspondance avec SUMO

- Relation avec SUMO : Utilisation de SUMO pour relier les concepts à des concepts logiques spécifiques.
- Synsets Anglais : Correspondance des synsets anglais avec leurs équivalents arabes.

3.4 Senti WordNet

Senti WordNet est une ressource développée spécifiquement pour faire l'analyse des sentiments. L'analyse des sentiments, parfois appelée extraction d'opinions, est le processus par lequel un texte est analysé pour en extraire des opinions, des sentiments, etc., et pour classer un document ou une phrase en différentes classes (positive, négative, neutre). Parfois, l'analyse essaie de donner à un document ou à une phrase (avis) un degré d'importance (la polarité). Le processus d'analyse des sentiments utilise de nombreux autres domaines, comme le traitement automatique du langage naturel (NLP), à travers une combinaison d'étapes de prétraitement et d'algorithmes d'apprentissage automatique [24].

3.5 Conception et Réalisation d'un Senti WordNet (DZ-SentiWordNet)

Notre objectif étant de réaliser un Senti WordNet pour faire l'analyse des sentiments des textes postés dans les réseaux sociaux algériens, nous allons décrire dans cette partie les algorithmes de base nous permettant de réaliser cette tâche. Il faut noter qu'il existe deux approches pour la réalisation, la première utilise le machine learning qui à partir d'un corpus de départ peut extraire les concepts avec leurs polarité. Par ailleurs la seconde approche utilise directement Senti WordNet de l'anglais et fait une conversion (mapping) de tous ses concepts vers l'arabe ensuite faire les ajustements spécifiques au dialecte Algériens. Nous avons choisi la seconde approche, coûteuse en effort humain dans sa partie vérification et validation, mais plus sûr. En effet nous allons nous baser sur deux linguistes pour vérifier et valider les résultats de conversion. Le SentiWord ainsi construit est nommé DZSentiWordNet.

3.5.1 Caractéristiques du dialecte algérien :

L'arabe algérien, appelé darja ou darjà par ses locuteurs, est la langue maternelle de 75 à 85 % de la population en Algérie et est maîtrisé par 95 à 100 % des Algériens. Contrairement à l'arabe littéraire, la darja est un dialecte faisant partie du groupe des dialectes arabes nord-africains. Ce dialecte est enrichi par l'influence de plusieurs langues, en raison de l'histoire linguistique complexe de la région. Ainsi, il intègre des mots et des expressions provenant du berbère, du français, de l'italien, entre autres. Le résultat est un langage riche et varié, reflétant le multiculturalisme de l'Algérie [21].

3.5.2 Algorithmes de construction du DZSentiWordNet

Pour la réalisation de nos programmes nous avons utilisé le langage Python et l'API `nltk`.

(a) Algorithme de parcours de tous les synsets de Senti WordNet

Début

- Importation des bibliothèques
 - Importer la librairie `nltk`, qui est nécessaire pour travailler avec WordNet.
- Accéder aux synsets (`get_tous_synsets`)
 - Cette fonction stocke tous les synsets de WordNet dans une liste.
 - Elle parcourt toutes les catégories grammaticales (*POS tags*) en utilisant `nltk.corpus.wordnet.POS_TAGS`.
 - Pour chaque catégorie grammaticale, elle récupère les synsets correspondants en utilisant `nltk.corpus.wordnet.synsets(pos)` et les ajoute à la liste `synsets`.
 - Enfin, elle renvoie la liste `synsets` contenant tous les synsets de WordNet.
- Accéder aux détails de chaque synset (`get_details_synsets`)
 - Cette fonction prend une liste de synsets en entrée et affiche les détails de chaque synset.
 - Elle parcourt chaque synset dans la liste `synsets`.
 - Pour chaque synset, elle retourne :
 - * Nom du synset en utilisant `synset.name()`.
 - * La définition du synset en utilisant `synset.definition()`.
 - * Des exemples de phrases utilisant le mot du synset en utilisant `synset.examples()`.
- Fin

(b) Algorithmes de traduction de SentiWordNet

Début

- Importation des bibliothèques
 - `nltk` : nécessaire pour travailler avec WordNet.
 - `nltk.corpus.sentiwordnet` : Importer les fonctionnalités de SentiWordNet.
 - `googletrans` : fonctions de traduction automatique utilisant l'API Google Translate.
- Fonction `translate_sentiwordnet_synsets`
 - Cette fonction prend en entrée une liste de `sentiwordnet_synsets` (des tuples contenant les noms des synsets et leurs scores de polarité).
 - Elle effectue les étapes suivantes :
 - (1) Initialisation d'un objet `Translator` en utilisant `googletrans.Translator` avec la `target_language` spécifiée, ici l'Arabe.
 - (2) Initialisation de listes vides : `translated_synsets` pour stocker les noms des synsets traduits et `polarity_scores` pour stocker les valeurs de polarité.
 - (3) Boucle sur chaque synset et son score de polarité dans la liste `sentiwordnet_synsets`.
 - (4) Pour chaque nom de synset :
 - * Traduction du nom de synset vers la langue cible en utilisant `translator.translate(synset_name).text`.
 - * Ajout du nom de synset traduit et du score de polarité d'origine à leurs listes respectives.
 - (5) Création de la liste de tuples (`sentiwordnet_synsets_traduit`) en combinant les synsets traduits et leurs scores de polarité.
 - (6) Renvoi la liste `sentiwordnet_synsets_traduit`.
- Fin

- (c) **Adaptation de SentiWordNet traduit en Arabe pour prendre charge le dialecte algérien** Cet algorithme utilise un dictionnaire contenant l'équivalent des mots arabe en dialecte algériens.

Début

- Importation des bibliothèques
 - `nltk` : nécessaire pour travailler avec WordNet.
 - `nltk.corpus.sentiwordnet` : Importer les fonctionnalités de SentiWordNet.
- Fonction `adapter_sentiwordnet_synsets`
 - Cette fonction prend en entrée une liste de `sentiwordnet_synsets` (des tuples contenant les noms des synsets et leurs scores de polarité).

- Elle effectue les étapes suivantes :
 - (1) Initialisation de listes vides : `synsets_adapté` pour stocker les noms des synsets adaptés et `polarity_scores` pour stocker les valeurs de polarité.
 - (2) Boucle sur chaque synset et son score de polarité dans la liste `sentiwordnet_synsets`.
 - (3) Pour chaque nom de synset :
 - * Recherche l'équivalent en dialecte algérien du nom de synset.
 - * Ajout du nom de synset adapté et du score de polarité d'origine à leurs listes respectives.
 - (4) Création de la liste de tuples (`sentiwordnet_synsets_adapté`) en combinant les synsets adaptés et leurs scores de polarité.
 - (5) Renvoie la liste `sentiwordnet_synsets_adaptés`.
- Fin

3.5.3 Vérification et validation du DZSentiWordNet

Pour vérifier et valider DZSentiWordNet, l'effort humain est inévitable, il nous permet d'avoir une ressource sûre. En effet deux linguistes vont se charger de cette partie un peu longue en temps. Cette étape est en cours de réalisation.

3.6 Conclusion

Dans ce chapitre nous avons décrit la partie réalisation de notre projet. En effet nous avons utilisé le langage Python et la bibliothèque open source nltk pour réaliser les fonctions de conversion (mapping) et le service Google Translate pour la traduction du Senti WordNet vers un SentWordNet qui prend en charge la langue Arabe ensuite nous avons utilisé un dictionnaire Arabe -dialecte Algérien pour adapter le Senti WordNet traduit. Pour la vérification et la validation l'apport des linguistes est incontournable.

Conclusion

Conclusion générale

Cette étude représente une étape importante dans la compréhension et le développement des domaines de l'intelligence artificielle et de l'analyse des données, en offrant une vue d'ensemble sur l'utilisation des techniques d'ontologie et d'analyse des sentiments dans le contexte des réseaux sociaux. Elle a débuté par l'analyse et la clarification des concepts fondamentaux de l'ontologie et de l'analyse des sentiments, contribuant ainsi à une compréhension plus approfondie du contexte social et culturel.

Des jalons importants ont été franchis au cours de cette recherche, avec un examen des concepts fondamentaux de l'ontologie et de l'analyse des sentiments, ainsi que du contexte entourant les réseaux sociaux. Nous avons commencé par analyser l'évolution et la définition de l'ontologie, ainsi que ses objectifs, puis nous avons abordé les outils disponibles pour travailler avec l'ontologie et le rôle du web sémantique en tant que contexte important pour l'étude.

Dans la deuxième partie de la recherche, la compréhension et l'analyse ont été approfondies lors de l'exploration des fondements de l'analyse des sentiments et de son importance dans plusieurs domaines. Nous avons commencé par une définition précise de l'analyse des sentiments et avons souligné son importance dans divers domaines, puis nous avons exploré le processus d'analyse des sentiments, de la collecte des données textuelles à l'interprétation des résultats. Nous avons également examiné les différents types et techniques d'analyse des sentiments, ainsi que les défis auxquels sont confrontés les chercheurs et les praticiens dans ce domaine.

Enfin, nous avons présenté un travail pratique de construction d'une ontologie pour l'analyse des sentiments dans le contexte des réseaux sociaux des Algériens. Nous avons utilisé des ressources telles que WordNet et SentiWordNet et avons développé une méthodologie pour concevoir et mettre en œuvre l'ontologie (DZSentiWordNet) en nous appuyant sur SentiWordNet. Nous avons décrit en détail les méthodes et les outils utilisés dans ce travail, y compris l'utilisation de Python et de la bibliothèque nltk, ainsi que le recours au service Google Translate pour traduire SentiWordNet en SentWordNet prenant en charge la langue arabe. De plus, nous avons souligné l'importance de l'intervention des linguistes pour la vérification et la validation afin d'assurer l'exactitude des résultats.

En résumé, cette étude montre comment la compréhension de l'ontologie et de l'analyse des sentiments peut contribuer de manière significative au développement des technologies et des outils bénéfiques pour la société dans son ensemble. Avec la poursuite des efforts et des investissements dans la recherche et le développement dans les domaines de l'intelligence artificielle et de l'analyse des données, nous pouvons nous attendre à un avenir prometteur et à un développement durable dans tous les aspects de la vie quotidienne.

Bibliographie

- [1] Bodon, Charles. "De l'Ontologie philosophique aux ontologies informatiques : implications sémantiques d'un passage du singulier au pluriel." Implications philosophiques (2023).
- [2] ounes, ASKANE, EL-OUCHI Younes, EL MESSBAHI Salima, EL JADIDI ALAOUI Sadik, ALI Younes, et MounaHamdane. « L'Ontologie (Informatique) », Université Abdelmalek Essadi (2010).
- [3] HACINE GHERBI Ahcine, Mémoire Présenté pour l'obtention du Diplôme de Magister Construction d'une Ontologie pour le WEB Sémantique, École doctorale d'informatique Option Ingénierie des Systèmes Informatiques (2014).
- [4] Wikipedia; Tim Berners-Lee; https://fr.wikipedia.org/wiki/Tim_Berners-Lee; Consulté le 01 Avril 2024.
- [5] « Chapitre1.pdf ». Les ontologies. Université Abou Bekr Belkaid – Tlemcen- Consulté 3 avril 2024. <http://dspace.univ-tlemcen.dz/bitstream/112/5308/4/Chapitre1.pdf>.
- [6] Wikipédia; « Ontologie (informatique) ». [https://fr.wikipedia.org/w/index.php?title=Ontologie_\(informatique\)&oldid=214191659#Outils](https://fr.wikipedia.org/w/index.php?title=Ontologie_(informatique)&oldid=214191659#Outils). Consulté le 17 Mars 2024.
- [7] EL MASSARI, HAKIM; Proposition d'un modèle de prédiction basé sur Machine Learning et le web sémantique; Mémoire de Fin d'études Master; Université de 8 Mai 1945 – Guelma, 2023.
- [8] BECHIR, Abderrahim Arslane. Un système de Catégorisation des Acteurs dans un Réseau Social Educatif. Mémoire de Fin d'études Master. Université de 8 Mai 1945 – Guelma - 2023.
- [9] « Recherche Google ». Consulté le 24 mai 2024. [الاجتماعي التواصل مواقع عن صور](#)
- [10] Nedioui Med Abdelhamid; Techniques d'apprentissage automatique pour l'analyse et la fouille des sentiments dans les réseaux sociaux. Présentée pour l'obtention du diplôme de Doctorat en sciences en Informatique; Université Mohamed Khider – BISKRA. 2021.
- [11] image « analyse des sentiments - Recherche Google ». Consulté le 01 Avril 2024.
- [12] Amazon Web Services, Inc. « Qu'est-ce que l'opinion mining ? - L'opinion mining expliqué – AWS ». <https://aws.amazon.com/fr/what-is/sentiment-analysis/>, Consulté le 07 MARS 2024.

- [13] Eva. « Analyse des sentiments : guide des bonnes pratiques et exemples ». Oscar Black (blog), <https://oscar-black.com/blog/marketing-digital/analyse-des-sentiments/>. 2 juillet 2023.
- [14] Cherif, Firas. « Exploration des approches d'analyse des sentiments à base d'aspects ciblés », UNIVERSITÉ DU QUÉBEC À MONTRÉAL. AVRIL 2022.
- [15] « Qu'est-ce que l'analyse des sentiments? | IBM », 26 décembre 2023. <https://www.ibm.com/fr-fr/topics/sentiment-analysis>. Consulté le 01 Avril 2024.
- [16] Sigrid Maurel et Luca Dini; Exploration de corpus pour l'analyse de sentiments. DEFT'09 « DÉfi Fouille de Textes », Atelier de clôture, Paris, 22 juin 2009.
- [17] Syed Zeeshan Haider. AN ONTOLOGY BASED SENTIMENT ANALYSIS. Master Degree Project in Informatics. Spring term Year 2012.
- [18] BEKADDOURI Nadia et MOHAMMED SEGHIR Nadia. Modèles d'apprentissage automatique pour l'analyse des sentiments. UNIVERSITE ABDELHAMID IBN BADDIS MOSTAGANEM. 2013.
- [19] Mosteghanemi Samira et Froukhi Afaf. Système de recommandation multilingue basé sur l'analyse des sentiments des opinions dans les commentaires en ligne. Université saad d'Alger Blida-1. 06/12/2020.
- [20] « Qu'est-ce que l'analyse des sentiments? | Un guide complet de l'analyse des sentiments ». <https://www.elastic.co/fr/what-is/sentiment-analysis>. Consulté le 28 mai 2024.
- [21] Elkateb, Sabri, William Black, Piek Vossen, David Farwell, Horacio Rodríguez, Adam Pease, Musa Alkhalifa, and Christiane Fellbaum. "Arabic WordNet and the Challenges of Arabic." In Proceedings of the International Conference on the Challenge of Arabic for NLP/MT, 15–24. London, UK, 2006. <https://aclanthology.org/2006.bcs-1.2>.
- [22] <https://www.iitp.ac.in/ai-nlp-ml/course/cscl/WordNet-WSD-iitp-28Mar17-By%20Sudha.pdf>
- [23] Benjamin, Martin. "Problems and Procedures to Make Wordnet Data (Retro)Fit for a Multilingual Dictionary." In Proceedings of the 8th Global WordNet Conference (GWC), edited by Christiane Fellbaum, Piek Vossen, Verginica Barbu Mititelu, and Corina Forascu, 26–33. Bucharest, Romania : Global Wordnet Association, 2016. <https://aclanthology.org/2016.gwc-1.5>.
- [24] Madani, Youness, Mohammed Erritali, Jamaa Bengourram, and Françoise Sailhan. "A Hybrid Multilingual Fuzzy-Based Approach to the Sentiment Analysis Problem Using SentiWordNet." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 28, no. 03 (June 2020) : 361–90. <https://doi.org/10.1142/S0218488520500154>.

Annexes

Annexe A

Définitions

Euro WordNet : Projet européen pour créer des dictionnaires sémantiques pour plusieurs langues européennes. Les dictionnaires sont interconnectés par des concepts de base communs pour faciliter la traduction entre les langues.

Balka Net : Projet similaire à Euro WordNet, mais pour les langues des Balkans. Comprend des langues comme le bulgare, le roumain, le turc, le grec et le serbe, avec des dictionnaires interconnectés par des concepts communs.

CBCs : La technologie CPCs est un processus en plusieurs étapes qui vise à créer une base de données de connaissances multilingue et sémantiquement interconnectée qui permet la compatibilité et l'intégration entre différentes langues, et garantit une traduction précise et appropriée des concepts en langue arabe, ce qui contribue à faire de l'arabe WordNet un puissant outil et efficace dans le traitement du langage naturel et d'autres applications linguistiques.

SUMO : Nom : Suggested Upper Merged Ontology. Objectif : Fournir une structure conceptuelle complète utilisée pour relier et unifier les concepts de diverses ontologies et domaines. Origine : Développée par une organisation appelée Teknowledge (Recherche sur les systèmes d'information cognitive).