



HAL
open science

Emergsem : une approche d'annotation collaborative et de recherche d'images basée sur les sémantiques émergentes

Damien Esse Zomahoun

► To cite this version:

Damien Esse Zomahoun. Emergsem : une approche d'annotation collaborative et de recherche d'images basée sur les sémantiques émergentes. Web. Université de Bourgogne, 2015. Français. NNT : 2015DIJOS019 . tel-01230075

HAL Id: tel-01230075

<https://theses.hal.science/tel-01230075>

Submitted on 17 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SPIM

Thèse de Doctorat



école doctorale **sciences pour l'ingénieur et microtechniques**
UNIVERSITÉ DE BOURGOGNE

Emergsem: une approche d'annotation collaborative et de recherche d'images basée sur les sémantiques émergentes

■ DAMIEN E. ZOMAHOUN

SPIM

Thèse de Doctorat



école doctorale sciences pour l'ingénieur et microtechniques
UNIVERSITÉ DE BOURGOGNE

N° Y Y Y

THÈSE présentée par

DAMIEN E. ZOMAHOUN

pour obtenir le

Grade de Docteur de
l'Université de Bourgogne

Spécialité : **Informatique**

Emergsem: une approche d'annotation collaborative et de
recherche d'images basée sur les sémantiques émergentes

Soutenue publiquement le 05 Juin 2015 devant le Jury composé de :

DJAMAL BENSLIMANE	Professeur à l'Université de Lyon 1	Rapporteur
NADINE CULLOT	Professeur à l'Université de Bourgogne	Examinatrice
ERNESTO DAMIANI	Professeur à l'Université de Milan	Rapporteur
ALBERT DIPANDA	Professeur à l'Université de Bourgogne	Examineur
AHMED LBATH	Professeur à l'Université de Grenoble 1	Rapporteur
DAVY MONTICOLO	Maître de Conférences à l'Université de Lorraine	Examineur
KOKOU YÉTONGNON	Professeur à l'Université de Bourgogne	Directeur de thèse

REMERCIEMENTS

Je tiens à exprimer mes remerciements et toute ma gratitude à *Kokou Yétongnon*, Professeur à l'Université de Bourgogne, pour avoir accepté de m'encadrer pendant cette thèse ainsi que pour la confiance qu'il m'a accordée. Mr *Yétongnon*, je vous suis reconnaissant pour le temps que vous m'avez consacré et surtout pour votre patience. Je vous dis un sincère merci.

Je remercie les membres du Jury qui ont accepté d'examiner mon travail :

- *Ernesto Damiani*, Professeur à l'Université de Milan, *Djamal Benslimane*, Professeur à l'Université de Claude Bernard - Lyon 1 et *Ahmed Lbath*, Professeur à l'Université Joseph Fourier Grenoble 1 d'avoir accepté d'être rapporteur de ma thèse.
- *Albert Dipanda*, Professeur à l'Université de Bourgogne, *Nadine Cullot* Professeur à l'Université de Bourgogne et *Davy Monticolo*, Maître de Conférences à l'Université de Lorraine d'avoir accepté d'examiner mon travail.

Je remercie le personnel enseignant et administratif de l'Université de Bourgogne. Un merci à *Pierre Gouton*, Professeur à l'Université de Bourgogne pour leurs précieux conseils, *Christelle Caillot* et *Véronique Magnin*, Secrétaires à l'Université de Bourgogne pour leurs disponibilités.

Merci à tous les membres du laboratoire Le2i qui ont toujours su m'ouvrir leur porte quand j'en avais besoin, je les en remercie sans oublier mes collègues de la GS 13 et de la GS 14. Merci à *Anna Maria Roxin*, à *Pélagie Hounguè* et à *Cheikh Emani*.

Je tiens à remercier tous les membres de ma famille et plus particulièrement mon père *Albert*, ma mère *Madeleine*. Je leur dis toute ma gratitude pour leur soutien inconditionnel et permanent. Un sincère merci à mes frères *Euloge*, *Maurice*, *Nicéphore*, *Magloire* et à mes sœurs *Laetitia* et *Félicité* pour leurs prières. Ils sont mon inépuisable source d'encouragements et de soutien. Je tiens particulièrement à remercier *Sabine* et *Ferdinand Mêmêvègny* de m'avoir donné la chance de poursuivre mes études en France. Merci pour leurs aides et leurs précieux conseils.

Bien sûr je ne peux terminer sans un merci à *Basile Zomahoun*, à *Rodrigue Zomahoun* à tous mes amis qui n'ont cessé de me soutenir et de m'encourager. Merci *Jean Bruno Abilé*, *Ogonna Josephine Uche*, *Jean – Paul Yélomè*, *Armel Houéhomègan*, *Alain Alona*, sans oublier la société *Leto Services*.

DÉDICACE

A Cosme,

RÉSUMÉ

L'extraction de la sémantique d'une image est un processus qui nécessite une analyse profonde du contenu de l'image. Elle se réfère à leur interprétation à partir d'un point de vue humain. Dans ce dernier cas, la sémantique d'une image pourrait être générique (par exemple un véhicule) ou spécifique (par exemple une bicyclette). Elle consiste à extraire une sémantique simple ou multiple de l'image afin de faciliter sa récupération. Ces objectifs indiquent clairement que l'extraction de la sémantique n'est pas un nouveau domaine de recherche.

Cette thèse traite d'une approche d'annotation collaborative et de recherche d'images basée sur les sémantiques émergentes. Il aborde d'une part, la façon dont les annotateurs pourraient décrire et représenter le contenu des images en se basant sur les informations visuelles, et d'autre part comment la recherche des images pourrait être considérablement améliorée grâce aux récentes techniques, notamment le clustering et la recommandation. Pour atteindre ces objectifs, l'exploitation des outils de description implicite du contenu des images, des interactions des annotateurs qui décrivent la sémantique des images et celles des utilisateurs qui utilisent la sémantique produite pour rechercher les images seraient indispensables.

Dans cette thèse, nous nous sommes penchés vers les outils du Web Sémantique, notamment les ontologies pour décrire les images de façon structurée. L'ontologie permet de représenter les objets présents dans une image ainsi que les relations entre ces objets (les scènes d'image). Autrement dit, elle permet de représenter de façon formelle les différents types d'objets et leurs relations. L'ontologie code la structure relationnelle des concepts que l'on peut utiliser pour décrire et raisonner. Cela la rend éminemment adaptée à de nombreux problèmes comme la description sémantique des images qui nécessite une connaissance préalable et une capacité descriptive et normative.

La contribution de cette thèse est focalisée sur trois points essentiels : La représentation sémantique, l'annotation sémantique collaborative et la recherche sémantique des images.

- La représentation sémantique permet de proposer un outil capable de représenter la sémantique des images. Pour capturer la sémantique des images, nous avons proposé une ontologie d'application dérivée d'une ontologie générique.
- L'annotation sémantique collaborative que nous proposons consiste à faire émerger la sémantique des images à partir des sémantiques proposées par une communauté d'annotateurs.
- La recherche sémantique permet de rechercher les images avec les sémantiques fournies par l'annotation sémantique collaborative. Elle est basée sur deux techniques : le clustering et la recommandation. Le clustering permet de regrouper les images similaires à la requête d'utilisateur et la recommandation a pour objectif de proposer des sémantiques aux utilisateurs en se basant sur leurs profils statiques et dynamiques. Elle est composée de trois étapes à savoir : la formation de la communauté des utilisateurs, l'acquisition des profils d'utilisateurs

x

et la classification des profils d'utilisateurs avec l'algèbre de Galois. Des expérimentations ont été menées pour valider les différentes approches proposées dans ce travail.

Mots-clés : Annotation, collaboration, Images, Recommandations, Annotateurs, Ontologie, Sémantiques Émergentes.

ABSTRACT

The extraction of images semantic is a process that requires deep analysis of the image content. It refers to their interpretation from a human point of view. In this latest case, the image semantic may be generic (e.g., a vehicle) or specific (e.g., a bicycle). It consists in extracting single or multiple images semantic in order to facilitate its retrieval. These objectives clearly show that the extraction of semantic is not a new research field.

This thesis deals with the semantic collaborative annotation of images and their retrieval. Firstly, it discusses how annotators could describe and represent images content based on visual information, and secondly how images retrieval could be greatly improved thank to latest techniques, such as clustering and recommendation. To achieve these purposes, the use of implicit image content description tools, interactions of annotators that describe the semantics of images and those of users that use generated semantics to retrieve the images, would be essential.

In this thesis, we focus our research on the use of Semantic Web tools, in particular ontologies to produce structured descriptions of images. Ontology is used to represent image objects and the relationships between these objects. In other words, it allows to formally represent the different types of objects and their relationships. Ontology encodes the relational structure of concepts that can be used to describe and reason. This makes them eminently adapted to many problems such as semantic description of images that requires prior knowledge as well as descriptive and normative capacity.

The contribution of this thesis is focused on three main points : semantic representation, collaborative semantic annotation and semantic retrieval of images.

- Semantic representation allows to offer a tool for the capturing semantics of images. To capture the semantics of images, we propose an application ontology derived from a generic ontology.
- Collaborative semantic annotation that we define, provides emergent semantics through the fusion of semantics proposed by the annotators.
- Semantic retrieval allows to look for images with semantics provided by collaborative semantic annotation. It is based on clustering and recommendation. Clustering is used to group similar images corresponding to the user's query and recommendation aims to propose semantics to users based on their profiles. It consists of three steps : creation of users community, acquiring of user profiles and classification of user profiles with Galois algebra. Experiments were conducted to validate the approaches proposed in this work.

Keywords : Annotation, Collaboration, Images, Recommendations, Annotators, Ontology, Emergent Semantic.

SOMMAIRE

1	Introduction	1
1.1	Contexte général et problématique	2
1.2	Objectifs et contributions	6
1.3	Organisation du manuscrit	7
2	État de l'art	9
2.1	Introduction	10
2.2	Représentation des images	10
2.2.1	Représentation physique	10
2.2.1.1	La couleur	11
2.2.1.2	La texture	11
2.2.1.3	La forme	12
2.2.1.4	La localisation spatiale	12
2.2.2	Segmentation des images	13
2.2.2.1	Segmentation basée sur les régions	13
2.2.2.2	Segmentation basée sur le contour	14
2.2.2.3	Segmentation basée sur la classification ou le seuillage	14
2.2.3	Réduction du fossé sémantique	14
2.2.4	Représentation sémantique	15
2.2.4.1	Niveau syntaxique	16
2.2.4.2	Niveau sémantique	17
2.3	Annotation d'images	20
2.3.1	Les techniques d'annotation	21
2.3.1.1	Annotation manuelle	21
2.3.1.2	Annotation automatique	22

2.3.1.3	Annotation semi-automatique	26
2.3.2	Autres approches d'annotation	27
2.3.2.1	Approches collaboratives	27
2.3.2.2	Approches sémantiques	29
2.4	Recherche d'images	32
2.4.1	La technique syntaxique	33
2.4.1.1	Le langage de représentation	33
2.4.1.2	Le pouvoir discriminant	34
2.4.2	La technique sémantique	34
2.5	Système de recommandation	35
2.5.1	Utilisation des systèmes de recommandation	35
2.5.2	Les approches de systèmes de recommandation	35
2.5.2.1	L'approche personnalisée et non personnalisée	36
2.5.2.2	L'approche démographique	36
2.5.2.3	L'approche collaborative	36
2.5.2.4	L'approche basée sur le contenu	37
2.5.2.5	L'approche hybride	37
2.6	Conclusion	38
3	Synthèse : Aperçu général d'<i>Emergsem</i>	41
3.1	Définition et concepts	42
3.2	Problématique générale	43
3.3	Description des couches d' <i>Emergsem</i>	44
3.3.1	Représentation sémantique d'images	44
3.3.2	Annotation sémantique collaborative et émergence des sémantiques	46
3.3.3	Recherche d'images basée sur le clustering et la recommandation	49
3.3.3.1	Extraction des sémantiques représentatives	50
3.3.3.2	Extraction des sémantiques voisines	51
3.3.3.3	Clustering	51
3.3.3.4	Recommandation	52
3.3.4	Expérimentation	54

4	Annotation sémantique collaborative d'images	55
4.1	Introduction	56
4.2	Définitions et concepts de base	56
4.2.1	Annotation sémantique	56
4.2.2	La collaboration	57
4.3	Représentation des images dans <i>Emergsem</i>	59
4.3.1	Descripteurs d'images	59
4.3.2	Ontologie générique	60
4.3.2.1	Méta concept abstrait <i>Image</i>	60
4.3.2.2	Descripteur sémantique	60
4.3.3	Ontologie d'application	63
4.3.4	Description de l'ontologie d'application	64
4.3.4.1	Les concepts	64
4.3.4.2	Les relations hiérarchiques	69
4.3.4.3	Les relations spatiales	70
4.3.5	Instanciation de l'ontologie d'application	71
4.3.5.1	Instanciation des concepts concrets	72
4.3.5.2	Instanciation des relations spatiales	73
4.4	Annotation collaborative	74
4.4.1	La communauté d'annotateurs	74
4.4.2	Dictionnaire lexical	76
4.4.3	Génération des sémantiques	77
4.4.4	Émergence des sémantiques	77
4.4.4.1	La fonction d'évaluation f_{emerg}	79
4.4.4.2	La sémantique émergente	80
4.5	Conclusion	81
5	Recherche sémantique d'images : Une approche basée sur le clustering et la recommandation	85
5.1	Introduction	86
5.2	Aperçu général et caractéristiques	86

5.3	Recherche d'images basée sur le clustering	88
5.3.1	Notation et définition	89
5.3.2	Exemple d'illustration	89
5.3.3	Mesures de similarité	90
5.3.4	Extraction des sémantiques représentatives	92
5.3.5	Extraction des sémantiques voisines	96
5.3.5.1	Rappel sur le clustering	96
5.3.5.2	Voisinage de sémantiques	98
5.3.6	Formation de groupes de sémantiques : Granularité	99
5.3.6.1	Similarité sémantique par structure hiérarchique	99
5.3.6.2	Rappel du parcours de graphes	100
5.3.6.3	Matrice de confusion	103
5.3.6.4	Détermination de conflit entre les groupes	106
5.3.6.5	Fonction de correspondance	107
5.3.7	Cluster d'images	108
5.3.8	Résumé	109
5.4	Recommandation de sémantiques	110
5.4.1	Introduction	110
5.4.2	Recommandation des sémantiques	110
5.4.3	Acquisition des profils	111
5.4.3.1	Formation de la communauté d'utilisateurs	111
5.4.4	Classification des profils : Algèbre et treillis de Galois	114
5.4.4.1	Algèbre de Galois	115
5.4.4.2	Treillis de Galois	117
5.4.5	Génération de la liste des recommandations	118
5.4.5.1	Le coefficient de recommandation	119
5.4.6	Résumé	120
5.5	Conclusion	120
6	Expérimentation	123
6.1	Introduction	124

6.2	Architecture du prototype	124
6.2.1	Création de l'ontologie d'application sous Protégé	124
6.2.2	Implémentation du système <i>Emergsem</i>	126
6.2.2.1	Les profils d'annotateurs et d'utilisateurs	127
6.2.2.2	Annotation des images	127
6.2.2.3	Interface d'instances des sémantiques proposées	128
6.2.2.4	Interface d'évaluation des instances	129
6.2.2.5	Interface de recherche	129
6.2.3	Stockage des informations	130
6.3	Évaluation des résultats	131
6.3.1	Comparaison d' <i>Emergsem</i> avec deux approches classiques	132
6.3.1.1	Annotation manuelle	132
6.3.1.2	Annotation automatique	134
6.3.1.3	Recherche d'images	137
6.3.2	Comparaison d' <i>Emergsem</i> avec CLUE et SIMPLlcity	139
6.4	Conclusion	140
7	Conclusion et perspectives	143
7.1	Conclusion	144
7.2	Perspectives	147
8	Annexes	149
8.1	Les Ontologies	151
8.1.1	Ontologie d'application pour annoter les images	151
8.1.2	Ontologie des profiles	157
8.2	Quelques outils de construction d'ontologie	161
8.3	Extrait du code d'enregistrement des annotations	163
8.4	Extrait du dictionnaire lexical	169
8.5	Publications	173
	Bibliographie	175

TABLE DES FIGURES

1.1	Les caractéristiques globales et locales	3
1.2	Limites de description d'images (1)	4
1.3	Limites de description d'images (2)	5
2.1	Un exemple d'espace couleur	11
2.2	Quelques exemples de textures	12
2.3	Quelques exemples de formes	13
2.4	Architecture du Web sémantique. Source : [45]	16
2.5	Différents types d'ontologies	19
2.6	Processus d'annotation manuelle	22
2.7	Processus d'annotation automatique	23
2.8	Architecture de l'annotation semi-automatique	27
2.9	Ontologie d'objets	32
2.10	Techniques de recherche d'images	33
3.1	Les couches d' <i>Emergsem</i>	43
3.2	Schéma de l'ontologie générique	44
3.3	Schéma de l'ontologie d'application	45
3.4	Sémantique : une ontologie d'application instanciée	46
3.5	Ontologie d'application proposée aux annotateurs	47
3.6	Image proposée	48
3.7	Émergence de la sémantique	49
3.8	Les couches de l'approche de recherche d'images	50
3.9	Extraction des sémantiques représentatives	50
3.10	Extraction des sémantiques voisines	51
3.11	La recommandation	52

3.12	Architecture d' <i>Emergsem</i>	53
4.1	Le travail collaboratif (Source :[165])	57
4.2	Le travail coopératif (Source :[165])	58
4.3	Représentation des images	59
4.4	L'ontologie générique	61
4.5	Concepts et relations de l'ontologie générique	62
4.6	Ontologie d'application proposée	63
4.7	Ontologie d'application et sa traduction en OWL	64
4.8	Exemple d'une relation spatiale	70
4.9	Modèle des relations spatiales cognitives	71
4.10	Description d'une image	72
4.11	Interface de proposition des instances	72
4.12	Mapping des instances avec les concepts d'ontologie	73
4.13	Interface de proposition des relations spatiales	74
4.14	Communauté d'annotateurs	74
4.15	Annotation sémantique collaborative	75
4.16	Génération des sémantiques	78
4.17	Calcul des poids	79
4.18	Proposition des relations spatiales	81
4.19	Évaluation des instances et des relations spatiales	82
4.20	Émergence des sémantiques	83
5.1	Recherche sémantique d'images dans <i>Emergsem</i>	87
5.2	Sémantique d'image	88
5.3	Étapes de la recherche sémantique d'images	92
5.4	Correspondance d'instances	97
5.5	Voisinage d'une sémantique	98
5.6	Granularité locale	100
5.7	Exemple de trois graphes et les poids des relations des concepts	103
5.8	Formation des groupes de sémantiques	103

5.9	Les deux ensembles et leurs groupes respectifs	104
5.10	La matrice de confusion	105
5.11	Correspondance des groupes les plus similaires	108
5.12	Clusters de sémantiques	109
5.13	Approche de recommandation	111
5.14	Formation de la communauté des utilisateurs	113
5.15	Acquisition des profils d'utilisateurs	114
5.16	Apport de l'algèbre de Galois dans <i>Emergsem</i>	116
5.17	Treillis de Galois	118
6.1	Architecture du prototype	124
6.2	Processus de description de l'ontologie sous Protégé	125
6.3	Modèle de l'ontologie	126
6.4	Relation entre les concepts	127
6.5	Formulaire d'enregistrement des annotateurs et des utilisateurs	128
6.6	Annotation collaborative des images	128
6.7	Un aperçu des instances des sémantiques proposées	129
6.8	Un extrait d'évaluation des instances	129
6.9	Interface de recherche	130
6.10	Ontologie de stockage des profils d'annotateurs et d'utilisateurs	131
6.11	Plateforme de sondage	132
6.12	Tableau du sondage	132
6.13	Interface d'annotation manuelle	133
6.14	Extrait de description d'images	133
6.15	Interface d'annotation automatique	134
6.16	Affichage des résultats de la recherche	135
6.17	Graphe des durées d'annotation	136
6.18	Graphe des durées de recherche	136
6.19	Graphe des taux de satisfaction	137
6.20	Graphe de comparaison des durées de recherche d'images	139
6.21	Comparaison des résultats entre <i>Emergsem</i> avec Clue et SIMPLcity	141

6.22 Graphe de comparaison des approches 142

LISTE DES TABLES

3.1	Liste des instances proposées	48
4.1	Descripteur du concept <i>Acteur</i>	65
4.2	Descripteur du concept <i>Objet</i>	65
4.3	Descripteur du concept <i>Construction</i>	65
4.4	Descripteur du concept <i>Etre</i>	66
4.5	Descripteur du concept <i>Humain</i>	66
4.6	Descripteur du concept <i>Animal</i>	66
4.7	Descripteur du concept <i>Oiseau</i>	66
4.8	Descripteur du concept <i>Insecte</i>	67
4.9	Descripteur du concept <i>Environnement</i>	67
4.10	Descripteur du concept <i>Temps</i>	67
4.11	Descripteur du concept <i>Espace</i>	68
4.12	Descripteur du concept <i>Vegetation</i>	68
4.13	Descripteur du concept <i>Place</i>	68
4.14	Descripteur du concept <i>Evenement</i>	68
4.15	Descripteur du concept <i>Organise</i>	69
4.16	Descripteur du concept <i>Inattendu</i>	69
4.17	Descripteur de la relation <i>Contient</i>	69
4.18	Descripteur de la relation <i>Décrit</i>	69
4.19	Descripteur de la relation <i>Est-un</i>	70
4.20	Exemple de description d'une image	71
5.1	Sémantiques du corpus	89
5.2	Table de vecteurs	90
5.3	Représentation du vecteur des mots-clés de la requête	91

5.4	Représentation du vecteur d'une sémantique (S_1)	91
5.5	Similarité par appariement	93
5.6	Quelques mesures de similarité extraites de [170]	94
5.7	Présentation des vecteurs de caractéristiques et calcul des similarités	95
5.8	Sémantiques <i>résultat</i> selon les seuil α	96
5.9	Exemple de similarité cosinus	99
5.10	Calcul de similarité entre les groupes de sémantiques	106
5.11	Liste des conflits entre les groupes	107
5.12	Calcul de la fonction de correspondance f	108
5.13	Table de contexte	116
5.14	Formation ce concept formel	117
5.15	$(a) \neq (b)$: Absence de concept formel	117
5.16	Occurrences des profils dynamiques	119
5.17	Tableau de calcul des coefficients de recommandation	120
6.1	Moyenne des durées de recherche d'images	134
6.2	Quelques exemples de clusters d'images	138
6.3	Durées de recherche d'images	138

INTRODUCTION

Dans ce chapitre, nous introduisons la thèse à travers trois points principaux. Le chapitre présente dans un premier temps le contexte général et la problématique de la thèse. Dans un deuxième temps, nous énumérons les points saillants, c'est-à-dire les objectifs et les contributions autour desquels nos travaux sont menés et dans un troisième temps, nous donnerons l'organisation du manuscrit.

Sommaire

1.1	Contexte général et problématique	2
1.2	Objectifs et contributions	6
1.3	Organisation du manuscrit	7

1.1/ CONTEXTE GÉNÉRAL ET PROBLÉMATIQUE

De nos jours, le développement de l'internet et des réseaux sociaux génèrent un volume sans cesse croissant d'images acquises à travers des dispositifs de capture comme par exemple les appareils photo, les caméras numériques et les webcams. Ces dispositifs augmentent considérablement le besoin de gérer les images, raison pour laquelle les systèmes de recherche d'images par le contenu (CBIR) ont développé des techniques qui facilitent la recherche d'images de grandes collections.

Traditionnellement, les images sont recherchées grâce aux mesures de similarité calculées sur les caractéristiques physiques [1], [2], [3] connues sous le nom de caractéristiques de bas niveau telles que la couleur, la texture, la forme, etc. Le principe est le suivant. Les images de la base de données sont au préalable décrites avec des caractéristiques de bas niveau extraites automatiquement de leur contenu. La requête image et les images de la base de données sont considérées comme un ensemble de vecteurs de caractéristiques de bas niveau. La pertinence entre la requête image et les images des bases de données est déterminée en fonction de leur similarité calculée sur les caractéristiques extraites.

Les mesures de similarité peuvent être locales ou globales. Les mesures de similarité calculées à base des caractéristiques globales décrivent l'image entière tandis que les mesures de similarité calculées à base des caractéristiques locales décrivent les objets ou les régions d'images qui sont généralement obtenues après la segmentation des images. La figure 1.1 présente les deux caractéristiques (globales et locales). Comme nous pouvons le constater, les deux images de cette figure présentent la manière dont les caractéristiques peuvent être extraites. Dans (a), les caractéristiques sont extraites de la totalité de l'image alors que dans (b), l'image est segmentée en plusieurs régions et de chaque région, des caractéristiques locales peuvent être extraites. Les mesures de similarité à base de ces caractéristiques (locales et globales) sont déterminées via plusieurs méthodes parmi lesquelles on peut citer les histogrammes de textures et les histogrammes de couleurs.

De nos jours, la sémantique devient prépondérante dans les systèmes de recherche d'images basé sur le contenu. De plus en plus d'applications tentent de manipuler les concepts capables d'interpréter les images du point de vue humain en proposant des méthodes de description de haut niveau pour faciliter la recherche des images. Or, les mesures de similarité calculées sur les caractéristiques n'impliquent qu'une forme de sémantique. Elles rendent compte de certains aspects de la sémantique de l'image, c'est-à-dire de la sémantique induite. Mais la sémantique d'une image est rarement évidente. La sémantique induite par la similarité ne correspond pas avec le concept de haut niveau. Par exemple la dominance d'une couleur peut laisser *supposer* la présence d'une verdure dans une image mais cela ne représente pas la description sémantique de l'image.

Le problème de fossé sémantique se pose. Le fossé sémantique est l'écart entre les limites relatives des descriptions réalisées par les caractéristiques de bas niveau et les concepts de haut niveau. C'est le manque de coïncidence entre l'information que l'on peut extraire à partir des caractéristiques visuelles et l'interprétation qu'ont ces mêmes données pour un utilisateur



FIGURE 1.1 – Les caractéristiques globales et locales

dans une situation donnée. Réduire cet écart sémantique est naturel pour un être humain, il est loin d'être évident à la machine. Une solution pour résoudre ce problème est le retour de pertinence. Il s'agit de mettre l'utilisateur dans la boucle de recherche en lui permettant de signaler au système les images pertinentes parmi les images trouvées par les mesures de caractéristiques physiques. Malgré l'intervention humaine dans les systèmes de recherche par le contenu, la capture de la sémantique des images n'est pas garantie et par conséquent, le problème du fossé sémantique persiste.

Pour capturer la sémantique des images, des approches d'annotation d'images sont proposées. L'annotation est définie comme le processus qui consiste à décrire et à caractériser les images en leur associant des labels. Elle constitue une manière possible d'associer aux images des informations, représentées par les caractéristiques de bas niveau ou de haut niveau. Plusieurs approches d'annotation sont proposées parmi lesquelles nous pouvons citer l'annotation manuelle et l'annotation automatique.

- L'annotation manuelle est une approche qui permet aux humains de décrire les images avec du texte ou des mots-clés [4], [5], [6]. Cette approche est confrontée à plusieurs problèmes liés à la description des images. Une description textuelle des images est une opération longue, coûteuse et difficile à élaborer car l'information est manuellement attachée par l'utilisateur. Elle ne décrit pas fidèlement le contenu de l'image parce qu'elle se fait à partir des termes, de la légende ou du texte qui entoure l'image. Une description textuelle nécessite plusieurs intermédiaires et surtout de gros moyens. Il faut en effet créer une nomenclature permettant de décrire l'ensemble des images et sélectionner pour chaque image un ensemble de mots-clés. Une description textuelle suppose à la fois une connaissance suffisante de la part de l'annotateur pour représenter correctement le contenu d'une image en un ensemble de mots-clés et de la part de l'utilisateur pour traduire une question, qui peut être plus ou moins précise, en un ensemble de mots-clés. Un mot-clé annotant une image peut avoir plusieurs significations. Par exemple, *avocat* peut désigner à la fois un arbre ou un métier. De plus, le contenu d'une image peut être décrit par un ensemble de mots-clés plus ou moins précis parce que le même mot-clé peut désigner plusieurs objets : *rivière* ou *étendue d'eau* sont différentes manières de décrire un même objet dans une image.
- Une autre manière de décrire les images est l'annotation automatique [7], [8], [9]. Elle est basée sur des algorithmes de classification. L'objectif est de regrouper les images sur la



FIGURE 1.2 – Limites de description d'images (1)

base de caractéristiques physiques dans des classes qui représentent une forme de sémantique. Plusieurs approches d'annotation d'images peuvent être distinguées parmi lesquelles nous pouvons citer : l'annotation mono-classe, l'annotation multi-classes et l'annotation contextuelle. L'annotation mono-classe est une classification binaire des images basée sur des méthodes linéaires comme par exemple le k plus proches voisins, le SVM, etc. L'annotation multi-classes est une classification basée sur plusieurs classes qui utilisent (ou non) des échantillons d'apprentissage. Elle est basée sur des méthodes probabilistes comme par exemple les classificateurs bayésiens. L'annotation contextuelle utilise des contextes textuels ou des méta-données couplés avec des caractéristiques de contenu visuel pour définir la sémantique des images.

Ces approches ne parviennent pas à mieux capturer la sémantique de haut niveau associée à une image. Les annotations ne sont souvent ni suffisantes, ni assez précises pour décrire le contenu des images. Pour illustrer le fait que les annotations existantes ne parviennent pas à représenter la sémantique des images et que l'utilisation des mots-clés donne bien souvent des résultats qui ne reflètent pas les besoins exprimés dans les requêtes, considérons l'exemple d'un internaute qui recherche Jaguar. Le résultat peut être une voiture ou un animal comme le montre la figure 1.2. Prenons également un autre exemple où l'internaute veut rechercher des images dans lesquelles il veut voir apparaître devant une maison, un chien assis sous un arbre. La figure 1.3 présente un résultat possible retourné par un système de recherche d'images. Ce résultat montre que le système extrait les images ayant au moins un mot-clé de la requête parce qu'il n'arrive pas à interpréter la requête de l'utilisateur qui veut des images *de chiens assis sous un arbre devant une maison*. Le système ne se préoccupe pas de la sémantique des images parce que cette sémantique est inexistante comme le montre la figure .

Comme nous pouvons le constater avec les deux exemples précités, la sémantique des images proposée ne représente pas fidèlement les différents objets présents dans une image et les relations qui existent entre ces objets. L'utilisateur est certes contraint à l'accès à des collections d'images similaires qui partagent quelques attributs ; cependant, il est regrettable de représenter les images et de ne pas utiliser les relations sémantiques entre concepts pour présenter les relations entre les objets présents dans l'image.

La problématique générale de cette thèse est l'annotation sémantique des images du point de vue humain. Qui dit annoter une image du point de vue humain, dit qu'il faut décrire non seulement les objets présents dans une image, mais également les relations qui existent entre

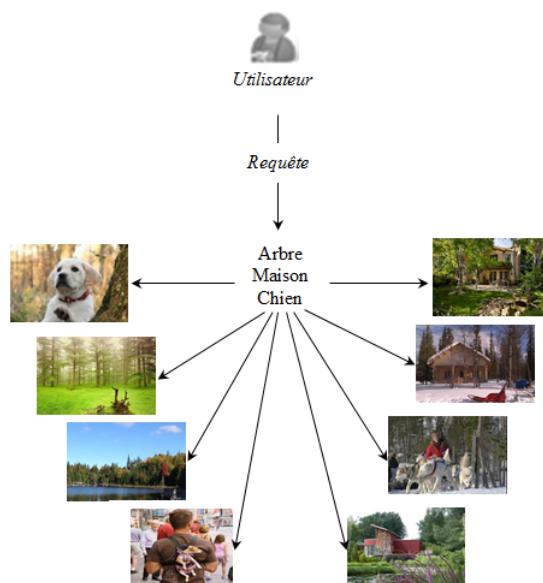


FIGURE 1.3 – Limites de description d'images (2)

ces objets. Il s'agit en effet de définir des modèles et des outils de description fidèle du contenu des images. Pour décrire la sémantique des images, le cadre méthodologique des systèmes de recherche d'images par le contenu doit évoluer profondément pour tirer parti des nouvelles possibilités offertes par les technologies du Web Sémantique.

Récemment, plusieurs travaux se sont intéressés à l'utilisation des avancées scientifiques dans le domaine de l'intelligence artificielle pour capturer la sémantique des images. Les outils sémantiques les plus connus sont : les graphes aussi appelés hiérarchies de concepts (ou ontologies légères) et les ontologies lourdes [10], [11], [12],[13]. Ces outils permettent d'identifier et de représenter les concepts et leurs relations et offrent une compréhension commune, structurée et partagée d'un domaine ou d'une tâche, qui peut être utilisée pour la communication entre les personnes et les machines. D'une part, les premiers outils sémantiques introduits dans les CBIR sont les graphes RDF (*Resource Description Framework*) qui sont des hiérarchies de concepts. Ils permettent de représenter les objets présents dans une image et les relations *hiérarchiques* qui existent entre ces objets. Il est caractérisé par est un ensemble de triplets *Sujet, Prédicat, Objet* où le *Sujet* représente la ressource à décrire, le *Prédicat* représente un type de propriété applicable à cette ressource et *Objet* représente une donnée ou une autre ressource : c'est la valeur de la propriété. Malgré leur performance, les graphes RDF sont limités à la description des objets et de leurs relations hiérarchiques, c'est-à-dire des relations de type *Est – un* aussi appelées relations de *spécification / généralisation*. Les graphes ne permettent pas de représenter les relations spatiales entre les objets (scènes d'images) comme par exemple *un homme sur un cheval* ce qui ne règle pas le problème du fossé sémantique. D'autre part, pour représenter la sémantique des images, les ontologies lourdes sont proposées dans la littérature. Elles permettent de décrire de façon formelle les objets et les relations qui

existent entre les objets (hiérarchiques et spatiales) du point de vue humain et peuvent être interprétées par l'homme et par la machine. L'apport significatif des ontologies lourdes résout le problème de la sémantique des images.

Dans la littérature, une autre approche d'annotation fournit une alternative pour une description fidèle de la sémantique d'images : l'annotation collaborative. Elle s'entend en fait par la description collective des images où tâche (description d'images) et but (déduction de la description la plus représentative) sont communs. C'est une approche qui permet aux annotateurs d'une communauté de décrire d'une part les mêmes images et de fusionner les descriptions individuelles pour obtenir la description qui représente mieux l'image [14], [15] et [16] et d'autre part de faire coopérer une communauté d'annotateurs pour décrire une collection d'images. Dans ce dernier cas, la collection d'images est partitionnée selon le nombre d'annotateurs et chaque annotateur annoté une partition donnée. L'annotation collaborative peut engager ou non une communication régulière entre les annotateurs et une connaissance précise de la progression de l'annotation.

L'idée principale de cette thèse est l'émergence de la sémantique des images à travers une approche d'annotation collaborative. L'émergence de la sémantique d'images est une réponse au problème de la subjectivité auquel les annotations sont confrontées lorsqu'elles sont réalisées par un annotateur parce que bien que les ontologies lourdes soient utiles pour réaliser des annotations sémantiques, structurées et significatives des images, les sémantiques peuvent être subjectives pour la simple raison que les annotateurs n'ont pas souvent la même perception de la même image. La pertinence de ces sémantiques peut être remise en cause parce qu'une image reçoit plusieurs interprétations et l'interprétation fournie par un annotateur peut être erronée, imprécise ou insuffisante. L'émergence de la sémantique d'images à travers l'annotation collaborative va permettre d'avoir un consensus par l'alignement ou la fusion des différentes sémantiques proposées.

1.2/ OBJECTIFS ET CONTRIBUTIONS

Dans cette thèse, nous proposons une approche d'annotation collaborative et de recherche d'images basée sur la sémantique émergente pour définir un cadre sémantique d'annotation et de recherche d'images. L'objectif de l'approche est de capturer la sémantique émergente des images à partir d'un ensemble de descriptions proposées par une communauté d'annotateurs et d'exploiter les relations entre ces descriptions pour fournir des réponses pertinentes aux requêtes d'utilisateurs. L'approche *Emergsem* prend en compte les propriétés suivantes :

- Un modèle de description qui va au-delà de l'annotation traditionnelle basée sur les mots-clés. Dans cette partie, deux niveaux d'ontologies sont proposés : une ontologie générique et une ontologie d'application. L'ontologie générique est composée d'un méta concept abstrait de haut niveau qui contient tous les concepts pouvant décrire les images quel que soit le domaine auquel elles appartiennent et d'un descripteur sémantique qui contient un ou plusieurs concepts (abstrait et concrets) et des relations et une ontologie d'application dérivée de l'ontologie générique pour décrire la sémantique des images d'un domaine bien précis. C'est une ontologie personnalisée dans laquelle on peut créer autant de concepts abstraits

de base qui peuvent à leur tour contenir un ou plusieurs concepts (abstraits ou concrets) et des relations reliant les concepts.

- Une approche d'annotation sémantique collaborative qui est hybride et se situe entre l'annotation manuelle et l'annotation automatique. Une telle description nécessite une analyse du contenu des images, correspondant à leur interprétation du point de vue humain. Pour décrire la sémantique des images, un dictionnaire lexical sera proposé aux annotateurs. Dans cette approche, nous n'allons pas nous arrêter seulement à la description sémantique d'images, mais nous allons mettre en place un processus d'émergence des sémantiques d'images vers la sémantique la plus représentative : la sémantique émergente. Pour cela, il faut un groupe ou une communauté d'annotateurs qui proposent la description sémantique des images.
- Une recherche d'images basée sur la sémantique et qui ne se limite pas aux résultats obtenus grâce à la comparaison des requêtes avec la base de données, mais qui exploite les relations sémantiques entre le résultat de la requête et les autres images de la base de données. Le but, c'est de proposer aux utilisateurs dans un premier temps des clusters d'images pertinentes formés à partir du voisinage du résultat des requêtes. Dans un deuxième temps, nous proposons la recommandation. Son objectif est de proposer des sémantiques aux utilisateurs en se basant sur leurs profils (statiques et dynamiques). Elle est basée sur l'acquisition des profils d'utilisateurs, la classification des profils qui permet de regrouper des profils en des groupes selon leur similarité (basée sur l'algèbre de Galois) et la génération de la liste des recommandations.

1.3/ ORGANISATION DU MANUSCRIT

Le reste de ce document est organisé comme suit :

- Le chapitre 2 présente l'état de l'art. Il décrit la représentation d'images, les techniques d'annotation et de recherche d'images et les systèmes de recommandation.
- Le chapitre 3 présente la synthèse de notre travail. Il est constitué de deux couches principales : la problématique générale et les contributions.
- Le chapitre 4 présente la technique d'annotation sémantique collaborative. Il est composé de la représentation sémantique et de l'annotation sémantique collaborative des images.
- Le chapitre 5 présente notre approche de recherche d'images. Il contient deux grandes sections : la recherche d'images basée sur le clustering et la recommandation des sémantiques.
- Le chapitre 6 présente l'expérimentation. La première partie décrit l'implémentation d'*Emergsem* et la deuxième partie présente les résultats de validation.
- Le chapitre 7 conclut notre travail et dresse quelques perspectives de recherche en liaison avec le sujet traité.
- Le chapitre 8 présente les annexes. Elle décrit les ontologies implémentées, quelques outils de construction d'ontologie, des extraits des codes et les publications.

ÉTAT DE L'ART

Les systèmes de recherche d'images ont accordé une attention particulière à deux techniques principales : les techniques d'annotation et les techniques de recherche. Dans ce chapitre, nous présentons dans un premier temps la représentation des images par les caractéristiques de bas niveau et de haut niveau. Nous étudions dans la deuxième partie de ce chapitre les techniques d'annotation existantes (manuelle, semi-automatique, automatique et collaborative) sans oublier les méthodes qui sont utilisées. La troisième partie du chapitre est consacrée aux techniques de recherche et la quatrième partie à une approche de gestion d'informations qui consiste à effectuer des recommandations d'informations basées sur les préférences des utilisateurs.

Sommaire

2.1	Introduction	10
2.2	Représentation des images	10
2.2.1	Représentation physique	10
2.2.2	Segmentation des images	13
2.2.3	Réduction du fossé sémantique	14
2.2.4	Représentation sémantique	15
2.3	Annotation d'images	20
2.3.1	Les techniques d'annotation	21
2.3.2	Autres approches d'annotation	27
2.4	Recherche d'images	32
2.4.1	La technique syntaxique	33
2.4.2	La technique sémantique	34
2.5	Système de recommandation	35
2.5.1	Utilisation des systèmes de recommandation	35
2.5.2	Les approches de systèmes de recommandation	35
2.6	Conclusion	38

2.1/ INTRODUCTION

La recherche d'images par le contenu (en anglais : *Content Based Image Retrieval-CBIR*) est une technique permettant de rechercher des images à partir de leurs caractéristiques visuelles. Avant de pouvoir rechercher les images, il faut d'abord les extraire leurs caractéristiques. L'extraction des caractéristiques contenues dans une image s'appelle la description structurelle. Celle-ci peut prendre la forme d'une image ou de toute structure de données permettant une description des entités contenues dans l'image. Essentiellement, l'analyse de l'image fait appel à la segmentation où l'on va tenter d'associer à chaque région de l'image un label en s'appuyant sur l'information portée et la distribution spatiale. Les images sont classiquement décrites comme rendant compte de leurs caractéristiques de bas niveau telles que la texture, couleur, forme, etc. Un cas typique d'utilisation des caractéristiques de bas niveau est la recherche d'images visuellement similaires à un exemple donné en requête (image). Cette technique s'oppose à la recherche d'images par des mots-clés, qui fut historiquement proposée par les moteurs de recherche où les images sont retrouvées en utilisant le texte qui les accompagne plutôt que le contenu de l'image elle-même.¹

Dans ce chapitre, nous présentons les techniques de représentation d'images, les techniques d'annotation, les techniques de recherche et les techniques de recommandation.

2.2/ REPRÉSENTATION DES IMAGES

Dans cette section, nous présentons la représentation des images. La représentation d'une image consiste à décrire une image via des descripteurs (ou caractéristiques). Deux types de représentation sont détaillés dans cette section : la représentation par les caractéristiques physiques (les caractéristiques de bas niveau) et la représentation par les caractéristiques sémantiques (les caractéristiques de haut niveau).

2.2.1/ REPRÉSENTATION PHYSIQUE

Une approche traditionnelle pour rechercher les images est de les représenter avec des caractéristiques de bas niveau (voir les articles [17], [18], [19], [20]) : il s'agit de la représentation physique des images. Les caractéristiques physiques sont la base des systèmes de recherche d'images par le contenu. Les caractéristiques extraites de l'image entière sont appelées des caractéristiques globales et les caractéristiques obtenues à partir d'une région de l'image sont appelées des caractéristiques locales. Ces dernières caractéristiques (locales) se sont avérées plus proche de la perception et sont basées sur la segmentation des images en des régions [21] et [22].

Les systèmes de recherche d'images basés sur les caractéristiques de bas niveau permettent aux utilisateurs de représenter leurs besoins par une ébauche de la forme, du contour de la couleur, de la texture des images qu'ils recherchent ou par la proposition d'une image similaire aux images qu'ils recherchent. L'image peut être fournie par l'utilisateur ou choisie dans une base de données image.

Nous avons quatre caractéristiques de bas niveau : la couleur, la texture, la forme et la

1. http://fr.wikipedia.org/wiki/Recherche_d'image_par_le_contenu

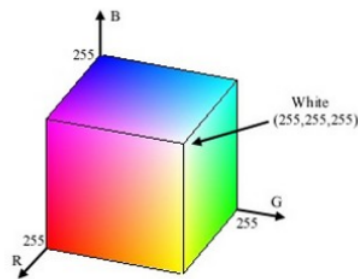


FIGURE 2.1 – Un exemple d'espace couleur

localisation spatiale comme le présentent Ying Liu et al. dans [23].

2.2.1.1/ LA COULEUR

Les couleurs sont l'une des caractéristiques les plus utilisées dans les systèmes de recherche d'images par le contenu. Elles sont riches en information et très utilisées pour la représentation des images. Elles forment une partie significative de la vision humaine. La perception de la couleur dans une image par l'être humain est un processus complexe et subjectif [24]. En effet, cette donnée varie considérablement avec l'orientation des surfaces, la caméra et l'illumination (positions et longueur d'onde des sources lumineuses), par exemple. Il est possible de représenter la couleur dans différents espaces à savoir RGB, HVS, LAB, etc. Le plus répandu est sans aucun doute l'espace R, V, B (RGB en anglais) qui code la couleur d'un pixel sur un vecteur en trois dimensions : rouge, vert, bleu. La figure 2.1 est un exemple d'espace couleur utilisé pour décrire les images.

2.2.1.2/ LA TEXTURE

La texture est la seconde caractéristique visuelle largement utilisée dans la recherche d'image par le contenu. Elle permet de combler un vide que la couleur est incapable de faire, notamment lorsque les distributions de couleurs sont très proches. La texture est généralement définie comme la répétition d'un motif créant une image visuellement homogène. Elle peut être vue comme un ensemble de pixels spatialement agencés, créant ainsi une région homogène. Dans les travaux consacrés aux caractéristiques de bas niveau, la texture joue un rôle très important parce qu'elle fournit des informations importantes dans la classification des images puisqu'elle décrit le contenu de nombreuses images comme par exemple la peau des fruits, des arbres, des briques et le tissu, etc. Par conséquent, la texture est un élément important dans la définition de sémantique de haut niveau. De nombreuses approches et modèles sont proposés dans [25] pour la caractérisation de la texture. Parmi les méthodes les plus connues, nous pouvons citer :

- les méthodes statistiques basées sur les évaluations quantitatives de la distribution du niveau de gris. Elles sont focalisées sur l'étude des relations entre un pixel et ses voisins et sont utilisées pour caractériser des structures sans une régularité apparente.
- les méthodes fréquentielles concernent les propriétés fréquentielles et s'appuient sur la transformée de Fourier [26], [27], le filtre de Gabor [28] et [29], les ondelettes [30] et [31]
- les méthodes géométriques permettent de décrire une forme à l'aide de propriétés statistiques

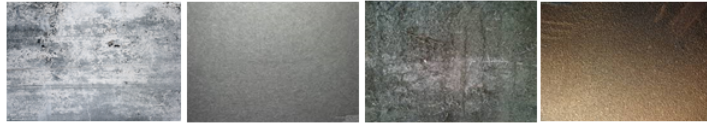


FIGURE 2.2 – Quelques exemples de textures

[32]. Elles représentent les propriétés spatiales de la distribution des pixels dans l'image.

La figure 2.2 présente quelques exemples de textures utilisées pour décrire les images.

2.2.1.3/ LA FORME

La forme permet d'extraire des caractéristiques particulièrement robustes et discriminantes. Nous les appellerons caractéristiques d'objets car elles présupposent une segmentation de l'image et manipulent donc des informations liées aux objets réels. Au même titre que la texture, la forme est une caractéristique complémentaire de la couleur. Les systèmes de recherche d'images par le contenu se sont focalisés sur l'extraction d'attributs géométriques ces dernières années. Deux méthodes ont été proposées pour représenter les formes : les descripteurs basés sur les régions et les descripteurs basés sur les contours (frontières). Les descripteurs basés sur les régions sont utilisés pour caractériser l'intégralité de la forme d'une région et les descripteurs basés sur les frontières se focalisent sur les contours des formes [24]. Les formes sont beaucoup utiles dans certains domaines tels que dans les objets fabriqués par l'homme. Les formes sont difficilement applicables sur des images en couleur utilisées dans la plupart des documents. Néanmoins, elles ont prouvé leur efficacité dans certains systèmes de recherche d'images comme le montrent les travaux menés dans [33] et dans [34]. La figure 2.3 présente quelques exemples de formes utilisées pour décrire les images.

2.2.1.4/ LA LOCALISATION SPATIALE

Outre les caractéristiques précitées, il est également possible de considérer l'organisation spatiale des différentes primitives comme une description en tant que telle. Il est évident qu'une telle description constitue un niveau intermédiaire entre les caractéristiques de bas niveau brutes et l'interprétation des images, et que ce niveau peut se révéler très expressif. Ainsi, lorsqu'un utilisateur recherche une image représentant un objet complexe, par exemple une voiture, la reconnaissance de chacune de ses parties (roues, carrosserie...) est une étape indispensable. La localisation spatiale n'est rien d'autre que la position des régions dans l'image. La position des régions est définie par rapport à la région du centre de l'image. Elle est généralement représentée par des expressions comme *Supérieur*, *Inférieur*, *Gauche*, *Bas* etc. Plusieurs approches utilisent la localisation spatiale. Nous pouvons citer les travaux de Manjunath et al. dans [35], les travaux de Song et al. dans [36] et les travaux de Mojsilovic et al. dans [37].

En résumé, nous pouvons dire que les caractéristiques de bas niveau de l'image peuvent être soit extraites de la totalité de l'image ou des régions de l'image. La plupart des systèmes actuels sont focalisés sur les régions d'images parce que les recherches basées sur les caractéristiques globales de l'image sont relativement simples mais les recherches basées sur les régions sont



FIGURE 2.3 – Quelques exemples de formes

efficaces car elles sont plus proches de la perception humaine.²

2.2.2/ SEGMENTATION DES IMAGES

Afin de localiser automatiquement les objets dans les images, les chercheurs ont d'abord proposé la segmentation des images en plusieurs régions. De manière générale, *segmenter* une image consiste à délimiter dans l'image des plages *régulières* ou *cohérentes*, c'est-à-dire des régions dans lesquelles les informations de l'image suivent un modèle d'organisation : zones relativement homogènes en intensité, en texture ou en couleur, zones relativement planes, lisses (de variations lentes), etc. De nombreux systèmes d'interprétation d'images s'appuient sur la segmentation des images en des régions, qui permet d'extraire des constituants élémentaires, ou *primitives*, qui serviront de base pour l'identification ou la reconstruction des images [38].

De nombreuses techniques de segmentation des images sont proposées dans la littérature. Les algorithmes de segmentation se basent généralement sur les caractéristiques de bas niveau. Ces techniques fonctionnent bien pour les images contenant uniquement des zones de couleur homogènes, tels que les méthodes de classification directes dans l'espace de couleur comme le présentent les travaux menés dans [39], [40] et [41].

Nous avons trois types de techniques de segmentation : la segmentation basée sur les régions, la segmentation basée sur les contours et la segmentation basée sur la classification ou le seuillage³.

2.2.2.1/ SEGMENTATION BASÉE SUR LES RÉGIONS

Les algorithmes de type décomposition/fusion exploitent les caractéristiques propres de chaque région (surface, intensité lumineuse, colorimétrie, texture, etc.). On cherche des couples de régions candidates à une fusion et on les note en fonction de l'impact que cette fusion aurait sur l'apparence générale de l'image. On fusionne alors les couples de régions les mieux notés, et on réitère jusqu'à ce que les caractéristiques de l'image remplissent une condition prédéfinie : nombre de régions, luminosité, contraste ou texture, ou alors jusqu'à ce que les meilleures notes attribuées aux couples de régions atteignent un certain seuil.

Les algorithmes de croissance de régions partent d'un premier ensemble de régions qui peuvent être calculées automatiquement (par exemple, les minima de l'image), ou fournies par un utilisateur de manière interactive. Les régions grandissent ensuite par incorporation des pixels les plus similaires suivant un critère donné, comme par exemple la différence entre le niveau de gris du pixel considéré et le niveau de gris moyen de la région.

2. http://glotin.univ-tln.fr/MCBIR/Segmentation_images_principes.pdf

3. http://fr.wikipedia.org/wiki/Segmentation_d'image

Les algorithmes fondés sur une modélisation statistique conjointe des régions et des niveaux de gris reposent sur la minimisation d'une fonction de vraisemblance. Cette fonction prend simultanément en compte la vraisemblance de l'appartenance du pixel à une région considérant son niveau de gris, et les régions auxquelles appartiennent les pixels voisins. Cette fonction effectue un compromis entre la fidélité à l'image initiale et la régularité des régions segmentées.

2.2.2.2/ SEGMENTATION BASÉE SUR LE CONTOUR

Cette approche cherche à exploiter le fait qu'il existe une transition détectable entre deux régions connexes. Les méthodes les plus anciennes utilisent des opérateurs de traitement d'images, tels que le filtre de Canny, pour mettre en évidence les pixels qui semblent appartenir à un contour. On peut aussi faire intervenir des modèles déformables à l'aide de courbes paramétriques (courbe de Bézier, spline...) ou de polygones (par exemple algorithme à bulle).

Pour commencer le processus, on recherche les points remarquables de l'image, tels que les points à l'intersection de trois segments au moins. De tels points sont appelés des graines (seeds en anglais). L'intérêt principal des méthodes de segmentation selon l'approche frontières est de minimiser le nombre d'opérations nécessaires en cas d'itération du processus sur des séries d'images peu différentes les unes des autres (cas des images vidéo notamment). En effet, une fois que les contours des régions aient été trouvés dans la première image, l'application du modèle déformable à l'image suivante est plus efficace que de tout recalculer, si la différence entre les images est peu importante.

2.2.2.3/ SEGMENTATION BASÉE SUR LA CLASSIFICATION OU LE SEUILLAGE

Le seuillage a pour objectif de segmenter une image en plusieurs classes via des histogrammes. On suppose donc que l'information associée à l'image permet à elle seule la segmentation, i.e. qu'une classe est caractérisée par sa distribution de niveaux de gris. A chaque pic de l'histogramme est associée une classe. Il existe de très nombreuses méthodes de seuillage. La plupart de ces méthodes s'appliquent correctement si l'histogramme contient réellement des pics séparés. De plus, ces méthodes ont été développées pour traiter le cas particulier de la segmentation en deux classes (i.e. passage à une image binaire) et leur généralité face aux cas multi-classes n'est que très rarement garantie.

Si de nombreuses méthodes de segmentation ont vu le jour, il ne faut pas oublier qu'elles restent extrêmement limitées lorsqu'elles s'appuient sur des descriptions de type couleur ou texture. En effet, ces dernières sont trop sensibles aux conditions de prise de vue et à d'autres artefacts. En outre, de telles méthodes restent limitées pour extraire des objets du monde 3D, souvent composés de différentes parties, avec pour chacune des descriptions de bas niveau très différentes.

2.2.3/ RÉDUCTION DU FOSSÉ SÉMANTIQUE

Les CBIR permettent d'annoter et de rechercher les images en se basant sur les caractéristiques citées ci-dessus. Bien qu'ils soient performants, ils souffrent aussi d'un problème. Les interprétations que les utilisateurs ont d'une image n'est pas la même que la compréhension de la machine. Il y a un écart significatif entre la perception de haut niveau et l'interprétation

des images par les utilisateurs et les descriptions de bas niveau. Cet écart est appelé fossé sémantique. Comblé ce fossé sémantique est naturel pour un être humain, mais loin d'être évidente à la machine ([42]). Smeulders et al. dans [43] donne la définition suivante : le fossé sémantique est le manque de coïncidence entre l'information que l'on peut extraire à partir des données visuelles et l'interprétation qu'ont ces mêmes données pour un utilisateur dans une situation donnée. La façon dont les objets et les concepts exprimés par le langage naturel sont stockés en mémoire est moins claire que cela pouvait l'être pour un humain. Par conséquent, l'imitation est inaccessible. Le cerveau humain est capable de reconnaître presque instantanément un grand nombre d'objets, et de les écrire en langue avec peu d'effort pour un être humain. À ce jour, de telles possibilités incroyables pourraient ne pas être atteintes par des machines. Pour réduire ce fossé sémantique, il faut que les machines soient dotées des outils leur permettant de raisonner comme les hommes.

Pour réduire le fossé sémantique entre la perception de haut niveau et l'interprétation des images par les utilisateurs, il y a une tendance actuelle qui consiste à utiliser des concepts et des outils sémantiques pour décrire et traiter les images parce que la représentation sémantique des images nécessite une analyse du contenu de l'image. Elle correspond à l'interprétation de l'image d'un point de vue humain. Plusieurs outils sémantiques existants peuvent être utilisés pour décrire les caractéristiques globales ou locales d'images.

La notion de sémantique a été largement utilisée par les chercheurs pour désigner les processus de description ou de représentation des objets dans un langage naturel qui peut être à la fois interprété par l'humain et la machine. Pour représenter cette sémantique, le Web Sémantique propose plusieurs outils dans la littérature. Parmi ces outils, nous pouvons citer les ontologies légères ou les ontologies lourdes. Une ontologie légère, aussi appelée hiérarchie de concepts, est composée de concepts liés par des relations de spécialisation ou de généralisation (les relations de type *Est – un* sont utilisées pour représenter les relations entre les objets) tandis qu'une ontologie lourde est composée de concepts liés par des relations plus complexes.

2.2.4/ REPRÉSENTATION SÉMANTIQUE

Le Web Sémantique est un mouvement collaboratif mené par le World Wide Web Consortium (W3C) qui favorise des méthodes communes pour échanger des données. Il est défini comme le Web de données qui peuvent être traitées directement et indirectement par des machines pour aider les utilisateurs à créer de nouvelles connaissances. Il vise à aider l'émergence de nouvelles connaissances en s'appuyant sur les connaissances déjà présentes sur Internet [44]. Son but principal est de permettre aux utilisateurs de trouver, de partager et de combiner l'information plus facilement⁴. Les humains sont capables d'utiliser le web pour effectuer des tâches telles que trouver le mot *Paris* pour réserver un livre à la bibliothèque, trouver un plan et réserver son billet de transport. Cependant, les machines ne peuvent pas accomplir toutes ces tâches sans aucune intervention humaine.

Comme prévu initialement, le Web Sémantique est un système qui permet aux machines de

4. http://fr.wikipedia.org/wiki/Web_sémantique

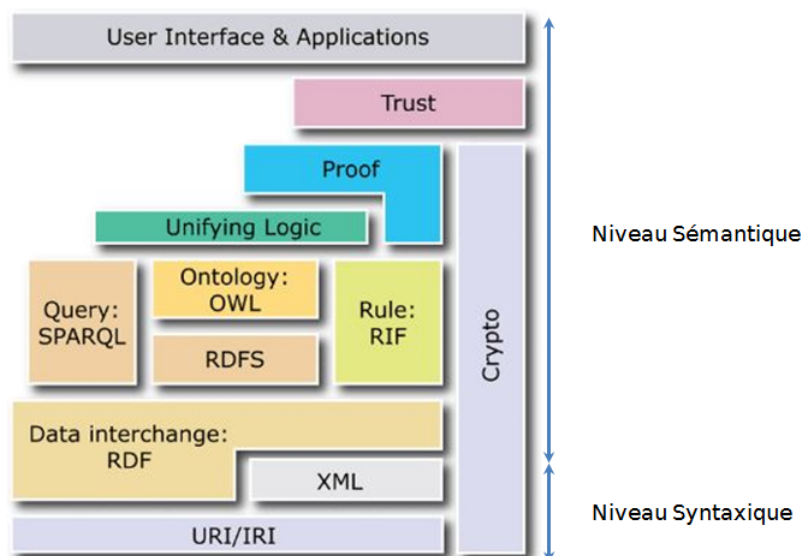


FIGURE 2.4 – Architecture du Web sémantique. Source : [45]

comprendre et de répondre aux demandes complexes de l'homme en fonction du sens de leurs demandes. C'est une approche qui permet à l'information d'être lisible par les humains et par les machines. Cela permet d'effectuer des travaux fastidieux et répétitifs dans le domaine de la recherche d'information par des machines [46]. Une telle compréhension exige que les sources d'information pertinentes soient sémantiquement structurées au préalable. Pour y parvenir, des outils sont mis à la disposition des utilisateurs comme le présente l'architecture de la figure 2.4.

2.2.4.1/ NIVEAU SYNTAXIQUE

- URI- Uniform Resource Identifier (Identifiant Uniforme de Ressource), est une courte chaîne de caractères identifiant une ressource sur un réseau, et dont la syntaxe respecte une norme d'Internet mise en place pour le World Wide Web. Les URI sont la technologie de base du World Wide Web car tous les hyperliens du web sont exprimés sous forme d'URI. Un URL (Uniform Resource Locator) est un URI qui, outre le fait qu'il identifie une ressource sur un réseau, fournit les moyens d'agir sur une ressource ou d'obtenir une représentation de la ressource en décrivant son mode d'accès primaire (emplacement) sur le réseau. Par exemple, l'URL <http://www.wikipedia.org/> est un URI qui identifie une ressource (page d'accueil Wikipédia) et implique qu'une représentation de cette ressource (une page HTML en caractères encodés) peut être obtenue via le protocole HTTP depuis un réseau hôte appelé www.wikipedia.org⁵. Un URN (Uniform Resource Name) est un URI qui identifie une ressource par son nom dans un espace de noms. Un URN peut être employé pour parler d'une ressource sans que cela préjuge de son emplacement ou de la manière de la référencer. Par exemple l'URN `urn :isbn :0-395-36341-1` est un URI qui, étant un numéro de

5. http://www.wikipedia.org/wiki/Uniform_Ressource_Identifier

l'International Standard Book Number (ISBN), permet de faire référence à un livre, mais ne suggère ni où, ni comment en obtenir une copie réelle.

- IRI-Internationalized Resource Identifier (Identificateur de ressource internationalisé) est un type d'adresse informatique prenant en compte les divers alphabets utilisés dans les différentes langues du monde. Les adresses IRI sont une norme généralisant et internationalisant les adresses URI (Uniform Resource Identifier) qui sont elles-mêmes un sur-ensemble des plus connus URL utilisés pour les adresses Web ⁶.
- XML - L'Extensible Markup Language (langage de balisage extensible en français) est un langage informatique de balisage générique qui dérive du SGML. Cette syntaxe est dite « extensible » car elle permet de définir différents espaces de noms, c'est-à-dire des langages avec chacun leur vocabulaire et leur grammaire, comme XHTML, XSLT, RSS, SVG... Elle est reconnaissable par son usage des chevrons (< >) encadrant les balises. L'objectif initial est de faciliter l'échange automatisé de contenus complexes entre systèmes d'informations hétérogènes. XML fournit une syntaxe élémentaire, pour la structure du contenu dans les documents, mais il ne décrit pas la sémantique du document ⁷.

2.2.4.2/ NIVEAU SÉMANTIQUE

- RDF- Resource Description Framework est un graphe du Web Sémantique. Un document structuré en RDF est un ensemble de triplets. Un triplet RDF est une association : *Sujet*, *Prédicat*, *Objet*. Le *Sujet* représente la ressource à décrire, le *Prédicat* représente un type de propriété applicable à cette ressource et *Objet* représente une donnée ou une autre ressource : c'est la valeur de la propriété. Un graphe RDF contient des concepts abstraits (classes) et des concepts concrets (sous-classes). En les parcourant de la racine aux feuilles, on découvre nettement la notion de spécialisation et dans l'autre sens, la notion de généralisation. Les objectifs du graphe RDF sont multiples [47]. Parmi ceux-ci, nous pouvons citer la manipulation et classification des méta-données, afin de fournir des informations sur les ressources du web et les systèmes qui les utilisent. Il permet de faire avec l'information traitable par machine ce que le web a fait pour l'hypertexte, en permettant aux informations d'être manipulées en dehors de l'environnement particulier dans lequel elles ont été créées, éventuellement à l'échelle d'Internet. Il favorise l'optimisation de la coopération entre les applications, en permettant de combiner les données de plusieurs applications pour générer de nouvelles informations. Un modèle basé sur RDF peut être représenté à travers plusieurs syntaxes d'échanges, par exemple, RDF/XML, N3, Turtle, et RDFa. RDF est une norme fondamentale du Web Sémantique. C'est un modèle de graphe destiné à décrire les ressources et leurs méta-données, de façon à permettre leur traitement automatique. Dans un graphe RDF, tous les termes sont organisés en une structure hiérarchique. Ce type de structure est souvent désigné comme un arbre, avec un tronc, des branches principales et de petites branches hors des branches principales. RDF est affichée comme un arbre, un arbre à l'envers, avec plusieurs petites branches pour les termes spécifiques.

6. http://fr.wikipedia.org/wiki/Internationalized_Resource_Identifier

7. http://fr.wikipedia.org/wiki/Extensible_Markup_Language

- RDF Schema ou RDFS est un langage extensible de représentation des connaissances. Il appartient à la famille des langages du Web Sémantique publiés par le W3C. RDFS fournit des éléments de bases pour la définition des ontologies. Avec ces ressources structurées avec RDFS dans une base de données spécialement conçue pour le stockage et la récupération de données RDF, il est possible d'utiliser le langage de requête SPARQL pour les atteindre à travers le web. RDFS étend le RDF et son vocabulaire pour pouvoir structurer les propriétés et les classes au sein d'une ressource décrite en RDF⁸.
- SPARQL- SPARQL Protocol And RDF Query Language est un langage de requête et un protocole qui permet de rechercher, d'ajouter, de modifier ou de supprimer des données RDF disponibles à travers Internet. SPARQL est l'équivalent de SQL car comme on accède aux données d'une base de données via SQL, SPARQL permet d'accéder aux données du web. Cela signifie qu'en théorie, on pourrait accéder à toutes les données du web avec ce standard. L'ambition du W3C est d'offrir une interopérabilité non seulement au niveau des services, comme avec les services web, mais aussi au niveau des données structurées ou non qui sont disponibles à travers l'Internet⁹.
- RIF- Rule Interchange Format est une couche de règle logique. RIF permet d'échanger des règles entre plusieurs langages de règles. Rappelons qu'une règle est un moyen simple pour représenter une connaissance, et est une simplification radicale de logique du premier ordre pour lequel il est relativement facile à mettre en œuvre des moteurs d'inférence qui peuvent traiter les conditions et déduire de bonnes conclusions.
- Ontologie : une ontologie est une approche formelle, la spécification explicite d'une conceptualisation partagée [48]. *Formelle* reflète le fait que l'ontologie est lisible par machine et permet de raisonner sur son contenu à partir de l'humain et la machine. *Explicite* signifie que le type de concepts utilisés et les contraintes sur leur utilisation sont explicitement définis. *Partagée* se réfère à la connaissance commune, énoncée dans l'ontologie. *Conceptualisation* se réfère au modèle obtenu en faisant abstraction de certains phénomènes qui existent dans le monde réel en identifiant les concepts pertinents de ces phénomènes. Une ontologie est composée de concepts et des relations entre les concepts. Les concepts dans les ontologies sont divisés en deux parties : les concepts abstraits et les concepts concrets [49]. Un concept abstrait décrit la scène d'une image, tandis que le concept concret correspond à des objets spécifiques à une image. Un concept abstrait est composé d'un ou de plusieurs concepts concrets. Exemple : le concept *Nature* qui contient *Ciel* et *Forêt*. *Ciel* et *Forêt* sont des concepts concrets de *Nature* qui est à son tour un concept abstrait. Différents types d'ontologies peuvent être utilisés dans les systèmes de récupération des images. Nous avons en général quatre types d'ontologies. Les ontologies peuvent être classifiées en fonction de deux dimensions : leur niveau de détail et leur niveau de dépendance par rapport à une tâche particulière, un point de vue. Plus précisément Guarino [50] propose une classification des ontologies selon leurs niveaux de généralité.
 - Les ontologies de haut niveau : Les ontologies de haut niveau (top-level ontologies) dé-

8. http://www.wikipedia.org/wiki/RDF_Schema

9. <http://fr.wikipedia.org/wiki/SPARQL>

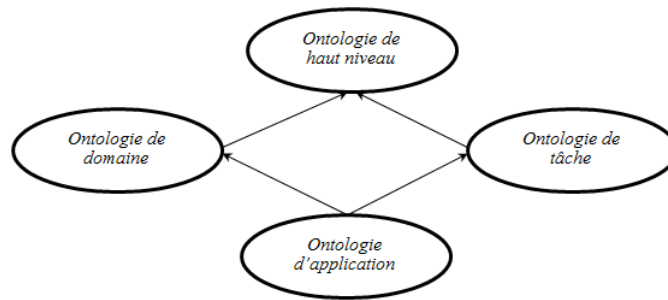


FIGURE 2.5 – Différents types d'ontologies

crivent les concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc., qui sont indépendants d'un problème ou d'un domaine d'application particulier.

- Les ontologies de domaine et de tâche : Les ontologies de domaine (domaine ontologies) et les ontologies de tâche (task ontologies) décrivent, respectivement, le vocabulaire lié à un domaine générique (comme la médecine, ou les automobiles) ou une tâche ou une activité générique (comme le diagnostic ou la vente), en spécialisant les concepts présentés dans les ontologies de hauts niveaux. Elles donnent une représentation formelle des concepts du domaine étudié ainsi que des différentes relations qui lient ces derniers ; elle ne contient pas les concepts pédagogiques, narratifs et structurels.
- Les ontologies d'application : Les ontologies d'application (application ontologies) décrivent des concepts dépendant à la fois d'un domaine et d'une tâche particulière, qui sont souvent des spécialisations des deux ontologies relatives. Ces concepts correspondent souvent aux rôles joués par des entités de domaine tout en exécutant une certaine activité, comme l'unité remplaçable ou le composant disponible. La figure 2.5 représente les différents types d'ontologies selon leur degré de dépendance vis-à-vis d'une tâche particulière ou d'un point de vue. Les flèches représentent des relations de spécialisation.
- Les langages : les langages utilisés sont généralement des langages déclaratifs, et presque toujours des généralisations de langages fondés sur la logique du premier ordre ([51] et [52]) ou la logique de description. Ils permettent l'encodage des connaissances sur des domaines spécifiques et contiennent souvent les règles de raisonnement qui supportent le traitement de ces connaissances [53]. RDFS est considéré comme un langage primitif de l'ontologie dans le Web Sémantique supportant la définition de classes (concepts), de ressources et des propriétés, et déterminant des relations de sous-classes et de sous-propriétés [53]. Plus complet et complexe que RDFS, OWL est une famille de langages de représentation des connaissances pour définir les ontologies. Basé sur les logiques de description, OWL (Ontology Web Language) enrichit les possibilités de description et de raisonnement offerts par RDF Schema. Par exemple, OWL permet d'exprimer des contraintes d'équivalence de classes, de cardinalités, de restriction de valeurs, etc. RDFS peut en effet être vu comme un moyen de définir de

simples taxonomies, tandis qu'OWL permet d'exprimer des connaissances plus expressives à travers les logiques de description et les ontologies [54]. Le langage OWL peut être employé pour permettre la représentation explicite des concepts et des relations entre les concepts. La conception d'OWL a été influencée par une variété des paradigmes de représentation de connaissance, par des langages d'ontologie existants et par des langages du Web. Une des influences les plus importantes sur la conception d'OWL est la spécification formelle des langages de LD (Logiques de Description). Il y a trois sous-langages d'OWL : OWL Lite, OWL DL et OWL Full. OWL Lite est le sous-langage d'OWL plus simple syntaxiquement. Il est adapté pour les cas où seule une hiérarchie de classes simples et des contraintes simples sont nécessaires. OWL DL est beaucoup plus expressive qu'OWL-Lite. Afin de rendre décidable OWL, certaines restrictions ont été introduites sur l'utilisation des concepts OWL. OWL-DL est basé sur logiques de description (d'où le suffixe DL). Les logiques de description sont un fragment décidable de la logique de premier ordre et sont donc ouverts au raisonnement automatisé [55]. OWL Full est le sous-langage d'OWL le plus expressive. Contrairement à OWL Lite et OWL DL, il est très expressif. OWL Lite et OWL DL sont des logiques de description fondamentalement très expressives avec une syntaxe de RDF. OWL Full fournit une intégration plus complète de RDF. Les ontologies basées sur les logiques de description peuvent être exploitées par des outils de raisonnement puissants afin de faciliter l'arrangement des ressources du Web [53].

2.3/ ANNOTATION D'IMAGES

L'annotation d'images consiste en la description des images. La description des images est basée soit sur les caractéristiques physiques ou sémantiques. Les techniques d'annotation d'images sont réparties en trois catégories. On distingue l'annotation manuelle, l'annotation automatique et l'annotation semi-automatique. Les outils utilisés pour l'annotation des images sont groupés en trois catégories : les caractéristiques physiques, les mots-clés et les outils du Web Sémantique (les ontologies légères et les ontologies lourdes). Les caractéristiques physiques permettent de décrire les images grâce à des caractéristiques appelées de bas niveau telles que la couleur, la texture, la forme, la localisation spatiales, etc. Les mots-clés comme nous venons de le mentionner un peu plus haut permettent de représenter les différentes régions d'une image. Le plus souvent, une annotation basée sur les mots-clés est une concaténation des différents termes représentant le contenu et /ou le contexte des images. Contrairement aux caractéristiques de bas niveau, les mots-clés sont classés dans la catégorie des outils de description sémantiques de l'image parce qu'ils permettent de la sémantique d'une région de l'image. Les ontologies permettent de représenter non seulement les objets présents dans une image, mais aussi les relations qui existent entre ces objets.

Cette section présente dans un premier temps les différentes techniques d'annotation : les techniques manuelles, les techniques automatiques et les techniques semi-automatiques. Dans un deuxième temps, elle présente deux nouvelles approches d'annotation : approches collaboratives et approches sémantiques.

2.3.1/ LES TECHNIQUES D'ANNOTATION

La description des images se fait grâce à des techniques d'annotation. Cette section sera consacrée à ces techniques d'annotation. Trois techniques d'annotation sont proposées dans la littérature : l'annotation manuelle, l'annotation automatique et l'annotation semi-automatique.

2.3.1.1/ ANNOTATION MANUELLE

L'annotation manuelle d'images consiste à faire annoter des bases d'images par un ou plusieurs annotateurs. Lors d'une annotation manuelle, l'annotateur assigne une description à chaque image en fonction de sa perception. Plusieurs systèmes d'annotation manuelle des images ont été présentés dans la littérature [56], [57] et [58].

L'annotation des images basée sur les mots-clés est un processus qui permet de décrire les images avec des termes dans le but de faciliter leur récupération. Agosti et al. dans [59] et Bradley dans [60] ont fourni une étude exhaustive sur les contours et la complexité des annotations.

Un mot-clé est un mot associé à une image qui, une fois indexé, permet d'identifier l'image dans une base de données. Il permet de faire une description textuelle du contenu des images. Comme tous les autres descripteurs, les mots-clés permettent d'une part de décrire les images, et d'autre part d'accéder aux images par des requêtes textuelles. Comparativement aux descripteurs de bas niveau (couleur, texture, forme, etc.), les mots-clés sont considérés comme des descripteurs efficaces. Grâce aux mots-clés, l'annotation, la récupération et même la comparaison des images deviennent possible. Plusieurs approches se sont focalisées sur la description et la récupération des images via les mots-clés. Les premiers systèmes apparus dans les années 90 sont basés sur le texte [4], [5], [6]. Ces derniers adoptent une approche qui consiste à décrire le contenu visuel sous forme textuelle. Les mots-clés utilisés servent d'index pour accéder aux données visuelles associées. L'avantage de cette approche est qu'elle permet de consulter les bases de données en utilisant les langages d'interrogation standard, par exemple SQL. Plusieurs approches sont basées sur la description du contenu visuel des images ([61], [62], [12], [63]), [64], [65], [66], [67], [68], [69] et [70].

Une autre méthode consiste à combiner le contenu visuel et le contenu textuel. Contrairement au contenu visuel décrit par les mots-clés, le contenu textuel représente l'ensemble des attributs spécifiques aux images, à savoir : les titres, les légendes, les commentaires, etc. qui entourent l'image. La combinaison du contenu visuel et textuel a été proposé par plusieurs chercheurs parmi lesquels nous pouvons citer Jin et al. dans [71], Gao et al. dans [72] et Ameesh et al. dans [73]. Jin et al. ont réalisé l'annotation des images en utilisant un modèle de langage cohérent parce que les mots-clés seuls seraient indépendants. Pour eux, il y a nécessité de prendre en compte d'autres informations comme les textes qui entourent les images, afin de produire une bonne description de l'image. Nous avons également les annotations à plusieurs niveaux proposées dans [72] qui visent non seulement à identifier les objets spécifiques dans une image, mais intègrent également les concepts pour regrouper des éléments similaires. Par exemple Ameesh et al. proposent que l'image d'un bureau peut être annotée non seulement

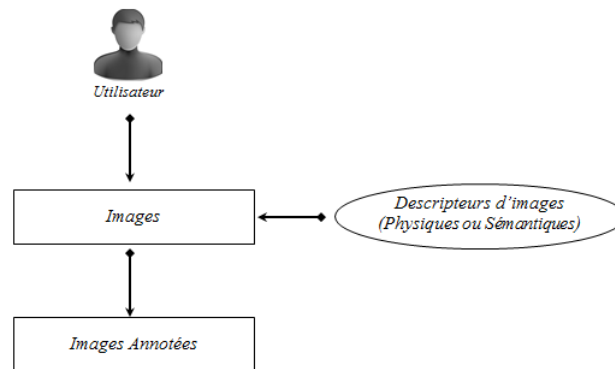


FIGURE 2.6 – Processus d'annotation manuelle

avec les éléments de l'image, comme un écran d'ordinateur et tapis de souris, mais aussi le bureau et l'intérieur du bureau.

Généralement, le processus de l'annotation manuelle comporte l'annotateur (ou les annotateurs) qui décrit l'image, l'image à décrire et les descripteurs fournis pour mener à bien la description comme le montre la figure 2.6.

2.3.1.2/ ANNOTATION AUTOMATIQUE

L'annotation automatique est un processus qui consiste à faire annoter les images par un système d'annotation. Le processus d'annotation automatique a généralement lieu lorsqu'une ou plusieurs nouvelles images sont ajoutées dans la base de données [74] et [75]. Le système utilise automatiquement chaque nouvelle image comme une requête et effectue une recherche d'images par le contenu. Pour un nombre n d'images similaires à une requête, les caractéristiques ayant servi à l'annotation de chacune d'elles (et classés par leur fréquence) sont analysées. Une liste de caractéristiques des images similaires à la nouvelle image est fournie et attribuée à la nouvelle image. La nouvelle image est ainsi annotée (bien que virtuellement et sans confirmation). Plusieurs travaux ont été déjà menés dans le domaine de l'annotation automatique. Parmi ces travaux, nous pouvons citer les méthodes proposées dans [74], [75], [76] et [77]. La figure 2.7 présente le processus d'annotation automatique.

Les annotations de nouvelles images seront faites à partir de cet échantillon. Les sections suivantes décrivent les méthodes d'annotation automatique d'images.

- La méthode basée sur la classification : elle permet de créer un très grand nombre de classes auxquelles les images sont associées [78], [7], [8], [9]. Elle permet de faire la description des caractéristiques des images sous forme de vecteurs ([79], [80] et [81]). Les images sont représentées par des vecteurs de leurs caractéristiques. Les graphes sont souvent utilisés par classer les images. Les vecteurs et les mots-clés des images sont représentés comme des nœuds d'un graphe $G = (V, E)$, où V est l'ensemble de tous les nœuds et E l'ensemble de tous arêtes comme le confirme les travaux menés dans [82] et dans [83]. La plus part des CBIR sont construits sur le modèle d'espace vectoriel. Le corpus d'images et de requête sont représenté comme des vecteurs de caractéristiques dans un espace vectoriel à n dimensions.

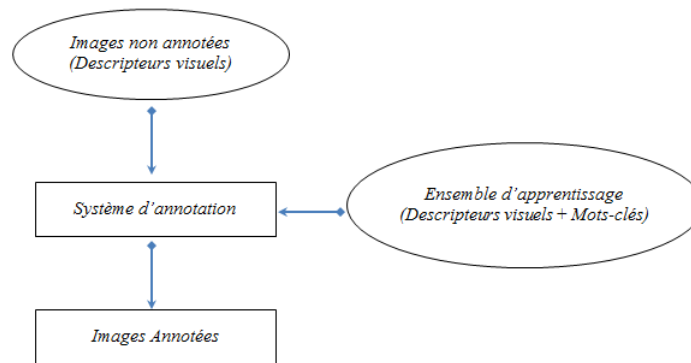


FIGURE 2.7 – Processus d'annotation automatique

[84].

Soit W l'ensemble des nœuds représentant des mots-clés uniques w utilisés pour toutes les images de la collection, F l'ensemble de nœuds représentant tous les vecteurs de caractéristiques extraites de toutes les images et I l'ensemble de tous les nœuds représentant les images dans la collection. Les sommets du graphe de l'image peuvent être définis comme $V = I \cup F \cup W$. La relation entre les images et leurs vecteurs caractéristiques peut être codée dans le graphe par une paire d'arêtes (i_n, f_j) et (f_j, i_n) reliant les nœuds d'image i_n et leurs vecteurs caractéristiques f_j . De la même manière, les relations entre les images et leurs mots-clés sont codées par une paire de d'arêtes (i_n, w_j) et (w_j, i_n) . Supposons maintenant que la fonction $\text{dist}(f_i, f_j)$ retourne une valeur réelle positive qui mesure la distance entre deux vecteurs de caractéristiques. Cette fonction peut être la distance euclidienne ou toute autre distance métrique sur des vecteurs de caractéristiques. Grâce à cette fonction, les k plus proches voisins de chaque vecteur caractéristique f_i sont sélectionnés, et une paire d'arêtes (f_i, f_k) , (f_k, f_i) pour tous les k plus proches voisins est utilisé pour désigner leur similitude dans le graphe. De même, une fonction $\text{dist}(w_i, w_j)$ ou $\text{sim}(w_i, w_j)$ renvoyant un nombre réel positif, retourne la distance entre deux mots-clés [85].

La méthode basée sur les graphes permet de représenter chaque image par un graphe. De façon concrète, considérons un ensemble α de n images $\alpha = I_1, I_2, \dots, I_n$. Chaque image est décrite par un ensemble de caractéristiques $\{f_1, f_2, \dots, f_m\}$ et une liste de mots-clés $W = w_1, w_2, \dots, w_k$. Une image peut alors être décrite par le vecteur $I_1 = \{f_1, f_2, \dots, f_m, W_i\}$. Si une nouvelle image I_x ($I_x = \{f_1, f_2, \dots, f_n\}$) non étiquetée est introduite dans ce système, le système trouvera un modèle capable de lui affecter des étiquettes. Autrement dit, l'image passe d'une représentation $I_x = \{f_1, f_2, \dots, f_n\}$ à une représentation $I_x = \{f_1, f_2, \dots, f_n, W_k\}$ [79], [80] et [85].

Typiquement, l'image est d'abord analysée par l'extraction de vecteurs de caractéristiques, et les mots pour l'apprentissage de l'annotation sont utilisés avec des techniques d'apprentissage automatique pour pouvoir être capable d'affecter automatiquement des annotations à de nouvelles images. Les premières méthodes dans le domaine apprenaient les corrélations

entre les descripteurs d'images et les annotations, puis des techniques utilisant des méthodes issues de la traduction automatique ont vu le jour, permettant de faire une traduction entre le vocabulaire textuel, et ce qu'on considérait comme un vocabulaire visuel, où chaque mot serait un groupe de régions similaires, obtenues par regroupement (clustering). La classification est basée sur l'apprentissage. L'apprentissage est un processus visant à améliorer les performances d'un système en se basant sur ses expériences passées. Trois grandes familles de techniques d'apprentissage sont répertoriées [86] :

- L'apprentissage supervisé est une technique d'apprentissage automatique qui produit de façon automatique une classification en se basant sur des exemples existants. L'apprentissage supervisé est possible grâce à certains classificateurs. C'est le cas, par exemple, des Machines à Vecteurs Supports SVM [87], des forêts aléatoires [88], de l'algorithme des k plus proches voisins (notés KPPV, ou KNN en anglais), et des classificateurs Bayésiens. Les SVM et les forêts aléatoires sont réputés pour être performants en présence d'un grand nombre de variables. Par contre, l'utilisation des SVM devient difficile à utiliser lorsque le nombre d'observations de la base d'apprentissage est important. Concernant les KNN, la procédure de classification est lourde car chaque image requête est comparée (sur la base de ses caractéristiques) à toutes les images stockées. Par contre cette méthode a l'avantage de ne pas nécessiter d'apprentissage : c'est l'échantillon qui constitue le modèle. Enfin, les classificateurs Bayésiens, quant à eux, sont sensibles à la dimensionnalité des données. Par contre, ils sont efficaces avec beaucoup de données d'apprentissage. Les classificateurs Bayésiens sont adaptés à la résolution de problèmes en présence de données manquantes, contrairement aux SVM [89]. D'autres techniques d'apprentissage telles que le réseau de neurones sont également utilisés pour l'apprentissage de concepts. Dans [90] les auteurs choisissent 11 catégories de concepts : brique, nuage, fourrure, herbe, glace, route, rocher, sable, la peau, l'arbre et de l'eau. Ensuite, une grande quantité de données d'apprentissage est introduite dans le classificateur de réseau neuronal pour établir le lien entre les caractéristiques de bas niveau de l'image et sa sémantique (catégorie d'étiquettes).
- L'apprentissage non supervisé (parfois dénommé « clustering ») est une méthode d'apprentissage automatique. Il s'agit pour d'une technique qui permet de diviser un groupe hétérogène de données, en sous-groupes de manière à ce que les données considérées comme les plus similaires soient associées au sein d'un groupe homogène et qu'au contraire les données considérées comme différentes se retrouvent dans d'autres groupes distincts.¹⁰ Pendant la phase de l'apprentissage non supervisé c'est le système qui doit déterminer ses sorties en fonction des similarités détectées entre les différentes entrées (règle d'auto organisation) [91]. Contrairement à l'apprentissage supervisé, l'apprentissage non supervisé permet de regrouper un ensemble de données d'image de façon à maximiser la similarité au sein des clusters et de minimiser la similitude entre les différents groupes. Chaque groupe résultant est associé à une étiquette de classe et les images dans une même grappe sont

10. http://fr.wikipedia.org/wiki/Apprentissage_non_supervisé

censés être semblables les uns aux autres.

- L'apprentissage par renforcement. C'est un apprentissage qui nécessite un superviseur qui dit à l'agent quelle action est correcte dans telle situation [92]. Dans l'apprentissage par renforcement, l'agent interagit avec l'environnement qui lui donne un retour quantitatif sur les valeurs de ses actions. L'objectif de l'apprentissage par renforcement est alors de générer à partir d'expériences (état courant, action, état suivant) une politique maximisant le rendement au cours d'une période donnée.
- La méthode probabiliste : elle permet de représenter les images par un vecteur caractéristique à n -dimensions. Ces vecteurs sont calculés grâce à des fonctions mathématiques ou des algorithmes. Un bon choix de caractéristiques pertinentes est indispensable pour atteindre un grand pouvoir de discrimination [79]. La méthode probabiliste est souvent dédiée à l'annotation d'images partiellement annotées, c'est-à-dire que des images ne possédant pas le nombre maximal de mots-clés. Lorsqu'une image est partiellement annotée, les mots-clés manquants sont considérés comme des valeurs manquantes. La méthode probabiliste est une méthode qui permet de calculer la distribution des mots-clés à une image [93], [94], [95]. Cette distribution représente une prédiction des mots-clés manquants d'une image. Pour chaque annotation manquante, le mot-clé du vocabulaire ayant la plus grande probabilité est retenu, si cette probabilité atteint un certain seuil [96]. Chaque image est annotée par les mots-clés ayant les plus grandes probabilités. C'est le cas de Jin et al. [71] qui proposent un modèle de langage pour annoter des images qui estime la probabilité d'un ensemble de mots-clés. L'ensemble de mots-clés ayant la plus grande probabilité est associé à l'image, si cette probabilité dépasse un certain seuil.

D'autres travaux d'annotation automatique des images du web ont été proposés par certains chercheurs. C'est l'exemple de Wen et al. dans [97] et de Feng et al. dans [98]. C'est une technique qui facilite la recherche d'images grâce aux informations complémentaires disponibles sur le web comme par exemple, l'URL du fichier image qui a souvent une structure hiérarchique claire, y compris des informations sur l'image comme la catégorie de l'image. Dans cette technique, les documents HTML sont également pris en compte parce qu'ils contiennent des informations utiles sur le titre de l'image, le texte descriptif qui entoure l'image, des hyperliens, etc. Plusieurs outils basés sur la description textuelle comme par exemple Google et AltaVista sont proposés pour rechercher des images sur le web. Bien que ces approches aident à trouver de nombreuses images pertinentes, la précision de récupération est faible, car elles ne peuvent pas confirmer si les images extraites contiennent vraiment les concepts de la requête. Les utilisateurs doivent parcourir toute la liste pour trouver les images désirées. C'est un processus de longue haleine parce que les résultats renvoyés sont des collections d'images.

Pour améliorer les performances de ces approches, les chercheurs ont tenté de fusionner l'information textuelle et le contenu visuel de l'image. C'est le but des travaux menés dans [98] qui combinent les descriptions textuelles (HTML) et les caractéristiques visuelles de l'image. Pour mener à bien l'annotation automatique, ils ont développé deux classificateurs indépendants. Le premier est basé sur le texte et le deuxième est basé sur les caractéristiques visuelles.

L'expérimentation a été faite sur un ensemble prédéfini de 15 concepts. Les résultats obtenus démontrent la performance substantielle du système. Toutefois, en raison de l'imprécision de l'extraction des informations textuelles, les performances de certains concepts ne sont pas satisfaisantes.

D'autres propositions ont été faites dans [99] et dans [100]. MSRA (Microsoft Research Asia) a pour objectif de regrouper les résultats de recherche des moteurs de recherche classique d'images Web, de sorte que les utilisateurs retrouvent rapidement les images. Tout d'abord, un algorithme de segmentation basé sur la vision intelligente est conçu pour segmenter une page web en blocs. Dans le bloc contenant l'image, les informations textuelles et les informations sur les liens d'une image peut être extraite de façon précise. Ensuite, un graphe de l'image est réalisé en utilisant des techniques d'analyse des liens entre les blocs. Ainsi, pour chaque image, on obtient trois types de représentations, la représentation visuelle basée sur les caractéristiques physiques, la représentation sémantique basée sur les caractéristiques textuelles et la représentation graphique. Les images de chaque catégorie sont ensuite réorganisées en fonction de leurs caractéristiques visuelles. La plupart de ces méthodes ont en commun le fait qu'elles nécessitent un échantillon d'apprentissage.

2.3.1.3/ ANNOTATION SEMI-AUTOMATIQUE

L'annotation semi-automatique résulte de la combinaison des deux annotations précédentes. Elle consiste à générer l'annotation des images à partir de l'annotation manuelle et de l'annotation automatique [101]. En général, les méthodes semi-automatiques consistent à faire intervenir l'utilisateur pour valider les décisions du système [102] et [11]. Cette collaboration peut se réaliser dans les deux sens : soit c'est l'humain qui vérifie et valide les annotations réalisées par la machine ou soit, c'est la machine qui achève le processus d'annotation entamé par l'humain. Les techniques d'annotation semi-automatique se basent généralement sur le retour de pertinence. Une variété d'interfaces d'utilisateur pour la recherche d'images et de retour de pertinence est utilisée pour la méthode d'annotation semi-automatique [75]. Généralement, une telle interface utilisateur comprend trois parties : l'interface de soumission de requêtes (soit une requête par des mots-clés, soit une requête d'images, ou une combinaison des deux), le navigateur d'images, et l'interface de retour de pertinence. Un scénario utilisateur typique est le suivant : lorsque de l'utilisateur soumet une requête, le système renvoie des résultats de recherche comme une liste de classement des images en fonction de leur similarité avec la requête soumise. Les images ayant une similarité supérieure ont des rangs plus élevés que celle ayant de faible similarité. Le navigateur d'image peut être une fenêtre déroulante des images, une fenêtre paginée. L'utilisateur peut parcourir les images dans le navigateur et utiliser l'interface de confirmation pour soumettre ses jugements de pertinence. Le système retourne de façon itérative les résultats de récupération raffinés basés sur le feedback de l'utilisateur et affiche les résultats dans le navigateur. Ce processus est illustré dans la figure 2.8.

Le retour de pertinence permet d'afficher les images plus pertinentes aux premiers rangs et offre à l'utilisateur plus de chance de les voir, de les confirmer, et donc de les annoter. Deux cas sont à prendre en compte à ce stade [75], [11]. Dans le premier cas, il n'y a pas d'images

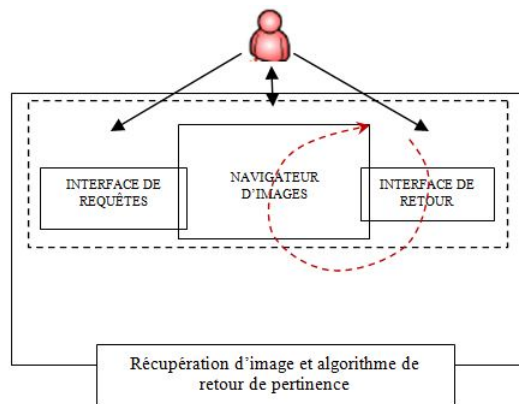


FIGURE 2.8 – Architecture de l'annotation semi-automatique

dans le système ayant été annotés avec les mots-clés de la requête soumise. Dans le second cas, certaines images sont déjà annotées avec les mots-clés correspondants à la requête. Dans le premier cas, le système retourne seulement une liste d'images au hasard car aucun mot-clé n'est matché et aucune image pertinente pour cette requête ne peut être trouvée. Dans le second cas, les images annotées avec les mots-clés de la requête sont récupérées et affichées à l'utilisateur. Il revient ensuite à l'utilisateur d'indiquer les images qui sont pertinentes. Pour chacune de ces images pertinentes, si l'image n'a pas encore été annotée avec tous les mots-clés de la requête, l'image est annotée avec les mots-clés avec un poids initial de 1. Si l'image a déjà été annotée avec un des mots-clés présents dans la requête, le poids de ce mot-clé pour cette image est augmenté de 1. Pour chacune des images non pertinentes, le poids de ce mot-clé est réduit d'un quart ($1/4$) de son poids d'origine. Si le poids devient très faible (par exemple inférieure à 1), le mot-clé est retiré de l'annotation de l'image. Le résultat est un ensemble de mots-clés et de leurs poids associés à chaque image et stockés dans une base de données.

2.3.2/ AUTRES APPROCHES D'ANNOTATION

Comme nous venons de l'observer, chacune de ces trois techniques présentées ci-dessus est réalisée de façon individuelle et avec des descripteurs tels que les caractéristiques de bas niveau et les mots-clés. Récemment, de nouvelles approches d'annotation ont vu le jour. Elle consiste à intégrer d'une part la sémantique et d'autre part la collaboration dans le système de d'annotation. Dans cette section, nous présentons l'approche d'annotation basée sur la collaboration et l'approche basée sur la sémantique.

2.3.2.1/ APPROCHES COLLABORATIVES

Certaines approches se sont penchées sur la collaboration. L'annotation collaborative consiste en la proposition de la description d'une image par plusieurs annotateurs comme l'expliquent les travaux menés dans [103], [15], [14], [104] et [105]. Elle peut être réalisée soit par l'utilisation des mots-clés. Généralement, plusieurs annotateurs procèdent à l'annotation de la même image et le système se charge d'en déduire la description la plus pertinente de l'image. Dans [14], un jeu est proposé aux internautes. L'exemple le plus connu est ESP Game.

La règle du jeu est la suivante : deux utilisateurs, connectés en même temps, se voient proposer par le système la même image. Chacun propose des mots-clés pour annoter cette image. Le système ne propose pas de vocabulaire. Une fois que les deux utilisateurs sont d'accord sur un mot-clé, ce mot-clé est validé, et les deux utilisateurs gagnent des points. Une autre image leur est alors proposée, et ainsi de suite. Quand un utilisateur atteint un certain nombre de points en une semaine, il se voit offrir un cadeau. Wang et al. quant à eux, ont proposé Image Webs dans [15]. C'est une approche qui privilégie la collaboration entre les annotateurs. Elle prend en compte plusieurs annotateurs et leur permet de partager non seulement les images annotées mais également des annotations (réalisées avec des mots-clés). Grâce à cette approche, les annotateurs décrivent les différentes régions des images par des mots-clés. Le système Image Webs relie l'ensemble des régions d'images contenant les mots-clés du même objet pour former un graphe. Ces graphes représentent les relations entre les images en fonction de leur contenu visuel partagé. L'approche [16] propose un système d'annotation collaboratif basé sur un serveur qui peut simplement être accessible via un navigateur Web. Cela permet de tenir des statistiques d'annotation et évite la distribution de grandes collections d'images aux utilisateurs. L'approche a pour objectif de simplifier et d'accélérer le processus d'annotation autant que possible, tout en maintenant en même temps les options de configuration et de personnalisation pour permettre différents styles d'annotation et préférences d'utilisateurs. Les utilisateurs pourront utiliser différents style d'annotation, comme par exemple, l'affichage et l'annotation de quelques images par page ; l'annotation d'images avec un seul terme avant de passer à l'annotation avec plusieurs termes simultanément, etc. Afin de maximiser la commodité et l'efficacité utilisateur, chaque utilisateur a alors la possibilité de personnaliser le nombre, la taille et la disposition des images affichées par page, et sélectionner un ou plusieurs concepts pour annoter à la fois. Dans [106], l'approche d'annotation collaborative d'image proposée permet aux utilisateurs de créer des balises qui fournissent un contexte hiérarchique permettant de définir les relations entre ces concepts. L'approche fournit également une technique visant à établir la crédibilité des utilisateurs qui annotent les images. C'est une technique qui permet de calculer la véracité et la fiabilité d'une déclaration particulière. Une approche similaire a été proposée par Kavasidis et al dans [104]. Ils proposent un environnement collaboratif d'annotation des séquences de vidéo par des mots-clés. L'environnement permet aux annotateurs de partager leurs propres annotations avec d'autres, accélérant ainsi donc le processus de génération des annotations. L'environnement dispose d'une interface qui permet au annotateurs d'annoter les séquences de vidéo et de partager les annotations avec d'autres annotateurs. L'objectif de cette approche est de soulager une grande partie de l'effort des annotateurs et d'accroître la qualité des annotations. Dans [105] par exemple, il est question de faire annoter les images par deux communautés de personnes : les experts et les amateurs. L'objectif est de réaliser une évaluation pour explorer l'influence entre les deux communautés qui ont différents niveaux d'appréciation. Pour cela, l'outil SeviAnno a été proposé. Il est basé sur une combinaison de normes de méta-données et des technologies Web 2.0 dans le domaine de la gestion du patrimoine culturel. Chaque région des images d'une vidéo peut être annotée, y compris les objets, les événements, les temps, les lieux, etc. La conception et la mise en

œuvre de l'outil, répond au transfert de connaissances entre experts du domaine et amateurs. Les experts et les amateurs collaborent sur les annotations des images pour gérer et partager des connaissances spécifiques du domaine. Les informations sur celui qui a annoté une image sont affichées à toute la communauté. Barthelmess et al. dans [107] présentent une nouvelle méthode d'annotation collaborative d'images. C'est une approche d'annotation d'images par un réseau de personnes. Il consiste à l'extraction des méta-données à partir de conversations entre des groupes de personnes discutant entre eux sur une image donnée. L'approche explore comment le langage peut aider dans l'annotation d'images. L'approche comprend trois étapes : la détection d'un ensemble d'étiquettes significatives qui peuvent être associées à chaque image par la variété et la richesse du langage employé par les utilisateurs lorsque ceux-ci se mettent à parler de leurs photos, la reconnaissance de la sémantique de ces termes par les interactions entre groupe et la propagation des étiquettes. C'est une approche qui explore une technique de fusion multimodale pour récupérer l'orthographe et la prononciation des termes redondants dans le discours et l'écriture.

2.3.2.2/ APPROCHES SÉMANTIQUES

Un enjeu actuel du Web Sémantique est de définir des techniques permettant de représenter les sémantiques des images et de pouvoir les utiliser pour recherches ces images comme le précisent [12],[13]. Dans cette section, nous présentons les approches sémantiques basées sur les ontologies légères et les ontologies lourdes.

- Les ontologies légères. Plusieurs travaux ont été consacrés à l'utilisation des graphes RDF. Dans [108] les auteurs ont proposé un classificateur hiérarchique sémantique basé sur WordNet pour intégrer les connaissances à priori sur les relations interclasses dans l'apprentissage des apparences. Dans le même esprit, ils ont proposé «ImageNET» [109], une hiérarchie de concepts à grande échelle des images construites sur l'épine dorsale de la structure WordNet. Naphade et al. ont proposé dans [110] une méthode qui vise à concevoir une taxonomie avec une couverture de près de 1 000 concepts pour la récupération des vidéos. Sivic et al. ont proposé dans [111], la descriptions textuelles des images grâce au graphe RDF. Dans cette méthode, un graphe de représentation connaissances visuelles est utilisé. Le graphe fournit un modèle graphique servant à la connaissance préalable pour représenter la diversité des motifs visuels. Certains chercheurs comme [110] ont préféré utiliser une hiérarchie multi-couche fondée sur des éléments visuels. D'autres par ailleurs ont proposé une méthode qui combine les caractéristiques visuelles et textuelles comme par exemple [77] et [112] pour construire des hiérarchies. Von-Wun Soo et al. [113] ont proposé une méthode d'annotation et d'indexation basée sur les graphes RDF. Dans cette approche, un analyseur (Case-Based Learning (CBL)) est utilisé. Il permet de convertir d'une part les mots-clés proposés sous forme d'un graphe RDF, et d'autre part convertir les requêtes écrites avec un langage naturel en requêtes capable d'indexer les graphes RDF (SPARQL). Les images sont récupérées en faisant correspondre les descriptions sémantiques et structurelles de la requête de l'utilisateur avec celles des métadonnées qui décrivent les images. N. Magueth et al. [114] ont présenté également une approche de description des caractéristiques visuelles basée sur le graphe

RDF et qui rejoint l'approche de Soo et al., ce qui permet de rechercher les images grâce au langage SPARQL. SPARQL est un langage de requête et un protocole qui permettra de rechercher, d'ajouter, de modifier ou de supprimer des données RDF disponibles à travers Internet. SPARQL est l'équivalent de SQL car comme en SQL, on accède aux données d'une base de données via ce langage de requête alors qu'avec SPARQL, on accède aux données du Web des données. Cela signifie qu'en théorie, on pourrait accéder à toutes les données du Web avec ce standard.

L'ambition du W3C est d'offrir une interopérabilité non pas seulement au niveau des services, mais aussi au niveau des données structurées ou non. SPARQL est considéré comme l'une des technologies clés du Web Sémantique. Il est adapté à la structure spécifique des graphes RDF et s'appuie sur les triplets qui les constituent. En cela, il est différent du classique SQL (langage de requête qui est adapté aux bases de données de type relationnelles), mais s'en inspire clairement dans sa syntaxe et ses fonctionnalités.

- Les ontologies lourdes. L'utilisation des ontologies a pour objectif la modélisation des connaissances dans un domaine donné comme le précisent travaux menés dans [10] et [11]. Les ontologies sont employées comme une forme de représentation de la connaissance d'un monde ou d'une certaine partie du monde. Concrètement, dans le contexte du Web Sémantique, les ontologies sont utilisées pour modéliser des ressources du web à partir de représentations conceptuelles des domaines concernés et de faire des inférences. Une fois construite et acceptée, une ontologie doit traduire un consensus explicite et certain niveau de partage des aspects essentiels pour permettre l'exploitation des ressources par différentes applications d'une part. D'autre part, la formalisation, autre facette des ontologies, est nécessaire pour qu'il soit possible de faire des raisonnements automatiques afin de décharger les utilisateurs d'une partie de leur tâche d'exploitation et de combinaison des ressources du Web. En résumé, les ontologies sont utilisées pour fournir le vocabulaire et la structure des méta-données associées aux ressources annotées, ou comme représentations pivot pour l'intégration de sources de données hétérogènes [53].

Comme dans les approches existantes, la gestion sémantique des images grâce aux ontologies se base généralement sur deux techniques : l'annotation et la recherche. La première technique permet de décrire formellement les images. La deuxième technique favorise la recherche sémantique des images. Pour mener à bien la gestion sémantique des images, l'extraction de la sémantique est primordiale. La tâche d'extraction sémantique peut être considérée comme une tâche de catégorisation, de classification, de reconnaissance et de compréhension. De la proposition faite par Biederman et al. [115], nous pouvons tirer trois types de contextes réels pour l'annotation d'une scène d'image : le contexte sémantique qui tient compte de la probabilité qu'un concept soit présent dans une scène (par exemple, *rue* a une forte probabilité de coexister avec *bâtiment* dans la même scène), le contexte spatial qui exprime les relations spatiales des concepts (par exemple, le ciel est généralement au-dessus des herbes dans une scène) et le contexte d'échelle qui tient compte de la taille de l'objet (un *humain* devrait occuper une petite région dans une scène où figure un *bâtiment*).

Un nombre considérable d'ontologies multimédias a été proposé, qui tente de définir une

norme pour la description du contenu des images [116], [117], [11], [118], [119], [120], [121], [122], [123] et [124]. D'autres domaines ont été proposés pour permettre une interprétation sémantique et un raisonnement sur les descriptions extraites [125] et [126]. Dans [127] par exemple, Avril Styrman crée des métadonnées et des systèmes qui les interprètent afin de faciliter la recherche des images pour un utilisateur moyen. L'objectif principal du processus est de créer une ontologie appropriée pour l'exposition des photos des événements organisés au sein des universités. Wang et al. dans [128] ont utilisé une ontologie multimodale pour l'annotation et la recherche des images d'animaux. Le travail effectué dans [129] se concentre sur l'annotation automatique des images. C'est un processus de création d'un modèle qui attribue automatiquement des termes aux images. L'objectif principal de cette approche est d'évaluer si l'information sous-jacente contenue dans une ontologie (créée à partir d'un vocabulaire de termes) peut être effectivement utilisée conjointement avec les informations visuelles extraites (des images) afin de produire des annotations plus précises. Saathoff et al. dans [1], Styrman et al. dans [127] et Jair et al. dans [130] ont également proposé une architecture pour l'annotation automatique des contenus images.

Pour améliorer la récupération des images, d'autres approches ont essayé de procéder à une combinaison de techniques comme la segmentation des images en régions et l'annotation des régions segmentées. C'est le cas de Wang et al. dans [128], de Ning Ruan dans [131] et de H. Jair et al. dans [130]. Les approches proposées dans [131] et [130] consistent à annoter les régions d'images grâce aux ontologies à la seule différence que dans [130], les images sont d'abord segmentées en régions alors que dans [131], les régions homogènes sont récupérées grâce à un algorithme non supervisé.

Comme nous pouvons le constater, plusieurs travaux existants ont été orientés vers la description sémantique des images. Pour cela, des ontologies ont été proposées. Les travaux menés dans [132] proposent une ontologie spécifique aux objets comme le montre la figure 2.9.

L'ontologie développée dans ces travaux représente les caractéristiques visuelles des objets de l'image par une description symbolique permettant une interprétation de haut niveau. Les chercheurs ont appliqué des algorithmes de segmentation non supervisée pour extraire les régions de l'image d'une part, et d'autre part d'autres algorithmes permettant le passage entre une description symbolique et une analyse de bas niveau de l'image. Les caractéristiques modélisées sont la luminosité (exprimée par cinq paramètres) et la distribution des couleurs dans la gamme vert-rouge et bleu-jaune (sept paramètres), la position de la région dans l'image : horizontale et verticale (trois paramètres par orientation), la taille de l'objet (trois paramètres) et la forme de l'objet (trois paramètres).

L'approche présentée dans [132] permet de décrire les caractéristiques de bas niveau. D'autres approches similaires sont proposées. Les images prises en compte dans cette approche sont des images satellitaires. La détection des objets se fait en effectuant la correspondance entre les objets de l'image avec les concepts de l'ontologie. L'algorithme développé recherche l'objet extrait de l'image dans l'ontologie en comparant ses descripteurs de bas niveau avec ceux des différents concepts de l'ontologie. Une fois cette comparaison faite, le système affiche

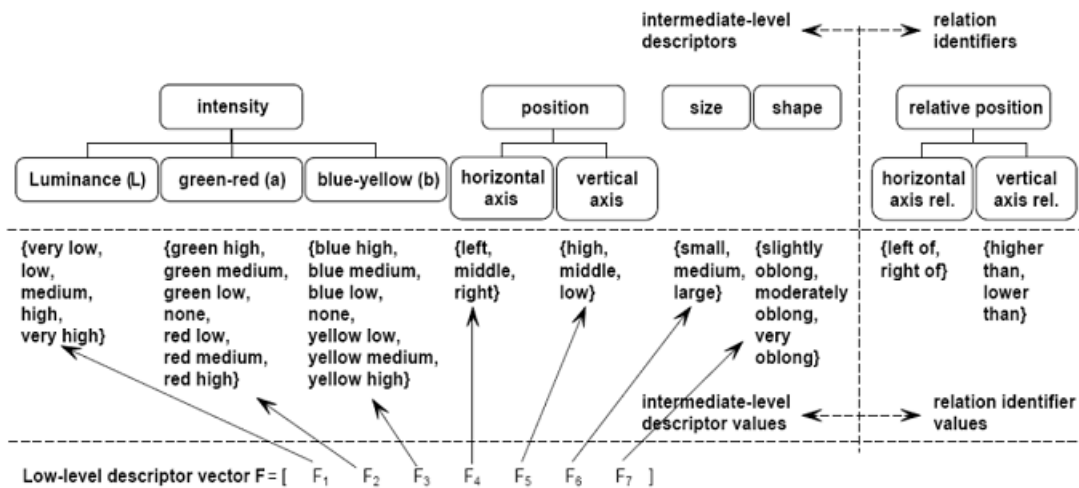


FIGURE 2.9 – Ontologie d'objets

le score de similarité avec chaque concept de l'ontologie.

L'annotation structurée des images a été l'un des principaux objectifs du travail mené dans [130]. Les auteurs de cet article ont d'abord procédé à l'organisation hiérarchique d'un vocabulaire. La hiérarchie et le vocabulaire d'annotation ont été définis manuellement par les auteurs après avoir analysé attentivement les images. Il faut noter que l'ontologie est proposée pour l'annotation et la représentation une collection de photos prises une agence de voyage.

L'approche présentée dans [130] n'est pas basée sur une ontologie lourde, mais une hiérarchie de concepts (ontologie légère). La hiérarchie de concept est principalement constituée des relations de type "est-un" entre les concepts.

2.4/ RECHERCHE D'IMAGES

La recherche d'une image est un processus qui consiste à retrouver une image dans une base de données grâce à ces descripteurs. Ce processus est intimement lié au processus d'annotation des images. Ainsi, une image, un ensemble de mots-clés, de noms, de phrases nominales sont mappés pour retrouver des images comme le confirment les travaux menés dans [133], [134], [135]. Dans la littérature, nous pouvons distinguer trois types de recherche d'images [136] comme le montre la figure 2.10.

- La recherche globale permet d'afficher une collection d'images respectant une mesure de similarité entre les caractéristiques de la requête image et les caractéristiques de chaque image du corpus. La recherche globale est composée de deux étapes : l'identification des groupes d'images grâce aux index pour déterminer les images qui sont susceptibles d'appartenir à l'ensemble des résultats, l'identification des images candidates qui vérifient la mesure de similarité.
- La recherche basée sur les sous-images consiste à comparer une requête image à toutes les régions d'image stockées dans la base de données, quelles que soient la taille de la requête image et celle des régions. La recherche basée sur les sous-images permet de considérer

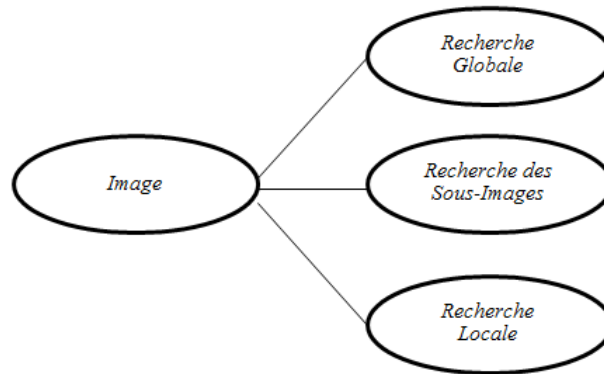


FIGURE 2.10 – Techniques de recherche d'images

chaque région d'image comme une image. Chaque image est représentée par un descripteur multi-niveau stockant les descripteurs des régions d'images. Chaque région pouvant être considérée comme une image à part entière.

- La recherche locale consiste à retrouver des images ayant une région similaire, la région étant composée d'objets. La recherche locale est un peu similaire à la recherche globale, excepté qu'au lieu d'accéder à toutes les régions, on accède une seule région. Elle est basée sur une méthode d'appariement des caractéristiques des images. Lorsque la région sélectionnée par l'utilisateur est composée de plusieurs parties d'image, la requête est décomposée en sous-requêtes, une sous-requête par partie. L'algorithme s'applique donc à chaque région, créant plusieurs ensembles de résultats d'identificateurs de l'image. L'identificateur d'une image apparaît dans le résultat lorsque la distance entre les caractéristiques de l'image et de la requête respecte un seuil.

Les systèmes de recherche d'images utilisent généralement deux techniques : la technique syntaxique et la technique sémantique.

2.4.1/ LA TECHNIQUE SYNTAXIQUE

La technique syntaxique d'une image repose soit sur l'analyse lexicale, soit l'analyse syntaxique du contenu des images. Elle consiste à déterminer automatiquement les termes représentatifs du contenu des images. Ces termes sont ensuite pondérés et leurs pouvoirs discriminants sont extraits. Deux paramètres sont pris en compte dans l'annotation classique : le langage de représentation et le pouvoir discriminant.

2.4.1.1/ LE LANGAGE DE REPRÉSENTATION

Le langage de représentation des images contient l'ensemble des termes susceptibles d'être retenus pour représenter les images. Ce langage peut être une sélection de termes choisis dans le contenu des images ou bien il peut être constitué a priori ; on parle alors de langage contrôlé. Un vocabulaire contrôlé, un glossaire, une hiérarchie informelle ou un thésaurus peut être utilisé. Cependant, l'utilisation d'un thésaurus pour l'indexation n'a pas donné de bons

résultats dans les expérimentations réalisées dans [137], [72], [111] et [138].

2.4.1.2/ LE POUVOIR DISCRIMINANT

Le pouvoir discriminant des termes est ensuite calculé. Un terme apparaissant dans peu d'images est très pertinent pour décrire le contenu de l'image s'il permet de discriminer les images entre elles. De plus, la fréquence d'un terme à l'intérieur de l'image permet de révéler l'importance de ce terme. Ce terme décrit a priori le contenu informationnel de l'image [137], [72], [139], [140] et [141]. Il est souvent déterminé par la formule *tf.idf*. La recherche classique repose donc généralement sur des mesures statistiques portant sur l'apparition des mots-clés dans les images. Une nouvelle génération de méthodes vise à prendre en compte les concepts plutôt que les mots-clés.

2.4.2/ LA TECHNIQUE SÉMANTIQUE

La technique sémantique repose sur une nouvelle intuition selon laquelle la sémantique des mots-clés qui composent les images, dépend des relations conceptuelles entre les objets auxquels elles font référence plutôt que des relations linguistiques et contextuelles trouvées dans leur contenu [137]. La technique sémantique repose alors sur l'utilisation d'ontologies modélisant la conceptualisation des objets.

Une hiérarchie de concepts est un ensemble de concepts ordonnés (des concepts abstraits aux concepts concrets) généralement reliés par des propriétés de type *Est-Un*. Rappelons qu'un concept est en effet défini à partir d'un ou plusieurs termes représentant les variantes lexicales que peuvent prendre les termes. Une hiérarchie de concepts permet de représenter les objets appartenant au même domaine comme par exemple le domaine animal, le domaine humain, etc.

La technique d'indexation des images basée sur les hiérarchies de concepts consiste à rechercher une image grâce aux concepts et/ou aux instances des concepts ou encore grâce aux hiérarchies. Plusieurs approches utilisant des hiérarchies de concepts sont proposées dans la littérature. L'approche proposée dans [48] permet de comprendre les exigences précises des utilisateurs afin de répondre à leurs besoins. Les utilisateurs proposent certaines propriétés dans leur requête comme par exemple les concepts par lesquels ils sont intéressés. L'approche permet d'exploiter des associations entre les concepts sélectionnés par l'utilisateur. Les concepts associés et leurs propriétés sont combinés sous forme de descripteur pour décrire les besoins de l'utilisateur. Cette approche a été utilisée dans [58], [61], [117], [142] et [118]. Elle permet de choisir les concepts représentant au mieux l'image.

La technique basée sur les ontologies repose alors sur l'utilisation de la sémantique des images. Elle n'est donc possible que par l'existence et l'utilisation de ressources décrivant explicitement l'information correspondant aux images [12] et [143]. La démarche orientée Web sémantique a donc un double objectif : annoter les images avec des ontologies mais aussi utiliser ces mêmes ontologies pour les récupérer.

Dans la littérature, plusieurs types d'ontologies sont utilisés dans le cadre de la recherche des images. WordNet est une ontologie souvent utilisée [144],[145], [146] et [108]. La raison principale qui motive son utilisation est qu'elle a pour objectif de représenter la langue

naturelle. Les approches présentées dans [144] et [145] visent à déterminer des parties des ontologies aussi appelées « régions » permettant de rechercher un ensemble d'images. Ces approches sont similaires à la méthode présentée dans [147] et consistent à créer un réseau sémantique pour chaque image [137]. D'autres méthodes de recherche, au contraire, s'appuient non pas sur des régions mais sur des ontologies de domaine, ce qui permet de mieux spécifier le langage recherche. Dans ces approches, une ontologie entière est utilisée pour rechercher les images. L'approche proposée dans [148] est un exemple de ce type de recherche des images. D'autres approches basées sur des ontologies visent à représenter l'ensemble des métadonnées qui peuvent être associées aux images. Ces ontologies sont formelles et permettent de mettre en place des inférences à partir de leurs axiomes. Les approches de recherches d'images proposées dans [149] et *RetrievOnto* [127] sont toutes basées sur l'utilisation d'ontologie. *SemRetriev* est l'environnement de recherche des images fourni par Popescu et al. dans [149] et *RetrievOnto* proposé par Styrman et al. dans [127]. La différence entre les deux environnements est que *RetrievOnto* utilise une hiérarchie de termes extraite de WordNet tandis que *SemRetriev* se base sur une ontologie dédiée aux images.

2.5/ SYSTÈME DE RECOMMANDATION

Le terme *Système de recommandation* décrit généralement un système qui produit des recommandations individualisées, et a pour effet de guider l'utilisateur de façon personnalisée aux articles intéressants ou utiles, dans un grand espace d'options possibles [150]. Ces systèmes suscitent l'intérêt des chercheurs en raison des difficiles questions en suspens [151]. C'est la raison pour laquelle plusieurs applications ont vu le jour dans le but d'aider les utilisateurs [152].

2.5.1/ UTILISATION DES SYSTÈMES DE RECOMMANDATION

Les systèmes de recommandation ont été identifiés comme étant les moyens d'aider les individus d'une communauté à trouver les informations ou des produits qui sont les plus susceptibles d'être intéressants pour eux ou pertinent à leurs besoins (voir [153] et [154]). Ils permettent également de prédire les préférences d'un utilisateur particulier, ou l'identification d'un ensemble de N éléments qui seraient utiles pour certains utilisateurs [155].

Dans un système de recommandation, les préférences de l'utilisateur sont représentées sous des formes diverses, ce qui peut impliquer une ou plusieurs variables. Lorsque les préférences et les besoins des utilisateurs sont représentés par une variable, le système est mono-critère. En particulier, dans les systèmes où les recommandations sont basées sur les opinions des autres, il est crucial d'intégrer plusieurs critères qui influent sur les opinions des utilisateurs. Les recommandations produites par bon nombre de ces systèmes sont basées sur plusieurs critères.

2.5.2/ LES APPROCHES DE SYSTÈMES DE RECOMMANDATION

Les systèmes de recommandation sont généralement classés en plusieurs catégories en fonction de leur approche à la recommandation.

2.5.2.1/ L'APPROCHE PERSONNALISÉE ET NON PERSONNALISÉE

L'approche personnalisée consiste à recommander des objets sur la base du comportement passé de l'utilisateur. Dans cette approche, les préférences de chaque utilisateur sont prises en compte ainsi qu'un certain nombre d'actions ou de décisions effectuées par l'utilisateur qui permettent de prédire de nouveaux produits susceptible de l'intéresser.

L'approche la plus simple est l'approche non personnalisée. Les recommandations sont soit sélectionnées manuellement par le commerçant en ligne, basé sur la popularité des articles (notes moyennes, les données de ventes, des visites au total) ou les recommandations peuvent être les produits les plus nouveaux. L'avantage de cette méthode est qu'elle est facile à réaliser, les données sont faciles à recueillir. Toutefois, les recommandations de ce système sont les mêmes pour tous les utilisateurs et la personnalisation manque et peut ne pas plaire à tout le monde [155].

2.5.2.2/ L'APPROCHE DÉMOGRAPHIQUE

La deuxième approche est basée sur la recommandation démographique. Les données démographiques peuvent être utilisées pour identifier le type d'utilisateurs qui aiment un certain objet. Cela permet de créer des *stéréotypes*. Par exemple, la présentation des informations sur l'âge, le sexe, l'éducation, etc. de personnes qui notent certains restaurants. Avec ces données, on peut apprendre le genre de personne qui aime un certain restaurant, ou tout autre type de produit [156].

L'avantage de l'approche démographique est que les évaluations des items de l'utilisateur ne sont pas utilisées, ainsi les nouveaux utilisateurs peuvent obtenir des recommandations avant de noter n'importe quel article. Des connaissances sur les produits et leurs caractéristiques ne sont pas nécessaires, donc la technique est indépendante du domaine.

L'inconvénient est que la collecte des données requises amène à des questions de vie privée [157]. La classification démographique est également trop brute pour les recommandations hautement personnalisés. Par exemple, toutes les femmes employées âgées de 20 ans ne jouissent des mêmes films [158]. Et les clients ayant des opinions différentes avec d'autres clients. Les recommandations pour eux sont très difficiles à trouver et ils provoquent aussi des recommandations incompatibles pour leurs utilisateurs corrélées, ce problème est appelé problème de moutons gris et est discuté dans [159]. Un autre défi réside dans la difficulté de modifier un profil créé sur un client une fois qu'il change d'opinion [160].

2.5.2.3/ L'APPROCHE COLLABORATIVE

L'approche collaborative est largement utilisée dans les systèmes de recommandation. Elle est destinée à filtrer des informations, la sélection de la bonne information à partir d'une collection de grande taille. La collaboration couvre le fait que l'information qui est utilisée pour filtrer la collection est alimentée par tous les utilisateurs du système. Par exemple si l'utilisateur actif aime A et B, l'approche compare ensuite les préférences des utilisateurs actifs aux préférences des autres utilisateurs et trouve les utilisateurs ayant des préférences similaires.

L'avantage de l'approche collaborative est qu'elle se comporte comme l'approche démographique car aucune connaissance n'est nécessaire sur les produits. Elle est également in-

dépendante du domaine. Les techniques de filtrage collaboratif sont en mesure de faire des recommandations "en dehors de la boîte" parce qu'ils regardent à l'extérieur des préférences de l'utilisateur individuel [150]. Le principal avantage de l'approche est qu'elle utilise peu de mémoire et de temps d'exécution.

Il y a encore plusieurs inconvénients, comme la taille de l'ensemble de données influe sur la qualité des recommandations. Et quand il y a de nouveaux utilisateurs et de nouveaux produits ne comportant aucune qualification, l'approche n'est pas en mesure de donner une recommandation, car elle n'est pas en mesure d'identifier les utilisateurs ayant la même préférence (voir [159], [159] et [160]).

2.5.2.4/ L'APPROCHE BASÉE SUR LE CONTENU

L'approche basée sur le contenu est une approche qui se base sur les qualités et propriétés intrinsèques de l'objet lui-même et en les corrélant avec les préférences et intérêts de l'utilisateur. Ce type de système va donc extraire un certain nombre de caractéristiques et attributs propres à un contenu, afin de pouvoir recommander à l'utilisateur des contenus additionnels possédant des propriétés similaires. Cette méthode crée un profil pour chaque objet ou contenu, c'est-à-dire un ensemble d'attributs/propriétés qui caractérisent l'objet.

La recherche sur les systèmes de recommandation basés sur le contenu prend place à l'intersection de différents domaines de recherche en informatique, notamment la Recherche d'Information (RI) et l'Intelligence Artificielle ([161]). En Recherche d'Information, il est considéré que les internautes voulant des recommandations sont engagés dans un processus de recherche d'information. Le processus principal réalisé par un système de recommandation basé sur le contenu consiste à faire correspondre les attributs d'un profil utilisateur (où les préférences et intérêts sont stockés) avec les attributs des items, dans le but de recommander à l'utilisateur de nouveaux objets intéressants. L'utilisateur exprime un besoin ponctuel en donnant une requête (habituellement une liste de mots-clés). Dans les systèmes de filtrage d'information, le besoin est représenté par le profil de l'utilisateur. Les items à recommander peuvent être très différents, en fonction du nombre et du type des attributs utilisés pour les décrire. Les systèmes de recommandation basés sur le contenu recommandent des objets similaires aux objets préférés par les utilisateurs par le passé. Ils comparent les utilisateurs entre eux sur la base de leurs évaluations passées pour créer les communautés, et chaque utilisateur reçoit les documents jugés intéressants par sa communauté ([162]).

2.5.2.5/ L'APPROCHE HYBRIDE

Les dernières approches sont les approches hybrides. Le but c'est de combiner plusieurs approches pour affiner les recommandations. Elles combinent généralement les systèmes de recommandation basés sur le contenu et les systèmes collaboratifs. Les systèmes hybrides sont généralement utilisés lorsqu'il est question de prendre en compte plusieurs critères. Il permet d'affiner les recommandations.

2.6/ CONCLUSION

Nous venons de présenter un état de l'art sur les techniques d'annotation d'images, les techniques de recherche d'images et les techniques de recommandation.

Les techniques d'annotation aussi diverses que variées permettent de décrire les images avec des caractéristiques de bas niveau ou des caractéristiques de haut niveau. Les techniques d'annotation automatique permettent d'annoter les images par une machine. Elles sont basées sur l'extraction automatique des caractéristiques physiques des images et elles sont relativement rapides et pas chères. Elles ne facilitent pas la compréhension des images d'où l'introduction de la description manuelle des images. L'annotation manuelle des images consiste à associer des mots-clés aux images. Cependant, annoter manuellement est un processus très long et coûteux parce qu'elles nécessitent beaucoup de moyens. L'annotation manuelle augmente la précision et baisse la productivité tandis que l'annotation automatique baisse la précision et augmente la productivité. Afin de faire un compromis entre ces deux tâches, leur combinaison est devenue nécessaire, d'où l'annotation semi-automatique. L'annotation semi-automatique, moins coûteuse que l'annotation manuelle et plus lente que l'annotation automatique permet aussi d'obtenir des résultats satisfaisants. Mais vu que les mots-clés ne parviennent pas représenter la sémantique des images du point de vue humain, les outils du Web Sémantique et notamment les ontologies légères et les ontologies lourdes viennent remédier à cette insuffisance.

Deux techniques de recherche d'images sont présentées dans cet état de l'art : la technique syntaxique et la technique sémantique. La technique syntaxique permet de rechercher les images avec des mots-clés tandis que la technique sémantique est basée sur l'utilisation des outils du Web Sémantique. La technique syntaxique permet d'accéder à une collection d'images à partir d'un (des) index proposés par l'utilisateur. Les images partageant un ou plusieurs index s'affichent et l'utilisateur sélectionne celles qui lui conviennent tandis qu'avec la technique sémantique, seule l'image recherchée par la sémantique est affichée. En résumé, la technique syntaxique se rapporte à la généralisation tandis que la technique sémantique n'est rien d'autre que la spécification.

En dernier, nous avons présenté un état de l'art sur les systèmes de recommandation qui permettent de faire des propositions aux utilisateurs en se basant sur leurs préférences. De nos jours, ces systèmes sont de plus en plus intégrés dans des systèmes de recherche pour accélérer la recherche. Les systèmes de recommandation se basent généralement sur plusieurs techniques comme nous venons de l'énumérer ci-dessus. Chacune des techniques a ses avantages et ses inconvénients. L'approche démographique n'utilise pas la préférence de chaque utilisateur mais sur la préférence d'un groupe d'individus. L'approche collaborative n'exploite pas ou ne demande aucune connaissance sur les contenus. L'approche basée sur le contenu n'a pas besoin d'une large communauté d'utilisateurs pour pouvoir effectuer des recommandations. Une liste de recommandations peut être générée même s'il n'y a qu'un seul utilisateur. L'approche personnalisée requière une étude approfondie des utilisateurs de leurs préférences pour effectuer des recommandations ciblées. Les méthodes hybrides sont de plus en plus utilisées, car elles permettent de résoudre des problèmes comme la rareté qu'on retrouve dans une approche de

recommandation uniquement collaborative. D'autre part, si par exemple on considère 2 utilisateurs avec les mêmes goûts mais qui n'ont pas évalué ou *raté* des objets en commun, un filtrage collaboratif pur ne les considérera pas comme similaires ou voisins.

SYNTHÈSE : APERÇU GÉNÉRAL

D'*Emergsem*

*Ce chapitre est une synthèse des approches que nous proposons dans cette thèse. La première partie de ce chapitre définit quelques concepts. La deuxième partie présente la problématique générale suivie des différentes propositions que nous avons faites pour améliorer l'annotation et la recherche des images à travers le système *Emergsem*. La troisième partie présente *Emergsem*. *Emergsem* est constitué de trois couches. La première couche décrit la représentation sémantique des images, la deuxième couche présente l'annotation collaborative et l'émergence de la sémantique des images et la troisième partie couche présente la recherche d'images basée sur le clustering et la recommandation. La quatrième partie de ce chapitre présente brièvement la phase expérimentale.*

Sommaire

3.1	Définition et concepts	42
3.2	Problématique générale	43
3.3	Description des couches d'<i>Emergsem</i>	44
3.3.1	Représentation sémantique d'images	44
3.3.2	Annotation sémantique collaborative et émergence des sémantiques	46
3.3.3	Recherche d'images basée sur le clustering et la recommandation	49
3.3.4	Expérimentation	54

3.1/ DÉFINITION ET CONCEPTS

Annotation collaborative des images : La collaboration s'entend en fait par une situation de travail collectif où tâches et buts sont communs. Chacun des membres du groupe travaille ainsi sur les mêmes points. Il s'agit en effet de fusionner les contributions individuelles (sur une tâche précise) d'un groupe d'individus. La responsabilité est donc ici collective. Ce mode de travail collectif engage par conséquent une communication régulière entre les membres du groupe et une connaissance précise de la progression de l'action collective (Wikipédia est exemple de site de description collaboratif). Partant de cette définition, nous pouvons dire que l'annotation collaborative des images est une technique qui consiste en la description du contenu des images de façon collaborative par un groupe d'annotateurs. Les annotateurs disposant des mêmes outils d'annotation proposent une description sémantique des images qui leur sont proposées.

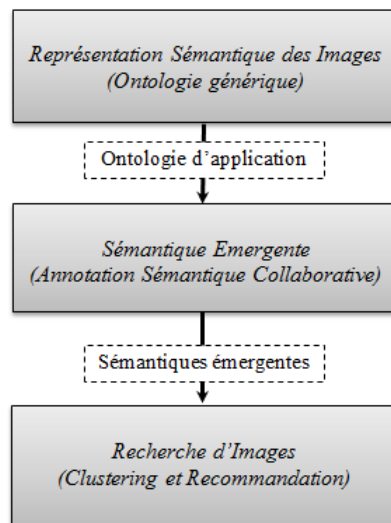
La recherche d'images par le contenu : Elle consiste, après analyse de tous les pixels ou d'une partie réduite (masque), de transformer l'information des pixels en un autre type d'information de telle sorte que la recherche d'images (à l'identique, ou de même catégorie) soit facilitée (en termes informatiques, taille compacte, vitesse, tout en conservant une sémantique proche de l'utilisateur). Les premiers systèmes d'indexation des images ont utilisé la couleur, plus tard l'analyse des couleurs par histogramme s'est à la fois améliorée et diversifiée. Plusieurs modèles de représentations des couleurs ont été utilisés, des raffinements de l'histogramme global primitif ont été introduits¹.

Descripteur d'images : Le descripteur d'une image est la représentation des caractéristiques visuelles de l'images. Les caractéristiques visuelles permettent généralement de mieux rendre compte de certaines propriétés visuelles de l'image, utilisées pour des traitements ultérieurs entrant dans le cadre d'applications telles que la détection d'objets ou la recherche d'images par le contenu². On distingue usuellement les caractéristiques globales qui sont calculées sur toute l'image et les caractéristiques locales qui sont calculées autour de points d'intérêt. On peut aussi parler des caractéristiques semi-locales quand celles-ci sont extraites dans des zones restreintes de l'image, résultant d'une segmentation de l'image en région ou simplement selon une grille arbitraire. Outre ces descripteurs qui sont calculés sur les images ou les régions d'images (les descripteurs physiques), il existe également les descripteurs sémantiques qui visent à mieux capturer la sémantique d'une image. Les descripteurs sémantiques seront présentés plus en détail dans la suite de la thèse.

Ontologie : Une ontologie est une spécification formelle et explicite d'une conceptualisation partagée. *Formelle* reflète le fait que l'ontologie est lisible par la machine et permet de raisonner sur son contenu à partir de l'humain et la machine. *Explicite* signifie que le type de concepts utilisés et les contraintes sur leur utilisation sont explicitement définis. *Partagée* se réfère à la connaissance commune, énoncée dans l'ontologie. *Conceptualisation* se réfère au modèle obtenu en faisant abstraction de certains phénomènes qui existent dans le monde réel en identifiant les concepts pertinents de ces phénomènes [48].

1. http://fr.wikipedia.org/wiki/Indexation_automatique_de_documents

2. http://fr.wikipedia.org/wiki/Extraction_de_caractéristique_en_vision_par_ordinateur

FIGURE 3.1 – Les couches d'*Emergsem*

3.2/ PROBLÉMATIQUE GÉNÉRALE

Malgré les apports considérables des systèmes existants dans les CBIR, nous retenons le que fossé sémantique reste un sujet préoccupant parce que d'une part, les caractéristiques de bas niveau comme le retour de pertinence ne parviennent pas à capturer la sémantique des images, et d'autre part, les approches d'annotation qui ont été proposées pour remédier au problème ne sont en mesure de décrire la sémantiques d'une image en tenant compte des différents objets qu'elle contient et les relations qui existent entre ces objets.

Dans ce travail, le sujet central est donc la proposition d'une nouvelle technique pour représenter et rechercher les images par leurs sémantiques. *Emergsem* est le système que nous proposons. Dans *Emergsem*, nous présentons une approche d'annotation sémantique collaborative et une approche de recherche sémantique d'images. Dans l'approche d'annotation sémantique collaborative, chaque annotateur utilise une ontologie et un dictionnaire lexical pour décrire le contenu sémantique des images. L'ontologie est utilisée pour identifier et décrire les objets contenus dans les images et les relations entre ces objets. Les descriptions fournies par les annotateurs sont combinés par un processus d'émergence sémantique qui utilise des descriptions fortement corrélés pour créer l'annotation finale d'une image. Dans l'approche de recherche sémantique d'images, deux méthodes sont utilisées : le clustering et la recommandation. Le clustering permet de regrouper les images similaires à la requête de l'utilisateur. Il est basé sur l'extraction des sémantiques les plus représentatives à partir d'une requête de de mots-clés, sur l'extraction des sémantiques voisines aux sémantiques les plus représentatives et sur un cluster de sémantiques qui permet d'afficher les images les plus pertinentes. La recommandation quant à elle, repose sur une classification des profils d'utilisateurs par l'algèbre de Galois pour proposer des sémantiques. Deux types de profils sont utilisés : profils statiques et dynamiques.

Emergsem a les caractéristiques suivantes :

- 1.) la représentation sémantique des images pour décrire les images ;

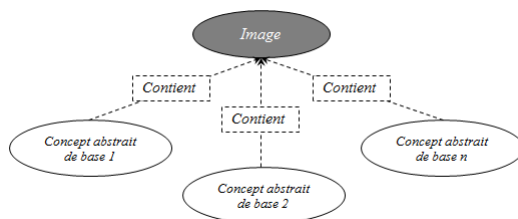


FIGURE 3.2 – Schéma de l'ontologie générique

2.) l'annotation sémantique collaborative pour faire émerger la sémantique des images ;
 3.) la recherche d'images basée sur le clustering et la recommandation des sémantiques et l'expérimentation.

Le but de ces approches proposées est d'améliorer la récupération des images. La figure 3.1 présente l'architecture du system *Emergsem*.

3.3/ DESCRIPTION DES COUCHES D'*Emergsem*

Dans cette section, nous présentons les différentes couches de l'approche *Emergsem*. Quatre couches sont décrites : la représentation sémantique des images, l'annotation sémantique collaborative et émergence des sémantiques, la recherche d'images basée sur le clustering et la recommandation et enfin l'expérimentation.

3.3.1/ REPRÉSENTATION SÉMANTIQUE D'IMAGES

La représentation sémantique des images est basée sur la description des images via les descripteurs sémantiques. Un descripteur sémantique a pour vocation de fournir une description d'images proche de la perception humaine dans un seul but, réduire le fossé sémantique. Qui dit fournir une description d'images proche de la perception humaine dit représenter non seulement le sens des différents objets présents dans les images, mais également les relations qui existent entre les objets. Par conséquent, un descripteur sémantique doit être en mesure de décrire les régions d'images et les liens qu'il y a entre ces régions dans un langage à la fois compréhensible par la machine que par les humains. Dans cette thèse, la représentation sémantique des images s'appuie sur plusieurs niveaux d'ontologie. Nous distinguons deux niveaux : une ontologie générique, regroupant un grand nombre de concepts communs et une ontologie d'application.

- Une ontologie générique d'images est une ontologie qui décrit des ensembles de domaines d'images. Une ontologie générique comporte plusieurs concepts généraux devant permettre de définir les principales distinctions entre n'importe quelles entités et des relations générales permettant, par exemple de représenter le fait qu'un concept est composé d'autres concepts. Elle est capable de générer des ontologies d'application qui à leur tour permettent de représenter les sémantiques d'images. L'ontologie générique est composée d'un méta concept abstrait *Image* et des concepts abstraits de base. Le méta concept abstrait est un concept de haut niveau qui décrit toutes les images quel que soit leurs domaines. Il est connu sous le nom d'*ancêtre* et permet la définition des concepts abstraits de base. Les concepts abs-

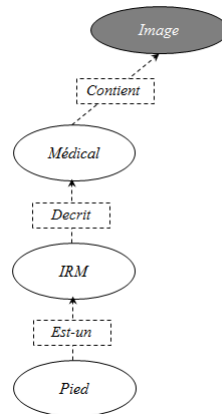


FIGURE 3.3 – Schéma de l'ontologie d'application

trait de base sont des concepts qui décrivent des entités globales d'images, c'est-à-dire que chacun des concepts abstraits de base englobe un domaine bien vaste de la représentation des images comme par exemple le domaine *Médical*. L'ontologie générique définit aussi une relation "*Contient*" qui permet de lier les concepts abstraits de base au méta concept abstrait *Image*. La relation *Contient* introduit les propriétés générales de l'image représentées par les concepts abstraits de base dans l'ancêtre *Image*. *Contient* est une relation pour signifier que tous les concepts abstraits de base sont contenus dans l'ancêtre *Image* comme le montre la figure 3.2. L'ontologie générique a donc pour vocation de générer autant d'ontologies adaptées à la description des images d'un domaine spécifique par la proposition des concepts et des relations qui décrivent ces domaines.

- L'ontologie d'application quant à elle est une ontologie dérivée de l'ontologie générique. Le principe de dérivation de l'ontologie d'application est basé sur la spécification des concepts abstraits de base et la proposition de sous concepts abstraits et concrets ainsi que des relations. La personnalisation est faite dans un but précis : décrire des images appartenant à des domaines bien spécifiques (Exemple : domaine médical). Les concepts abstraits d'un domaine d'images (représenté par un concept abstrait de base) définissent chacun un sous domaine de ce domaine. Par exemple pour le concept abstrait de base décrivant le domaine médical, un sous concept abstrait pourrait décrire uniquement les images *IRM*. Les concepts concrets du sous domaine pourront être aussi proposés. Ces concepts décrivent un objet ou une région précise. Ils sont la spécialisation des sous domaines représentés par les concepts abstraits. Un concept concret du sous concept *IRM* pourrait être consacré uniquement à la description des os du pied. Nous avons également proposé des relations pour relier les concepts : la relation *Decrit* et la relation *Est – un*. La relation *Decrit* est une relation qui peut être vue comme la relation *Propriété – de*. Elle permet de relier les sous concepts abstraits aux concepts abstraits de base parce que les sous concepts sont des "*propriétés de*" concepts de base, d'où la relation *Decrit* (Exemple : Les images médicales décrivent les images *IRM*, c'est-à-dire *Médical Decrit IRM*). Cette relation peut relier également deux

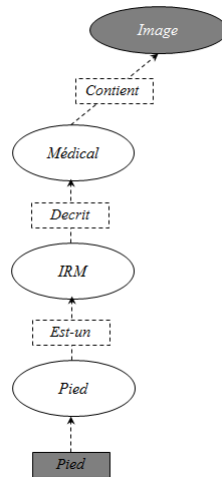


FIGURE 3.4 – Sémantique : une ontologie d'application instanciée

concepts abstraits (l'un étant le père et l'autre le fils). La relation *Est – un* est une relation de spécification qui permet désigner une information précise attribuée à un concept. *Est – un* relie les concepts abstraits aux concepts concrets (Exemple : Les images du pied sont des images d'IRM, alors *Pied Est – un IRM*). La figure 3.3 présente un schéma d'ontologie d'application. L'ontologie d'application est composée d'un méta concept abstrait *Image* qui représente toutes les images, d'un concept abstrait de base qui représente uniquement les images médicales représenté par le concept "Medical" et relié au méta concept abstrait par la relation "Contient". Elle est également composée d'un concept abstrait et d'un concept concret. Le concept abstrait représente les images "IRM" et relié au concept abstrait de base par la relation "Decrit" et le concept concret "Pied" qui représente les os du pied. Il est relié au concept abstrait "IRM" par la relation "Est – un".

La représentation finale est l'ontologie d'application instanciée. Elle est constituée de méta concept abstrait et des concepts abstraits de base provenant de l'ontologie générique, des concepts abstraits et des concepts concrets introduits dans la phase de dérivation et au niveau des feuilles de l'arbre ontologique (des concepts concrets) des instances, c'est-à-dire des mots-clés associés au concepts concrets. La figure 3.4 présente une ontologie d'application instanciée.

3.3.2/ ANNOTATION SÉMANTIQUE COLLABORATIVE ET ÉMERGENCE DES SÉMANTIQUES

L'annotation sémantique collaborative est la deuxième contribution. Rappelons que dans la section précédente, nous avons proposé une représentation sémantique des images grâce à une ontologie d'application instanciée par association de mots-clés. L'annotation sémantique collaborative est un processus qui consiste à faire décrire les images par une communauté d'annotateurs. Il permet de recueillir des opinions diverses sur la sémantique d'une image et de les faire converger vers une seule sémantique. Mais pourquoi une annotation collaborative ? La

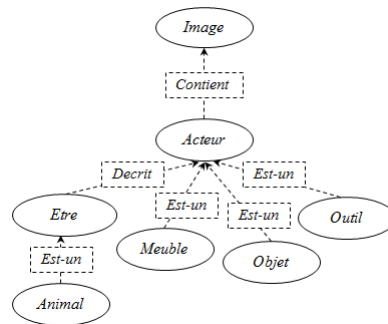


FIGURE 3.5 – Ontologie d'application proposée aux annotateurs

sémantique obtenue par un seul annotateur ne pourrait-elle pas être la bonne ? C'est à cette question que nous allons répondre.

Bien que les ontologies soient utiles pour réaliser une description sémantique, structurée et significative des images, les annotateurs n'ont pas souvent la même perception des images et par conséquent, les sémantiques associées aux images peuvent être différentes d'un annotateur à un autre. Comme nous venons de le rappeler un peu plus haut, la pertinence de ces sémantiques reste fort subjective parce qu'elles peuvent ne pas être en adéquation avec le contenu des images. Une image reçoit plusieurs interprétations et les annotations, même si elles sont sémantiques pourraient être remise en cause parce que l'interprétation peut être erronée ou imprécise ou encore insuffisante. Prenons l'exemple l'image de la figure 3.6 décrite par cinq utilisateurs avec une ontologie. Le but de la description est que chaque annotateur fournisse la sémantique de l'image en instanciant les concepts concrets d'une ontologie d'application avec des mots-clés d'un dictionnaire lexical. Une ontologie d'application a été mise à la disposition des annotateurs. Dans cette ontologie d'application, un concept abstrait de base "Acteur" décrivant les acteurs d'une image. *Acteur* est composé d'un concept abstrait "Etre" et de quatre concepts concrets "Animal", "Meuble", "Objet" et "Outil" comme le montre la figure 3.5.

Et l'image de la figure 3.6 est l'image proposée aux annotateurs.

Les mots-clés proposés sont présentés dans la table 3.1. Nous remarquons deux problèmes majeurs : la divergence des points de vue sur les mêmes objets et l'insuffisance des descriptions faites par certains annotateurs comme le montre la colonne 5 du tableau pour l'annotateur 4 et les colonnes 4 et 5 de tableau pour l'annotateur 5 qui sont restées vides. La raison pourrait être que les objets ne sont pas perçus par ces annotateurs ou bien qu'ils ne savent pas à quoi ils correspondent.

Dans cette section nous proposons l'émergence des sémantiques associées à une image vers une sémantique résultante par une annotation sémantique collaborative basée sur une communauté d'annotateurs. L'annotation collaborative consiste en la proposition de la description d'une image par plusieurs annotateurs. Des approches se sont penchées sur la collaboration comme nous l'avons évoqué et décrit dans l'état de l'art. Ces approches se basent sur des

mots-clés. Généralement, plusieurs annotateurs procèdent à l'annotation de la même image et le système se charge d'en déduire la description la plus pertinente de l'image. Nous pouvons citer en guise d'exemple, les sites d'annotation collaboratifs comme Wikipédia et des frameworks de description collaborative des images qui sollicite un public plus ou moins large.

L'annotation collaborative que nous proposons est basée sur une communauté d'annotateurs. Chaque annotateur de la communauté propose une représentation sémantique de la même image. Différents points de vue seront ainsi exprimés dans les annotations. Pour que les annotateurs puissent exprimer leurs points de vue, un dictionnaire lexical est proposé. WordNet est le dictionnaire lexical proposé. L'ontologie est instanciée par chaque annotateur avec des mots-clés provenant du dictionnaire lexical. Une fois que les différents points de vue sont exprimés par chaque annotateur à travers les sémantiques proposées, nous procédons ensuite à l'émergence de toutes les sémantiques. L'émergence consiste en effet à faire converger toutes les sémantiques proposées vers la sémantique la plus représentative comme le montre la figure 3.7.

Le principe est le suivant. Chaque annotateur propose une sémantique de la même image

Annotateurs N °	Concepts concrets			
	Animal	Meuble	Outil	Objet
1	Chien	Table	Ordinateur	Souris
2	Chien	Bureau	Téléphone	Bouteille
3	Kangourou	Bureau	Ordinateur	Souris
4	Chien	Bureau	Ordinateur	- - -
5	Loup	Table	- - -	- - -

TABLE 3.1 – Liste des instances proposées

en instanciant les concepts concrets de l'ontologie d'application grâce aux mots-clés provenant de WordNet. Ces mots-clés sont proposés via une interface d'annotation reliée aux concepts concrets de l'ontologie d'application. Une fois les mots-clés proposés (les sémantiques générées), nous procédons ensuite à la détermination de la sémantique la plus représentative à partir de toutes les sémantiques proposées, c'est-à-dire, la détermination de l'instance la plus pertinente parmi les instances de chaque concept concret des sémantiques. Ces instances (pertinentes) seront ensuite mappées dans l'ontologie d'application pour donner la sémantique émergente. Pour déterminer la sémantique émergente, nous avons introduit dans le système une fonction appelée f_{emerg} . La fonction f_{emerg} permet d'évaluer les instances proposées dans



FIGURE 3.6 – Image proposée

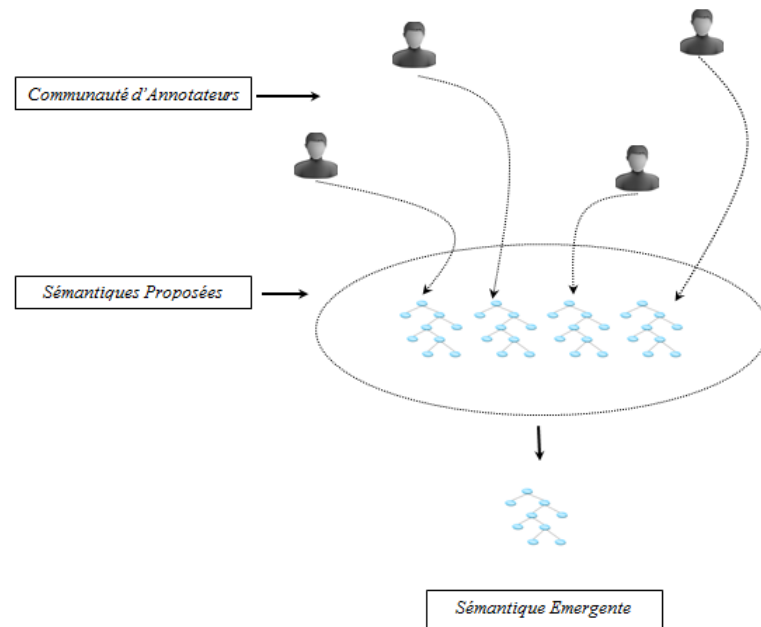


FIGURE 3.7 – Émergence de la sémantique

chaque concept concret des sémantiques. La fonction f_{emerg} combine deux techniques d'évaluation : la fréquence et le poids. La fréquence permet de déterminer les instances les plus fréquentes pendant le processus de description des images et le poids permet de déterminer la position des instances par rapport à la position des autres instances proposées par un annotateur (le rang qu'elles occupent pendant l'annotation).

3.3.3/ RECHERCHE D'IMAGES BASÉE SUR LE CLUSTERING ET LA RECOMMANDATION

La recherche d'images est une technique qui permet de retrouver les images avec des descripteurs. Pour rechercher les images, les systèmes de recherche comparent les descripteurs fournis dans la requête aux descripteurs ayant servi à annoter les images. La première technique de recherche est basée sur la similarité des caractéristiques physiques des images. Les utilisateurs proposent une requête image. Le système calcule la similarité entre les caractéristiques physiques de la requête image et les caractéristiques physiques des images de la base, puis sélectionne les meilleures images. C'est une technique qui utilise la relation d'*image à image*. Comme les techniques basées sur la similarité entre les caractéristiques physiques ne parviennent pas à satisfaire les attentes des utilisateurs, la technique syntaxique est proposée. La technique syntaxique est basée sur la relation de *mots – clés à image*. Elle consiste à comparer les mots-clés proposés dans la requête aux mots-clés décrivant les images. Mais vu les limites liées à l'utilisation de la technique syntaxique, comme nous l'avons annoncé dans la problématique générale, les recherches basées sur la sémantique des images sont proposées. La démarche orientée vers les sémantiques permet de rechercher les images à partir des ressources permettant d'extraire les concepts et les propriétés. La recherche sémantique repose sur

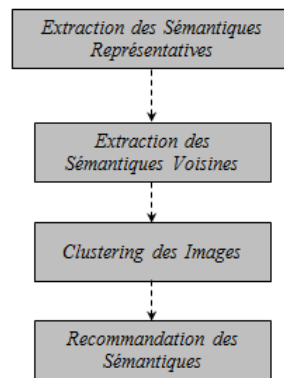


FIGURE 3.8 – Les couches de l'approche de recherche d'images

l'intuition suivant laquelle la sémantique des images dépend des relations conceptuelles entre les objets auxquels elles font référence plutôt que des relations linguistiques et contextuelles trouvées dans leur contenu. Partant de ce fait, la recherche sémantique d'une image ne serait possible que par l'existence et l'utilisation de ressources décrivant explicitement l'information correspondant aux objets.

Dans cette section, nous présentons une technique de recherche d'images basée sur le clustering et la recommandation. Elle est composée de quatre couches : l'extraction des sémantiques représentatives, l'extraction des sémantiques voisines, le clustering des images et la recommandation comme le montre la figure 3.8.

3.3.3.1/ EXTRACTION DES SÉMANTIQUES REPRÉSENTATIVES

L'extraction des sémantiques représentatives permet d'afficher les sémantiques les plus proches de la requête de mots-clés proposée par l'utilisateur. Le principe d'extraction des sémantique représentatives est le suivant. L'utilisateur propose des mots-clés. Les sémantiques représentatives sont extraites grâce à une mesure de similarité qui compare le vecteur des

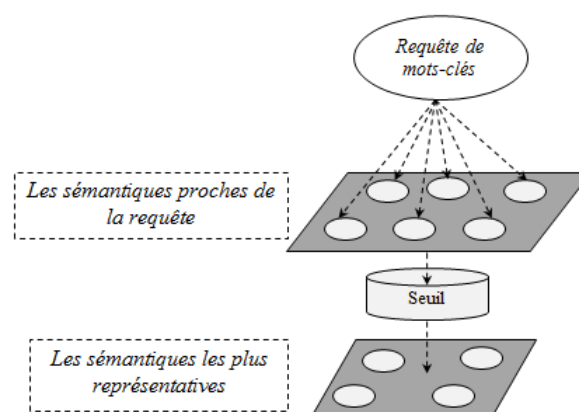


FIGURE 3.9 – Extraction des sémantiques représentatives

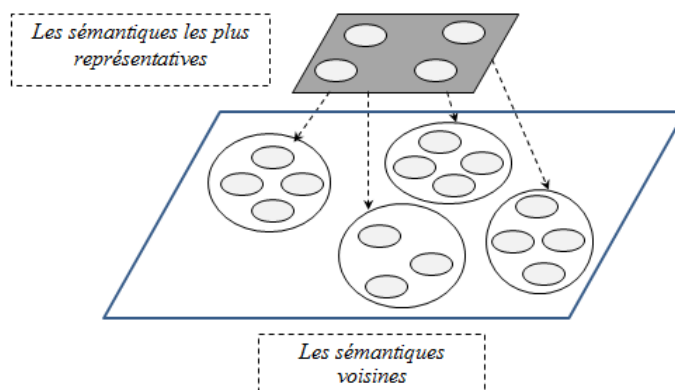


FIGURE 3.10 – Extraction des sémantiques voisines

mots-clés proposés au vecteur des instances de chaque sémantique du corpus. Les similarités des vecteurs des sémantiques du corpus par rapport au vecteurs de mots-clés sont affichés. Un seuil est fixé et seules les sémantiques qui respectent ce seuil deviennent les sémantiques les plus représentatives de la requête de l'utilisateur comme le montre la figure 3.10.

3.3.3.2/ EXTRACTION DES SÉMANTIQUES VOISINES

Une fois les sémantiques les plus représentatives sont obtenues, le système procède ensuite à l'extraction des sémantiques voisines de chaque sémantique représentative. L'extraction des sémantiques voisines est basée sur la même technique que celle utilisée pour extraire les sémantiques représentatives à la seule différence que cette fois-ci, ce sont les sémantiques représentatives qui sont comparées aux sémantiques du corpus. L'objectif est de créer des ensembles des sémantiques voisines à partir des sémantiques les plus représentatives de la requête de l'utilisateur comme le montre la figure 3.10.

3.3.3.3/ CLUSTERING

Le clustering est une technique de regroupement d'un ensemble d'objets de telle sorte que les objets du même cluster (groupe) soient plus proches les uns des autres que ceux des autres groupes. Dans les systèmes de recherches d'images par le contenu, le clustering permet de grouper les images similaires. Cette similarité peut être mesurée par la comparaison des descripteurs physiques ou sémantiques des images. Dans notre travail, nous proposons une technique de recherche d'images basée sur le cluster d'images. Le but est de grouper les images afin de permettre aux utilisateurs de disposer d'une collection d'images similaires. Contrairement au système CBIR, qui propose des clusters d'images de toute la collection d'une base de données à l'utilisateur, nous proposons des clusters d'images basés non pas sur la requête de l'utilisateur, mais sur les sémantiques voisines obtenues. La technique de clustering est composée méthodes : la formation des groupes de sémantiques au sein de chaque ensemble de sémantique voisines et la formation des clusters de sémantiques.

La première méthode est la formation des groupes de sémantiques similaires au sein de chaque ensemble de sémantiques voisines. Ces groupes de sémantiques voisines sont obtenus

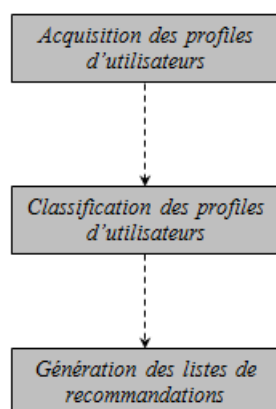


FIGURE 3.11 – La recommandation

grâce à la comparaison de la granularité des sémantiques. La granularité est une mesure qui permet de déterminer la similarité de la structure des ontologies ayant servi à décrire les sémantiques. L'objectif est de grouper les sémantiques au sein de chaque ensemble de sémantiques voisines en comparant leur granularité

La deuxième méthode est la formation des clusters de sémantiques. Elle permet d'évaluer la similarité entre les différents groupes de sémantiques pour observer les différences entre les regroupements. Elle est basée sur la matrice de confusion et la fonction de correspondance. La matrice de confusion est utilisée pour comparer deux groupes de sémantiques appartenant à deux ensembles et la fonction de correspondance permet de vérifier si un groupe de sémantique appartenant à un ensemble de sémantiques voisines est similaire à un autre groupe de sémantiques se trouvant dans un autre ensemble. Les images décrites par les sémantiques du cluster de sémantiques sont proposées à l'utilisateur. La figure présente deux ensembles de sémantiques voisines contenant chacun des groupes de sémantiques similaires et un cluster des sémantiques les plus pertinentes.

3.3.3.4/ RECOMMANDATION

Les systèmes de recommandation sont une forme spécifique de filtrage de l'information visant à présenter les éléments qui sont susceptibles d'intéresser l'utilisateur. Généralement, un système de recommandation permet de comparer le profil d'un utilisateur à certaines caractéristiques de référence, et cherche à prédire le comportement d'un utilisateur³. Le système de recommandation que nous proposons vise à recommander des sémantiques aux utilisateurs en leur suggérant des sémantiques basées sur leurs préférences. Ceci sous-entend que l'historique des sémantiques préférées par les utilisateurs doit être pris en compte. Le système de recommandation que nous proposons est basé sur trois points essentiels : l'acquisition des profils des utilisateurs, la classification des profils et la génération des listes de recommandation comme le montre la figure 3.11.

L'acquisition des profils est une technique qui consiste à stocker les profils de la com-

3. http://fr.wikipedia.org/wiki/Système_de_recommandation

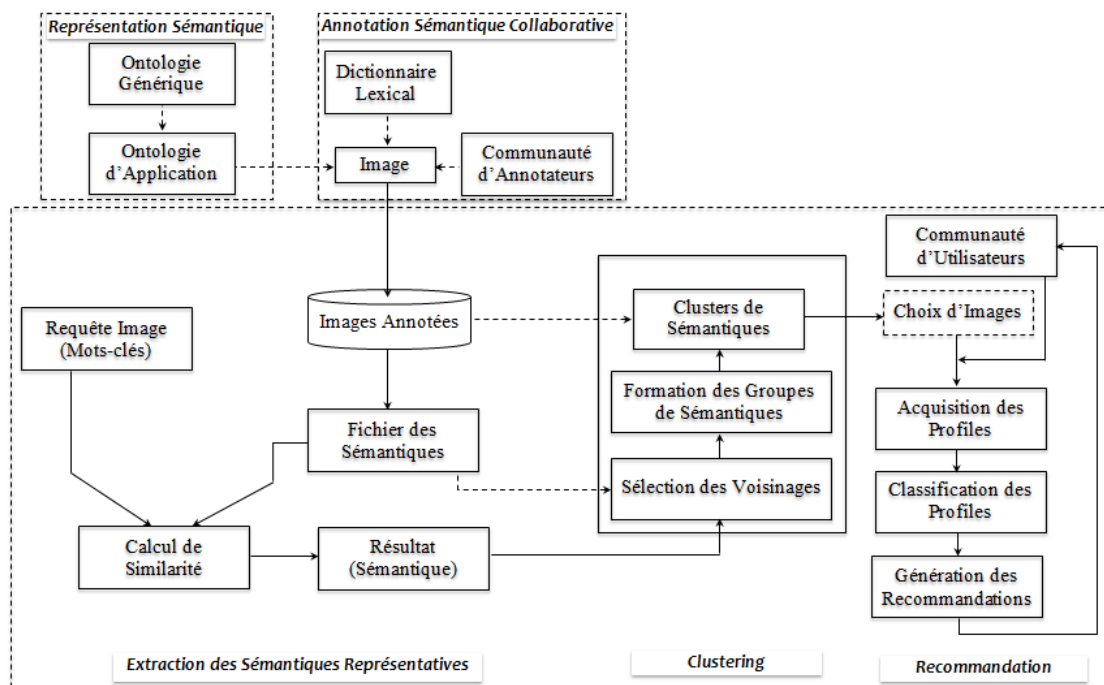


FIGURE 3.12 – Architecture d'Emergsem

munauté d'utilisateurs. Lorsqu'un utilisateur s'inscrit, il fournit des informations nécessaires à son identification (identifiant, email, mot de passe, etc.). Ces informations représentent les informations statiques de l'utilisateur. Statiques parce que ces informations sont fournies une seule fois (même si elles peuvent être modifiées) et n'ont aucune influence sur le système de recherche d'images. Lorsqu'un utilisateur sélectionne une image dans le cluster d'images, la sémantique associée à l'image sélectionnée est également stockée. C'est la préférence de l'utilisateur, son profil dynamique. Dynamique parce que la préférence de l'utilisateur peut changer, c'est-à-dire peut varier d'une recherche à une autre.

La classification des profils est une technique qui permet de regrouper les profils similaires. Elle est basée sur les profils dynamiques que les utilisateurs ont en commun. Pour créer les groupes de profils dynamiques, nous avons utilisé l'algèbre de Galois. L'algèbre de Galois permet de regrouper les objets de manière exhaustive dans les classes, appelé "concept formel", en utilisant leurs propriétés partagées. Il est généralement basé sur une matrice booléenne, appelé contexte de la matrice et noté C , dont les lignes représentent un ensemble d'objets O qu'on veut décrire et en colonnes, un ensemble d'attributs A que ces objets ont ou n'ont pas.

Lorsque les groupes de profils dynamiques sont constitués, un ou plusieurs d'entre eux peuvent être recommandés aux utilisateurs. Étant donné qu'un profil dynamique peut appartenir à deux groupes, la génération des recommandations permet d'éliminer les doublons et de faire des recommandations personnalisées aux utilisateurs.

3.3.4/ EXPÉRIMENTATION

Nous avons mené un certain nombre d'expérimentations pour évaluer la pertinence des approches que nous avons proposées. Dans un premier temps, nous avons implémenté *Emergsem*. Dans un deuxième temps, nous avons évalué les résultats de nos expérimentations. L'implémentation du système *Emergsem* est rendu possible grâce à la modélisation d'une ontologie d'application sous *Protégé* que nous avons intégrée dans le prototype que nous avons développé. La première phase d'expérimentation a porté sur les techniques d'annotations d'images. Elle compare les résultats des techniques d'annotation classiques aux résultats d'*Emergsem* d'une part, et d'autre part l'influence de la recommandation sur le système de recherche d'images. La deuxième phase d'expérimentation compare les résultats d'*Emergsem* avec les résultats de deux autres approches. La figure 3.12 présente l'architecture de *Emergsem*.

ANNOTATION SÉMANTIQUE COLLABORATIVE D'IMAGES

Ce chapitre définit une nouvelle approche d'annotation d'images. Elle se base sur deux principaux concepts : la sémantique et la collaboration. La sémantique des images est obtenue grâce à une ontologie et un dictionnaire lexical qui permettent aux annotateurs de décrire les images. Dans cette approche, la description sémantique n'est pas faite par un seul annotateur, mais de façon collaborative, c'est-à-dire par une communauté d'annotateurs. L'architecture utilisée est constituée de deux couches : une couche de génération des sémantiques et une couche d'émergence des sémantiques. Les composantes des différentes couches sont décrites dans ce chapitre et les différents processus permettant la génération et l'émergence des sémantiques sont expliqués.

Sommaire

4.1	Introduction	56
4.2	Définitions et concepts de base	56
4.2.1	Annotation sémantique	56
4.2.2	La collaboration	57
4.3	Représentation des images dans <i>Emergsem</i>	59
4.3.1	Descripteurs d'images	59
4.3.2	Ontologie générique	60
4.3.3	Ontologie d'application	63
4.3.4	Description de l'ontologie d'application	64
4.3.5	Instanciation de l'ontologie d'application	71
4.4	Annotation collaborative	74
4.4.1	La communauté d'annotateurs	74
4.4.2	Dictionnaire lexical	76
4.4.3	Génération des sémantiques	77
4.4.4	Émergence des sémantiques	77
4.5	Conclusion	81

4.1/ INTRODUCTION

L'annotation d'images a une longue histoire de recherche, et sans surprise, les perspectives de recherche et d'interprétations de ce qu'est une annotation sont censées être variables, comme nous venons de le préciser dans l'état de l'art. L'annotation d'une image est un processus de descriptions des différents objets présents dans cette image. Elle constitue une manière possible de donner un sens à une image ou d'associer aux images des informations, représentées par les caractéristiques de bas ou de haut niveau [163].

Ce chapitre sera consacrée à une forme d'annotation des images : il s'agit de l'annotation sémantique collaborative. Pour décrire la sémantique des images, nous orientons nos travaux vers l'utilisation des ontologies parce qu'elles présentent des avantages considérables. Les ontologies fournissent un cadre formel qui peut contenir des définitions formelles [116] et [117]. Elles peuvent être directement traitées par une machine et peuvent permettre en même temps d'extraire des connaissances implicites par inférence automatique (voir [61] et [117]).

Rappelons qu'une ontologie est une spécification formelle et explicite d'une conceptualisation partagée. Ainsi, une ontologie permet de capturer la connaissance pertinente d'un domaine, de donner une compréhension commune de cette connaissance du domaine, de déterminer le vocabulaire reconnu de ce domaine, et de donner la définition explicite du vocabulaire (termes) et les relations entre ces vocabulaires dans des modèles formels à différents niveaux.

Dans notre démarche, plusieurs sémantiques peuvent être attribuées à une image. Les sémantiques réelles seront les sémantiques effectivement acceptées par les annotateurs. Pour composer une telle sémantique, il faut notamment des processus d'organisation et de sélection des informations les plus pertinentes.

Mais en réalité, c'est quoi une annotation sémantique collaborative? La réponse à cette question nous amène à faire la lumière d'une part sur ce qu'est l'annotation sémantique, et d'autre part sur ce qu'est la collaboration.

4.2/ DÉFINITIONS ET CONCEPTS DE BASE

Cette section a pour objectif de définir les concepts de base liés à l'approche d'annotation sémantique collaborative que nous proposons. Nous définissons dans un premier temps ce qu'est l'annotation sémantique et dans un deuxième temps la notion de la collaboration.

4.2.1/ ANNOTATION SÉMANTIQUE

Amardeih et al. dans [164] définissent l'annotation sémantique, comme une représentation formelle d'un contenu, exprimée à l'aide des concepts, des relations et des instances décrits dans une ontologie, et reliée à la ressource documentaire. Elle se réfère à l'interprétation des images à partir de la perception humaine. Partant de ce constat, nous pouvons dire que la sémantique d'une image consiste à représenter non seulement des objets qu'elle contient, mais aussi les relations qui existent entre ces différents objets (les scènes d'image). Deux enjeux majeurs sont à considérer : la prise en compte du contenu mais également du contexte. Cela pose donc un problème capital. Le problème de la traduction de ce que l'œil humain perçoit dans un langage compréhensible par la machine. Il s'agit donc de l'extraction d'une sémantique

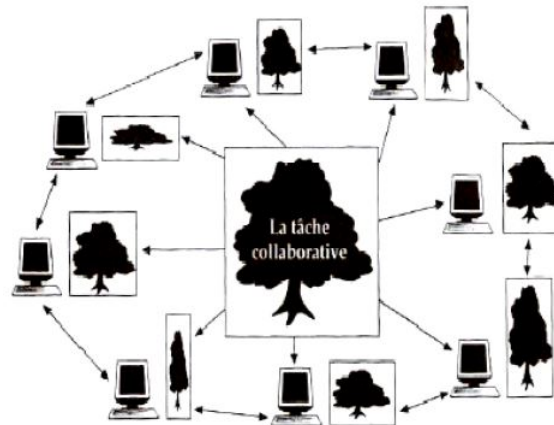


FIGURE 4.1 – Le travail collaboratif (Source :[165])

simple ou multiple de l'image entière, ou l'extraction de la sémantique des différents objets d'une image. Une telle analyse doit intégrer une notion de syntaxe qui est considérée comme régissant le domaine d'intérêt de sorte que l'explication la plus probable des objets observés puisse être trouvée.

L'idée générale est que la reconnaissance d'un objet ou d'un événement exige que l'on rapporte les représentations symboliques vaguement définies par des concepts à des cas concrets de l'objet référencé. Cela est mieux abordé dans une structure hiérarchique en associant différentes parties de l'image à chaque niveau de la hiérarchie selon des règles bien précises. Ainsi les règles syntaxiques peuvent être utilisées pour conduire à la reconnaissance des objets ou des événements, basée sur la détection des composants individuels correspondant aux caractéristiques détectées dans le temps et l'espace. L'analyse visuelle revient alors à l'analyse d'un flux d'objets présents dans l'image pour en trouver l'interprétation la plus probable afin d'établir une correspondance entre descriptions symboliques et numériques de l'information. Partant de ce fait, l'utilisation des ontologies pour l'annotation des images est devenue une voie très explorée comme nous venons de le montrer dans les sections précédentes. Les images ne seront donc plus représentées par un ensemble de mots-clés, mais par leur sémantique. Une différence importante entre l'utilisation des mots-clés et l'utilisation d'une ontologie est qu'avec les mots-clés, l'information visuelle est intrinsèquement plus ambiguë et sémantiquement pauvre. Les règles qui sous-tendent la définition des concepts tels que les scènes et les événements observés sont encore difficiles à interpréter avec les mots-clés comme le langage naturel par un système informatique.

4.2.2/ LA COLLABORATION

La distinction entre le travail coopératif et le travail collaboratif peut s'effectuer en différenciant les relations existantes entre les membres du groupe (obligation ou liberté), la responsabilité engagée ou non de chacun par rapport aux actions (responsabilité déléguée au coordinateur ou constamment partagée), la capacité de chacun à influencer sur la définition et la succession

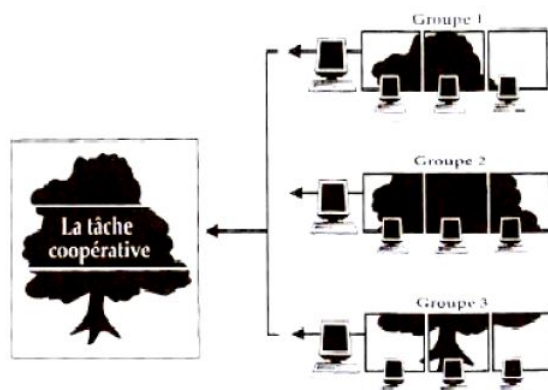


FIGURE 4.2 – Le travail coopératif (Source :[165])

des actions permettant d'atteindre l'objectif assigné au groupe.

Le travail collaboratif ne relève pas d'une répartition a priori des rôles. La collaboration s'entend en fait par une situation de travail collectif où tâches et buts sont communs. Chacun des membres du groupe travaille ainsi sur les mêmes points. Il va en effet plutôt s'agir ici de fusionner les contributions individuelles dans l'action. La responsabilité est donc ici collective et incombe au groupe en tant que totalité. Ce mode de travail collectif engage par conséquent une communication régulière entre les membres du groupe et une connaissance précise de la progression de l'action collective [166]. Les interactions interpersonnelles sont donc permanentes afin d'assurer une cohérence globale, condition nécessaire de l'efficacité de l'action et donc de l'atteinte de l'objectif fixé.

Le travail coopératif, par contre, peut être défini comme une organisation collective du travail dans laquelle la tâche à satisfaire est fragmentée en sous-tâches. Il est un travail de groupe hiérarchiquement organisé et planifié impliquant des délais et un partage des tâches selon une coordination précise. Chaque intervenant sait ainsi ce qu'il doit faire dès le début et communique, échange ou partage des éléments uniquement pour arriver à son objectif individuel. A la fin, le travail de chacun est réuni pour créer un objet unique de travail. En d'autres termes, c'est la succession progressive et coordonnée des actions de chacun qui aboutit au résultat escompté.

En résumé, nous pouvons dire que le travail coopératif permet de partitionner un travail W en w_1, w_2, \dots, w_n et d'attribuer chaque w_i à chaque participant tandis que le travail collaboratif consiste à soumettre la même tâche W à chaque participant avec les mêmes consignes. Ceux-ci appliquent les mêmes consignes à la tâche qui leur est confiée pour obtenir le résultat escompté. Le travail collaboratif est la méthode que nous utiliserons. Puisque les annotateurs ont chacun une interprétation des images et qu'il y a une forte chance qu'une image ne soit pas interprétée de la même façon, il est indispensable que les images soient soumises à un groupe d'annotateurs avec tous les moyens nécessaires. Les annotateurs auront donc pour mission d'attribuer des sémantiques aux images qui leur sont confiées tout en décrivant intégralement ce qu'ils perçoivent, c'est-à-dire les objets présents dans l'image, les relations entre les objets,

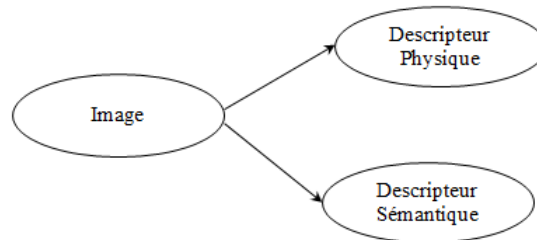


FIGURE 4.3 – Représentation des images

l'évènement décrits par l'image, etc.

4.3/ REPRÉSENTATION DES IMAGES DANS *Emergsem*

Cette section a pour objectif de présenter dans un premier temps la représentation des images à travers les descripteurs physiques et sémantiques. Dans un deuxième temps, nous allons nous focaliser sur la description sémantique des images basée sur une ontologie générique, une ontologie d'application dérivée de l'ontologie générique et un dictionnaire lexical pour attribuer des mots-clés aux concepts de l'ontologie. Dans un troisième temps, nous allons détailler le processus d'instanciation de l'ontologie.

4.3.1/ DESCRIPTEURS D'IMAGES

Une image peut être représentée par les descripteurs physiques ou par les descripteur sémantiques comme le montre la figure 4.3.

Plusieurs techniques utilisant le *Descripteur Physique* se basent généralement sur les caractéristiques de bas niveau des images telles que la couleur, la forme, la texture, etc. Ces techniques permettent de décrire les régions d'image manuellement, automatiquement ou de façon semi-automatique comme nous l'avons présenté dans l'état de l'art.

Le *Descripteur Sémantique* quant à lui permet de fournir une description identique à la perception humaine afin de faciliter la recherche des images. Il permet de décrire les images au niveau sémantique, c'est-à-dire au niveau de jugement que l'humain porte en regardant les images dans un but bien précis : la réduction du fossé sémantique. Rappelons que le fossé sémantique est l'écart qu'il y a entre l'interprétation du point de vue des humains (le haut niveau) et la description des machines (le bas niveau). Nous en avons pour preuve qu'à notre connaissance, nous voyons qu'une tache verte observée dans une image correspond à de la végétation, ce que l'ordinateur ne peut savoir. Pour mener à bien la description de la sémantique d'une image, il faut rechercher des moyens pour connecter (lier) la connaissance sémantique à l'image. Pour atteindre ces objectifs, deux techniques de description sémantique des images sont proposées : la description globale et la description locale. La description sémantique globale consiste à proposer une sémantique de l'image entière sans détailler de façon implicite les différents objets qu'elles contiennent comme par exemple une image de sport, une image de manifestation ou encore une image de plage. Les expériences ont montré l'inefficacité de la sémantique globale parce qu'elle ressemble beaucoup à la description des images avec des

mots-clés. Contrairement à la description sémantique globale, la description sémantique locale consiste à décrire la sémantique des images grâce à la représentation des objets qu'elles contiennent et aussi des relations qui existent entre ces objets. C'est une technique un peu plus complexe que la première mais les résultats sont bien meilleurs. Qui dit décrire les objets et les relations entre les objets, dit qu'il faut représenter les objets et les relations dans un langage compréhensible par la machine. Pour atteindre ces objectifs, plusieurs outils sont proposés dans la littérature pour représenter la sémantique des images. Parmi ceux-ci, nous pouvons citer les outils du Web Sémantique et notamment les ontologies. L'utilisation des ontologies dans le domaine de la récupération d'images est généralement axée sur plusieurs objectifs parmi lesquels nous pouvons citer :

- Ontologie pour la description visuelle : où les ontologies sont utilisées pour représenter les différents types de relations entre les caractéristiques de l'image tels que les régions [118] ;
- Ontologie pour la description des connaissances : les ontologies sont utilisées pour modéliser les concepts (objets) et les relations entre eux. Typiquement, ce sont les approches qui utilisent un raisonnement sur les concepts ou des informations contextuelles. Ces approches abordent le plus souvent le problème de l'interprétation des images [138] ;
- Ontologie pour le mapping sémantique : les ontologies sont utilisées pour aider à la mise en correspondance entre les caractéristiques visuelles et le niveau sémantique.

Tout ceci contribue généralement à améliorer la recherche d'images en y intégrant les connaissances de base. La connaissance intégrée peut aider dans diverses tâches du processus de recherche d'images : allant de l'analyse d'image, à l'enrichissement d'annotation, en passant par le mapping des caractéristiques visuelles dans les concepts sémantique.

4.3.2/ ONTOLOGIE GÉNÉRIQUE

Une ontologie est une spécification qui rend compte d'une conceptualisation. Ce qui veut signifier que l'on y définit des concepts et des relations pour que l'ontologie soit par la suite utilisée dans un artefact informatique dont on veut spécifier le comportement.

Nous proposons une ontologie générique pour la description sémantique des images. Comme le montre la figure 4.4, l'ontologie est constituée d'un méta concept abstrait *Image* et d'un descripteur sémantique relié au méta concept abstrait.

4.3.2.1/ MÉTA CONCEPT ABSTRAIT *Image*

Le méta concept abstrait *Image* est un concept de très haut niveau qui contient tous les concepts pouvant décrire les images quelque soit leurs domaines, comme par exemple les concepts décrivant les images dans le domaine médical, dans le domaine aéronautique, etc. Il est en quelque sorte l'ancêtre de tous les concepts.

4.3.2.2/ DESCRIPTEUR SÉMANTIQUE

Le descripteur sémantique contient un ou plusieurs concepts abstraits, un ou plusieurs concepts concrets et des relations. Ceci se traduit par le fait que le ou les concepts proposés sont des entités du méta concept abstrait *Image*. Le but du descripteur sémantique est de permettre la définition d'autres concepts (abstraites et concrets) et des relations entre ces

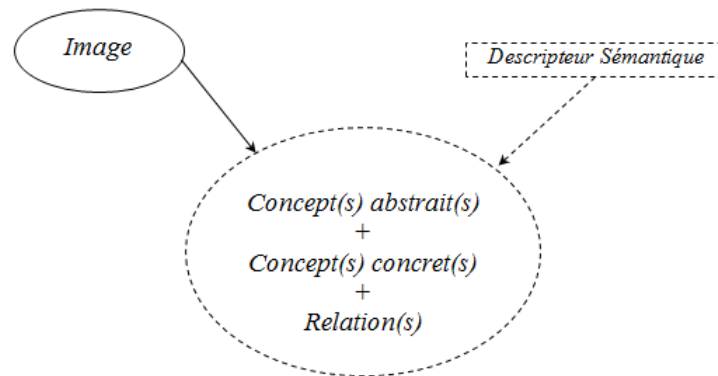


FIGURE 4.4 – L'ontologie générique

concepts.

- Un concept abstrait est un concept qui ne peut pas être instancié, c'est-à-dire un concept auquel on ne peut affecter aucune instance décrivant un objet réel, mais qui peut contenir d'autres concepts abstraits ou concrets capables de décrire les images. C'est en quelque sorte des concepts *génériques*. Prenons un exemple palpable : lorsqu'on nous demande de donner le nom d'un engin, la première question que nous posons est : "Quel type d'engin?". Cette question cache plusieurs réponses. S'agit-il d'un engin maritime, spatial, terrestre, etc. Et s'il s'agissait d'un engin terrestre, une nouvelle question se pose : "Est-ce une voiture ou une moto?" Nous comprenons donc qu'il serait un peu "vague" de répondre simplement par "Citroen DS3" à la question qui est posée sans demander le type d'engin. "Engin" est alors un concept abstrait qui contient d'autres concepts abstraits "Engin spatial", "Engin maritime" et "Engin terrestre". "Engin terrestre" contient à son tour d'autres concepts "Voiture" et "Moto".
- A la différence du concept abstrait qui ne peut pas être instancié, le concept concret est un concept qui peut recevoir des instances qui décrivent les objets. C'est un concept relié à un concept abstrait qui permet de décrire de façon précise un objet réel. Exemple : "Citroen DS3" est une "Voiture". "Voiture" est un concept concret relié au concept abstrait "Engin terrestre".
- Une relation est un mot ou groupe de mots qui précise le lien qui existe entre deux concepts. Elle se définit par les concepts qu'elle relie. Ces concepts constituent la signature sémantique de la relation. Une relation se définit en outre par un contenu sémantique intrinsèque articulant les deux concepts. Les relations traduisent les associations (pertinentes) existant entre les concepts. Ces relations incluent les associations suivantes : *sous-classe* de (généralisation-spécialisation), *partie de* (agrégation ou composition); *Associé à*, *Instance de*, etc. Les relations (aussi connues comme les propriétés) entre les objets dans une ontologie précisent comment les objets sont liés entre eux. Elles peuvent également inclure les relations spatiales comme par exemple : à côté de, touche, éloigné de, disjoint, etc. entre différence entre les

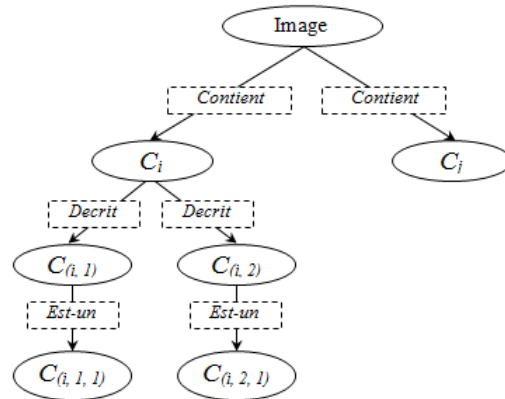


FIGURE 4.5 – Concepts et relations de l'ontologie générique

objets. Les relations spatiales entre les objets d'une image sont très importantes. La disposition spatiale des objets fournit des informations pour les tâches de reconnaissance et d'interprétation, en particulier lorsque les objets sont intégrés dans un environnement complexe. Les êtres humains font un usage intensif des relations spatiales pour décrire, détecter et reconnaître des objets. Elles permettent de résoudre l'ambiguïté entre les objets ayant une apparence similaire. Prenons encore notre exemple à propos des engins. Dans le paragraphe décrivant les concepts concrets, nous avons dit que *Voiture Est un Engin terrestre*. Le groupe de mots "Est un" représente une relation qui relie donc le concept concret *Voiture* au concept abstrait *Engin terrestre*. Trois relations de base ont été proposées dans l'ontologie générique : *Contient*, *Decrit* et *Est – un*. La première relation *Contient* est une relation qui relie les concepts abstraits de base au méta concept abstrait *Image* parce que *Image* est le concept abstrait de haut niveau contient tous les concepts pouvant décrire les images. Tous les concepts abstraits de base seront systématiquement reliés à l'ancêtre *Image* par le relation *Contient*. La deuxième relation *Decrit* relie un concept abstrait de base à un sous concept abstrait. Cette relation est aussi valable pour deux concepts abstraits (un des deux concepts étant le *fil*s de l'autre) même si aucun d'eux n'est un concept de base. La troisième relation *Est – un* quant à elle relie un concept abstrait à un concept concret. Tous les concepts concrets seront donc reliés à un concept abstrait par cette relation.

Soient deux concepts abstraits de base C_i et C_j représentant respectivement deux entités globales des domaines i et j , $C_{i,1}$ et $C_{i,2}$ deux sous-concepts abstraits de C_i et $C_{i,1,1}$, $C_{i,2,1}$ respectivement deux concepts concrets de $C_{i,1}$ et de $C_{i,2}$. Tous les concepts abstraits de base seront reliés au méta concept abstrait par la relation *Contient*, les sous-concepts des concepts abstraits de base seront reliés aux concepts abstraits de base par la relation *Decrit* et les concepts concrets seront reliés aux concepts abstraits par la relation *Est – un* comme le montre la figure 4.5.

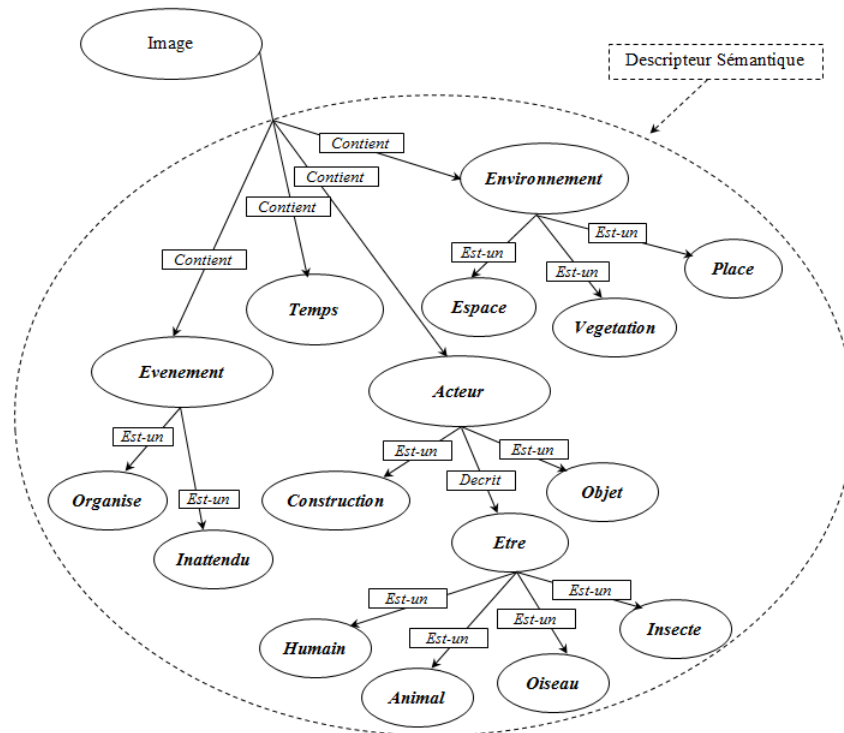


FIGURE 4.6 – Ontologie d'application proposée

4.3.3/ ONTOLOGIE D'APPLICATION

Une ontologie d'application est une ontologie personnalisée pour décrire les images d'un domaine précis. Elle consiste à définir (nommer) les concepts de l'ontologie générique afin de l'adapter à un besoin spécifique. On peut donc créer autant de concepts abstraits de base qui peuvent à leur tour contenir un ou plusieurs concepts (abstrait et concrets). L'ontologie générique a donc pour vocation de générer autant d'ontologies d'application, c'est-à-dire des ontologies capables de décrire toute sorte d'images dans n'importe quel domaine comme par exemple le domaine de la médecine, de l'armée, de l'agriculture, etc.

En nous basant sur l'ontologie générique, nous avons proposé une ontologie d'application capable de décrire les images. Nous avons proposé quatre types de concepts de base : un concept temporaire, un concept événementiel, un concept environnemental et un concept décrivant les acteurs.

- Le concept événementiel permet de comprendre et de décrire les événements véhiculés dans une image comme par exemple les événements organisés (la célébration d'un mariage, un pique-nique, etc.) ou des événements imprévus (un accident, une tempête, etc.)
- Le concept environnemental qui permet de décrire l'environnement dans lequel les scènes d'images se produisent. Il peut s'agir de l'emplacement où se déroule une scène (une place, une végétation ou un espace).
- Le concept représentant les acteurs est un concept qui décrit tous les acteurs qui interviennent

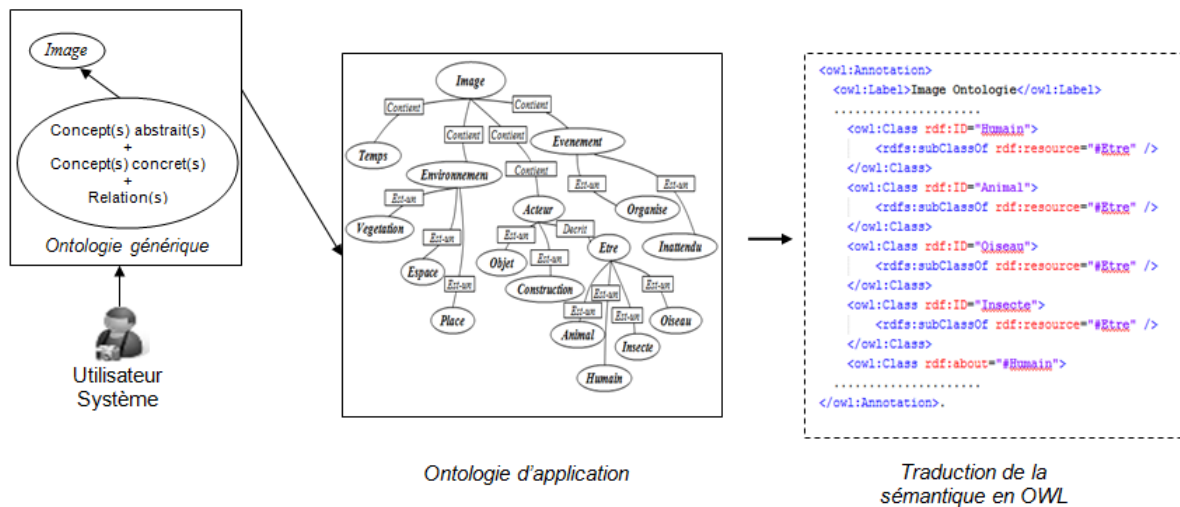


FIGURE 4.7 – Ontologie d'application et sa traduction en OWL

dans une image. Nous pouvons citer les êtres (les humains, les animaux, les oiseaux et les insectes), les objets (les véhicules, les matériels, les outils de travail, etc.) et les constructions (toutes les infrastructures immobilières, routières, etc.).

- Le concept temporaire est un concept qui décrit le moment du déroulement des scènes d'image (le matin, la nuit, etc.).

La figure 4.6 est une personnalisation *possible* de l'ontologie générique. L'exemple que nous proposons contient quatre concepts : *Temps*, *Environnement*, *Evènement* et *Acteur*. Les quatre concepts sont reliés au concept ancêtre *Image* par la relation *Contient*. Ceci indique que ces classes sont contenues dans l'ancêtre *Image*.

4.3.4/ DESCRIPTION DE L'ONTOLOGIE D'APPLICATION

La figure 4.7 présente le modèle de l'ontologie d'application générée à partir de l'ontologie générique et l'extrait de sa traduction en OWL. Dans cette section, nous allons procéder à la description des concepts et des relations de l'ontologie d'application.

4.3.4.1/ LES CONCEPTS

Les concepts correspondent aux abstractions pertinentes d'un segment de la réalité (le domaine du problème) retenus en fonction des objectifs qu'on se donne et de l'application envisagée pour l'ontologie.

- La classe *Acteur* : le concept *Acteur* est un concept qui décrit les acteurs qui interviennent dans les images. Elle est un sous-concept du méta concept abstrait *Image*. Le concept *Acteur* a des sous-concepts *Etre*, *Construction* et *Objet* qui décrivent respectivement tous les êtres vivants, tous les ouvrages humains ainsi que tous les objets sans exception (voir 4.1).
- La classe *Objet* : le concept *Objet* est un concept qui décrit les objets qui apparaissent dans les images (les engins, les matériaux, les ordinateurs, etc.). Il est un sous-concept du

<i>Composantes</i>	<i>Concept Auteur</i>
<i>Description</i>	<i>Acteur</i> est un concept qui permet de décrire tous les acteurs qui interviennent dans l'image. Acteur est un sous-concept de l'ancêtre Image
<i>Sous – concepts</i>	Objet Etre Construction
<i>Commentaire</i>	Acteur est un concept abstrait

TABLE 4.1 – Descripteur du concept *Acteur*

concept Acteur (voir 4.2).

<i>Composantes</i>	<i>Concept Objet</i>
<i>Description</i>	<i>Objet</i> est un concept qui permet de décrire les objets qui apparaissent dans les images. Objet est un sous-concept du concept Acteur
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Objet est un concept concret

TABLE 4.2 – Descripteur du concept *Objet*

- Le concept *Construction* : le concept *Construction* est un concept qui décrit les infrastructures qui apparaissent dans les images (les maisons, les tours, les ponts, les musées, les monuments, etc.). Il est un sous-concept du concept Acteur (voir 4.3).

<i>Composantes</i>	<i>Concept Construction</i>
<i>Description</i>	<i>Construction</i> est un concept qui permet de décrire toutes les infrastructures qui apparaissent dans les images. Construction est un sous-concept du concept Acteur
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Construction est un concept concret

TABLE 4.3 – Descripteur du concept *Construction*

- Le concept *Etre* : le concept *Etre* est un concept qui décrit les êtres qui apparaissent dans les images. Il est un sous-concept du concept Acteur (voir 4.4).
- Le concept *Humain* : le concept *Humain* est un concept qui décrit les humains qui apparaissent dans les images. Grâce à ce concept, non seulement les hommes peuvent être représentés individuellement ainsi que les groupes d'hommes et les couple d'hommes, mais aussi les différentes parties du corps humain tel que le visage, la tête et les mains d'homme. Elle est un sous-concept du concept *Etre* (voir 4.5).
- Le concept *Animal* : le concept *Animal* permet de représenter tous les animaux sans exception. Le concept *Animal* comme la plus part des autres concepts, peut être qualifié de

<i>Composantes</i>	<i>Concept Etre</i>
<i>Description</i>	<i>Etre</i> est un concept qui permet de décrire tous les être qui apparaissent dans une image. C'est un sous-concept du concept Acteur de l'ancêtre Image
<i>Sous – concepts</i>	Humain, Animal, Oiseau, Insecte
<i>Commentaire</i>	Etre est un concept abstrait

TABLE 4.4 – Descripteur du concept *Etre*

<i>Composantes</i>	<i>Concept Humain</i>
<i>Description</i>	<i>Humain</i> est un concept qui permet de décrire tous les être humains qui apparaissent dans les images. Humain est un sous-concept du concept Etre
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Humain est un concept concret

TABLE 4.5 – Descripteur du concept *Humain*

concepts générique des animaux parce qu'il prend en compte la description de toute sorte d'animaux (marins, terrestres, etc.) qui apparaissent dans les images. Il est un sous-concept du concept Etre (voir 4.6).

<i>Composantes</i>	<i>Concept Animal</i>
<i>Description</i>	<i>Animal</i> est un concept qui permet de décrire tous les animaux qu'on peut retrouver dans les images. Animal est un sous-concept du concept Etre
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Animal est un concept concret

TABLE 4.6 – Descripteur du concept *Animal*

- Le concept *Oiseau* : le concept *Oiseau* est un concept qui décrit les oiseaux qui apparaissent dans les images. Il est un sous-concept du concept *Etre* (voir 4.7).

<i>Composantes</i>	<i>Concept Oiseau</i>
<i>Description</i>	<i>Oiseau</i> est un concept qui permet de décrire tous les oiseaux qu'on peut retrouver dans les images. Oiseau est un sous-concept du concept Etre
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Oiseau est un concept concret

TABLE 4.7 – Descripteur du concept *Oiseau*

- Le concept *Insecte* : le concept *Insecte* est un concept qui décrit les insectes qui apparaissent dans les images. Il est un sous-concept du concept *Etre* (voir 4.8).

<i>Composantes</i>	<i>Concept Insecte</i>
<i>Description</i>	<i>Insecte</i> est un concept qui permet de décrire tous les insectes qu'on peut retrouver dans les images. <i>Insecte</i> est un sous-concept du concept <i>Etre</i>
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	<i>Insecte</i> est un concept concret

TABLE 4.8 – Descripteur du concept *Insecte*

- Le concept *Environnement* : le concept *Environnement* est un concept qui décrit l'environnement où se déroulent les scènes d'images. Il est un sous-concept du méta concept abstrait *Image* (voir 4.9).

<i>Composantes</i>	<i>Concept Environnement</i>
<i>Description</i>	<i>Environnement</i> est un concept qui permet de décrire l'environnement dans lequel les acteurs de l'image interviennent. <i>Environnement</i> est un sous-concept de l'ancêtre <i>Image</i>
<i>Sous – concepts</i>	Vegetation, Espace, Place
<i>Commentaire</i>	<i>Environnement</i> est un concept abstrait

TABLE 4.9 – Descripteur du concept *Environnement*

- Le concept *Temps* : le concept *Temps* est un concept qui décrit le moment où se déroulent les scènes d'images (le matin, le soir, la nuit, etc.). Il est un sous-concept du méta concept abstrait *Image* (voir 4.10).

<i>Composantes</i>	<i>Concept Temps</i>
<i>Description</i>	<i>Temps</i> est un concept qui permet de décrire la période au cours duquel les acteurs interviennent dans une image. <i>Temps</i> est un sous-concept de l'ancêtre <i>Image</i>
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	<i>Temps</i> est un concept concret

TABLE 4.10 – Descripteur du concept *Temps*

- Le concept *Espace* : le concept *Espace* est un concept qui décrit l'espace (les nuages, les astres, le ciel, etc.). Il est un sous-concept du concept *Environnement* (voir 4.11).
- Le concept *Vegetation* : le concept *Vegetation* est un concept qui décrit les végétations qui apparaissent dans les images (une prairie, un jardin, etc.). Il est un sous-concept du concept *Environnement* (voir 4.12).
- Le concept *Place* : le concept *Place* est un concept qui décrit les emplacements qui apparaissent dans les images (un parking, un square, une place publique, etc.). Il est un sous-concept du concept *Environnement* (voir 4.13).
- Le concept *Evenement* : le concept *Evenement* est un concept qui décrit les événements

<i>Composantes</i>	<i>Concept Espace</i>
<i>Description</i>	<i>Espace</i> est un concept qui permet de décrire l'espace dans lequel les acteurs de l'image interviennent. Espace est un sous-concept d'Environnement
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Espace est un concept concret

TABLE 4.11 – Descripteur du concept *Espace*

<i>Composantes</i>	<i>Concept Vegetation</i>
<i>Description</i>	<i>Vegetation</i> est un concept qui permet de décrire la végétation. C'est un sous-concept du concept Environnement
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Vegetation est un concept concret

TABLE 4.12 – Descripteur du concept *Vegetation*

<i>Composantes</i>	<i>Concept Place</i>
<i>Description</i>	<i>Place</i> est un concept qui permet de décrire les emplacements. C'est un sous-concept du concept Environnement
<i>Sous – concepts</i>	Pas de sous-concept
<i>Commentaire</i>	Place est un concept concret

TABLE 4.13 – Descripteur du concept *Place*

qui se déroulent dans les images. Il est un sous-concept du méta concept abstrait *Image* (voir 4.14).

<i>Composantes</i>	<i>Concept Evenement</i>
<i>Description</i>	<i>Evenement</i> est un concept qui permet de décrire les événements. C'est un sous-concept de l'ancêtre <i>Image</i>
<i>Sous – concepts</i>	Organise, Inattendu
<i>Commentaire</i>	Evenement est un concept abstrait

TABLE 4.14 – Descripteur du concept *Evenement*

- Le concept *Organise* : le concept *Organise* est un concept qui décrit les événements organisés qui apparaissent dans les images (Une promenade, un baptême, un mariage, etc.). Il est un sous-concept du concept *Evenement* (voir 4.15).
- Le concept *Inattendu* : le concept *Inattendu* est un concept qui décrit les événements inattendus qui apparaissent dans les images. Il peut s'agir de tempête, d'incendie, d'accident, d'un séisme, etc. Il est un sous-concept du concept *Evenement* (voir 4.16).

<i>Composantes</i>	<i>Concept Organise</i>
<i>Description</i>	<i>Organise</i> est un concept qui permet de décrire les évènements organisés. C'est un sous-concept du concept Evènement
<i>Sous – concepts</i>	Pas de sous concept
<i>Commentaire</i>	Organise est un concept concret

TABLE 4.15 – Descripteur du concept *Organise*

<i>Composantes</i>	<i>Concept Inattendu</i>
<i>Description</i>	<i>Inattendu</i> est un concept qui permet de décrire les évènements inattendu. C'est un sous-concept du concept Evènement
<i>Sous – concepts</i>	Pas de sous concept
<i>Commentaire</i>	Inattendu est un concept concret

TABLE 4.16 – Descripteur du concept *Inattendu*

4.3.4.2/ LES RELATIONS HIÉRARCHIQUES

Les tableaux présentés ci-dessous décrivent les relations (propriétés) de l'ontologie d'application.

- La propriété *Contient* : *Contient* relie *Image* aux classes *Environnement*, *Evènement*, *Acteur* et *Temps*. La formalisation de cette propriété est présentée dans le tableau 4.17.

<i>Composantes</i>	<i>Relation Contient</i>
<i>Description</i>	<i>Contient</i> est une relation qui relie <i>Image</i> aux concepts de base : <i>Image Contient Acteur</i> , <i>Image Contient Evènement</i> , <i>Image Contient Environnement</i> , <i>Image Contient Temps</i>
<i>Inverse de la relation</i>	<i>Est – contenu – dans</i> permet de lire la relation dans le sens opposé (<i>Acteur Est – contenu – dans Image</i>)
<i>Commentaire</i>	<i>Contient</i> est une relation qui relie l'ancêtre <i>Image</i> aux principaux concepts abstraits (concepts de base)

TABLE 4.17 – Descripteur de la relation *Contient*

- La propriété *Décrit* : *Décrit* relie *Acteur* à la classe *Etre*. La formalisation de cette propriété est présentée dans le tableau 4.18.

<i>Composantes</i>	<i>Relation Décrit</i>
<i>Description</i>	<i>Décrit</i> est une relation qui relie les concepts abstraits entre eux : Exemple : <i>Acteur</i> (concept abstrait) <i>Décrit</i> un <i>Etre</i> (concept Abstrait)
<i>Inverse de la relation</i>	<i>Est – décrit – dans</i> permet de lire la relation dans le sens opposé Exemple : <i>Etre Est – décrit – dans Acteur</i>
<i>Commentaire</i>	<i>Décrit</i> est une relation qui relie les concepts abstraits entre eux

TABLE 4.18 – Descripteur de la relation *Décrit*

- La propriété *Est-un* : *EstUn* relie *Etre* aux classes *Humain*, *Animal*, *Oiseau* et *Insecte*. La formalisation de cette propriété est présentée dans le tableau 4.19. Dans cet exemple, nous avons décrit deux relations (entre *Etre* et *Humain* et entre *Etre* et *Animal*).

<i>Composantes</i>	<i>Relation Est – un</i>
<i>Description</i>	<i>Est – un</i> est une relation qui relie les concepts abstraits aux concepts concrets. Exemple : <i>Evenement Est – un Organise</i> , <i>Evenement Est – un Inattendu Acteur Est – un Construction</i> , <i>Acteur Est – un Objets</i> , <i>Environnement Est – un Vegetation</i> , <i>Environnement Est – un Place</i> , <i>Environnement Est – un Espace</i> , <i>Etre Est – un Humain</i> , <i>Etre Est – un Animal</i> , <i>Etre Est – un Oiseau</i> , <i>Etre Est – un Insecte</i>
<i>Inverse de la relation</i>	<i>Est – un</i> permet de lire la relation dans le sens opposé. C'est une relation équivalente à la première relation : <i>Humain Est – un Etre</i>
<i>Commentaire</i>	<i>Est-un</i> est une relation qui relie les concepts abstraits aux concepts concrets

TABLE 4.19 – Descripteur de la relation Est-un

4.3.4.3/ LES RELATIONS SPATIALES

Les relations spatiales sont des relations qui relient les concepts concrets entre eux. A la différence des relations hiérarchiques, les relations spatiales permettent de représenter les relations qui existent entre les objets d'une image comme le montre la figure 4.8.

Trois concepts sont représentés dans cette figure : Un concept abstrait *Acteur* et deux concepts concrets *Objet* et *Construction* qui sont des sous-concepts du concept abstrait *Acteur*. Les concepts concrets sont reliés par la relation spatiale *Est – garée – Dans* représentée par le rectangle coloré en gris.

Pour définir les relations spatiales, nous allons nous inspirer des précédents travaux sur la proposition des relations spatiales entre les objets présents dans une image. Dans ces travaux les auteurs ont examiné les relations spatiales reconnues par les utilisateurs lorsque les utilisateurs

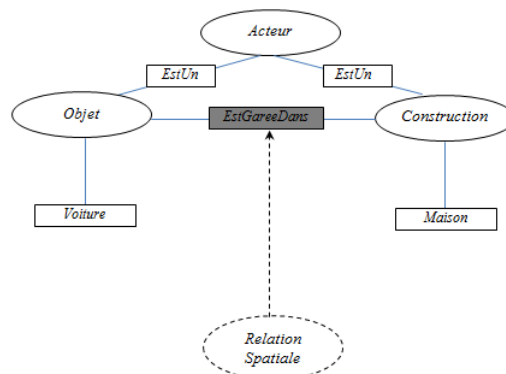


FIGURE 4.8 – Exemple d'une relation spatiale

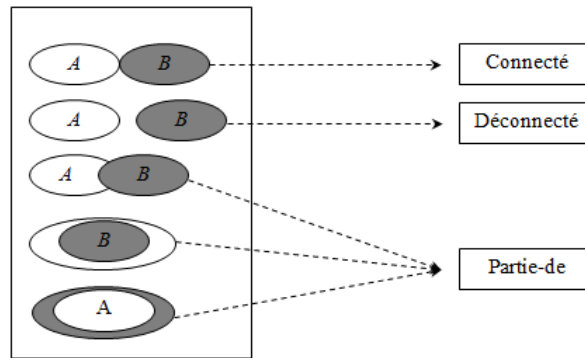


FIGURE 4.9 – Modèle des relations spatiales cognitives

regardent les images. Trois relations cognitives sont extraites : "Partie – de", "Connecté" et "Déconnecté". "Connecté" si les objets de l'image sont accolés, "Déconnecté" si les objets sont distants et "Partie – de" si les objets font partie d'autres objets ou s'ils s'entremêlent comme le montre la Figure 4.9.

La disposition des objets de toutes les images respecte ces trois relations spatiales. Dans chacune de ces trois relations, des verbes, des mots-clés et des groupes de mots-clés décrivant les relations seront enregistrés. Par exemple, "Connecté" pourrait contenir *toucher, tenir, saluer, Réparer, Sur, Dans* etc. *Partie – de* pourrait contenir *associer, composer – de*, etc. et la relation *Déconnecté* pourrait contenir *éloigner, rapprocher, loin de, enseigner, sous, Près de* etc. comme le montre la table 4.20.

Partie-de	Connecté	Déconnecté
<i>Composer – de, Insérer</i>	<i>Tenir, Saluer</i>	<i>Rapprocher, Distancer</i>
<i>Se – déplacer – dans</i>	<i>Sur, Dans</i>	<i>Eloigner de, Près de</i>
<i>Dans, Conduire</i>	<i>Soulever, Embrasser</i>	<i>Loin de, Proche de</i>
<i>Asseoir, Au milieu de</i>	<i>Scier, Réparer</i>	<i>A côté de, Au – dessus de</i>
<i>Nager dans</i>	<i>Acoller, Relier</i>	<i>Derrière, Devant</i>

TABLE 4.20 – Exemple de description d'une image

La figure 4.10 est un exemple de description des concepts et des relations spatiales. Elle présente une image que nous avons décrite avec les concepts (abstraites et concrets) et les relations (hiérarchiques et spatiales). L'inverse des relations peut être décrit. La section suivante présente le processus d'instanciation des concepts et des relations.

4.3.5/ INSTANCIATION DE L'ONTOLOGIE D'APPLICATION

L'instanciation de l'ontologie d'application consiste à instancier les concepts concrets et les relations spatiales. Les sous-sections suivantes présentent respectivement l'instanciation des concepts concrets et l'instanciation des relations spatiales.

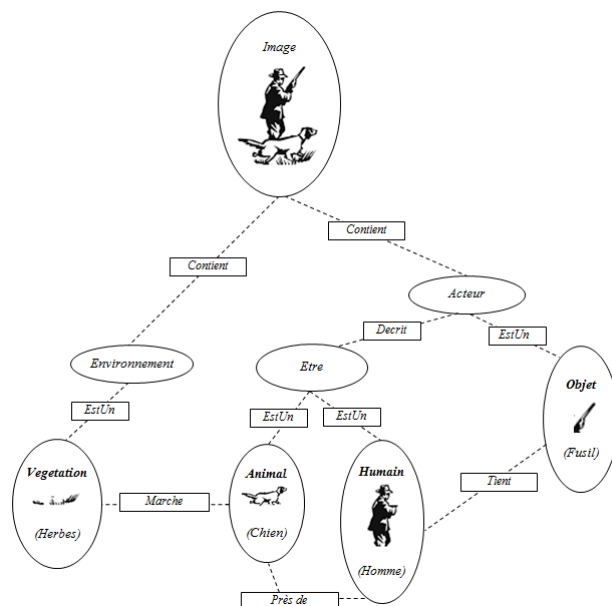


FIGURE 4.10 – Description d'une image

4.3.5.1/ INSTANCIATION DES CONCEPTS CONCRETS

L'annotation sémantique que nous proposons passe inévitablement par le processus d'instanciation de l'ontologie. L'instanciation des ontologies est un processus par lequel l'utilisateur attribue des instances (décrivant les régions d'images) aux concepts de l'ontologie via une interface comme le montre la figure 4.11. La proposition d'une interface d'instanciation a pour objectif de faciliter l'attribution des mots-clés.

Comme nous pouvons le constater sur cette image, nous avons des menus déroulants directement liés aux concepts de l'ontologie et permettant de choisir les instances qui représentent les différentes régions de l'image. Les utilisateurs proposent des mots-clés qui décrivent les régions de l'image qu'ils souhaitent annoter. Dans l'exemple de la figure 4.11, l'utilisateur a

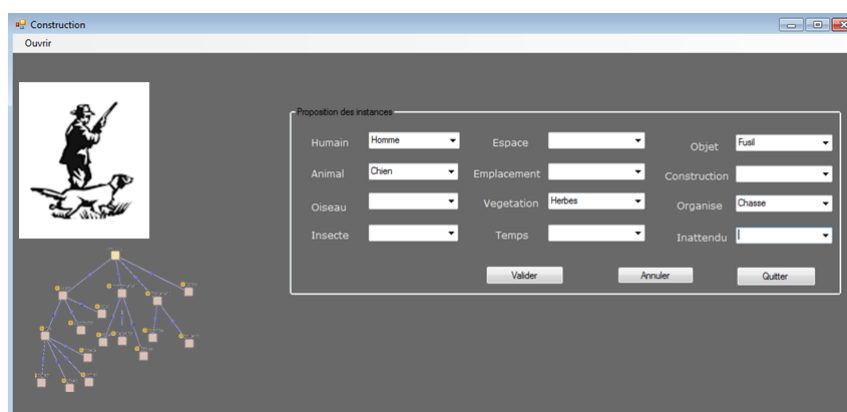


FIGURE 4.11 – Interface de proposition des instances

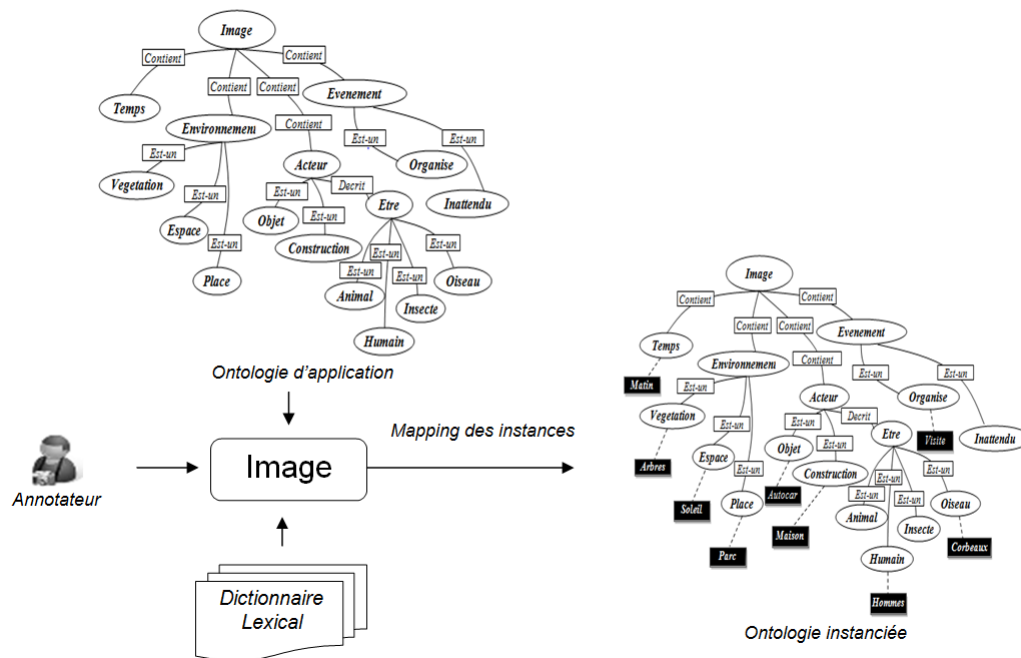


FIGURE 4.12 – Mapping des instances avec les concepts d'ontologie

proposé un certain nombre de mots-clés. Dans le concept *Vegetation*, l'utilisateur propose *Herbes*, dans le concept *Objet*, il choisit *Fusil* comme instance, etc. Seuls les concepts représentatifs des régions d'une l'image sont instanciés. Lorsque l'annotateur valide sa description, les instances sont sauvegardées dans l'ontologie, c'est le mapping des instances avec les concepts de l'ontologie.

4.3.5.2/ INSTANCIATION DES RELATIONS SPATIALES

L'instanciation des relations spatiales consiste à proposer entre deux concepts concrets un mot ou un groupe de mots qui décrit la relation qui existe entre eux. Entre deux concepts concrets de l'ontologie d'application, des trois relations spatiales cognitives *PartieDe*, *Connecté* et *Déconnecté*. Chacune de ces relations spatiales cognitives contient des verbes, des mots-clés et même des groupes de mots-clés qui décrivent les liens qui existent entre les objets d'une image.

Pour proposer une relation, les utilisateurs choisissent une relation spatiale cognitive qui convient au type de relations (*Partie – de*, *Connecté* ou *Déconnecté*) qu'ils souhaitent décrire. Une fois que la relation cognitive est choisie, une liste de mots-clés, de groupe de mots-clés ou de verbes lui sont proposés. Le tableau représenté sur la table 4.20 présente un exemple d'ensemble de mots-clés, de groupe de mots-clés ou de verbes mappés dans les trois relations spatiales cognitives. L'instanciation des relation spatiales se fait également via une interface comme le montre la figure 4.13.

La description fournie et stockée dans l'ontologie représente la sémantique de l'image qui servira à indexer l'image. La sémantique des images est enregistrée dans des fichiers de format xml.

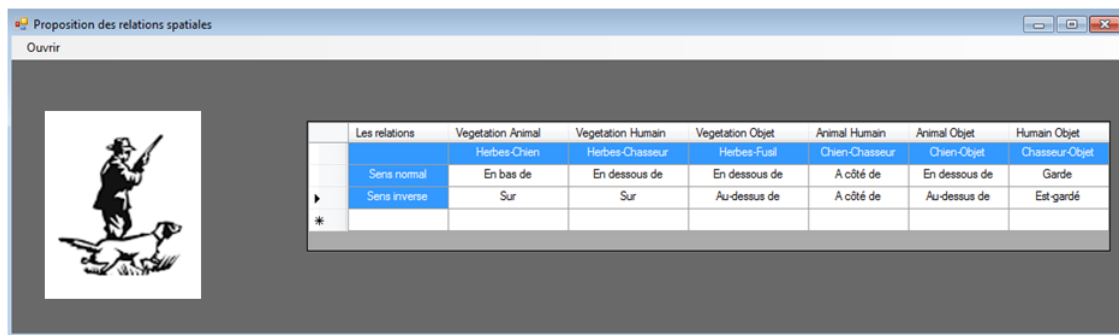


FIGURE 4.13 – Interface de proposition des relations spatiales

4.4/ ANNOTATION COLLABORATIVE

Cette section est consacrée à un processus d'annotation d'images basé sur la collaboration. Le processus consiste à faire annoter une image par une communauté d'annotateurs. Il consiste à l'instanciation des concepts concrets de l'ontologie d'application par des termes qui décrivent les régions de l'image. Dans l'approche que nous proposons, l'annotation sémantique ne sera pas faite par un seul annotateur, mais par une communauté d'annotateurs. La figure 4.15 représente le processus de l'annotation sémantique collaborative. Trois étapes se distinguent dans cette architecture : la formation de la communauté d'annotateurs, la génération des sémantiques et l'émergence des sémantiques.

4.4.1/ LA COMMUNAUTÉ D'ANNOTATEURS

Dans le processus d'annotation collaborative, la formation des communautés d'annotateurs (qui ont pour mission de produire les annotations) s'avère indispensable. Elle devient désormais possible grâce à Internet et au développement des sites collaboratifs comme mentionné dans [74]. Plusieurs études ont souligné l'importance de la formation des communautés et de leurs impacts sur les performances des systèmes collaboratifs. Une communauté met en relation un groupe d'individus qui possède des caractéristiques communes. Deux expressions se dégagent cette définition : *groupe d'individus* et *caractéristiques communes*. Ceci signifie qu'avant

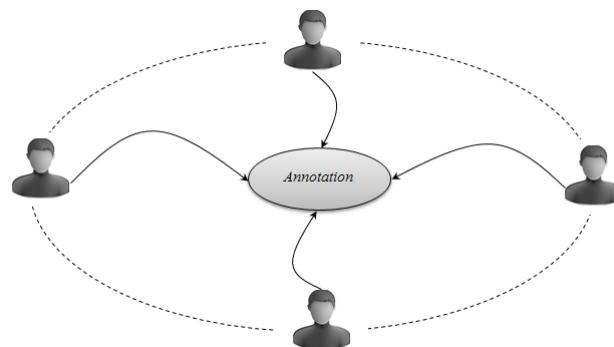


FIGURE 4.14 – Communauté d'annotateurs

}, où chaque annotateur peut être considéré comme un nœud unique dans un réseau social. Elle permet de déterminer non seulement les interactions directes entre les différents acteurs mais aussi les contributions apportées par chacun d'eux à une tâche bien précise.

Pour mener à bien le processus d'annotation collaborative des images, un dictionnaire lexical est mis à la disposition de la communauté d'annotateurs. Ce dictionnaire permet aux annotateurs de la communauté de proposer de façon collaborative des mots-clés décrivant les régions d'images qu'ils souhaiteraient annoter.

4.4.2/ DICTIONNAIRE LEXICAL

Un dictionnaire lexical est un vocabulaire contrôlé dans lequel les mots sont catégorisés. Les dictionnaires lexicaux aussi appelés *thésauri*¹ contiennent que des termes, sous la forme de noms communs ou de groupes nominaux. Les termes figurant dans un dictionnaire lexical sont censés être représentatifs de certains domaines et les termes sont liés entre eux par des relations. On considère classiquement trois types de relations possibles :

1.) La relation "*générique/spécifique*" est une relation d'ordre : elle fournit la structure globale du dictionnaire mais ne le contraint pas pour autant à être représentable sous la forme d'un unique arbre : il n'y a pas nécessairement unicité du terme le plus général ;

2.) La relation de "*synonymie*" est symétrique et relie deux termes pouvant être échangés l'un avec l'autre ;

3.) La relation "*terme associé*" associe deux termes du même *champ sémantique* sans être pour autant synonymes : elle sert à étendre une requête à des termes sémantiquement proches.

L'avantage de l'utilisation d'un dictionnaire lexical est qu'il contient des mots-clés prédéfinis et groupés par domaine et hiérarchiquement classés. Il permet aux utilisateurs de parcourir les hiérarchies et de sélectionner les mots-clés qui conviennent. La proposition d'un dictionnaire aux annotateurs leur permet de ne pas proposer des instances qui n'ont rien à voir avec le contenu des images, même si cela n'évite pas que l'on assiste parfois à l'attribution des mots-clés qui n'ont rien à voir avec les objets concernés. Un autre avantage est que le dictionnaire lexical permet de gagner du temps pendant la phase d'annotation des images.

WordNet est le dictionnaire lexical que nous utilisons dans ce travail. Il aura pour but de permettre aux utilisateurs de proposer des mots-clés. *WordNet* est développé à l'université de Princeton. Sa conception est inspirée par les théories psycholinguistiques actuelles sur la mémoire lexicale humaine (c'est-à-dire sur l'organisation de termes par des relations linguistiques ou sémantiques). *WordNet* associe des racines de mots (mots simples ou composés) à des ensembles de synonymes qui représentent leurs sens sous-jacents, et inversement, chaque ensemble de synonymes est lié aux racines de mots dont ils représentent le sens. Plus précisément, un ensemble de synonymes est composé de ces racines de mots synonymes. Chaque racine de mot peut avoir différents sens dans différents contextes, et peut donc appartenir à plusieurs ensembles de synonymes. Un ensemble de synonymes représente un seul sens et est donc unique. Deux ensembles de synonymes différents peuvent n'être composés que d'une

1. http://www.lattice.cnrs.fr/sites/itellier/poly_info_ling/linguistique007.html

unique et même racine de mot (dans chacun de ces deux ensembles, une seule racine de mot est connue pour le sens représenté). Cependant, deux ensembles de synonymes ne partagent jamais le même *identificateur de sens* (un numéro d'identification interne à WordNet), ce qui en assure l'unicité. Les ensembles de synonymes sont liés par quelques relations lexicales (exemple l'antonymie) et diverses relations sémantiques². La relation sémantique servant de critère pour l'agrégation d'un groupe de concepts définira le type de l'ontologie. WordNet répertorie ainsi une grande variété de relations sémantiques permettant d'organiser le sens des mots (et donc par extension les mots eux-mêmes) en des systèmes de catégories qu'on peut consulter de manière cohérente et uniforme. On pourra ainsi interroger le système quant aux hyperonymes d'un mot particulier. À partir par exemple du sens le plus commun du mot car (correspondant au synset 1. car, auto...) la relation d'hyperonymie définit un arbre de concepts de plus en plus généraux³.

4.4.3/ GÉNÉRATION DES SÉMANTIQUES

La génération des sémantiques d'une image est un processus d'instanciation des concepts et des relations spatiales de l'ontologie par une communauté d'annotateurs. Puisque les annotateurs ont chacun une interprétation des images et qu'une image ne peut être interprétée de la même façon (comme nous l'avons précisé un peu plus haut), il est indispensable que les images soient soumises à un groupe d'annotateurs. Il s'agit donc de mettre les images à la disposition des annotateurs avec tous les moyens nécessaires afin que ceux-ci les annotent et que le système en déduise les sémantiques résultantes des différentes annotations grâce à la technique de sélection des instances.

La technique de sélection des instances et des relations spatiales est utilisée pour classer les instances de chaque concept de l'ontologie et les relations spatiales selon leur pondération. L'objectif est de déterminer la sémantique la plus représentative d'une image donnée. Rappelons-nous que le nombre de sémantique proposées à une image est égal au nombre d'annotateurs qui instancient les concepts et les relations spatiales de l'ontologie. Soit O l'ontologie et n le nombre de sémantiques obtenues. $S_{(O,n)}$ est l'ensemble de sémantiques générées telles que $S_{(O,n)} = \{S_{(O,1)}, S_{(O,2)}, \dots, S_{(O,n)}\}$. La figure 4.16 illustre la génération des sémantiques. Sur cette figure, nous avons donné l'exemple de trois sémantiques générées par trois annotateurs.

4.4.4/ ÉMERGENCE DES SÉMANTIQUES

L'émergence de la sémantique des images consiste à faire converger toutes les annotations produites par les annotateurs vers une et une seule annotation. Il s'agit en d'autre terme de la résultante de toutes les sémantiques produite pour la même image. Dans le système que nous avons mis en place, l'annotation d'une image passe par l'instanciation des concepts de l'ontologie. Ce qui nous fait obtenir à la fin de la phase d'annotation, plusieurs sémantiques pour une image. Le nombre de sémantiques obtenues pour image donnée est égal au nombre d'annotateurs ayant fourni la description de l'image. Autrement dit, chaque concept de l'onto-

2. <http://www.phmartin.info/PhD/these-phmartin.2.TravailRealise.pdf>

3. <http://fr.wikipedia.org/wiki/WordNet>

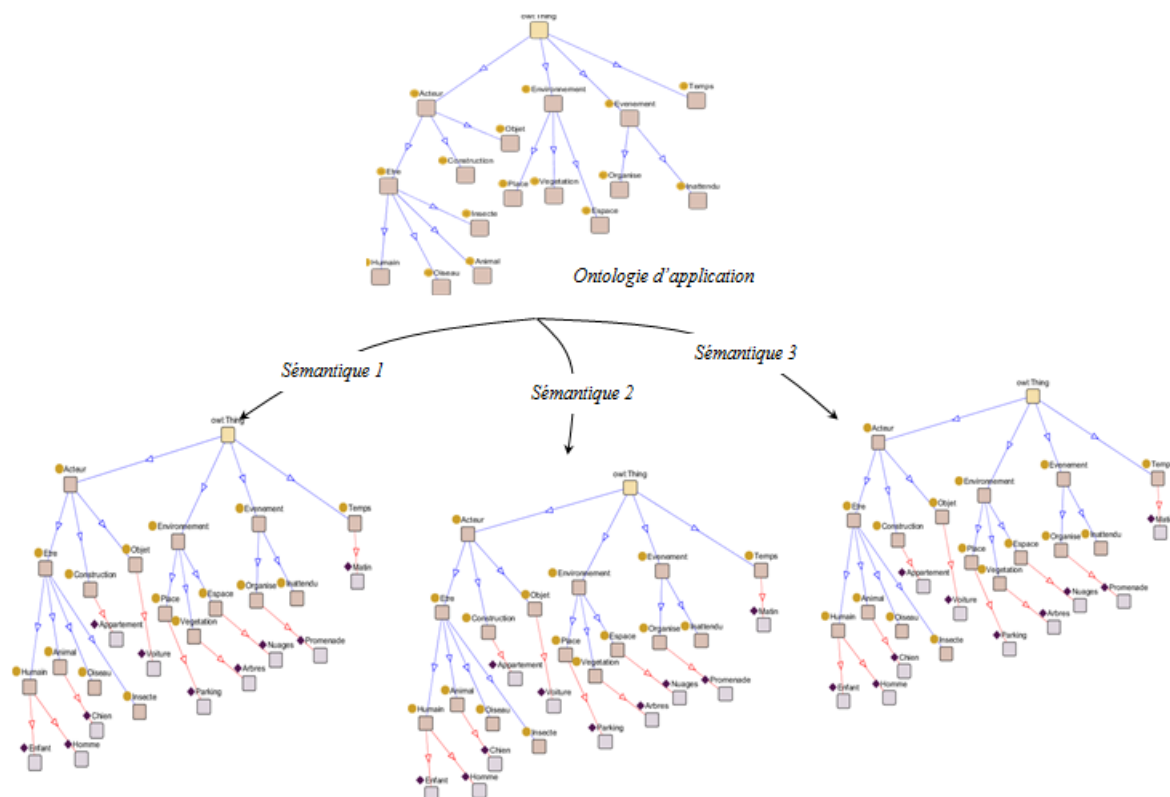


FIGURE 4.16 – Génération des sémantiques

logie peut contenir plusieurs instances identiques ou non à la fin de l'annotation d'une image. On a des instances identiques si les différents annotateurs attribuent le même mot-clé au même objet de l'image. Dans le cas contraire, cela signifie que le même objet est différemment décrit par les annotateurs, idem pour les relations spatiales. Les sémantiques proposées par les annotateurs sont évaluées par la fonction d'évaluation que nous avons mise en place à cet effet, ce qui permet d'extraire les instances et les relations spatiales les plus représentatives. Rappelons au passage qu'une sémantique émergente regroupe des instances et des relations *émergentes*. Cette sémantique peut évoluer dans le temps si et seulement si les instances et les relations spatiales évoluent. La sémantique émergente, c'est-à-dire la sémantique la plus représentative devient systématiquement la sémantique *réelle* de l'image.

Pour une image i , nous évaluons l'ensemble des sémantiques S_i proposées en nous basant sur les instances et les relations spatiales. Les relations hiérarchiques ne sont pas évaluées parce que c'est la même ontologie d'application qui est mise à la disposition de tous les annotateurs. Les instances et les relations spatiales entre les concepts sont évaluées par une technique que nous avons mise en place pour déterminer les instances et les relations spatiales les plus représentatives. Il s'agit de la technique basée sur la fonction f_{emerg} que nous présentons dans la section suivante.

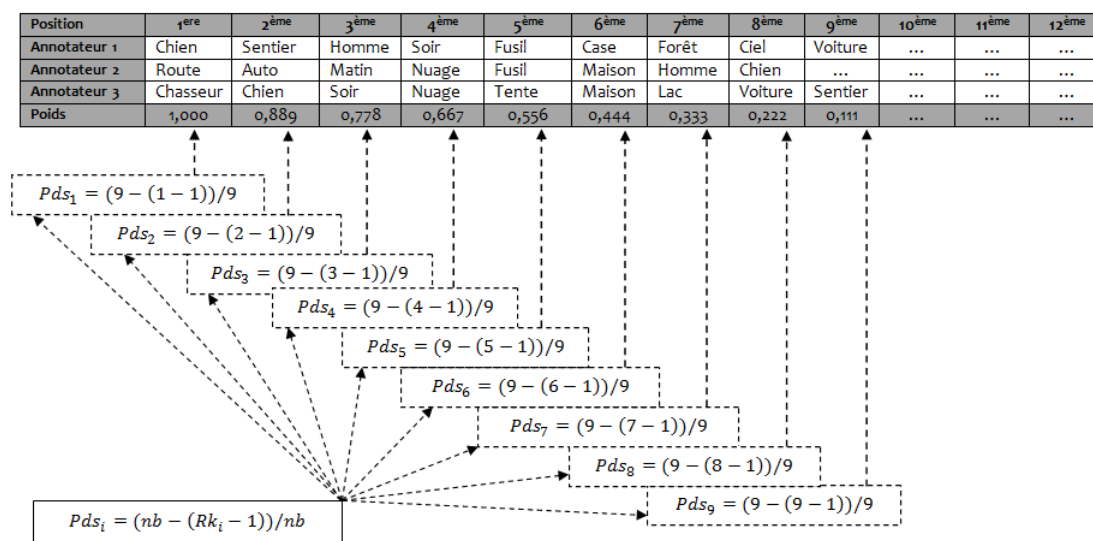


FIGURE 4.17 – Calcul des poids

4.4.4.1/ LA FONCTION D'ÉVALUATION f_{emerg}

L'évaluation des instances permet de ressortir les instances les plus représentatives. Cette étape permet d'en déduire une étude statistique des termes proposés pour une image. La pondération est une composante très importante dans le processus de recherche d'information. Elle est souvent calculée à partir des considérations et interprétations statistiques. Le calcul de la pertinence d'un terme dépend des poids attribués aux termes représentant l'image [167].

La fréquence est l'approche que nous utiliserons. A cette mesure va s'ajouter la moyenne des poids correspondant à la position à laquelle se trouve le terme concerné. Un poids est attribué aux termes selon l'ordre du choix. L'avantage de cette méthode est qu'elle considère que l'ordre du choix des termes joue un rôle très important parce que le terme placé en premier devra avoir une évaluation plus importante qu'un terme vers la fin de la liste des termes. Dans notre approche, la moyenne des poids correspondant à la position à laquelle se trouve le terme concerné est ajoutée à la mesure de la fréquence. Une fois les sémantiques proposées par les annotateurs, le système procède à l'évaluation des instances mappées dans chaque sémantique. Le but de cette évaluation est d'obtenir les instances les plus utilisées pour chaque concept lors de l'annotation. La fonction f_{emerg} est déterminée grâce à la combinaison de la fréquence et de la pondération des instances des sémantiques proposées par la communauté.

$$f_{emerg} = Frq_i * Pds_i, \text{ avec}$$

Frq_i , la fréquence du terme i , Pds_i : le poids du terme i .

- La fréquence des instances : Notre approche consiste en la pondération des instances contenues dans des concepts identiques des différents sens proposés. La pondération des sémantiques dépend de la pondération des instances représentant chaque concept de l'ontologie. Elle est généralement basée sur la fréquence d'apparition des instances [167].

La fréquence du terme i est donnée par le rapport entre le nombre d'occurrence(s) de i (N_i) et le nombre total d'occurrences ($\Sigma(N)$) comme le présente la formule suivante :

$$Frq_i = \frac{N_i}{\Sigma(N)}$$

- Le poids des instances : Même si la fréquence (moyenne ou faible) d'une instance qui apparaît souvent dans une image est susceptible d'être plus importante que celle qui se produit rarement, des expériences ont montré que ce n'est pas toujours vérifié. Pour améliorer l'évaluation des instances, nous avons ajouté à la fréquence, la moyenne du poids correspondant à la position des instances. Un poids est donné aux instances suivant l'ordre de leur choix. L'avantage de cette méthode est qu'elle considère que l'ordre du choix des instances joue un rôle très important parce que l'instance placée en premier devra avoir une évaluation plus importante qu'une instance vers la fin de la liste. Le poids du terme i est calculé par :

$$Pds_i = \frac{(Nb - (Rk_i - 1))}{Nb}, \text{ où}$$

Nb est le nombre maximum de concepts qui pourrait être instancié par un annotateur et Rk_i la position de l'instance pendant l'annotation.

Considérons une ontologie ayant douze concepts "instanciables" et au maximum neuf(9) concepts instanciés pour une image donnée comme le montre le tableau de la figure 4.17.

La fonction f_{emerg} s'applique aussi aux relations spatiales proposées pour déterminer les relations les plus représentatives entre deux concepts concrets. Dans la pratique, les concepts de l'ontologie sont proposés deux à deux aux annotateurs. Les annotateurs proposent des relations en choisissant le type de relation cognitive adapté aux deux concepts (*Connecté*, *Déconnecté* ou *Partie - de*). Une liste de mots-clés, de groupes de mots-clés ou des verbes leur est proposée compte tenu de la relation cognitive choisie.

Les annotateurs proposent des relations dans le sens des concepts que leur sont proposés et également dans le sens inverse. Une fois le tableau obtenu, on pourra donc appliquer la fonction f_{emerg} pour déterminer les relations les plus pertinentes. La figure 4.18 présente la proposition des relations spatiales. Le tableau a.) présente un extrait de la liste des relations mappées dans les relations cognitives et le tableau b.) l'instanciation des relations spatiales.

4.4.4.2/ LA SÉMANTIQUE ÉMERGENTE

La sémantique émergente n'est rien d'autre que la sémantique la plus représentative obtenue grâce à la fonction émergente f_{emerg} qui combine la fréquence et le poids des instances et des relations spatiales proposées. La figure 4.19 présente le tableau de l'évaluation des instances et le tableau d'évaluation des relations spatiales proposées pour une image par les annotateurs.

De chaque concept de l'ontologie, émerge une instance ω parmi les instances proposées par les annotateurs pour une image donnée. Et des paires de relations spatiales (relation avec son inverse) émerge une paire de relations μ pour relier les objets. Les instances et les relations émergées et mappées dans l'ontologie donnent la sémantique émergente de l'image.

	PartieDe	Connecté	Déconnecté
Verbes, mots et groupe de mots	Composer de, Insérer dans, Conduire, Asseoir, Nager, Se déplacer dans, Marcher sur, etc.	Tenir, Saluer, Embrasser, Soulever, Scier, Réparer, Remplir, Accoler, Nager, courir, Dans, Sur, etc.	Eloigner, Rapprocher, Distancer, Loin, Proche, A côté, Regarder, Sauter, Derrière, Devant, A gauche, A droite, Sous, etc.

a.) Relations mappées dans les relations cognitives

		Vegetation Animal	Vegetation Humain	Vegetation Objet	Animal Humain	Animal Objet	Humain Objet
Annotateur 1	Sens Normal	Sous	Sous	Sous	Près de	Loin de	Tient
	Sens Inverse	Sur	Sur	Sur	Près de	Loin de	Est tenu
Annotateur 2	Sens Normal	En bas de	En dessous de	En dessous de	A côté de	En dessous de	Garde
	Sens Inverse	Sur	Sur	Au-dessus de	A côté de	Au-dessus de	Est gardé
Annotateur 3	Sens Normal	Sous	Sous	En dessous de	Près de	En bas de	Tient
	Sens Inverse	Sur	Sur	Au-dessus de	Près de	En haut de	Est tenu
Annotateur 4	Sens Normal	Sous	Sous	En dessous de	Près de	En dessous de	Tient
	Sens Inverse	Sur	Sur	Au-dessus de	Près de	Au-dessus de	Est tenu

b.) Instanciation des relations spatiales

FIGURE 4.18 – Proposition des relations spatiales

L'évaluation commence dès la proposition de la première sémantique et se termine lorsque le dernier annotateur actif valide sa sémantique. Au fur et à mesure que les sémantiques sont proposées par les annotateurs, une sémantique émergente s'obtient. La sémantique émergente initialement obtenue peut évoluer compte tenu de nouvelles sémantiques proposées par de nouveaux annotateurs. Il peut arriver que tous les annotateurs ne proposent pas les mêmes instances pour un concept donné (idem pour les relations). Dans ce cas, une notification est envoyée à chaque annotateur avec la liste des différentes instances. Les annotateurs pourront donc choisir une instance parmi la liste qui leur est fournie par le système.

La figure 4.20 en est une illustration. Elle représente trois sémantiques *Sémantique 1*, *Sémantique 2* et *Sémantique 3* qui émergent vers une sémantique la *Sémantique émergente*.

4.5/ CONCLUSION

Dans ce chapitre, nous avons présenté la représentation d'images à travers les descripteurs sémantiques. Deux descripteurs sémantiques sont proposés : une ontologie générique et une ontologie d'application. L'ontologie d'application est dérivée de l'ontologie générique et est composée de concepts génériques ou abstraits provenant de l'ontologie générique qui contiennent d'autres concepts qui peuvent être abstraits ou concrets. Aux concepts concrets ou concepts de spécialisation sont attachées des instances (termes) décrivant les régions ou les objets présents dans une image. Les concepts (abstrait ou concrets) sont reliés entre eux par des relations qui représentent le lien qui existe entre les objets de l'image. Grâce à l'ontologie d'application et à un dictionnaire lexical, une communauté d'annotateurs décrit la sémantique des images. Les différentes sémantiques proposées par chaque membre de la communauté sont matchées grâce à une fonction f_{emerg} qui combine la fréquence et le poids des instances. Cette fonction permet de faire émerger les différentes sémantiques vers une et une seule sémantique, la sémantique émergente qui va annoter l'image.

Ce chapitre permet de fournir les sémantiques dont le but est de substituer à une image un

Concepts	Instances	Fréquence	Poids	f_{emerg}	Instances retenues
Animal	Chien	0,3200	0,6937	0,2220	Chien
	Cheval	0,1200	0,2133	0,0256	
	Léopard	0,2000	0,3200	0,0640	
	Loup	0,2000	0,4267	0,0853	
	Chat	0,1600	0,5333	0,0853	
Humain	Jeune-homme	0,2000	0,3733	0,0747	Homme
	Personne	0,2000	0,6667	0,1333	
	Individu	0,1200	0,0533	0,0064	
	Homme	0,3200	0,5867	0,1877	
	Adolescent	0,1200	0,1867	0,0224	
Objet	Balle	0,2800	0,6133	0,1717	Ballon
	Ballon	0,3600	0,6667	0,2400	
	Bille	0,2000	0,3200	0,0640	
	Boule	0,1200	0,1600	0,0192	
	Pomme	0,0400	0,0267	0,0011	
Construction	Maison	0,2000	0,2933	0,0587	Habitation
	Pavillon	0,3600	0,1867	0,0672	
	Immeuble	0,0400	0,6667	0,0267	
	Habitation	0,2400	0,6400	0,1536	
	Logement	0,1600	0,1067	0,0171	
Vegetation	Forêt	0,4000	0,5867	0,2347	Forêt
	Arbre	0,2400	0,2400	0,0576	
	Plantes	0,2000	0,4800	0,0960	
	Fleur	0,0800	0,2933	0,0235	
	Prairie	0,0800	0,5333	0,0427	

(a) Tableau d'évaluation des instances

		Relations	Fréquence	Poids	f_{emerg}	Relations retenues
Vegetation Animal	Sens Normal	Sous	0,750	0,790	0,593	Sous
		En bas de	0,250	0,240	0,060	
	Sens Inverse	Sur	1,000	0,360	0,360	
Vegetation Humain	Sens Normal	Sous	0,750	0,480	0,360	Sous
		En dessous de	0,250	0,600	0,150	
	Sens Inverse	Sur	1,000	0,420	0,420	
Vegetation Objet	Sens Normal	Sous	0,250	0,750	0,188	Sous
		En dessous	0,750	0,060	0,045	
	Sens Inverse	Sur	0,250	0,660	0,165	Sur
		Au-dessus de	0,750	0,210	0,158	
Animal Humain	Sens Normal	Près de	0,750	0,690	0,518	Près de
		A côté de	0,250	0,750	0,188	
	Sens Inverse	Près de	0,750	0,360	0,270	Près de
		A côté de	0,250	0,180	0,045	
Animal Objet	Sens Normal	Loin de	0,250	0,030	0,008	En dessous de
		En dessous de	0,500	0,330	0,165	
		En bas de	0,250	0,210	0,053	
	Sens Inverse	Loin de	0,250	0,750	0,188	Au-dessus de
		Au-dessus de	0,500	0,720	0,360	
		En haut de	0,250	0,120	0,030	
Humain Objet	Sens Normal	Tient	0,750	0,660	0,495	Tient
		Garde	0,250	0,270	0,068	
	Sens Inverse	Est tenu	0,750	0,540	0,405	Est tenu
		Est gardé	0,250	0,330	0,083	

(b) Tableau d'évaluation des relations spatiales

FIGURE 4.19 – Évaluation des instances et des relations spatiales

représentant (ou descripteur) sémantique compréhensible par l'humain que par les machines, moins encombrant qui la caractérise le mieux possible et de ne travailler que sur ce modèle lors de la recherche. Cela permettra une meilleure organisation des données.

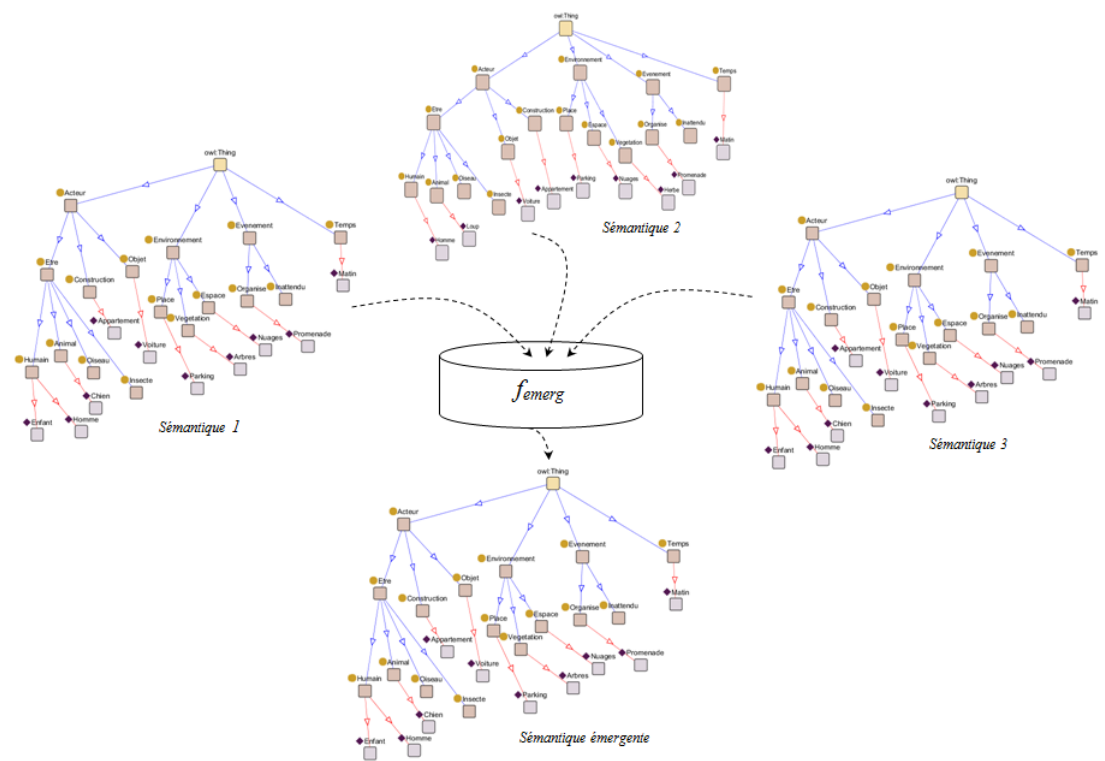


FIGURE 4.20 – Émergence des sémantiques

RECHERCHE SÉMANTIQUE D'IMAGES : UNE APPROCHE BASÉE SUR LE CLUSTERING ET LA RECOMMANDATION

Dans ce chapitre, nous présentons une approche de recherche sémantique d'images. L'approche se base sur deux techniques : le clustering des images et la recommandation des sémantiques. Le clustering permet de regrouper les sémantiques similaires aux requêtes des utilisateurs et la recommandation a pour objectif de proposer des sémantiques aux utilisateurs en se basant sur leurs préférences.

Sommaire

5.1	Introduction	86
5.2	Aperçu général et caractéristiques	86
5.3	Recherche d'images basée sur le clustering	88
5.3.1	Notation et définition	89
5.3.2	Exemple d'illustration	89
5.3.3	Mesures de similarité	90
5.3.4	Extraction des sémantiques représentatives	92
5.3.5	Extraction des sémantiques voisines	96
5.3.6	Formation de groupes de sémantiques : Granularité	99
5.3.7	Cluster d'images	108
5.3.8	Résumé	109
5.4	Recommandation de sémantiques	110
5.4.1	Introduction	110
5.4.2	Recommandation des sémantiques	110
5.4.3	Acquisition des profils	111
5.4.4	Classification des profils : Algèbre et treillis de Galois	114
5.4.5	Génération de la liste des recommandations	118
5.4.6	Résumé	120
5.5	Conclusion	120

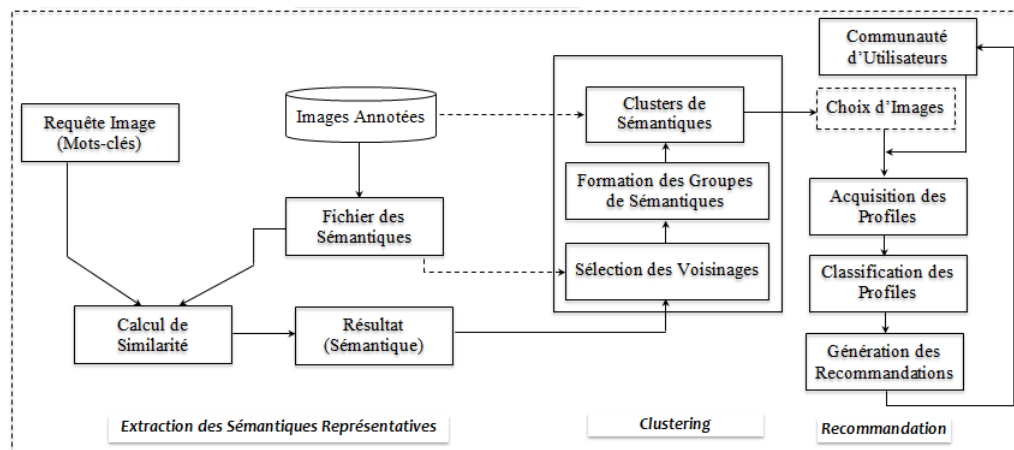
5.1/ INTRODUCTION

Les recherches d'images sont généralement axées sur deux techniques : la recherche séquentielle d'un ensemble d'images dans le cas où l'utilisateur n'a vraiment pas une idée de ce qu'il veut et la recherche d'images basée sur la description du contenu, lorsque l'utilisateur sait exactement ce qu'il veut. Mais vu les difficultés auxquelles étaient confrontés les utilisateurs, les chercheurs ont fait évoluer les techniques de recherche d'images. Comme difficulté, nous pouvons citer le fossé sémantique. Rappelons que le fossé est l'écart significatif entre la perception de haut niveau et l'interprétation des images par les utilisateurs et les descriptions de bas niveau. Une des attentes des utilisateurs dans le domaine de la recherche d'images se situe au niveau de la sémantique des images. Ces derniers temps, plusieurs études consacrées aux systèmes de recherche d'images par le contenu (*Content-Based Image Retrieval Systems*) s'appuient sur l'utilisation de la sémantique de l'image comme nous l'avons décrit dans le chapitre précédent.

5.2/ APERÇU GÉNÉRAL ET CARACTÉRISTIQUES

Dans le cadre de l'approche *Emergsem*, nous proposons une technique de recherche d'images basée sur la sémantique et la recommandation. La technique qui s'appuie sur des descripteurs sémantiques représentés par l'ontologie d'application que nous avons développée dans le chapitre précédent et qui est constituée de concepts génériques et d'instanciations de ces concepts.

La principale différence entre les CBIR et la proposition que nous faisons se résume en deux principaux points. Le premier point réside dans le fait que notre approche étend la technique de recherche sur le voisinage des images retrouvées grâce aux informations de la requête. Dans un CBIR typique, la requête de l'utilisateur est comparée aux images du corpus pour déterminer un ensemble d'images pertinentes qui sont retournées à l'utilisateur. Le résultat initial de la recherche peut être amélioré par un processus de feedback (retour de pertinence) qui s'appuie sur l'utilisateur pour affiner la précision de la recherche. Un système CBIR contourne l'étape du voisinage et transmet directement les résultats triés aux utilisateurs et ne prend pas en compte les liens de similarité entre les images sélectionnées par la comparaison de la requête initiale aux images de la collection. Notre technique récupère les sémantiques voisines de chaque sémantique représentative de la requête de l'utilisateur au lieu d'un ensemble d'images commandées : les images voisines des images représentatives sont sélectionnées selon une mesure de similarité, regroupées et renvoyées à l'utilisateur. Ceci nous permet de créer des clusters d'images les plus pertinentes. De cette façon, les relations entre les images récupérées sont prises en considération par les clusters. Le deuxième point réside dans le fait que les CBIR n'arrivent pas à faire des recommandations d'images aux internautes parce que les préférences des internautes ne sont pas stockées. Dans *Emergsem*, nous intégrons une approche de recommandation des images. L'objectif est de stocker les préférences des utilisateurs, c'est-à-dire les sémantiques utilisées pour rechercher les images. Des recommandations pourront être ainsi faites aux utilisateurs en se basant sur leurs préférences.

FIGURE 5.1 – Recherche sémantique d'images dans *Emergsem*

Emergsem s'appuie sur : (1) les descripteurs sémantiques pour rechercher les images, (2) une recherche de liens de voisinage entre les images en vue de créer des clusters pour une recherche plus précise et plus pertinente et (3) sur une phase de recommandation des sémantiques aux utilisateurs qui permet de proposer des sémantiques aux utilisateurs en se basant sur leurs préférence. Les contributions majeures de notre approche sont présentées dans l'architecture illustrée sur la figure 5.1.

- L'idée centrale de la sémantique est d'identifier les objets ou acteurs composant une image ainsi que les relations complexes entre les objets. De nos jours, la recherche d'images basée sur la sémantique est introduite dans des systèmes de recherche des images parce que les expériences ont montré que la performance des résultats de recherche dépend de la qualité et de la consistance des sémantiques associées à chaque image (voir [168], [169]). Comme nous l'avons déjà souligné précédemment, la technique de recherche dans *Emergsem* est basée sur la sémantique des images, c'est-à-dire qu'elle repose sur une prise en compte des relations conceptuelles entre les objets constituant de l'image. Rappelons que la sémantique d'une image est une description de l'image obtenue grâce à une ontologie d'application. Elle est en effet le résultat du processus d'instanciation de concepts et des relations de l'ontologie d'application avec des mots-clés ou des termes. La figure 5.2 présente la sémantique d'une image (l'ontologie d'application avec quelques concepts concrets instanciés) composée de concepts génériques tels que le concept Humain est instancié par *Homme* et *Enfant*, le concept Animal par *Chien*, le concept Construction par *Appartement*, le concept Place par *Parking*, Vegetation par *Arbre*, Espace par *Nuage*, Organise par *Promenade* etc.
- L'approche de recherche sémantique basée sur le clustering est constituée de deux étapes. La première étape présente l'extraction des sémantiques représentatives de la requête de l'utilisateur et la deuxième étape permet d'extraire les sémantiques voisines de chaque sémantique représentative. L'extraction des sémantiques représentatives est basée sur la similarité entre la requête de mots-clés de l'utilisateur et les instances des sémantiques du corpus. L'extrac-

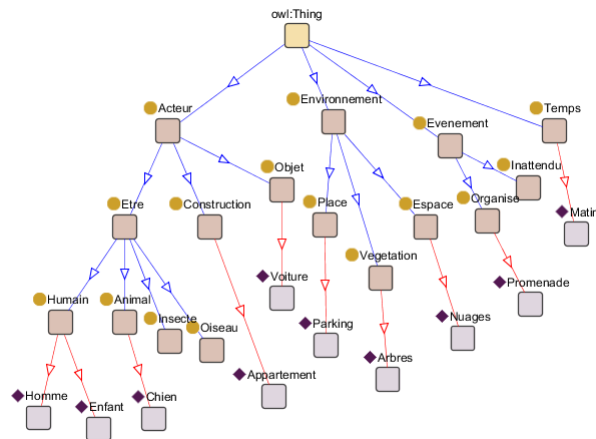


FIGURE 5.2 – Sémantique d'image

tion des sémantiques voisines quant à elle est composée de trois niveaux. Le premier niveau est la sélection du voisinage des sémantiques représentatives. La sélection du voisinage est basée sur la comparaison des vecteurs d'instances des sémantiques du corpus. Le deuxième niveau est la formation des groupes de sémantiques au sein de chaque ensemble de voisinage. Elle est basée sur la granularité des sémantiques. Le troisième niveau est la génération des clusters de sémantique. La génération des clusters d'images consiste à l'appariement des groupes formés au sein de chaque ensemble voisinage pour déterminer les groupes de sémantiques les plus pertinentes. L'appariement est basé sur la matrice de confusion, le conflit et la fonction de correspondance.

- L'approche de recommandation a pour objectif de recommander des sémantiques aux utilisateurs en se basant sur leurs préférences, c'est-à-dire les sémantiques utilisées pour rechercher des images. Cela sous-entend que pour qu'il y ait recommandations personnalisées, il faut impérativement que les différentes sémantiques utilisées par les utilisateurs soient stockées. L'approche de recommandation comprend trois parties. La première partie est consacrée à l'acquisition des profils des utilisateurs. Deux profils se distinguent : les profils statiques et les profils dynamiques qui sont détaillés un peu plus bas dans la section de recommandation des sémantiques. La deuxième partie est la classification des profils utilisateurs. Elle permet en quelque sorte de regrouper dans des petits groupes des profils (statiques et dynamiques) similaires. Le regroupement est basé sur l'algèbre de Galois qui permet de classer les profils d'utilisateurs de façon exhaustive. La troisième partie est la génération de la liste des recommandations qui sera adressée aux utilisateurs selon leurs préférences.

5.3/ RECHERCHE D'IMAGES BASÉE SUR LE CLUSTERING

La recherche sémantique d'une image est une approche qui permet de rechercher les images non seulement avec les mots-clés qui décrivent leurs régions respectives, mais également avec les relations entre les régions d'images. C'est une techniques qui a pour but de cibler l'image

recherchée en se basant sur la requête de l'utilisateur. L'avantage de cette technique est qu'elle peut être utilisée pour rechercher les images similaires via la comparaison de leur sémantiques respectives. Plus les sémantiques sont proches, plus proches aussi sont les images qu'elles décrivent. Les recherches sémantiques d'images sont généralement axées sur les méthodes d'appariement des sémantiques d'images. Pour rechercher une image, les utilisateurs proposent généralement des requêtes décrivant les images qu'ils souhaitent rechercher. Les informations de la requête sont comparées avec les informations stockées dans des bases de données grâce à des techniques d'appariement. Ces techniques se basent sur des mesures de similarité.

Nous présentons d'abord la notation que nous allons adopter dans ce chapitre. Nous rappelons ensuite les mesures de similarité. Enfin, nous présentons la définition.

5.3.1/ NOTATION ET DÉFINITION

- $Sem = \{S_i\}$ l'ensemble des sémantiques du corpus ;
- $C(S_i)$ l'ensemble des concepts de S_i ;
- $Inst(S_i)$ l'ensemble des instances de S_i ;
- $R(S_i) = \{(c_i, c_j) \mid c_i \in C \text{ et } c_j \in C\}$ l'ensemble des relations entre les concepts c_i et c_j ;
- $S_r = \{S_{r(n)} \mid S_r \in Sem\}$ l'ensemble des sémantiques *résultat* ;
- $e_j = \{S_v^{(e_j)} \mid S_v^{e_j} \in Sem\}$ l'ensemble des sémantiques voisines de S_r ;
- $G^{(l)} = \{G_1^{(l)}, G_2^{(l)}, \dots, G_{e_j}^{(l)}\}$ l'ensemble des l groupes de e_j ensembles de sémantiques voisines.
- $\lambda_{p,q}^{(l,k)} = |G_p^{(l)} \cap G_q^{(k)}|$ le nombre de sémantiques communes aux groupes $G_p^{(l)}$ et $G_q^{(k)}$.

5.3.2/ EXEMPLE D'ILLUSTRATION

Pour illustrer nos discussions dans le reste du chapitre, nous allons utiliser un exemple comprenant un corpus Sem de 10 sémantiques $S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9$ et S_{10} dont les instances sont représentées dans la table 5.1 et d'une requête utilisateur k^{Rq} représentée par un ensemble de 4 mots-clés $k^{Rq} = \{Chien, Chasseur, Fusil, Forêt\}$.

Sem	Instances des Sémantiques S_i
S_1	{Chien, Forêt, Chasseur, Fusil, Gibier, Tente, Soir}
S_2	{Arbre, Chien, Chasseur, Forêt, Fusil, Matinée}
S_3	{Forêt, Chasseur, Fusil, Gibier, Chien, Feu}
S_4	{Cabane, Forêt, Chasseur, Fusil, Montagne, Après – midi}
S_5	{Aigle, Rivière, Chasseur, Fusil, Chien, Maison, Forêt}
S_6	{Chien, Brousse, Chasseur, Bâton, Forêt}
S_7	{Perdrix, Arbres, Chasseur, Fusil, Brousse, Gibecière}
S_8	{Chien, Maison, Prairie, Etang, Gourde}
S_9	{Gardien, Arbres, Maison, Sac}
S_{10}	{Femme, Calebasse, Prairie, Fleuve, Enfant}

TABLE 5.1 – Sémantiques du corpus

5.3.3/ MESURES DE SIMILARITÉ

Différent types de mesures peuvent être utilisés pour définir la similarité d'images. Dans les approches traditionnelles (classiques) de recherche d'images, une requête et les images du corpus sont représentées par des vecteurs de caractéristiques physiques. Les images du corpus sont classées ou triées par ordre de pertinence (similarité) selon une distance définie par les caractéristiques physiques.

Dans les approches de recherche d'informations textuelles, la requête utilisateur et les documents sont typiquement représentés par des vecteurs de mots-clés. Le but du traitement de requête est de sélectionner les documents d'une collection contenant les termes de la requête.

La similarité sémantique est fréquemment utilisée en tant que mesure de ressemblance entre deux documents. Il s'agit de comparer les textes issus d'un corpus dans une optique de classification (regrouper tous les documents relatifs à une thématique particulière), ou de recherche d'information (dans ce cas, un document vectorisé est constitué par l'ensemble de mots-clés et est comparé par mesure de cosinus de l'angle avec des vecteurs correspondant à tous les documents présents dans le corpus. On évalue ainsi lesquels sont les plus proches)¹.

Dans nos travaux, le calcul de similarité entre deux sémantiques d'images se base sur deux niveaux de caractéristiques. D'une part, nous prenons en compte les termes du vocabulaire utilisé dans les instances de l'ontologie d'application pour définir (1) des vecteurs de caractéristiques et (2) des mesures ou distances (exemple : fréquence/entropie, mesure de cosinus) pour définir les sémantiques les plus proches d'une sémantique donnée. D'autre part, nous prenons en compte la structure et la hiérarchie entre concepts de l'ontologie d'application pour comparer niveau par niveau les éléments les éléments d'instanciation de deux sémantiques.

Comme dans la recherche textuelle de documents, le calcul de mesure de similarité sémantique dans *Emergsem* est basé sur les vecteurs de caractéristiques construits à partir d'ensemble de termes qui peuvent être des mots-clés ou des groupes de mots-clés liés par des relations. Par exemple dans Wordnet, les éléments sont représentés par des synset qui définissent des groupes de mots-clés liés par synonymie ou autres relations.

$$SemV(S) \begin{array}{|c|c|c|c|c|c|} \hline Label & m_1 & m_2 & m_3 & \dots & m_n \\ \hline V & w_1 & w_2 & w_3 & \dots & w_n \\ \hline \end{array}$$

TABLE 5.2 – Table de vecteurs

Définition : Les instances du corpus $Sem_Instances = \cup_{S_i \in Sem} Inst(S_i)$ où S_i est une sémantique d'image et $Inst(S_i)$ est l'ensemble des mots-clés ou instances de la sémantique S_i . Un vecteur de caractéristiques sémantiques est représenté par $SemV(S)$ comme le montre la table 5.2 où *Label* représente une instance dont les valeurs $m_i \in Sem_Instances$ et $n = |Sem_Instances|$ représente le nombre total de termes (mots-clés) dans toutes les instances du corpus d'images *Sem*. w_i est la fréquence (le poids) du label dans le vecteur. Elle est égale à 1 si le mot-clé $m_i \in Inst(S_i)$ apparaît dans l'instanciation de la sémantique S_i . Les algorithmes de formation des vecteurs sémantiques qui comparent d'une part les mots-clés

1. http://fr.wikipedia.org/wiki/Similarité_cosinus

de la requête aux instances du corpus et d'autre part les instances des sémantiques avec les instances du corpus sont présentés dans 5.1.

Algorithme 5.1 Algorithme de formation de vecteurs de caractéristiques sémantiques

Soit k^{Rq} les mots-clés d'une requête Rq

```

1: for  $i := 1$  to  $n$ ;  $n = |Sem\_Instances|$  do
2:   if  $SemV(Rq).Label[i] \in k^{Rq}$  then
3:      $SemV(Rq).V[i] \leftarrow 1$ 
4:   else
5:      $SemV(Rq).V[i] \leftarrow 0$ 
6:   end if
7: end for

```

a.) Vecteur de caractéristiques sémantiques d'une requête

```

1: for  $i := 1$  to  $n$ ;  $n = |Sem\_Instances|$  do
2:   if  $SemV(Rq).Label[i] \in k^{Rq}$  then
3:      $SemV(S_i).V[i] \leftarrow 1$ 
4:   else
5:      $SemV(S_i).V[i] \leftarrow 0$ 
6:   end if
7: end for

```

b.) Vecteur de caractéristiques sémantiques d'une sémantique

La représentation du vecteur des mots-clés (voir 5.1) de la requête utilisateur est présenté dans la table 5.3. Dans cette table, tous les labels de l'union des instances du corpus et leur valeur ne pourront pas être représentés. Pour retrouver les labels remplacés par les pointillés, il faut se référer à l'union des instances du corpus ordonnées par ordre alphabétique $Ins = \{ Aigle, Après - midi, Arbre, Bâton, Brousse, Cabane, Calebasse, Chasseur, Chien, Enfant, Etang, Femme, Feu, Fleuve, Forêt, Fusil, Gardien, Gibecière, Gibier, Gourde, Maison, Matinée, Montagne, Perdrix, Prairie, Rivière, Sac, Soir, Tente \}$. Chacune de leur valeur équivaut à 0.

Label	Aigle	...	Chasseur	Chien	...	Fleuve	Forêt	Fusil	...	Soir	Tente
V	0	...	1	1	...	0	1	1	...	0	0

TABLE 5.3 – Représentation du vecteur des mots-clés de la requête

Nous proposons également la représentation du vecteur d'une sémantique du corpus : la sémantique S_1 comme le montre la table 5.4

Label	Aigle	...	Chasseur	Chien	...	Fleuve	Forêt	Fusil	...	Soir	Tente
V	0	...	1	1	...	0	1	1	...	1	1

TABLE 5.4 – Représentation du vecteur d'une sémantique (S_1)

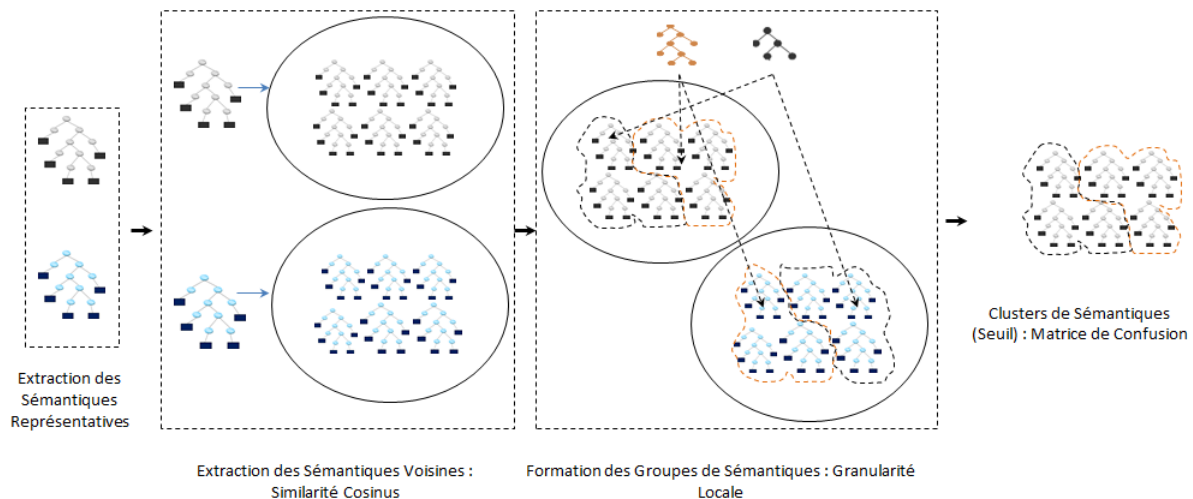


FIGURE 5.3 – Étapes de la recherche sémantique d'images

Dans le reste de la section, nous décrivons les différentes étapes de la recherche sémantique d'images comme le montre figure 5.3. Dans un premier temps, nous présenterons l'extraction des sémantiques représentative. Nous décrivons ensuite l'extraction des sémantiques voisines de chaque sémantique *résultat*. La formation des groupes de sémantiques similaires au sein de chaque ensemble de sémantiques voisines sera la troisième étape et en dernier la formation des clusters de sémantiques pertinentes.

5.3.4/ EXTRACTION DES SÉMANTIQUES REPRÉSENTATIVES

La détermination des sémantiques représentatives se fait se fait de deux méthodes :

- Par appariement des mots-clés de la requête et les instances des sémantiques du corpus. Pour cela, on associe un poids (une fréquence) à chaque sémantique du corpus. Le poids (une fréquence) représentant le nombre de fois qu'un mot-clé de la requête utilisateur est trouvé dans les instances de la sémantique. Il est une valeur comprise entre 0 et 1. Il est égal à 1 lorsque tous les mots-clés sont présents dans les instances d'une sémantique. La table ?? présente un algorithme pour le calcul des poids entre une requête et les sémantiques du corpus. Prenons l'exemple de section 5.3.2 pour illustrer l'extraction des sémantiques représentatives. Les mots-clés de la requête de l'utilisateur sont représentés dans la première colonne de la table 5.5 et le corpus d'images dans la deuxième colonne de la table. Le résultat de l'algorithme est un ensemble de poids respectifs correspondant à une comparaison des mots-clés de la requête aux instances des sémantiques du corpus (voir la troisième colonne de la table).
- Par définition de mesures de similarité entre les mots-clés de la requête de l'utilisateur et les instances des sémantiques du corpus. Plusieurs méthodes peuvent être utilisées pour définir

la similarité entre deux sémantiques d'images S_i et S_j ou entre une sémantique S_i et une requête Rq comme l'algorithme 5.6. Le tableau présente quelques mesures de similarité qui permet de calculer la similarité entre les termes de la requête et les sémantiques du corpus où la valeur $terme_{i,k}$ représente la valeur de $SemV(S_i).V[k]$.

Algorithme 5.2 Algorithme de calcul de similarité par appariement

Soit $S_i \in Sem \Rightarrow Sem$ étant l'ensemble des sémantiques du corpus

Soit $C(S_i)$ l'ensemble des concepts de S_i

Soit q_j^i, q_j^i est un instance de S_i

Soit $Inst(S_i)$ les instances de S_i , $Inst(S_i) = \{q_j^i \mid \exists c_j^i \in C \text{ et } k^{Rq} \text{ instance de } c_j^i\}$

Soit t le compteur des mots – clés identiques

```

1: for  $S_i \in Sem$  do
2:   for  $x \in k^{Rq}$  do
3:     for  $q_j^i \in Inst(S_i)$  do
4:       if  $x = q_j^i$  then
5:          $t \leftarrow t + 1$ 
6:       end if
7:     end for
8:   end for
9:    $P(S_i) \leftarrow t / |k^{Rq}|$ 
10:   $t \leftarrow 0$ 
11: end for
  
```

Mots – clés	Instances des Sémantiques (S_i)	$P(S_i)$
	$Inst(S_1) = \{Chien Forêt Chasseur Fusil Gibier Tente Soir\}$	1,00
	$Inst(S_2) = \{Arbre Chien Chasseur Forêt Fusil Matinée\}$	1,00
	$Inst(S_3) = \{Forêt Chasseur Fusil Gibier Chien Feu\}$	1,00
Chasseur	$Inst(S_4) = \{Cabane Forêt Chasseur Fusil Montagne Après – midi\}$	0,75
Fusil	$Inst(S_5) = \{Aigle Rivière Chasseur Fusil Chien Maison Forêt\}$	1,00
Chien	$Inst(S_6) = \{Chien Brousse Chasseur Bâton Forêt\}$	0,75
Forêt	$Inst(S_7) = \{Perdrix Arbres Chasseur Fusil Brousse Gibecière\}$	0,50
	$Inst(S_8) = \{Chien Maison Prairie Etang Gourde\}$	0,25
	$Inst(S_9) = \{Gardien Arbres Maison Sac\}$	0,00
	$Inst(S_{10}) = \{Femme Calebasse Prairie Fleuve Enfant\}$	0,00

TABLE 5.5 – Similarité par appariement

La table 5.7 présente des calculs de similarité entre une requête Rq et des sémantiques d'images représentées par vecteurs de caractéristiques sémantiques. Quatre mesures de similarité sont utilisées : Dice, Cosinus, Appariement, appelée *Appar* et Overlap. Les quatre dernières lignes de la table correspondent aux résultats des calculs de similarité. On constate que les méthodes de calcul de similarité basées sur Dice et sur Cosinus donnent des résultats presque similaires, de même que la méthode basée sur Overlap et *Appar*.

Méthodes	Mesures de similarité
<i>Dice</i>	$Similarité(S_i, S_j) = \frac{2 \sum_{k=1}^t terme_{ik} terme_{jk}}{\sum_{k=1}^t terme_{ik} + \sum_{k=1}^t terme_{jk}}$
<i>Cosinus</i>	$Similarité(S_i, S_j) = \frac{\sum_{k=1}^t terme_{ik} terme_{jk}}{\sqrt{(\sum_{k=1}^t terme_{ik}^2 + \sum_{k=1}^t terme_{jk}^2)}}$
<i>Overlap</i>	$Similarité(S_i, S_j) = \frac{\sum_{k=1}^t terme_{ik} terme_{jk}}{\min(\sum_{k=1}^t terme_{ik}, \sum_{k=1}^t terme_{jk})}$
<i>Asymmetric</i>	$Similarité(S_i, S_j) = \frac{\sum_{k=1}^t \min(terme_{ik}, terme_{jk})}{\sum_{k=1}^t terme_{jk}}$
<i>Minkowski</i>	$Similarité(S_i, S_j) = \sqrt[r]{\sum_{k=1}^t (terme_{ik} - terme_{jk})^r}$

TABLE 5.6 – Quelques mesures de similarité extraites de [170]

Dans les deux méthodes, les sémantiques pertinentes de la requête sont déterminées par rapport à un seuil α en dessous duquel les sémantiques ne pourront pas être considérées comme des sémantiques représentatives. Dans l'exemple de la table 5.7, nous considérons trois valeurs de seuil :

1. Si $\alpha = 1$ nous remarquons que pour les mesures de similarité Dice et Cosinus ne donnent aucune sémantique pertinente. Les mesures de similarité Overlap et Appart retiennent les sémantiques S_1, S_2, S_3 et S_5 comme le montre la figure 5.8. Ces deux mesures effectuent une recherche exacte sur l'ensemble des mots-clés fournis dans la requête. Les sémantiques obtenues contiennent tous les mots-clés de la requête. Notez que dans le cas de Overlap, si la valeur minimum (voir 5.7) correspond au nombre de mots-clés, la recherche correspond à une recherche exacte. En revanche, si le minimum correspond au nombre d'instance de la sémantique S_i , alors, bien que la valeur de similarité soit égale à 1, tous les mots-clés de la requête Rq n'ont pas de correspondance avec les instances de la sémantique S_i . Par exemple si on considère une sémantique S'_6 dont les instances sont $Inst(S'_6) = \{Forêt, Chien, Chasseur\}$, la similarité $Overlap(S'_6, Rq) = 1$.
2. Si $\alpha = 0$, les toutes sémantiques du corpus sont sélectionnées.
3. Si α est une valeur quelconque, comme par exemple la moyenne des similarités :
 - (a) Si α la moyenne des similarités de Dice est prise en compte pour déterminer les sémantiques représentatives, soit $\alpha = \frac{\sum_{i=1}^{10} Sim_{Dice}(S_i, Rq)}{10} = 0,49$, seules les sé-

Labels	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	k^{Rq}
Aigle	0	0	0	0	1	0	0	0	1	0	0
Après – midi	0	0	0	1	0	0	0	0	0	0	0
Arbre	0	1	0	0	0	0	1	0	1	0	0
Bâton	0	0	0	0	0	1	0	0	0	0	0
Brousse	0	0	0	0	0	1	1	0	0	0	0
Cabane	0	0	0	1	0	0	0	0	0	0	0
Calebasse	0	0	0	0	0	0	0	0	0	1	0
Chasseur	1	1	1	1	1	1	1	0	0	0	1
Chien	1	1	1	0	1	1	0	1	0	0	1
Enfant	0	0	0	0	0	0	0	0	0	1	0
Etang	0	0	0	0	0	0	0	1	0	0	0
Femme	0	0	0	0	0	0	0	0	0	1	0
Feu	0	0	1	0	0	0	0	0	0	0	0
Fleuve	0	0	0	0	0	0	0	0	0	1	0
Forêt	1	1	1	1	1	1	0	0	0	0	1
Fusil	1	1	1	1	1	0	1	0	0	0	1
Gardien	0	0	0	0	0	0	0	0	1	0	0
Gibecière	0	0	0	0	0	0	1	0	0	0	0
Gibier	1	0	1	0	0	0	0	0	0	0	0
Gourde	0	0	0	0	0	0	0	1	0	0	0
Maison	0	0	0	0	1	0	0	1	1	0	0
Matinée	0	1	0	0	0	0	0	0	0	0	0
Montagne	0	0	0	1	0	0	0	0	0	0	0
Perdrix	0	0	0	0	0	0	1	0	0	0	0
Prairie	0	0	0	0	0	0	0	1	0	1	0
Rivière	0	0	0	0	1	0	0	0	0	0	0
Sac	0	0	0	0	0	0	0	0	1	0	0
Soir	1	0	0	0	0	0	0	0	0	0	0
Tente	1	0	0	0	0	0	0	0	0	0	0
$Dice(Rq, S_i)$	0,72	0,80	0,80	0,60	0,72	0,66	0,40	0,22	0	0	–
$Cosinus(Rq, S_i)$	0,75	0,82	0,82	0,61	0,75	0,67	0,41	0,22	0	0	–
$Overlap(Rq, S_i)$	1	1	1	0,75	1	0,75	0,50	0,25	0	0	–
$Appar(Rq, S_i)$	1	1	1	0,75	1	0,75	0,50	0,25	0	0	–

TABLE 5.7 – Présentation des vecteurs de caractéristiques et calcul des similarités

mantiques S_1, S_2, S_3, S_4, S_5 et S_6 comme le montre la figure 5.8.

- (b) Si α la moyenne des similarités de Cosinus est prise en compte pour déterminer les sémantiques représentatives, soit $\alpha = \frac{\sum_{i=1}^{10} Sim_{Cosinus}(S_i, Rq)}{10} = 0,50$, seules les sémantiques S_1, S_2, S_3, S_4, S_5 et S_6 .
- (c) Si α la moyenne des similarités de Overlap est prise en compte pour déterminer les sémantiques représentatives, soit $\alpha = \frac{\sum_{i=1}^{10} Sim_{Overlap}(S_i, Rq)}{10} = 0,62$, seules les sémantiques S_1, S_2, S_3, S_4, S_5 et S_6 .
- (d) Si α la moyenne des similarités d'Appart est prise en compte pour déterminer les

sémantiques représentatives, soit $\alpha = \frac{\sum_{i=1}^{10} Sim_{Appart}(S_i, Rq)}{10} = 0,62$, seules les sémantiques S_1, S_2, S_3, S_4, S_5 et S_6 .

- (e) Si α la moyenne des similarités de toutes les quatre méthodes est prise en compte pour déterminer les sémantiques représentatives, soit $\alpha = \frac{\sum_{i=1}^{40} (Sim_{Dice}(S_i, Rq), Sim_{Cosinus}(S_i, Rq), Sim_{Overlap}(S_i, Rq), Sim_{Appart}(S_i, Rq))}{40} = 0,56$, seules les sémantiques S_1, S_2, S_3, S_4, S_5 et S_6 .

Seuil	Méthodes de Similarité	Résultats
$(Rq, \alpha = 1)$	Overlap, Appart	$\{S_1, S_2, S_3, S_5\}$
$(Rq, \alpha = 0)$	Overlap, Appart, Dice, Cosinus	$\{S_1, S_2, \dots, S_{10}\}$
$(Rq, \alpha = Moy(Sim))$	Overlap, Appart, Dice, Cosinus	$\{S_1, S_2, S_3, S_4, S_5, S_6\}$

TABLE 5.8 – Sémantiques résultat selon les seuil α

Cette étape nous a permis d'extraire les sémantiques représentative de la requête de l'utilisateur. L'objectif que nous nous sommes fixés est d'étendre la recherche à toutes les collections d'images du corpus. Ceci nous amène donc à créer des clusters de sémantiques autour de chacune des sémantiques représentatives. L'étape suivante va donc consister à réaliser des clusterings en nous basant sur le voisinage de chaque sémantique *résultats*.

5.3.5/ EXTRACTION DES SÉMANTIQUES VOISINES

L'extraction des sémantiques représentatives est essentiellement basée sur la comparaison des mots-clés de la requête utilisateur avec les instances du corpus. Ceci sous-entend que lorsque l'instance d'une sémantique n'est pas proposée, cette instance ne participe pas à l'extraction des sémantiques représentatives parce que seuls les mots-clés proposés dans la requête sont recherchés. L'extraction des sémantiques voisines va prendre en compte toutes les instances de chacune des sémantiques représentatives. Le but est de comparer les sémantiques représentatives avec toutes les sémantiques du corpus afin de déterminer leur voisinage comme le montre la figure 5.4.

5.3.5.1/ RAPPEL SUR LE CLUSTERING

L'objectif est de trouver des groupes d'images au sein d'un ensemble d'images. Les groupes d'images recherchés, communément appelés *clusters*, forment des ensembles homogènes d'images partageant des caractéristiques communes. Pour réaliser le clustering, on fait généralement appel à la notion de similarité entre les images parce qu'il est question d'évaluer à quelle point deux images sont similaires (ou dissimilaires) pour les regrouper ou les séparer. La mesure de comparaison peut être faite sur les caractéristiques physiques de l'image ou sur les caractéristiques sémantiques.

Il existe plusieurs techniques de clustering parmi lesquelles nous pouvons citer les méthodes basées sur les distances, les techniques basées sur une grille, les techniques probabilistes et les méthodes techniques. La technique basée sur les distances se focalise sur la distance entre les caractéristiques des images. Lorsque deux images sont proches suivant cette distance, ils doivent être groupés ensemble dans un même cluster. Les techniques basées sur une grille se

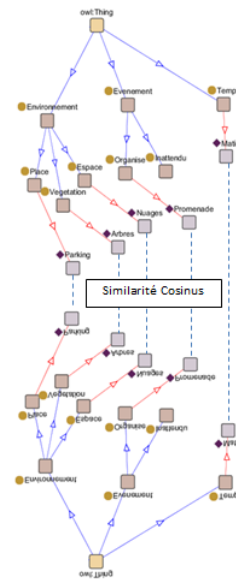


FIGURE 5.4 – Correspondance d'instances

basent sur la discrétisation de l'espace des données. L'espace est décomposé en un ensemble de cellule pour construire l'unité de la grille. L'objectif est de réduire l'explosion combinatoire des techniques basée sur la densité qui fait suite à l'augmentation du nombre d'objets. Les techniques probabilistes permettent d'estimer les paramètres de la loi de probabilité pour définir un modèle de mélange de lois pour représenter les différents clusters. Les techniques hiérarchiques permettent d'explorer des images à plusieurs niveaux de granularité. Elles sont basées sur la construction d'une hiérarchie de clusters, c'est-à-dire un arbre de clusters ou chaque nœud contient ses clusters *filis*, et les nœuds *frères* partitionne les données contenues dans leur parent.

Les techniques de clustering sont groupées dans deux approches : les approches supervisées et les approches non supervisées. L'approche supervisée est une technique d'apprentissage automatique où l'on cherche à produire automatiquement des règles à partir d'une base de données d'apprentissage contenant des « exemples » (en général des cas déjà traités et validés)². L'objectif de l'approche supervisée est principalement de définir des règles permettant de classer des images dans des classes à partir de variables qualitatives ou quantitatives caractérisant ces images. Dans l'approche non supervisée il y a en entrée un ensemble de données collectées. Ensuite le programme traite ces données comme des variables aléatoires et construit un modèle de « densités jointes » pour cet ensemble de données. L'approche non supervisée est une méthode d'apprentissage automatique. Il s'agit pour un logiciel de diviser un groupe hétérogène de données, en sous-groupes de manière à ce que les données considérées comme les plus similaires soient associées au sein d'un groupe homogène et qu'au contraire les données

2. http://fr.wikipedia.org/wiki/Apprentissage_supervisé

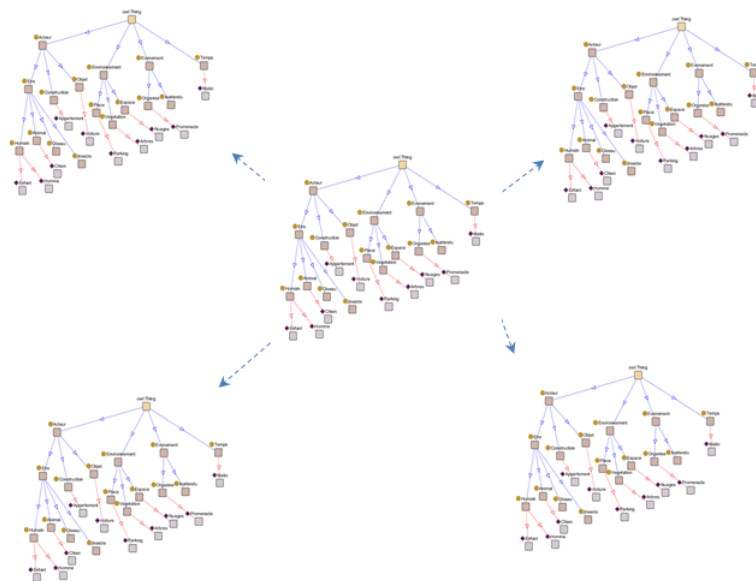


FIGURE 5.5 – Voisinage d'une sémantique

considérées comme différentes se retrouvent dans d'autres groupes distincts.

Le but de ces approches est de construire une partition d'un ensemble en des groupes données homogènes. Les objets de données au sein d'un groupe devraient être similaires, tandis que les objets de données affectés à différents groupes devraient être différents de façon significative. Une bonne méthode de regroupement permet de garantir une grande similarité intra-groupes et une faible similarité inter-groupes. La qualité d'un clustering dépend donc de la mesure de similarité.

Comme les opérations algébriques simples peuvent permettre d'obtenir une mesure de similarité, sans perte de généralité, nous allons utiliser la similarité cosinus pour créer des ensembles de sémantiques voisines au sémantiques *résultat* comme le montre la figure 5.5.

5.3.5.2/ VOISINAGE DE SÉMANTIQUES

L'extraction des sémantiques voisines est une étape de la technique de recherche d'images que nous proposons. Cette opération a pour but de rechercher les sémantiques voisines de chaque sémantique *résultat*. Pour déterminer les voisinages des sémantiques *résultat*, nous allons utiliser la similarité cosinus entre sémantiques.

Le calcul des similarités cosinus est fait entre le vecteur de chaque sémantique *résultat* $SemV(S_r)$ et le vecteur de chacune des sémantiques du corpus $SemV(S_i)$. La détermination des vecteurs est détaillée un peu plus haut (voir l'algorithme (b) de la table 5.1). Les similarités sémantiques entre les $SemV(S_i)$ et les $SemV(S_r)$ sont calculées et un seuil de similarité φ est défini pour sélectionner les sémantiques les plus proches de chaque $SemV(S_r)$. En dessous de ce seuil, la sémantique ne pourra donc pas être considérée comme voisine de la sémantique *résultat*. Le tableau 5.9 présente une liste de similarités sémantiques calculées par rapport à la

sémantique *résultat* S_1 . Dans cet exemple, le seuil φ est déterminé par rapport à la moyenne des similarités, soit $\varphi = Moy(Similarités) = 0,3907$.

Sémantiques	Similarités
S_6	0,755
S_4	0,245
S_3	0,484
S_8	0,577
S_2	0,042
S_5	0,132
S_{10}	0,387
S_7	0,894
S_9	0,001

TABLE 5.9 – Exemple de similarité cosinus

5.3.6/ FORMATION DE GROUPES DE SÉMANTIQUES : GRANULARITÉ

L'extraction des sémantiques voisines est essentiellement basée sur la comparaison des instances des sémantiques, ce qui ne prend pas en compte toutes les caractéristiques de l'ontologie d'application comme les relations entre les objets (concepts). Pour former les groupes de sémantiques au sein de chaque ensemble de sémantiques voisines, nous allons prendre la structure de l'ontologie d'application en compte.

La technique de granularité locale a pour but de comparer la structure des sémantiques (graphe) afin de générer des groupes $G^{(l)}$ de sémantiques similaires au sein de chaque ensemble tels que $G^{(l)} = \{G_1^{(l)}, G_2^{(l)}, \dots, G_t^{(l)}\}$. Elle consiste à déterminer les concepts des sémantiques voisines $S_v^{e_j}$ au sein de chaque ensemble e_j . C'est une technique qui permet de ne pas simplement se limiter à la comparaison des instances parce que même si les concepts peuvent ne pas avoir les mêmes instances, ils peuvent être identiques comme le montre la figure 5.6. Les deux flèches indiquent la similarité entre les concepts mais les instances qu'ils portent décrivent des objets différents.

5.3.6.1/ SIMILARITÉ SÉMANTIQUE PAR STRUCTURE HIÉRARCHIQUE

Plusieurs études ont été consacrées aux mesures de similarité entre des structures hiérarchiques. Ces études sont généralement axées sur une approche qui compare les arêtes des graphes. L'approche consiste à estimer la distance (la longueur des arêtes) entre les nœuds qui correspondent aux concepts qu'on veut comparer. La distance peut être mesurée par la distance géométrique entre les nœuds représentant des concepts. Nous pouvons citer entre autres :

- Les travaux de Budanitsky et al. dans [171] et les travaux de Fellbaum et al. dans [172]. Les mesures proposées utilisent la densité du réseau local (le nombre de liaisons de fils qui s'étend à partir d'un nœud parent), la profondeur d'un nœud dans la hiérarchie, le type de lien, etc ;
- Les travaux basés sur les mesures utilisant une ressource lexicale, comme WordNet. Ces mesures s'appuient souvent sur la structure d'un graphe orienté. Elles exploitent les vecteurs

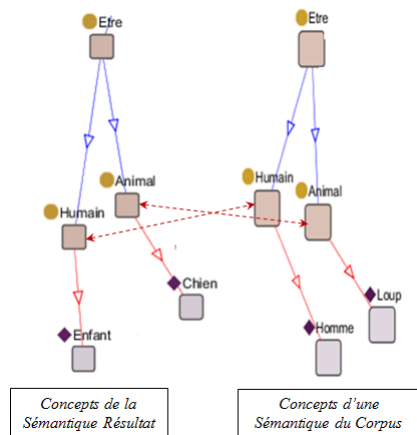


FIGURE 5.6 – Granularité locale

de cooccurrences entre les mots-clés contenus dans chaque concept. Plus précisément, dans une première étape, un espace de mots de taille P est constitué en prenant l'ensemble des mots significatifs utilisés pour définir l'ensemble des synsets de WordNet. Ensuite, chaque concept c_i est représenté par un vecteur $\vec{c_i w_i}$ de taille P , où chaque $i^{\text{ème}}$ élément de ce vecteur représente le nombre d'occurrence du $i^{\text{ème}}$ mot de l'espace des mots de la définition de c_i . La similarité sémantique entre les concepts c_i et c_j est mesurée en utilisant la similarité cosinus entre $\vec{c_i w_i}$ et $\vec{c_j w_j}$.

- Les travaux basés sur les mesures de l'emplacement d'un concept dans un groupe de nœuds sont également proposés dans la littérature. Ces mesures sont basées sur la granularité locale d'un concept et la spécificité commune des concepts concrets. La granularité est déterminée par le rapport entre la profondeur de ce concept dans une hiérarchie et la profondeur de la hiérarchie. La technique basée sur la spécificité commune est une mesure de l'information partagée entre deux concepts³.

Dans notre approche, nous optons pour la granularité des sémantiques. Elle consiste en effet à comparer la structure des sémantiques. La structure de chaque sémantique est parcourue et les relations entre les nœuds (concepts) sont extraites. Les structures seront parcourues grâce à un algorithme de parcours de graphe.

5.3.6.2/ RAPPEL DU PARCOURS DE GRAPHES

Un graphe c'est une structure de données, tout comme les tableaux. En effet, c'est une structure de données composée de nœuds inter-connectés entre eux par des arêtes. Chaque nœud contient une ou plusieurs données de n'importe quel type. Il existe deux façons classiques de représenter un graphe $g = (S, A)$: par un ensemble de listes d'adjacences, ou par une matrice d'adjacence. La représentation par listes d'adjacence est souvent préférée, car elle fournit un moyen peu encombrant de représenter les graphes peu denses. Cette représentation

3. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655943/>

est dans notre technique. Elle consiste en un tableau tab de $|S|$ listes, une pour chaque sommet de S . Pour chaque u de S , la liste d'adjacence $tab[u]$ est une liste chaînée des sommets v tel qu'il existe un arc (u, v) . Autrement dit, $tab[u]$ est constituée de tous les sommets adjacents de u dans G . Les sommets de chaque liste d'adjacence sont en général des chaînes selon un ordre arbitraire.

Il existe deux types principaux de parcours de graphes : le parcours en largeur et le parcours en profondeur.

- Le parcours en largeur est l'un des algorithmes de parcours les plus simple sur un graphe, et la base de nombreux algorithmes importants sur les graphes. Étant donné un graphe $g = (S, A)$ et un sommet origine s , le parcours en largeur emprunte systématiquement les arcs G pour "découvrir" tous les sommets accessibles depuis s . Il calcule la distance (le plus petit nombre d'arcs) entre s et tous les sommets accessibles. Il construit également une "arborescence en largeur" de racine s , qui contient tous les sommets accessibles depuis s . Pour tout sommet v accessible depuis s , le chemin reliant s à v dans l'arborescence en largeur correspond à un "plus court chemin" de s vers v dans G , autrement dit un chemin contenant le plus petit nombre d'arc. L'algorithme de parcours en largeur tient son nom au fait qu'il découvre d'abord tous les sommets situés à une distance k de s avant de découvrir tous les sommets situés à une distance $k + 1$ de s .
- Le parcours en profondeur est, comme son nom l'indique, de descendre plus "profondément" dans le graphe chaque fois que c'est possible. Lors d'un parcours en profondeur, les arcs sont explorés à partir du sommet s découvert le plus récemment et dont on n'a pas encore exploré tous les arcs incidents. Lorsque tous les arcs de s ont été explorés, l'algorithme "revient en arrière" pour explorer les arcs qui partent du sommet à partir duquel s a été découvert. Ce processus se répète jusqu'à ce que tous les sommets accessibles à partir du sommet origine initial aient été découverts. S'il reste des sommets non découverts, on en choisit un qui servira de nouvelle origine, et le parcours reprend à partir de cette origine. Le processus complet est répété jusqu'à ce que tous les sommets aient été découverts. Contrairement au parcours en largeur, pour lequel le sous-graphe de liaison forme une arborescence, le sous-graphe de liaison obtenu par un parcours en profondeur peut être composé de plusieurs arborescences, car le parcours peut être répété à partir de plusieurs origines⁴.

Dans le cadre de notre travail, nous avons choisi le parcours en profondeur. Le principe est le suivant. L'algorithme parcourt les graphes g_i et détermine les nœuds des graphes. Dans notre approche, tous les concepts sont liés par des relations. Cela s'explique par le fait que la présence d'une relation témoigne l'existence de deux concepts. Lorsque l'algorithme retrouve un concept (nœud), il parcourt les concepts reliés à ce concept. Le poids p_i de la relation $R(S_i) = \{(c_i, c_j)\}$ entre c_i et c_j du graphe G_i obtenu par la sémantique S_i est égal à 1 si les deux concepts c_i et c_j sont liés. Si un concept c_j du graphe n'est pas relié à c_i alors son poids p_i est égal à 0. Les p_i de chaque graphe sont stockés dans des tableaux à double entrées,

4. http://users.polytech.unice.fr/buffa/cours/java/internetEssi96_97/PROJETS/mangel/index.html

possédant en ligne comme en colonne les labels des concepts des graphes. La similarité entre deux graphes va consister à calculer le produit scalaire du poids des relations divisé par la moyenne du nombre d'occurrences de 1 des deux tableaux comme le montre la formule suivante.

$$Sim(g_i, g_j) = \frac{\sum(p_i \cdot p_j)}{Moy(n_i + n_j)};$$

n_i et n_j les occurrences de relations égales à 1 entre les graphes g_i et g_j .

L'algorithme 5.3 est un algorithme de parcours des graphes en profondeur.

Lorsque la similarité entre les deux graphes est égale à 1, alors ces deux graphes sont

Algorithme 5.3 Algorithme de parcours des graphes en profondeur : ParcoursGraphe

Soit G le graphe $g = (X, U)$

Soit x : sommet

Soit $\lambda(x) = \{y \in X; (x, y) \in U\}$, U l'ensemble des arcs

Soit $Visite[]$ tableau de sommet visité

Soit tab la table des poids

1: $y \leftarrow$ sommet; y est une variable

2: $tab[] \leftarrow 1$

3: $Visite[] \leftarrow x$

4: **for** ($y \in \lambda(x)$) **do**

5: **if** ($y \notin Visite[]$) **then**

6: $ParcoursGraphe(G, y)$

7: **end if**

8: **end for**

similaires et peuvent appartenir au même groupe. Dans le cas contraire, ils ne peuvent pas appartenir au même groupe.

Soit g_1, g_2 et g_3 représentant respectivement trois graphes des sémantiques S_1, S_2 et S_3 comme le montre la figure 5.7.

La similarité donnée entre g_1 et g_2 donne $Sim(g_1, g_2) = \frac{8}{(16/2)} = 1$, la similarité entre g_1, g_3 donne $Sim(g_1, g_3) = \frac{4}{(12/2)} = 0,66$ et la similarité entre g_2, g_3 donne $Sim(g_2, g_3) = \frac{4}{(12/2)} = 0,66$. Rappelons que pour que deux sémantiques S_i, S_j se retrouvent dans le même groupe, il faut qu'elle est la même granularité, c'est-à-dire que la similarité de leurs graphes respectifs soit égale à 1. Si nous considérons l'exemple de nos trois sémantiques S_1, S_2 et S_3 , nous constatons que les graphes g_1 et g_2 des sémantiques S_1 et S_2 ont une similarité égale à 1, alors ces deux sémantiques appartiendront au même groupe au sein de l'ensemble des sémantiques voisines.

La formation des groupes de sémantiques G_i au sein d'un ensemble de sémantiques voisines a pour objectif de regrouper les sémantiques qui partagent la même granularité locale. La figure 5.8 est un exemple de formation de groupe de sémantiques.

Nous introduisons ensuite une méthode pour déterminer les groupes de sémantiques les plus pertinentes. Déterminer les groupes de sémantiques pertinentes revient à définir

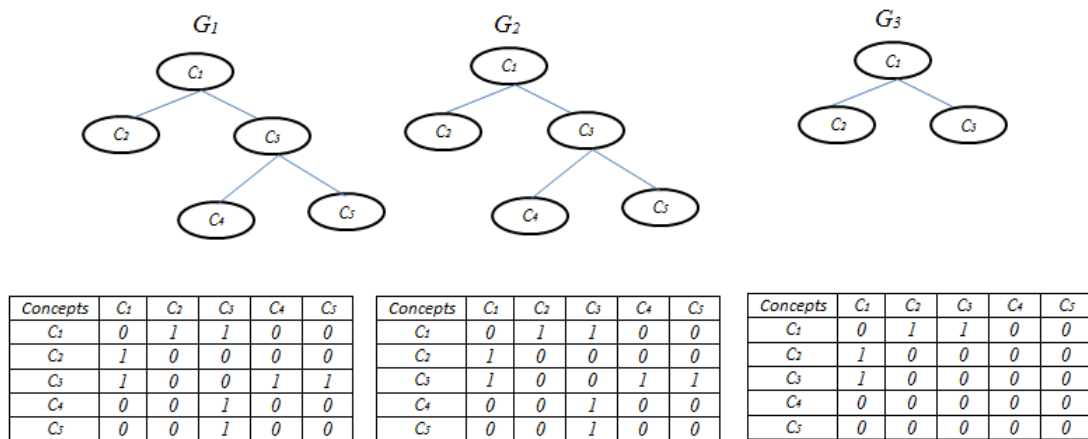


FIGURE 5.7 – Exemple de trois graphes et les poids des relations des concepts

un seuil en dessous duquel certaines sémantiques ne pourront pas être prises en compte. Dans notre démarche, la pertinence des groupes de sémantiques sera obtenue grâce à la matrice de confusion.

5.3.6.3/ MATRICE DE CONFUSION

La matrice de confusion, dans la terminologie de l'apprentissage supervisé, est un outil servant à mesurer la qualité d'un système de classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou de référence). Les

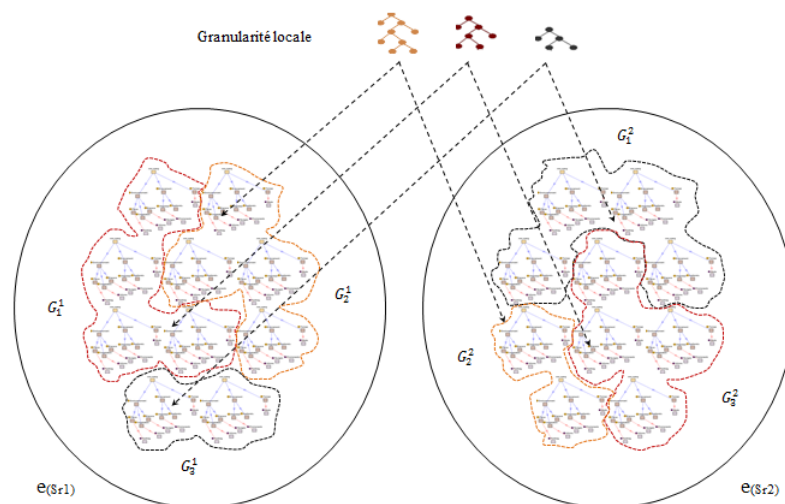


FIGURE 5.8 – Formation des groupes de sémantiques

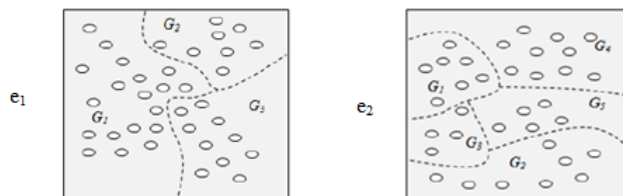


FIGURE 5.9 – Les deux ensembles et leurs groupes respectifs

données utilisées pour chacun de ces groupes doivent être différentes⁵. La matrice de confusion permet de se rendre compte de la répartition des sémantiques dans les ensembles de sémantiques voisines. Elle permet d'observer si les sémantiques d'un groupe d'un ensemble de sémantiques voisines ont été regroupées de manière similaire dans les autres ensembles de sémantiques voisines ou au contraire ont été réparties dans plusieurs groupes et dans quelle proportion. La matrice de confusion revient à comparer deux couples de groupes de sémantiques pour évaluer leur similarité et leur différence.

Dans notre travail, nous appliquons la matrice de confusion pour comparer les ensembles de sémantiques voisines à travers les groupes de sémantiques de chaque ensemble. Cette technique est inspirée des travaux menés dans [173] et [174]. La technique utilisée est basée sur une méthode collaborative appelée clustering collaboratif qui a pour objectif de raffiner mutuellement et de façon itérative plusieurs résultats de clustering pour obtenir des clusters plus fiables. L'avantage du clustering collaboratif est qu'il permet de comparer non seulement les objets entre eux, mais également les groupes auxquels ces objets appartiennent. La méthode utilise la matrice de confusion pour chaque couple de groupes d'objets provenant de deux ensembles d'objets. La matrice de confusion est une approche non supervisée. Pour mener à bien le clustering, nous sommes amenés à former à regrouper les sémantiques au sein de chaque ensemble de sémantiques voisines selon leurs granularité comme le montre la section précédente. Les groupes de sémantiques obtenus dans chaque ensemble de sémantiques voisines sont étiquetés comme le montre la figure 5.9).

La matrice de confusion δ entre deux groupes G^i et G^j de sémantiques est déterminée par :

5. http://fr.wikipedia.org/wiki/Matrice_de_confusion

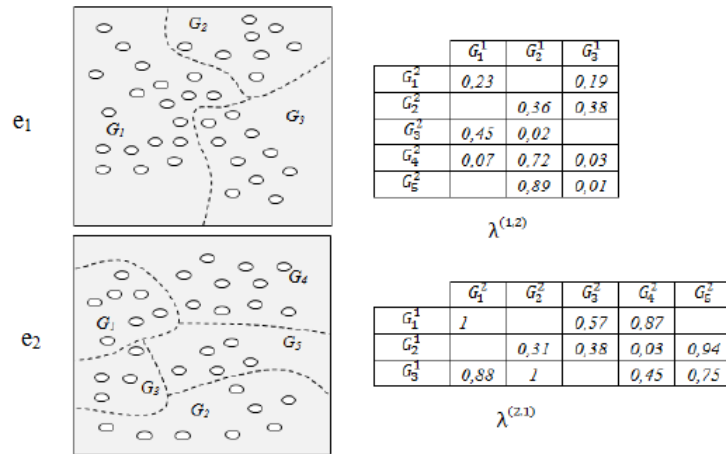


FIGURE 5.10 – La matrice de confusion

$$\delta^{(i, j)} = \begin{pmatrix} \lambda_{1, 1}^{(i, j)} & \dots & \lambda_{1, K^{(j)}}^{(i, j)} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \lambda_{K^{(i)}, 1}^{(i, j)} & \dots & \lambda_{K^{(i)}, K^{(j)}}^{(i, j)} \end{pmatrix} \quad \text{où } \lambda_{k, t}^{(i, j)} = \frac{|G_k^{(i)} \cap G_t^{(j)}|}{|G_k^{(i)}|}$$

La figure 5.10 présente deux ensembles de sémantiques voisines avec les groupes et le tableau du calcul de la matrice de confusion (e_1 et e_2 représente les deux ensembles de sémantiques voisines). La matrice de confusion compare deux groupes de sémantiques de deux ensembles de sémantiques voisines. Pour se faire deux matrices de confusion sont déterminées. Comme les ensembles de sémantiques voisines n'ont, d'une part, pas forcément le même nombre de groupes de sémantiques, et que d'autre part, le coefficient de la matrice sont normalisés par la taille des groupes de sémantiques d'un des deux ensembles, on obtient $\lambda^{(i, j)} \neq \lambda^{(j, i)}$ ([174]), ce qui oblige à calculer la matrice de confusion dans les deux sens $\lambda^{(i, j)}$ et $\lambda^{(j, i)}$ pour pouvoir comparer les deux résultats. La similarité peut donc être déterminée en utilisant les deux matrices. La similarité S est calculée par :

$$S(G_k^{(i)}, G_l^{(j)}) = \lambda_{k, l}^{(i, j)} \lambda_{l, k}^{(j, i)}$$

La similarité entre les groupes ne prend en compte que l'intersection entre les deux groupes considérés sans prendre en compte la répartition des sémantiques dans les autres groupes. Pour prendre en compte cette répartition, la similarité entre deux groupes est définie telle que :

$$S(G_k^{(i)}, G_l^{(j)}) = \varphi_k^{(i, j)} \lambda_{l, k}^{(j, i)}$$

où

$$\varphi_k^{(i, j)} = \sum_{r=1}^n (\lambda_{k, r}^{(j, i)})^2$$

La table 5.10 présente la similarité entre les groupes de sémantiques.

Groupes	G_1^1	G_2^1	G_3^1
G_1^2	1,321626		1,091778
G_2^2		2,068632	2,183556
G_3^2	2,585790	0,114924	
G_4^2	0,402234	4,137264	0,172386
G_5^2		5,114118	0,057462

a.) Similarité entre le groupe G_k^1 et l'ensemble e_2 ($S(G_k^1, e_2)$)

Groupes	G_1^2	G_2^2	G_3^2	G_4^2	G_5^2
G_1^1	1,882300		1,072911	1,637401	
G_2^1		0,583513	0,715274	0,056469	1,769362
G_3^1	1,656424	1,882300		0,847035	1,411725

b.) Similarité entre le groupe G_k^2 et l'ensemble e_1 ($S(G_k^2, e_1)$)

TABLE 5.10 – Calcul de similarité entre les groupes de sémantiques

5.3.6.4/ DÉTERMINATION DE CONFLIT ENTRE LES GROUPES

Si deux groupes ne sont pas en accord, il y a donc *conflit*. La détermination du conflit Cf consiste à rechercher tous les couples $(G_k^{(i)}, e_j)$, $i \neq j$ tel que $S(G_k^{(i)}, \mu(G_k^{(i)}, e_j)) < 1$ où μ est un coefficient de conflit qui est par défaut égal à 1, ce qui signifie que les groupes de sémantiques $G_k^{(i)}$ ne peuvent pas être exactement retrouvés dans l'ensemble des sémantiques voisines e_j . Cela ne veut pas dire que tous les éléments du groupe d'un ensemble seraient absents dans un autre ensemble. Lorsque les sémantiques d'un ensemble sont distinctes d'un groupe à un autre et d'un ensemble à un autre, on peut parler de conflit généralisé dans ces conditions, soit les images recherchées par l'utilisateur ne se trouvent pas dans la base ou la requête de l'utilisateur est à revoir. L'utilisateur doit reformuler sa requête. Les cellules colorées en gris du tableau de la figure 5.11 présente la liste des conflits entre les groupes de sémantiques (voir les cellules colorées en noir).

La détermination des conflits permet d'éliminer les sémantiques les moins pertinentes. Il s'agit des sémantiques ayant de très faible taux de similarité. Pour extraire les sémantiques les plus pertinentes, nous avons introduit la fonction de correspondance.

Groupes	G_1^1	G_2^1	G_3^1
G_1^2	1,321626		1,091778
G_2^2		2,068632	2,183556
G_3^2	2,585790	0,114924	
G_4^2	0,402234	4,137264	0,172386
G_5^2		5,114118	0,057462

a.) Conflit entre le groupe G_k^1 et l'ensemble e_2 ($Cf(G_k^1, e_2)$)

Groupes	G_1^2	G_2^2	G_3^2	G_4^2	G_5^2
G_1^1	1,882300		1,072911	1,637401	
G_2^1		0,583513	0,715274	0,056469	1,769362
G_3^1	1,656424	1,882300		0,847035	1,411725

b.) Conflit entre le groupe G_k^2 et l'ensemble e_1 ($Cf(G_k^2, e_1)$)

TABLE 5.11 – Liste des conflits entre les groupes

5.3.6.5/ FONCTION DE CORRESPONDANCE

La fonction de correspondance est une étape qui consiste à associer à un groupe de sémantique appartenant à un ensemble donné, un autre groupe qui lui est similaire dans un autre ensemble de sémantiques voisines. Il s'agit de la mise en correspondance d'un groupe de sémantiques $G_k^{(i)}$ d'un ensemble de sémantiques voisines e_i avec son groupe le plus similaire dans un autre ensemble e_j . La mise en correspondance est déterminée par la fonction f :

$$f(G_k^{(i)}, e_j) = \arg \max_{G_l^{(j)} \in e_j} S(G_k^{(i)}, G_l^{(j)})$$

avec $\arg \max$ est la valeur maximale de $S(G_k^{(i)}, G_l^{(j)})$ pour une valeur de $G_l^{(j)}$.

Les tableaux de la figure 5.12 présente le calcul de la similarité entre les groupes de sémantiques de deux ensembles de sémantiques voisines. Les valeurs qui seront retenues ($\arg \max$) sont les similarités se trouvant dans les colonnes colorées en noires. Le premier tableau présente le calcul de $f(G_k^{(1)}, e_2)$ et le deuxième tableau présente le calcul de $f(G_k^{(2)}, e_1)$.

La correspondance entre les deux ensembles de sémantiques voisines donne les diagrammes de la figure 5.11. Comme nous pouvons le remarquer dans la figure 5.11 qui présente les correspondances, tous les groupes de sémantiques $G_k^{(i)}$ de l'ensemble de sémantiques voisines e_1 n'ont pas de correspondance avec les groupes de sémantiques $G_1^{(2)}$ et $G_4^{(2)}$ de l'ensemble de sémantiques voisines e_2 . Par ailleurs, chacun des groupes de sémantiques de l'ensemble e_2 ont une correspondance avec les groupes de sémantiques de l'ensemble e_1 . Grâce à la fonction de correspondance, il est possible d'identifier si

Groupes	G_1^1	G_2^1	G_3^1
G_1^2	1,321626		1,091778
G_2^2		2,068632	2,183556
G_3^2	2,585790	0,114924	
G_4^2	0,402234	4,137264	0,172386
G_5^2		5,114118	0,057462

a.) Correspondance entre le groupe G_k^1 et l'ensemble e_2 ($f(G_k^1, e_2)$)

Groupes	G_1^2	G_2^2	G_3^2	G_4^2	G_5^2
G_1^1	1,882300		1,072911	1,637401	
G_2^1		0,583513	0,715274	0,056469	1,769362
G_3^1	1,656424	1,882300		0,847035	1,411725

b.) Correspondance entre le groupe G_k^2 et l'ensemble e_1 ($f(G_k^2, e_1)$)

TABLE 5.12 – Calcul de la fonction de correspondance f

un groupe de sémantiques d'un ensemble est accord avec un ensemble de sémantiques voisines.

5.3.7/ CLUSTER D'IMAGES

L'étape suivante consiste à proposer les groupes d'images à l'utilisateur, c'est-à-dire le cluster qui répond le mieux aux besoins de l'utilisateur. Proposer un cluster d'images revient à proposer un cluster des sémantiques similaires. La proposition de ce cluster passe par l'élimination des *conflicts*. Voyons ici *conflict* comme un groupe de sémantiques d'un ensemble donné qui n'a aucune correspondance dans un autre ensemble. Cela signifie que

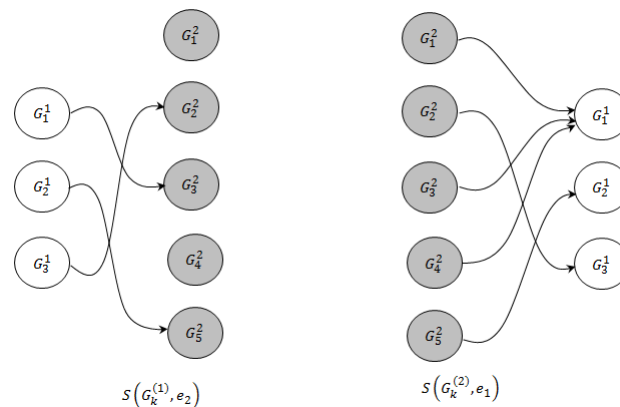


FIGURE 5.11 – Correspondance des groupes les plus similaires

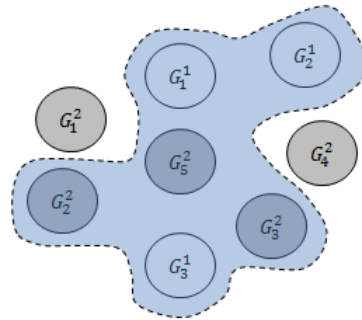


FIGURE 5.12 – Clusters de sémantiques

nous prendrons uniquement en compte dans le cluster tous les groupes de sémantiques sauf les conflits. La formation du cluster revient à rechercher l'intersection entre les deux fonctions $f(G_k^{(1)}, e_2)$ et $f(G_k^{(2)}, e_1)$. Les groupes les plus similaires sont donc représentés par $\sigma = f(G_k^{(1)}, e_2) \cap f(G_k^{(2)}, e_1)$. La figure 5.12 présente les clusters obtenus grâce à l'intersection entre les deux fonctions de correspondance. Les images représentées par ces sémantiques seront affichées à l'utilisateur.

5.3.8/ RÉSUMÉ

Le système de recherche d'images que nous avons proposé est basé sur une technique qui permet aux utilisateurs de disposer d'une collection d'images similaires à partir d'une requête de mots-clés et de relations. La technique combine trois techniques. La première technique permet d'extraire les sémantiques les plus représentatives de la requête de mots-clés de l'utilisateur. La deuxième technique permet de rechercher les sémantiques voisines à chaque sémantique représentative grâce à la spécificité commune. La troisième technique est le clustering. Le clustering est divisé en trois parties. La première partie est l'extraction des ensembles de sémantiques voisines. Elle est basée sur la similarité cosinus. La deuxième partie est le regroupement des sémantiques semblables au sein de chaque ensemble de sémantiques voisines. Elle est basée sur la granularité locale. La troisième partie est consacrée à la détermination des clusters des sémantiques les plus pertinentes. Deux méthodes sont utilisées pour former le cluster de sémantiques : la matrice de confusion et la fonction de correspondance.

Une fois les clusters d'images obtenus, ils sont proposés à l'utilisateur. Celui-ci n'a qu'à sélectionner l'image qui répond mieux à ses besoins. L'étape suivante consiste à mémoriser les sémantiques préférées par les utilisateurs dans le but de leur faire des recommandations personnalisées plus tard. La section suivante présente l'approche de recommandation des sémantiques.

5.4/ RECOMMANDATION DE SÉMANTIQUES

Dans cette section, nous présentons les différentes étapes de la technique de recommandation des sémantiques que nous avons mise en place.

5.4.1/ INTRODUCTION

Le système recommandations de sémantique permet de faire des recommandations de sémantiques personnalisées en se basant les préférences des utilisateurs. Ceci sous-entend que l'historique des recherches menées par les utilisateurs doit être pris en compte. Afin de mener à bien la recommandation des sémantiques, nos travaux se sont focalisés sur les profils des utilisateurs. Nous en avons deux : les profils dynamiques et les profils statiques. Le profils statiques et dynamiques sont détaillés dans les sections qui vont suivre. Ce chapitre est consacré à notre dernière approche : le système de recommandation des sémantiques. Il s'agit notamment de l'intégration d'un système de recommandation au sein d'un système de recherche sémantique des images.

Un système de recommandation est intégré dans *Emergsem* pour effectuer des recommandations personnalisées de sémantiques aux utilisateurs, tout en s'adaptant en permanence au besoin en information de chacun. L'approche proposée dans [175] semble la mieux adaptée à notre système. C'est une approche qui pondère un ensemble de concepts des ontologies de domaine. Cet ensemble est obtenu en collectant les concepts choisis par les utilisateurs pour la recherche des articles. Grâce à cette technique, les préférences des utilisateurs seront déterminées par le système et par conséquent, des recommandations personnalisées pourront être adressées à chaque utilisateur. Pour mettre en place cette technique dans *Emergsem*, la gestion des profils des utilisateurs est primordiale, ce qui permet de choisir les informations qu'il faut recommander. Ce que nous retenons des précédents travaux, c'est que la comparaison entre les profils donne lieu à la formation de regroupements d'utilisateurs proches les uns des autres, des regroupements qu'on appelle généralement "communautés" (voir [176] et [151]). Alors, on peut dire que la notion de communauté est un des facteurs clés dans un système de recommandation. Cependant, nous observons que le positionnement des utilisateurs est une tâche particulièrement difficile. On peut citer deux difficultés dans le positionnement d'utilisateurs : l'absence de valeur pour un ou plusieurs critères, et l'existence de valeur douteuse. Ces problèmes s'aggravent dans les systèmes de recommandation existants où les communautés sont produites comme une étape de calcul intermédiaire pour générer des recommandations (voir [154] et [176]). La formation de la communauté des utilisateurs est le but visé par certaines approches pour faciliter la tâche de positionnement [177] et à terme de produire de meilleures recommandations.

5.4.2/ RECOMMANDATION DES SÉMANTIQUES

La figure 5.13 présente l'architecture de l'approche de recommandation des sémantiques. Elle se compose des étapes suivantes, qui sont détaillées dans les sections ci-

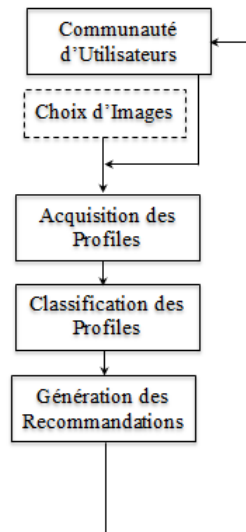


FIGURE 5.13 – Approche de recommandation

dessous :

- L'acquisition des profils est une étape qui consiste à extraire les profils des utilisateurs (les profils statiques et les profils dynamiques) ;
- La classification des profils est une étape qui permet de grouper les profils selon leur ressemblance. Pour la classification des profils, nous utilisons l'algèbre de Galois. L'avantage de l'utilisation de l'algèbre de Galois est qu'il permet de faire un groupage exhaustif.
- La génération de la liste des recommandations est une étape qui permet de mettre les sémantiques recommandées à la disposition des utilisateurs.

5.4.3/ ACQUISITION DES PROFILS

L'acquisition des profils est une étape de la recommandation qui permet d'extraire les informations concernant les utilisateurs. Elle a pour but de prendre connaissance des différentes interactions des utilisateurs avec le système et surtout de se faire une idée de leurs préférences. Son objectif est de faciliter le regroupement ou la classification des profils d'utilisateurs afin d'effectuer des recommandations ciblées à la communauté d'utilisateurs. Pour cela, il faut constituer une communauté d'utilisateurs.

5.4.3.1/ FORMATION DE LA COMMUNAUTÉ D'UTILISATEURS

Les systèmes collaboratifs se basent généralement sur la formation de communautés pour accomplir des tâches complexes. La collaboration a une influence sur la capacité qu'a un système de tirer parti des actions collectives de personnes. Comprendre l'impact de la conception, de la structure organisationnelle et de la dynamique peut aider à fournir une meilleure gestion des systèmes complexes. La formation de communauté est un type

de dynamique organisationnelle qui est affecté par la structure du réseau. Au cours des dernières années, deux tendances interdisciplinaires sont apparues et ont un grand impact sur l'étude des organisations : la modélisation informatique basée sur les internautes et l'étude de la structure et de la dynamique des réseaux complexes. La modélisation informatique des internautes a touché de nombreuses disciplines, notamment la physique, la chimie, la biologie, l'écologie, les sciences sociales, la médecine et l'informatique. L'impact des structures complexes de réseaux sur la compréhension du comportement des systèmes est tout aussi diversifié.

Dans *Emergsem*, nous ferons mention de la modélisation des internautes. Le travail présenté ici se concentre sur la formation de communautés des internautes. Pour former notre communauté d'internautes, nous avons choisi un modèle de système motivé par des travaux antérieurs de formation de communautés. Le modèle fournit un environnement de formation de communauté où la communauté se forme spontanément de manière totalement décentralisée. Il facilite la formation d'une communauté d'internautes par leur intégration dans un réseau social. Le but de cette formation est de produire des tâches partagées. Les internautes de la communauté sont intégrés dans un réseau social qui limite l'ensemble des communautés possibles d'internautes : en particulier, pour qu'un internaute soit dans une communauté, il doit avoir un lien social avec au moins un autre internaute de la communauté.

Lorsque les internautes s'inscrivent, ils fournissent des informations personnelles suivant des paramètres prédéfinis dans le système d'informations. Parmi ces renseignements, on retrouve bien souvent des informations relatives à l'identité des internautes (nom, prénom, année de naissance, etc.), des informations relatives à la profession, à la situation géographique et bien souvent des informations concernant les préférences des internautes. De toutes ces informations, une ou plusieurs informations sont utilisées pour grouper les internautes.

Nous constatons que les communautés dans la plupart des systèmes sont généralement formées sur un seul critère, par exemple les évaluations d'internautes. Cependant, il existe plusieurs critères sur lesquels la formation de communautés pourrait s'appuyer. Dans ce cas précis, si on crée un espace de communautés pour chacun de ces critères, les utilisateurs sont évidemment associés très différemment les uns aux autres selon le critère choisi. Dans ce contexte de variété de critères, la question du choix du critère pour former les communautés se pose. Nous pensons qu'il est réducteur de regrouper les utilisateurs sur un seul critère : l'action des internautes. Pour se faire une idée sur l'action des internautes, il faut absolument connaître les informations personnelles des utilisateurs auxquelles ces actions seront associées.

Dans *Emergsem* deux profils d'internautes se distinguent : les profils statiques et les profils dynamiques.

- Les profils statiques représentent toute donnée personnelle des utilisateurs. Il peut y

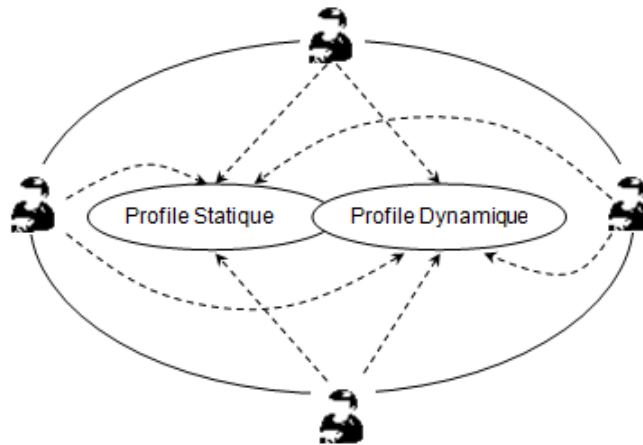


FIGURE 5.14 – Formation de la communauté des utilisateurs

être définies les informations personnelles des utilisateurs comme par exemple sa situation professionnelle. Cette partie ne nécessite pas de grands moyens puisque les utilisateurs, avant d'utiliser le système sont amenés à créer un compte et par conséquent à fournir ces informations. Ce n'est qu'après cette étape qu'ils pourront disposer d'un compte et bénéficier des services du système.

- Les profils dynamiques viennent en deuxième position. Ils sont constitués des données qui représentent les besoins, les intérêts et les buts de l'utilisateur. Autrement dit les préférences des utilisateurs. Cette partie sera construite par le système suite aux interactions des utilisateurs avec le système, c'est pourquoi on l'appelle aussi le modèle actif. Pour y arriver, le système a besoin notamment de collecter les données relatives aux choix sémantiques de chaque utilisateur. L'analyse de ces données permet ensuite de construire le modèle utilisateur qui va être utilisé par le système pour recommander les sémantiques jugées pertinentes à chaque utilisateur.

La communauté se compose de n utilisateurs et peut être représentée par $C_{U_n} = \{U_1(\text{Statique}, \text{Dynamique}), U_2(\text{Statique}, \text{Dynamique}), \dots, U_n(\text{Statique}, \text{Dynamique})\}$, où chaque utilisateur peut être considéré comme un nœud unique dans un réseau social comme le montre la figure 5.14.

L'acquisition des profils d'utilisateurs intervient après la phase de recherche des images, c'est-à-dire après que les profils statiques et les profils dynamiques soient disponibles dans le système comme le montre la figure 5.15. La disponibilité des profils statiques révèle que les utilisateurs se sont inscrits (les utilisateurs ont fourni des informations de connexion) et la présence des profils dynamique montre que des sémantiques ont été utilisées pour rechercher des images dans le système. Les profils statiques et les profils dynamiques de tous les utilisateurs sont récupérés.

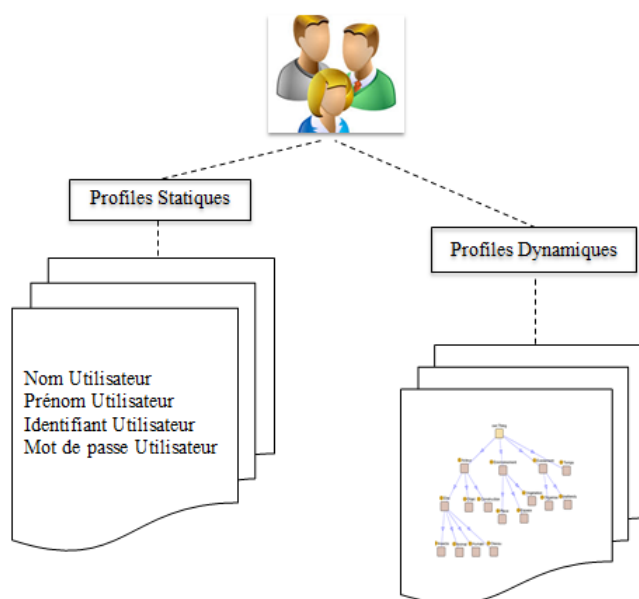


FIGURE 5.15 – Acquisition des profils d'utilisateurs

La section suivante présente la méthode de classification qui nous avons utilisée pour classifier les profils des utilisateurs.

5.4.4/ CLASSIFICATION DES PROFILS : ALGÈBRE ET TREILLIS DE GALOIS

La classification est une technique de regroupement d'informations similaire relevant de l'apprentissage (supervisé ou non supervisé). Plusieurs techniques sont proposées dans la littérature. Quelques unes sont présentées dans la chapitre 2.

La classification des profils d'utilisateurs est une étape du système de recommandation qui à pour objectif de regrouper les profils similaires au sein d'une même classe pour faciliter la recommandation des sémantiques. Rappelons que la recommandation des sémantiques aux utilisateurs se fait par rapport aux deux profils (statiques et dynamiques). Le système identifie les profils dynamiques (les sémantiques utilisées par l'utilisateur) associés à chaque profil statiques et recommande ces profils dynamiques aux utilisateurs toutes les fois qu'ils souhaitent rechercher une nouvelle image. Si les profils recommandés répondent à leurs besoins, les utilisateurs les utiliseront pour obtenir systématiquement les images, dans le cas contraire, ils devront suivre les différentes phases de recherche d'une image (présentées un peu plus haut dans ce chapitre). Ce que nous allons apprendre de la classification des profils est que la comparaison entre les profils de la communauté d'utilisateurs conduit à la formation de groupes de profils proches. Donc nous pouvons dire que la notion de communauté est un facteur clé dans un système de recommandation pour produire des recommandations généralisées ou ciblées [177].

Il est clair que le positionnement des utilisateurs dans le système dépend de façon cruciale de leur profils dynamiques. Les profils dynamiques de chaque utilisateur reflètent les différentes interactions du l'utilisateur avec le système et évoluent à chaque fois que celui-ci recherche une image.

Dans cette thèse, nous avons choisi l'algèbre de Galois pour classifier les profils d'utilisateurs.

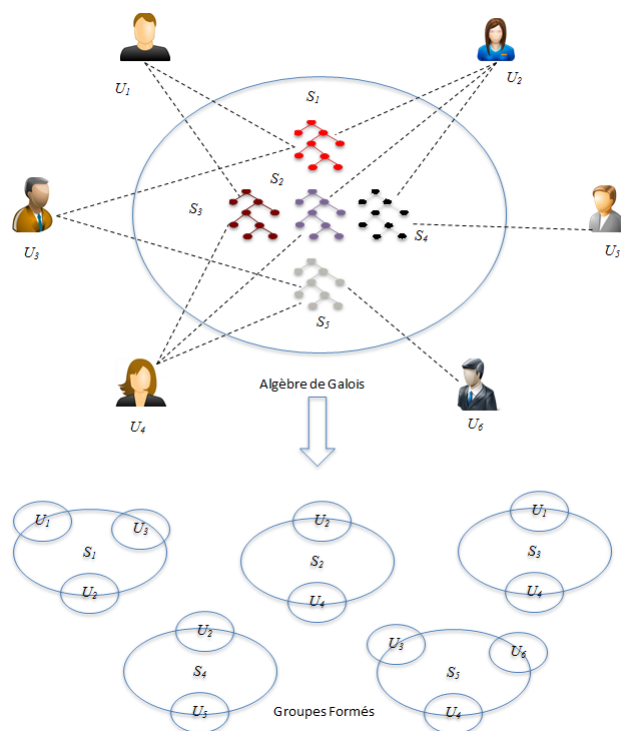
5.4.4.1/ ALGÈBRE DE GALOIS

L'algèbre de Galois est une méthode incrémentale de construction du treillis de concepts. L'algèbre de Galois présente en effet un énorme avantage : dès que l'on démontre une propriété relative à une structure donnée (un groupe, par exemple), on peut utiliser cette propriété sans la redémontrer chaque fois que l'on reconnaît cette structure (groupe de nombres, de fonctions, de vecteurs, etc.). Galois a ainsi ouvert la voie à une mathématique économe en démonstrations et unificatrice (les structures ne dépendent plus des objets, mais des relations entre les objets). Les structures, présentes dans tous les domaines des mathématiques, offrent une lisibilité sous la forme d'une nouvelle grille de classification⁶. L'utilisation de l'algèbre de Galois va permettre de regrouper les profils similaires afin de pouvoir faire des recommandations ciblées comme le montre l'image 5.16. L'image présente six (6) utilisateurs ($U_1, U_2, U_3, U_4, U_5, U_6$) avec l'historique des sémantiques (S_1, S_2, S_3, S_4, S_5) et les groupes de profils formés.

L'algèbre de Galois se base sur l'analyse de concepts formels [178], [179], [180], [181] et [182]. Un treillis de Galois peut regrouper, les objets de manière exhaustive dans les classes, appelé "concept formel", en utilisant leurs propriétés partagées. Un treillis est généralement basé sur une matrice booléenne, appelé contexte de la matrice et noté C , dont les lignes représentent un ensemble d'objets O qu'on veut décrire et en colonnes, un ensemble d'attributs A que ces objets ont ou n'ont pas. Le principe est le suivant. La possession de l'attribut $a \in A$ par l'objet $o \in O$ prouve qu'il existe une relation I entre eux : aIo . Le triplet $K = (O, A, I)$ est appelé *contexte formel*. A l'ensemble $X \subset O$ est l'ensemble des attributs possédés par tous les objets X et est donné par la fonction : $f(X) = \{a \in A \mid \forall o \in X, oIa\}$. Inversement l'ensemble $Y \subset A$ est l'ensemble des objets possédés par tous les attributs Y et est donné par la fonction : $g(Y) = \{o \in O \mid \forall a \in Y, oIa\}$. Le couple (f, g) est appelé "Connexion de Galois". Un concept formel est alors défini par tout couple $C = (X, Y) \subset (O * A)$, tel que les objets de X soient les seuls à posséder tous les attributs de Y , en d'autres termes $(X * Y)$ forme, à deux permutations près de O et de A , un rectangle maximal dans C , c'est-à-dire que $f(X) = Y$ et $g(Y) = X$.

Considérons la table 5.13 pour illustrer cette cette approche. La table présente une description des profils de quatre utilisateurs. Cette description est basée sur un extrait de la liste des sémantiques choisies par les utilisateurs.

6. http://www.pourlascience.fr/ewb_pages/a/article-l-algebre-apres-galois-28891.php



Légende : $S \rightarrow$ Sémantique, $U \rightarrow$ Utilisateur

FIGURE 5.16 – Apport de l'algèbre de Galois dans *Emergsem*

Profils Dynamiques	Profils Statiques			
	Utilisateur 1	Utilisateur 2	Utilisateur 3	Utilisateur 4
Sémantique 1	x			x
Sémantique 2		x	x	
Sémantique 3	x	x	x	
Sémantique 4				x
Sémantique 5			x	
Sémantique 6		x		
Sémantique 7		x		x

TABLE 5.13 – Table de contexte

Soit l'ensemble $X = \{\text{Sémantique 2}, \text{Sémantique 3}\}$ donne un concept car $f(X) = \{\text{Utilisateur 2}, \text{Utilisateur 3}\} = Y$ et $g(Y) = X$, et ce concept formel est donc $(\{\text{Sémantique 2}, \text{Sémantique 3}\}, \{\text{Utilisateur 2}, \text{Utilisateur 3}\})$ comme le montre la figure de la table 5.14.

Lorsque nous considérons l'ensemble $X' = \{\text{Sémantique 1}, \text{Sémantique 4}\}$, nous constatons qu'il ne donne pas un concept formel car $f(X') = \{\text{Utilisateur 4}\} = Y'$

Profils Dynamiques	Profils Statiques			
Sémantiques	Utilisateur 1	Utilisateur 2	Utilisateur 3	Utilisateur 4
Sémantique 1	x			x
Sémantique 2 \Rightarrow		\Leftarrow x	x	
Sémantique 3 \Rightarrow	x	\Leftarrow x	x	
Sémantique 4				x
Sémantique 5			x	
Sémantique 6		x		
Sémantique 7		x		x

TABLE 5.14 – Formation ce concept formel

et $g(Y') = \{Sémantique\ 1, Sémantique\ 4, Sémantique\ 7\}$, mais par contre ce dernier ensemble, lui, donne le concept $\{Sémantique\ 1, Sémantique\ 4, Sémantique\ 7\}, \{Utilisateur\ 4\}$ comme le montre la table 5.15.

Profils Dynamiques	Profils Statiques			
Sémantiques	Utilisateur 1	Utilisateur 2	Utilisateur 3	Utilisateur 4
Sémantique 1 \Rightarrow	x			x
Sémantique 2		x	x	
Sémantique 3	x	x	x	
Sémantique 4 \Rightarrow				x
Sémantique 5			x	
Sémantique 6		x		
Sémantique 7		x		x

$$(a) f(X') = f(Sémantique\ 1, Sémantique\ 4) = \{Utilisateur\ 4\} = Y'$$

Profils Dynamiques	Profils Statiques			
Sémantiques	Utilisateur 1	Utilisateur 2	Utilisateur 3	Utilisateur 4
Sémantique 1	x			\Leftarrow x
Sémantique 2		x	x	
Sémantique 3	x	x	x	
Sémantique 4				\Leftarrow x
Sémantique 5			x	
Sémantique 6		x		
Sémantique 7		x		\Leftarrow x

$$(b) g(Y') = \{Sémantique\ 1, Sémantique\ 4, Sémantique\ 7\}$$

TABLE 5.15 – (a) \neq (b) : Absence de concept formel

5.4.4.2/ TREILLIS DE GALOIS

Un treillis de Galois permet de regrouper des objets en classes, appelées "concepts", en utilisant leur propriétés partagées. La figure 5.17 est le treillis réalisé grâce au tableau

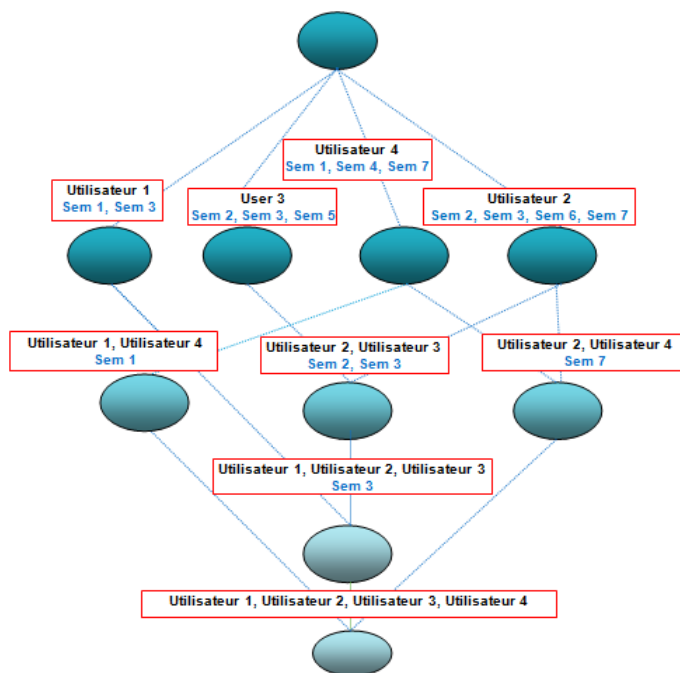


FIGURE 5.17 – Treillis de Galois

de contexte 5.13.

En lisant le treillis de Galois du bas vers le haut, nous distinguons plusieurs concepts. Il s'agit des notamment des profils statiques auxquels un certain nombre de profils dynamiques pourront être recommandés. Le premier concept $\{ \text{Sémantique 3} \}$, $\{ \text{Utilisateur 1, Utilisateur 2, Utilisateur 3} \}$ montre que l'objet "Sémantique 3" pourra être recommandé à la fois aux attributs "Utilisateur 1", "Utilisateur 2" et "Utilisateur 3". De même, les attributs "Utilisateur 2", "Utilisateur 4" peut se voir recommander l'objet "Sémantique 3", les attributs "Utilisateur 2", "Utilisateur 3" se voient recommander les objets "Sémantiques 2", "Sémantiques 3" et les attributs "Utilisateur 1", "Utilisateur 4" se voient recommander l'objet "Sémantique 1", etc.

L'un des avantages de la classification sur la base de treillis de Galois est que pour une table de contexte formel donné, le réseau résultant est unique et exhaustif. Cette classification nous permettra de trouver tous les groupes de profils statiques en relation avec les groupes de profils dynamiques et de les représenter de manière similaire.

5.4.5/ GÉNÉRATION DE LA LISTE DES RECOMMANDATIONS

La génération de la liste des recommandations est un processus par lequel des recommandations de sémantiques sont adressées de façon personnalisée aux utilisateurs. Lorsque les groupes sont constitués, un ou plusieurs groupes de sémantiques peuvent être recommandés aux utilisateurs. On peut avoir une ou plusieurs sémantiques appartenant à des groupes différentes comme par exemple la *Sémantique 3* qui peut être

recommandée aux utilisateurs *Utilisateur 1*, *Utilisateur 2* et *Utilisateur 3* et un peu plus haut du treillis, les sémantiques *Sémantique 2* et *Sémantique 3* sont recommandées aux utilisateurs *Utilisateur 2* et *Utilisateur 3*. La liste de recommandations est constituée de sorte que la même sémantique ne soit pas doublement recommandée à un utilisateur au même moment. Une fois les doublons éliminés, les sémantiques peuvent être recommandées aux utilisateurs auxquels elles sont associées. Mais une question se pose. Suffit-il qu'un utilisateur choisisse une sémantique pour rechercher une image pour que cette sémantique lui soit recommandée ? Si tel est le cas, des recommandations se feraient à chaque fois que les utilisateurs auront choisi une sémantique. Pour affiner les recommandations, nous proposons un coefficient de recommandation.

5.4.5.1/ LE COEFFICIENT DE RECOMMANDATION

Le coefficient de recommandation est un coefficient qui permet d'évaluer les profils dynamiques. Il consiste à comparer les occurrences de chaque profil dynamique associé à un utilisateur donné à la moyenne des occurrences de ce même profil dynamique associé aux utilisateurs de la communauté. L'objectif, c'est de recommander les profils dynamiques les plus fréquents. Cela sous-entend qu'un travail préliminaire est nécessaire : il s'agit d'une part de la classification des profils dynamiques (l'algèbre de Galois nous aide à réaliser cette classification), et d'autre part le calcul des moyennes des occurrences de chaque profil dynamique. Le tableau 5.16 présente les occurrences de profils dynamiques et le calcul de leur moyenne.

Le calcul des coefficients de recommandation se fait par le rapport entre les oc-

<i>Utilisateurs</i>	U_1	U_2	U_3	U_4	U_5	U_6	<i>Moyenne</i>
<i>Sémantique 1</i>	08	07	06	06	02	01	5,000
<i>Sémantique 2</i>	06	10	07	08	13	06	8,500
<i>Sémantique 3</i>	08	09	11	07	05	10	8,333
<i>Sémantique 4</i>	06	04	07	09	02	09	6,167

TABLE 5.16 – Occurrences des profils dynamiques

currences du profil dynamique (sémantique) de chaque utilisateur et la moyenne des occurrences du profil dynamique du groupe auquel appartient chaque utilisateur. Si le coefficient d'une occurrence du profil dynamique est supérieur ou égal à 1, cette sémantique peut être recommandée à l'utilisateur. Dans le cas contraire, elle ne pourra donc pas être recommandée à l'utilisateur. Cette méthode nous permet de faire des recommandations basées sur les préférences réelles des utilisateurs et non sur les choix des utilisateurs. Il ne suffit donc pas qu'un utilisateur choisisse une sémantique pour se voir recommander cette sémantique. La recommandation dépend de l'instant t où l'utilisateur recherche une image avec une sémantique. L'utilisateur qui se voit recommander une sémantique à un instant t_1 peut ne pas se voir recommander la même sémantique à un instant t_2 . Le contraire est aussi possible. Autrement dit, ce n'est pas parce qu'il

ne se voit pas recommander une sémantique à un instant t_1 (après avoir utilisé cette sémantique pour rechercher une image) qu'il ne se verra pas recommander cette sémantique quelques jours après. Au fur et à mesure qu'il choisira cette sémantique, son coefficient de correspondance pourrait s'améliorer. Cette méthode permet aussi d'avoir une idée précise sur le comportement des utilisateurs. La figure 5.17 donne un aperçu du calcul du coefficient de correspondance. Les cellules du tableau colorées en noir sont celles qui représentent les coefficients qui respectent le critère de correspondance (coefficient ≥ 1). La *Sémantique 1* ne sera donc plus recommandée à tous les utilisateurs l'ayant choisi. Elle sera recommandée aux utilisateurs U_1 , U_2 , U_3 et U_4 . La *Sémantique 2* sera recommandée aux utilisateurs U_2 et U_5 , la *Sémantique 3* recommandée aux utilisateurs U_2 , U_3 et U_6 et la *Sémantique 4* aux utilisateurs U_3 , U_4 et U_6 .

Utilisateurs	U_1	U_2	U_3	U_4	U_5	U_6
<i>Sémantique 1</i>	1,600	1,400	1,200	1,200	0,400	0,200
<i>Sémantique 2</i>	0,700	1,176	0,824	0,941	1,529	0,824
<i>Sémantique 3</i>	0,960	1,080	1,320	0,840	0,600	1,200
<i>Sémantique 4</i>	0,973	0,649	1,135	1,459	0,324	1,459

TABLE 5.17 – Tableau de calcul des coefficients de recommandation

5.4.6/ RÉSUMÉ

Dans cette section, nous avons proposé une approche de recommandation basée sur quatre points essentiels. Le premier point est la formation de la communauté d'utilisateurs qui permet de représenter les utilisateurs au sein du système. Le deuxième point est l'acquisition des profils des utilisateurs. Deux profils se distinguent : les profils statiques et les profils dynamiques. Le troisième point est la classification des profils. Les profils sont regroupés grâce à l'algèbre de Galois. Le dernier point est la génération des listes de recommandations qui permet d'adresser des recommandations personnalisées aux utilisateurs.

5.5/ CONCLUSION

Nous venons de présenter dans ce chapitre une technique de recherche sémantique d'images basée sur le clustering et la recommandation. La recherche d'images basée sur le clustering comporte trois étapes. La première étape est l'extraction des sémantiques représentatives de la requête de l'utilisateur. Elle permet de calculer la similarité entre les vecteurs des caractéristiques sémantiques de la requête et des sémantiques du corpus en comparant les mots-clés de la requête aux instances des sémantiques du corpus. Pour cela, deux méthodes générales sont proposées : la méthode basée sur l'appariement et la méthode basée sur les mesures de similarité comme nous les avons présentés dans la section 5.3.4. Le but de l'extraction des sémantiques représentatives est d'extraire du

corpus les sémantiques qui répondent mieux à la requête de l'utilisateur. La deuxième étape est l'extraction des sémantiques voisines de chaque sémantique représentative. L'extraction du voisinage de chaque sémantique représentative est basée sur la comparaison des vecteurs des caractéristiques des instances des sémantiques. La mesure de similarité cosinus est utilisée dans ce cas. La troisième étape est la formation des clusters de sémantiques. Le clustering des sémantiques repose sur deux méthodes, la formation des groupes de sémantiques au sein de chaque sémantique voisine basée sur la granularité et sur la matrice de confusion. Ceci permet de fournir à l'utilisateur des clusters des sémantiques les plus pertinentes. La recommandation des sémantiques quant à elle comporte trois étapes. La première partie de la recommandation est l'acquisition des profils d'utilisateurs. Deux types de profils sont distingués : les profils statiques et les profils dynamiques. Dans la deuxième partie, la classification des profils, les profils sont regroupés en des groupes similaires grâce à l'algèbre de Galois. Le but est de proposer des recommandations personnalisées à chaque utilisateur. La dernière partie est la génération des recommandations aux utilisateurs.

EXPÉRIMENTATION

*Ce chapitre présente le prototype de l'application que nous avons développée afin de concrétiser les idées présentées dans le cadre du système *Emergsem* et par la suite les résultats de l'évaluation des techniques proposées.*

Sommaire

6.1	Introduction	124
6.2	Architecture du prototype	124
6.2.1	Création de l'ontologie d'application sous Protégé	124
6.2.2	Implémentation du système <i>Emergsem</i>	126
6.2.3	Stockage des informations	130
6.3	Évaluation des résultats	131
6.3.1	Comparaison d' <i>Emergsem</i> avec deux approches classiques	132
6.3.2	Comparaison d' <i>Emergsem</i> avec CLUE et SIMPLIcity	139
6.4	Conclusion	140

6.1/ INTRODUCTION

Dans ce chapitre, nous présentons dans un premier temps l'architecture du prototype du système *Emergsem*. Elle est composée de la création de l'ontologie d'application qui permet d'annoter les images, de l'implémentation du système à travers l'interface de l'annotation collaborative, l'interface d'instances des sémantiques proposées, l'interface d'évaluation des instances et l'interface de recherche et enfin du stockage des données. La deuxième partie décrit l'évaluation des résultats obtenus. Elle est basée sur la comparaison d'une part les résultats entre *Emergsem* et deux approches basées sur les techniques classiques que nous avons développées dans le cadre de notre expérimentation, et d'autre part les résultats entre *Emergsem* et deux approches existantes.

6.2/ ARCHITECTURE DU PROTOTYPE

L'architecture du prototype que nous proposons est composée de deux parties : l'implémentation de l'ontologie d'application sous *Protege* et le système *Emergsem*. La figure 6.1 présente l'architecture du prototype.

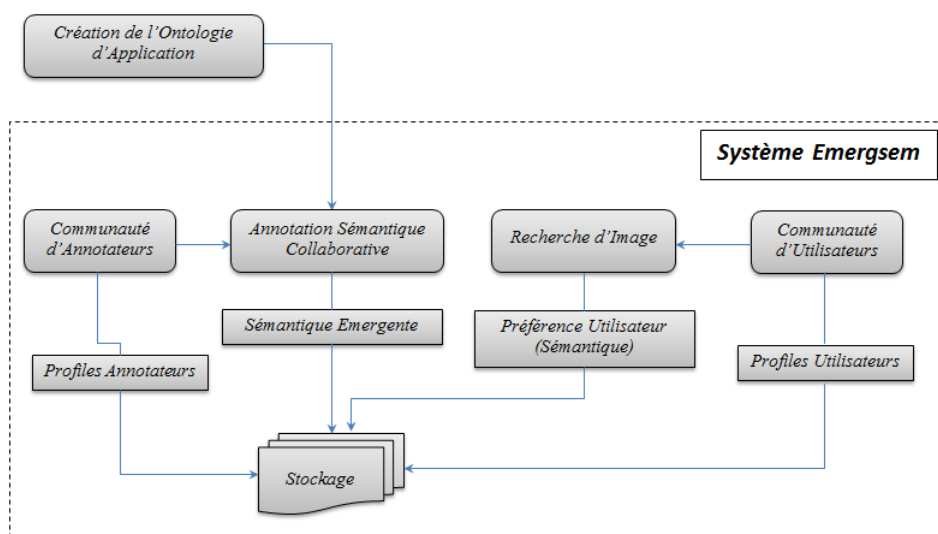


FIGURE 6.1 – Architecture du prototype

6.2.1/ CRÉATION DE L'ONTOLOGIE D'APPLICATION SOUS PROTÉGÉ

Dans notre système, l'ontologie est créée grâce au logiciel *Protégé*. *Protégé* est un logiciel qui a été créé à l'université Stanford et est très populaire dans le domaine du Web sémantique et au niveau de la recherche en informatique. Il est gratuit et son code source est publié sous une licence libre (la Mozilla Public License). La plateforme *Protégé* est un éditeur d'ontologies basé sur Java. Une ontologie, en informatique, est un ensemble structuré de savoirs dans un domaine de connaissance particulier. *Protégé* offre 2 approches pour développer une ontologie : l'éditeur *Protégé-OWL* permet de construire des ontologies pour le web sémantique et l'édi-

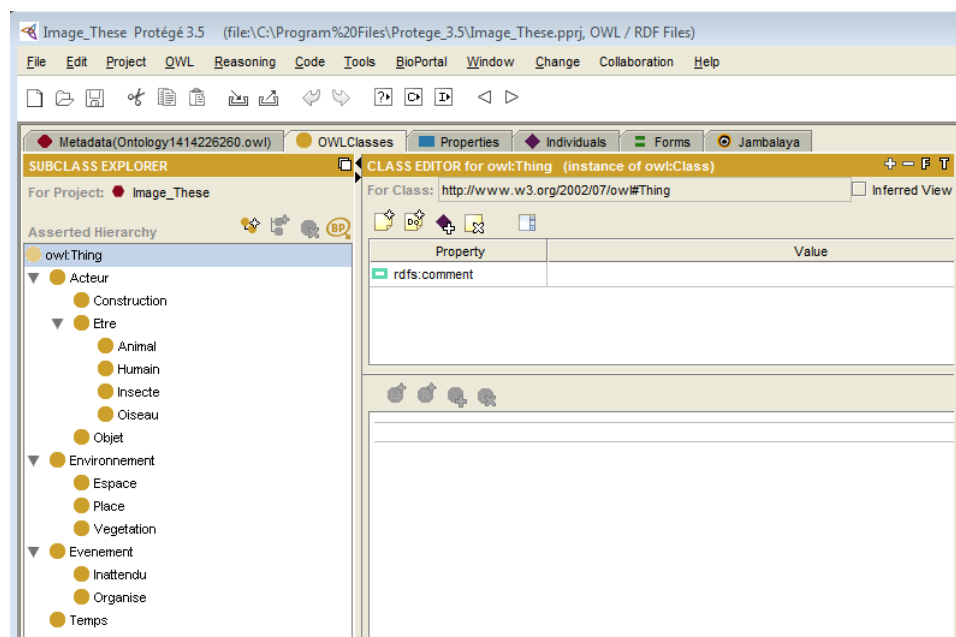


FIGURE 6.2 – Processus de description de l'ontologie sous Protégé

teur Protégé-Frames permet de construire et peupler des bases de connaissances basées sur des frames. Protégé permet d'utiliser le langage OWL, qui est le langage informatique utilisé pour modéliser des ontologies. OWL est un dialecte XML qui étend RDFS. Il permet un fonctionnement en mode client/serveur avec une gestion des droits ajustables en fonction des profils souhaités. Protégé est surtout utilisé dans le domaine de la recherche médicale, biologique ou informatique, notamment pour le cas de modélisations conceptuelles de connaissances biomédicales. Néanmoins tout domaine de connaissance peut-être modélisé à l'aide de Protégé. Les fonctionnalités de Protégé peuvent être étendues grâce à une architecture en plugin et à l'aide de l'API Java fournie. Ce qui permet, en outre, de générer automatiquement du code Java¹. Protégé peut lire et sauvegarder des ontologies dans la plupart des formats d'ontologies : RDF, RDFS, OWL, etc.².

Grâce au logiciel Protégé, nous avons décrit notre ontologie d'application et le sauvegarder sous le format OWL. La figure 6.2 présente le processus de description de l'ontologie d'application.

La figure 6.3 présente la représentation de l'ontologie d'application avec Protege.

Pour réaliser notre expérimentation, l'ontologie que nous avons créée contient quatre concepts abstraits de base (Acteur, Environnement, Evenement et Temps). Chacun de ces concepts contient des concepts abstraits et/ou des concepts concrets comme le montre la figure 6.3. Les relations entre les concepts sont aussi créées mais le déploiement du modèle de l'ontologie ne permet pas d'afficher les relations. Les relations sont bien visible dans la repré-

1. <https://www.projet-plume.org/fiche/protege>

2. [http://fr.wikipedia.org/wiki/Protégé_\(logiciel\)](http://fr.wikipedia.org/wiki/Protégé_(logiciel))

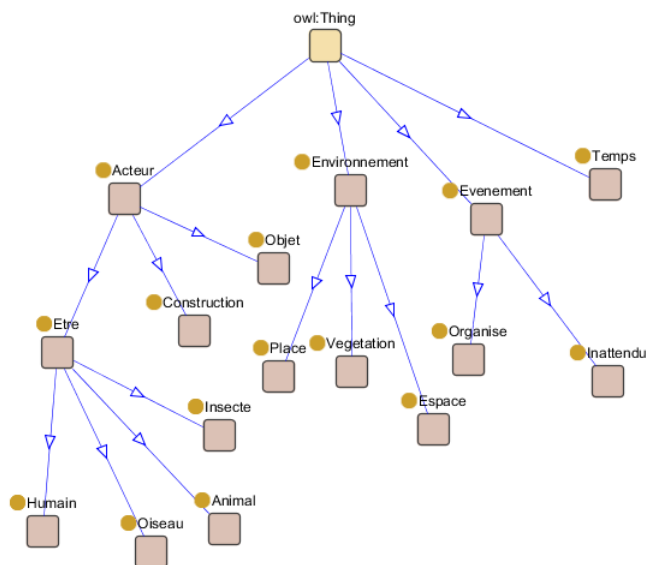


FIGURE 6.3 – Modèle de l'ontologie

sentation sous le format xml comme le montre la figure 6.4. Dans cette figure, nous présentons la relation *Contient* entre le méta concept abstrait *thing* et les quatre concepts abstraits de base.

Une fois créée, l'ontologie d'application est intégrée dans le système que nous avons mis en place et pourra donc être utilisée pour annoter les images.

6.2.2/ IMPLÉMENTATION DU SYSTÈME *Emergsem*

Emergem est le système que nous avons mis en place sous l'environnement *Visual Studio*. Il est développé sous le langage C (C Sharp). Le langage C (C Sharp) est un langage objet créé spécialement pour le framework Microsoft .NET. L'équipe qui a créé ce langage a été dirigée par Anders Hejlsberg, un informaticien danois qui avait également été à l'origine de la conception du langage Delphi pour la société Borland (évolution objet du langage Pascal). Le Framework .NET est un environnement d'exécution (CLR Common Language Runtime) ainsi qu'une bibliothèque de classes (plus de 2000 classes). L'environnement d'exécution (CLR) de .NET est une machine virtuelle comparable à celle de Java. Le runtime fournit des services aux programmes qui s'exécutent sous son contrôle : chargement/exécution, isolation des programmes, vérification des types, conversion code intermédiaire (IL) vers code natif, accès aux métadonnées (informations sur le code contenu dans les assemblages .NET), vérification des accès mémoire (évite les accès en dehors de la zone allouée au programme), gestion de la mémoire (Garbage Collector), gestion des exceptions, adaptation aux caractéristiques nationales (langue, représentation des nombres), compatibilité avec les DLL et modules COM qui sont en code natif (code non managé). Les classes .NET peuvent être utilisées par tous les langages prenant en charge l'architecture .NET. Les langages .NET doivent satisfaire certaines spécifications : utiliser les mêmes types CTS (Common Type System), les compilateurs doivent



The screenshot shows a web browser window titled "Création d'Annotateurs / d'Utilisateurs". The main heading is "Création d'Annotateurs et d'Utilisateurs". The form includes a dropdown menu for "Type d'internaute" set to "Annotateurs". Below this are input fields for "Nom", "Prénom(s)", "Date Nais", "Adresse", "Email", "Téléphone", "Identifiant", "Mot de passe", and "Confirmation". A "Soumettre" button is located at the bottom right.

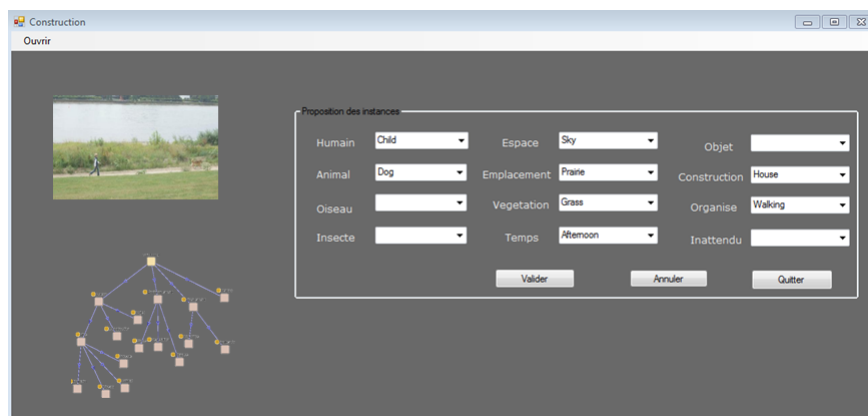
FIGURE 6.5 – Formulaire d'enregistrement des annotateurs et des utilisateurs

lexical. La même image est mise à la disposition de tous les annotateurs et ces derniers déroulent les menus déroulants pour proposer des termes décrivant les régions de l'image proposée. Lorsqu'une image ne contient pas un objet, le menu déroulant de ce concept n'est pas instancié comme par exemple le menu déroulant des événements imprévus (voir la figure 6.6). Grâce à cette interface, la communauté d'annotateurs peut annoter toutes les images qui lui seront proposées.

Lorsque les termes sont proposés, la deuxième étape est proposition des relations spatiales. Pour proposer les relations spatiales, une petite interface est proposée aux annotateurs

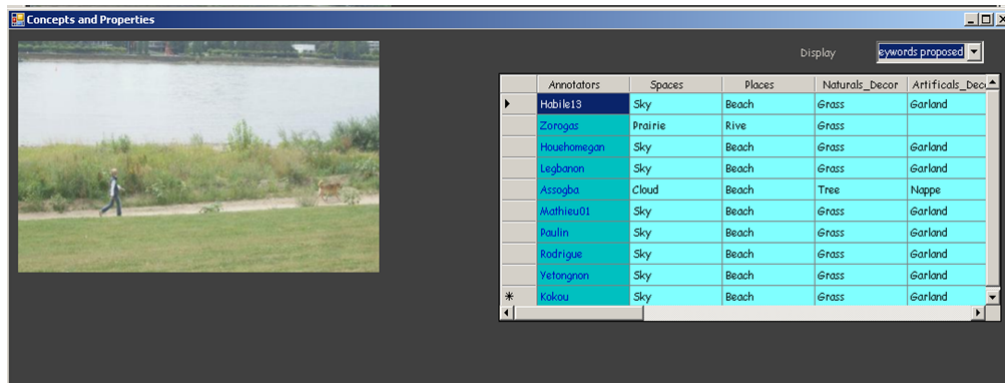
6.2.2.3/ INTERFACE D'INSTANCES DES SÉMANTIQUES PROPOSÉES

La figure 6.7 présente les sémantiques proposées. La première colonne du tableau de la figure présente la liste des annotateurs ayant fourni une sémantique de l'image affichée. Les autres colonnes détaillent les instances de la sémantique proposée par chacun des annotateurs.



The screenshot shows a web browser window titled "Construction" with a sub-header "Ouvrir". On the left, there is a small image of a landscape with a person walking. Below the image is a hierarchical tree diagram. On the right, there is a "Proposition des instances" form with dropdown menus for "Humain" (Child), "Animal" (Dog), "Oiseau", "Insecte", "Espace" (Sky), "Emplacement" (Prairie), "Vegetation" (Grass), "Temps" (Afternoon), "Objet", "Construction" (House), "Organise" (Walking), and "Inattendu". At the bottom of the form are buttons for "Valider", "Annuler", and "Quitter".

FIGURE 6.6 – Annotation collaborative des images



Annotators	Spaces	Places	Naturals_Decor	Artificials_Deo
Habile13	Sky	Beach	Grass	Garland
Zorngas	Prairie	Rive	Grass	Garland
Houhamegan	Sky	Beach	Grass	Garland
Legbanon	Sky	Beach	Grass	Garland
Assogba	Cloud	Beach	Tree	Nappe
Wathieu01	Sky	Beach	Grass	Garland
Paulin	Sky	Beach	Grass	Garland
Rodrigue	Sky	Beach	Grass	Garland
Yatangnon	Sky	Beach	Grass	Garland
* Kokou	Sky	Beach	Grass	Garland

FIGURE 6.7 – Un aperçu des instances des sémantiques proposées

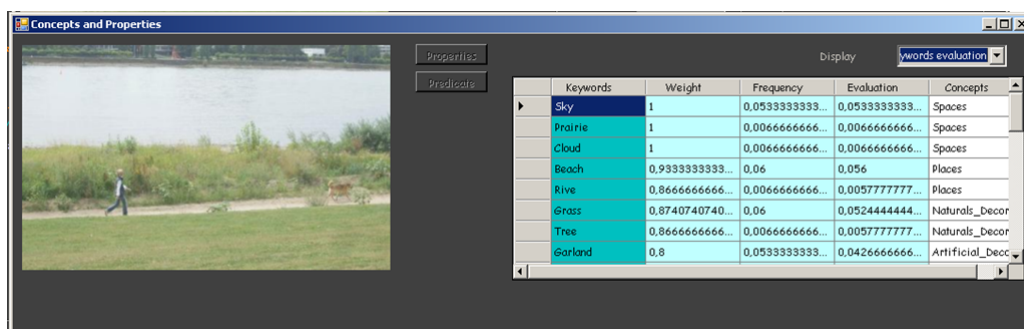
6.2.2.4/ INTERFACE D'ÉVALUATION DES INSTANCES

L'interface des évaluations des instances permet de déterminer les instances les plus représentatifs. Comme nous l'avons détaillé dans le chapitre 4 en sa section *Emergence des sémantiques*, cette évaluation est basée sur la fonction f_{emerg} qui tient compte de la fréquence et du poids des instances (voir la figure 6.8). S'il arrivait que deux instances du même concept aient la même évaluation, nous comparons le poids de chacune d'elles et l'instance ayant le poids le plus élevé est retenu pour la sémantique émergente.

6.2.2.5/ INTERFACE DE RECHERCHE

Dans le système *Emergsem*, lorsqu'un utilisateur propose une requête de mots-clés (si les sémantiques recommandées ne répondent pas à ses attentes), un cluster d'images lui est proposé comme le montre la figure 6.9 qui présente un cluster de 28 images similaires à la requête de l'utilisateur.

L'utilisateur pourra donc sélectionner l'image qu'il recherche. La sémantique de l'image est reliée à l'utilisateur et stockée dans une base de données, ce qui permettra de faire des



Keywords	Weight	Frequency	Evaluation	Concepts
Sky	1	0.05333333333...	0.05333333333...	Spaces
Prairie	1	0.00666666666...	0.00666666666...	Spaces
Cloud	1	0.00666666666...	0.00666666666...	Spaces
Beach	0.93333333333...	0.06	0.056	Places
Rive	0.86666666666...	0.00666666666...	0.00577777777...	Places
Grass	0.8740740740...	0.06	0.05244444444...	Naturals_Decor
Tree	0.86666666666...	0.00666666666...	0.00577777777...	Naturals_Decor
Garland	0.8	0.05333333333...	0.04266666666...	Artificial_Decc

FIGURE 6.8 – Un extrait d'évaluation des instances



FIGURE 6.9 – Interface de recherche

recommandations à l'utilisateur plus tard.

Lorsque l'image recherchée par un utilisateur avait déjà fait l'objet de recherche par le même utilisateur, la sémantique de l'image est recommandée à l'utilisateur. L'utilisateur peut donc sélectionner la sémantique pour afficher l'image. Dans ce cas précis, l'utilisateur ne se voit pas afficher un cluster d'images, mais l'image décrite par la sémantique recommandée.

6.2.3/ STOCKAGE DES INFORMATIONS

Dans *Emergsem*, le stockage des profils d'utilisateurs et d'annotateurs se fait grâce à une ontologie comme se fit le cas avec la description des images. Cette ontologie stockant les informations telles que les profils d'annotateurs et d'utilisateurs est composée de quatre niveaux de concepts. Le premier niveau est la racine *Thing* que nous appelons l'*ancêtre*. Le deuxième niveau contient un seul concept *profiles* qui est relié à la racine par une relation *Contient*. Le concept *profiles* est un concept abstrait et permet de décrire tous les profils d'utilisateurs et les profils d'annotateurs. Il possède deux concepts au troisième niveau. Il s'agit du concept concret *Dynamiques* qui décrit les profils dynamiques d'utilisateurs et du concept abstrait *Statiques* qui permet de décrire les profils statiques des annotateurs comme des utilisateurs. Ces deux concepts sont reliés au concept *profiles* par la relation *Possède* et dont l'inverse est la relation *Est – possédé – par*. Le concept *Statiques* est composé de 5 concepts concrets *Nom*, *Prenom*, *Adresse*, *Pseudo*, *Mot_de_passe* qui décrivent respectivement le nom, le prénom, l'adresse, le pseudonyme et le mot de passe des utilisateurs comme des annotateurs. La figure 6.10 présente l'ontologie de stockage des profils d'annotateurs et d'utilisateurs.

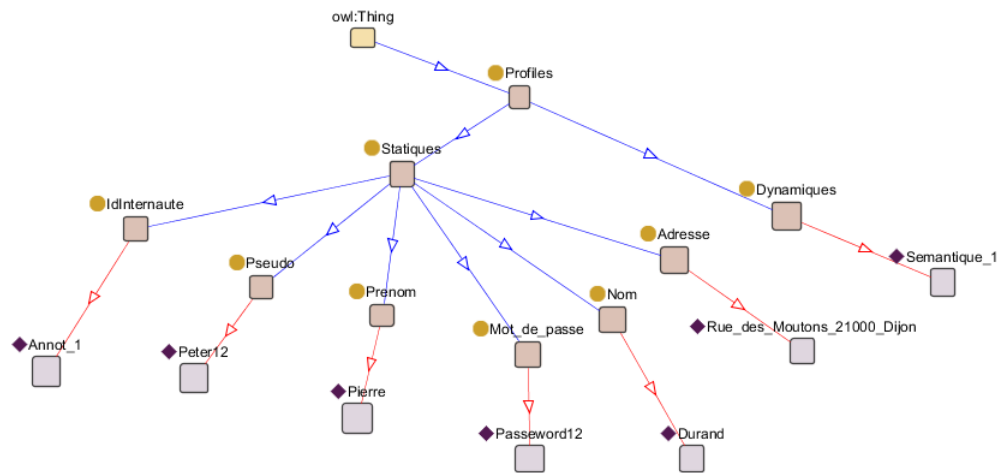


FIGURE 6.10 – Ontologie de stockage des profils d'annotateurs et d'utilisateurs

6.3/ ÉVALUATION DES RÉSULTATS

Les expérimentations portent dans un premier temps sur les systèmes d'annotation classique d'images. Dans un deuxième temps nous allons comparer les systèmes d'annotation existants et à notre système *Emergsem*. Nos expérimentations sont faites sur une collection de 60.000 images provenant de la base de données COREL⁴. Le choix de cette base de données est motivé par le fait que les approches auxquelles nous voulons comparer *Emergsem* l'ont utilisée et les résultats obtenus sont disponibles. Nos expérimentations sont réalisées grâce à une communauté de 89 annotateurs et d'une communauté de 135 utilisateurs.

Au départ, nous avons fait émerger la sémantique des images grâce à la fonction f_{emerg} basée uniquement sur la fréquence des mots-clés et des relations spatiales. Mais après le processus d'annotation, nous avons mené un sondage auprès de la communauté d'annotateurs (89 membres au total) via la plateforme *Sondage Online* qui est une plateforme de création et d'évaluation de sondages comme le montre la figure 6.11.

Il en ressort que 91,91 % des annotateurs ont avoué qu'ils ont d'abord pris soin de proposer les instances des objets qui leur paraissaient plus évidents avant de se consacrer aux objets les moins évidents, contre 08,09 % comme le montre le graphe de la figure 6.12. *Pourcentage 1* représente le pourcentage d'annotateurs qui ont confirmé avoir annoté les objets qui maîtrisent le mieux avant de se pencher sur les autres et *Pourcentage 2* représente les annotateurs qui ont instancié les objets et les relations spatiales de façon aléatoire.

Cette révélation nous a poussés à repenser la fonction d'évaluation en y ajoutant le poids des instances, c'est-à-dire le rang occupé par chacune d'elles lors de l'annotation.

Pour évaluer la performance de notre approche, nous sommes amenés à développer deux applications basées sur les techniques classiques d'annotation : manuelle et automatique. Nous

4. <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>



FIGURE 6.11 – Plateforme de sondage

avons ensuite comparé l'annotation sémantique collaborative aux deux autres techniques d'annotation : l'annotation manuelle et l'annotation automatique que nous avons développées.

6.3.1/ COMPARAISON D'*Emergsem* AVEC DEUX APPROCHES CLASSIQUES

Deux systèmes d'annotation classique sont développés dans le cadre de notre expérimentation : un système d'annotation manuelle et un système d'annotation automatique.

6.3.1.1/ ANNOTATION MANUELLE

Nous avons mis en place un système d'annotation manuelle qui permet d'annoter les images. Ce système comporte deux interfaces importantes : une interface de description d'images

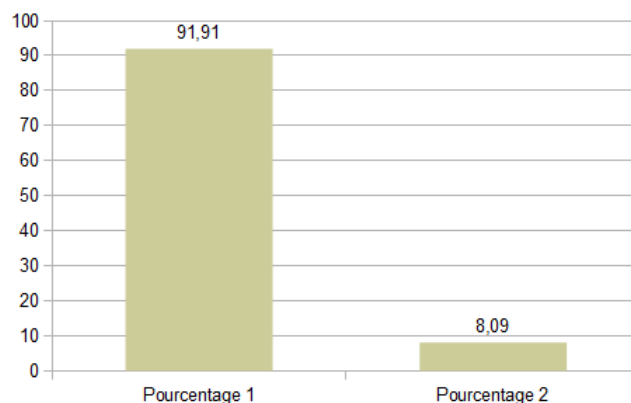


FIGURE 6.12 – Tableau du sondage

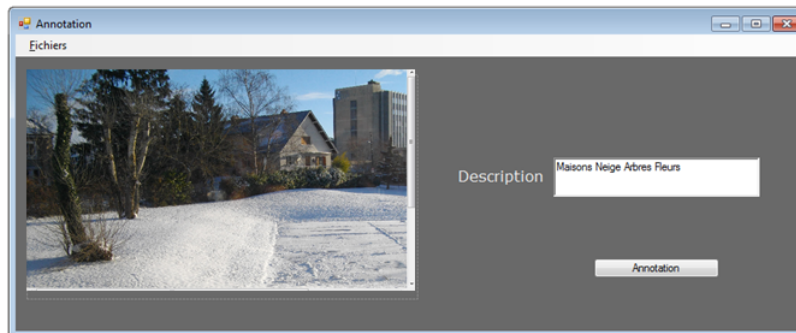


FIGURE 6.13 – Interface d'annotation manuelle

où les annotateurs proposent des termes (mots-clés) qui décrivent les objets ou les régions d'images et une interface de recherche qui permet aux utilisateurs de proposer des mots-clés ou des groupes de mots-clés pour retrouver les images. Comme nous pouvons le constater sur la figure de l'interface d'annotation 6.13, seules les régions d'images sont décrites avec les mots-clés. L'interface de la figure 6.14 présente l'affichage des plusieurs images annotées avec des mots-clés. L'ensemble des mots-clés décrivant les régions d'une image représente l'image. Avec l'utilisation des mots-clés, les utilisateurs mettent beaucoup de temps pour retrouver les images parce que l'indexation se base sur la généralisation, c'est-à-dire que toutes les images qui auront en commun un des mots-clés proposés dans la requête seront affichées. Il revient donc à l'utilisateur de parcourir les images affichées pour déterminer l'image qui répond à ses attentes.



FIGURE 6.14 – Extrait de description d'images

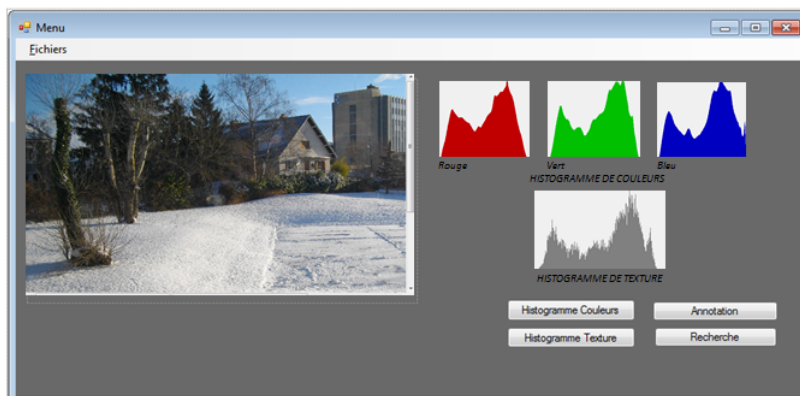


FIGURE 6.15 – Interface d'annotation automatique

6.3.1.2/ ANNOTATION AUTOMATIQUE

Le système d'annotation automatique que nous proposons est un système basé sur l'apprentissage non supervisé et la classification des images via le vecteur de leurs caractéristiques. Deux vecteurs sont utilisés dans le cadre de notre expérimentation : l'histogramme de couleur (RGB) et l'histogramme de texture. Lorsqu'une image est proposée, le système extrait ces caractéristiques et les compare aux caractéristiques des images déjà annotées et stockées dans des classes. Si les caractéristiques de l'image sont similaires aux caractéristiques des images d'une classe, l'image est affectée à cette classe. Dans le cas contraire, une nouvelle classe est créée. La figure 6.15 l'interface d'annotation automatique des images. La recherche d'image se fait également de la même manière, c'est-à-dire que lorsqu'une requête image est proposée, seules les classes auxquelles l'image appartient sont affichées à l'utilisateur comme le montre la figure 6.16.

Nous avons évalué le temps mis pour annoter les images, le temps mis pour rechercher les images et le taux d'acceptation des résultats des requêtes proposées par les internautes. Le tableau de la figure 6.1 présente les résultats obtenus.

Les résultats obtenus montrent que l'annotation collaborative prend plus de temps que

Paramètre de Comparaison	Annotation Automatique	Annotation Collaborative	Annotation Manuelle
Durée moyenne d'annotation(secondes)	6,31	251,60	199,81
Durée moyenne de recherche (secondes)	378,23	195,41	217,57
Taux de satisfaction (%)	49,03	98,39	67,23

TABLE 6.1 – Moyenne des durées de recherche d'images

l'annotation manuelle et l'annotation automatique. Ceci est dû au fait que les sémantiques obtenues grâce à l'annotation collaborative sont des sémantiques qui se forment progressivement. Les sémantiques fournies par chaque annotateur ne sont pas des sémantiques définitives des images mais contribuent à l'émergence de la vraie sémantique des images, alors que dans le

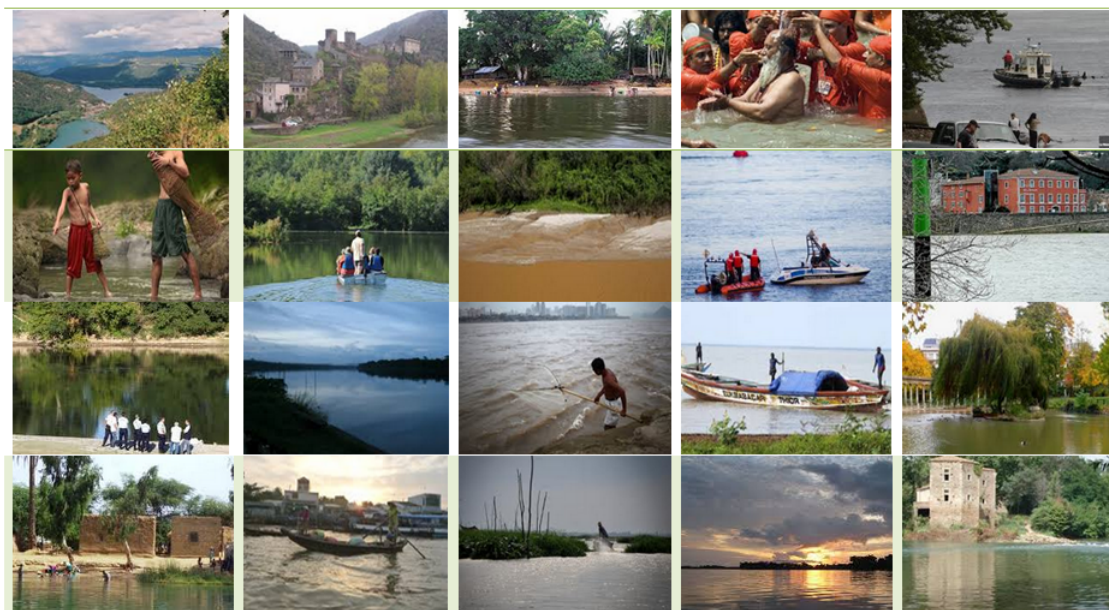


FIGURE 6.16 – Affichage des résultats de la recherche

cadre où l'annotation est produite par un seul annotateur, les descriptions obtenues deviennent immédiatement les descriptions des images. C'est la raison pour laquelle le temps mis pour fournir la sémantique des images par l'annotation manuelle est largement inférieur au temps mis par l'annotation collaborative, ce qui justifie la différence de 81,79 *secondes* entre les deux techniques d'annotations. Mais en comparant l'annotation automatique aux deux autres annotations (manuelle et collaborative), nous remarquons que la moyenne des temps mis pour décrire les images est très inférieure à la moyenne des temps mis par l'annotation manuelle et l'annotation collaborative, soit une moyenne de 06,31 *secondes* pour l'annotation automatique, une moyenne de 251,6 *secondes* pour l'annotation collaborative et une moyenne de 199,81 *secondes* pour l'annotation manuelle comme le montre le graphe de la figure 6.17

Les annotations produites sont ensuite soumises à une communauté d'utilisateurs (135 utilisateurs au total). Les utilisateurs se sont servis de ces annotations pour rechercher les images. La moyenne des durées de recherche est calculée et présentée dans la tableau de la figure 6.1. Le temps mis par le système basé sur l'annotation automatique est pratiquement le double du temps mis par le système basé sur l'annotation manuelle et le temps mis par le système basé sur l'annotation manuelle est supérieur au temps mis par le système basé sur l'annotation collaborative. Cette différence est due au fait que les descripteurs physiques utilisés pour annoter les images dans système d'annotation automatique ne parviennent pas à représenter les images du point de vue des humains et que les descripteurs textuelles, malgré qu'ils parviennent à représenter les régions d'images n'arrivent pas à rendre fidèlement compte de la sémantique des images. Le système de recherche basé sur l'annotation manuelle et automatique permettent d'afficher une collection d'images qui ont une ou plusieurs caractéristiques en commun. Les

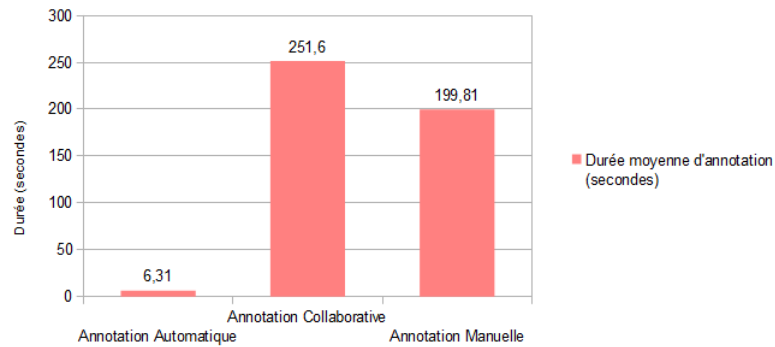


FIGURE 6.17 – Graphe des durées d'annotation

utilisateurs sont donc contraints à parcourir la collection afin de retrouver l'image qui répond à leur besoin. Ceci leur fait perdre énormément de temps. La figure 6.19 présente le graphe de cette évaluation.

Le taux d'acceptation se mesure par rapport à la satisfaction des utilisateurs. Il permet de déterminer si l'image recherchée répond aux attentes de l'utilisateur ou pas. D'après les résultats que nous avons obtenus, 67,23 % des utilisateurs ayant utilisé les descriptions résultant de l'annotation manuelle sont satisfaits, 49,03 % sont satisfaits après avoir utilisé les descriptions obtenues grâce à l'annotation automatique et 98,39 % pour les utilisateurs ayant utilisé les descriptions provenant de l'annotation collaborative.

En résumé, nous pouvons dire que malgré que l'annotation collaborative requière plus de temps que l'annotation manuelle et l'annotation automatique, elle nous donne des résultats très satisfaisants. Les descriptions obtenues sont généralement acceptées par les utilisateurs parce qu'elles sont fournies par une communauté d'annotateurs via des outils sémantiques.

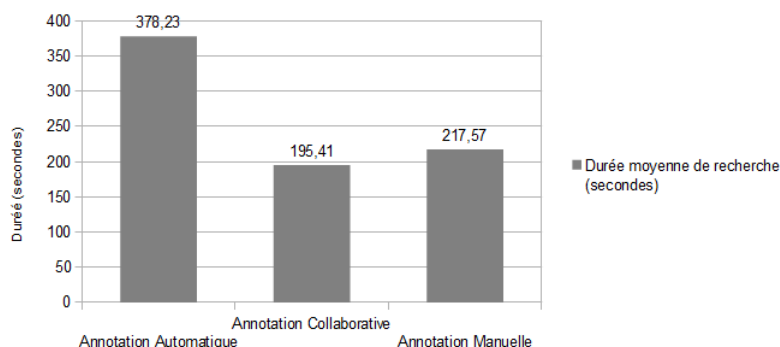


FIGURE 6.18 – Graphe des durées de recherche

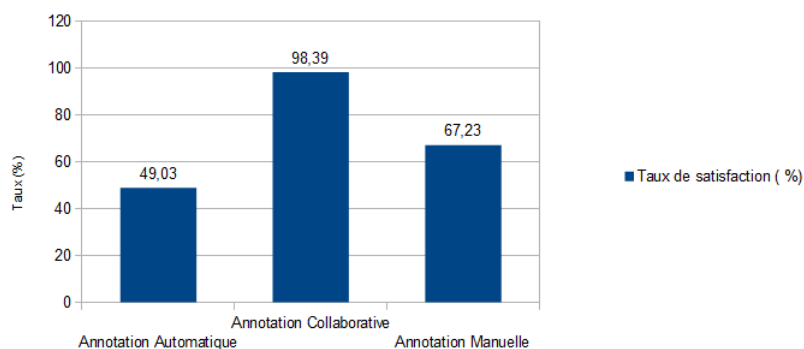


FIGURE 6.19 – Graphe des taux de satisfaction

6.3.1.3/ RECHERCHE D'IMAGES

La technique de recherche d'images que nous avons présentée permet non seulement de rechercher les images, mais aussi effectue des recommandations de sémantiques aux utilisateurs. La technique de recherche d'images et la technique de recommandation des sémantiques sont indissociables dans le système *Emergsem*. La communauté d'utilisateurs utilise les sémantiques pour retrouver les images via le système de recherche et le système de recommandation se base sur les sémantiques ayant servi à indexer des images pour effectuer des recommandations aux utilisateurs. Cette techniques facilite la récupération des images parce que les internautes peuvent visualiser plusieurs images similaires grâce la même requête puis que les images sont reliées entre elles. Dans la méthode que nous proposons, les internautes ont la possibilité de fournir des instances ou des termes pour rechercher les images.

Comparativement aux résultats obtenus avec les techniques traditionnelles, nous avons remarqué que l'annotation sémantique collaborative donne des résultats bien meilleurs. Les images du cluster sont sémantiquement similaires les unes des autres. L'utilisateur pourra donc facilement retrouver l'image ou les images qui répondent à ses attentes. La figure 6.2 présente un résultat de cluster d'images obtenu suite à la requête de mots-clés d'un utilisateur qui souhaiterait obtenir des images de groupe de personnes se déplaçant dans une pirogue.

Nous avons également évalué le temps mis par les internautes pour retrouver les images. Il s'agit du temps mis avant et après la recommandation des sémantiques. Les résultats de ces expérimentations sont présentés dans le tableau 6.3.

Le test est fait sur une vingtaine d'images recherchées par les utilisateurs. Les résultats que nous avons obtenus ont montré que les graphes suivent pratiquement la même allure. Au temps T_0 aucune recherche d'image n'a encore enregistrée. Au temps $T_0 + 1$, les utilisateurs ont recherché des images en fournissant des description avec une moyenne de $5,344.10^{-2}$ heure, soit 3,2064 minutes. Au temps $T_0 + 2$ nous remarquons que le temps mis pour indexer les images au temps $T_0 + 1$ est pratiquement divisé par 20. Ces résultats justifient le fait que les utilisateurs n'ont plus fourni les descriptions des images, mais ces descriptions leur ont été



TABLE 6.2 – Quelques exemples de clusters d'images

Images	T_0	T_{0+1}	T_{0+2}	T_{0+3}	T_{0+4}
Image 1	0,0500	0,0650	0,0041	0,0033	0,0032
Image 2	0,0497	0,0417	0,0027	0,0029	0,0033
Image 3	0,0602	0,0672	0,0028	0,0031	0,0030
Image 4	0,0475	0,0375	0,0032	0,0032	0,0029
Image 5	0,0595	0,0550	0,0042	0,0032	0,0033
Image 6	0,0520	0,0520	0,00351	0,0033	0,0029
Image 7	0,0652	0,0652	0,0034	0,0030	0,0031
Image 8	0,0478	0,0388	0,0029	0,0029	0,0032
Image 9	0,0655	0,0550	0,0042	0,0032	0,0033
Image 10	0,0652	0,0520	0,0035	0,0033	0,0029
Image 11	0,0472	0,0652	0,0024	0,0031	0,0031
Image 12	0,0488	0,0388	0,0029	0,0030	0,0032
Image 13	0,0575	0,0550	0,0042	0,0032	0,0033
Image 14	0,0592	0,0520	0,0035	0,0033	0,0030
Image 15	0,0652	0,0652	0,0034	0,0030	0,0030
Image 16	0,0578	0,0550	0,0042	0,0032	0,0033
Image 17	0,0656	0,0620	0,0035	0,0033	0,0029
Image 18	0,0672	0,0652	0,0034	0,0030	0,0031
Image 19	0,0478	0,0388	0,0029	0,0029	0,0032
Image 20	0,0662	0,0672	0,0028	0,0031	0,0030

TABLE 6.3 – Durées de recherche d'images

proposées par le système de recommandation. La figure 6.20 présente le graphe résultant de ces expérimentations.

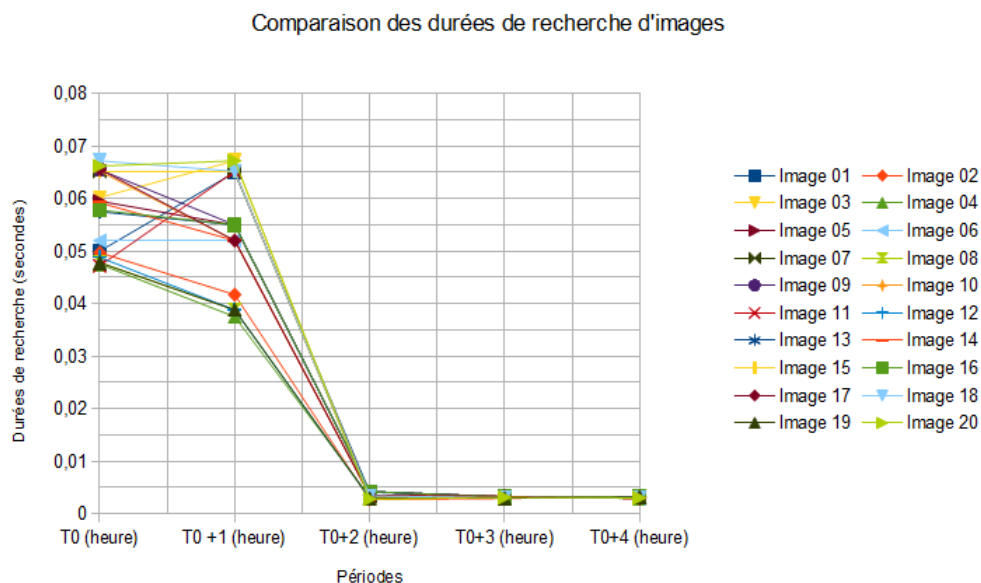


FIGURE 6.20 – Graphe de comparaison des durées de recherche d'images

6.3.2/ COMPARAISON D'*Emergsem* AVEC CLUE ET SIMPLiCITY

Les dernières expérimentations que nous avons menées sur la recherche nous ont permis de comparer notre approche à d'autres approches existantes. La première approche est une approche de clustering basées sur la recherche d'images avec des caractéristiques physiques (CLUE) dans [183]. C'est une approche de recherche d'images basée sur des clusters d'images par apprentissage non supervisé. La deuxième approche est SIMPLiCITY [?] qui utilise les méthodes de classification sémantiques basée sur les ondelettes pour l'extraction des caractéristiques des images. La comparaison va porter la précision. le rappel n'a pas été calculé, car la base de données est grande et il est difficile d'estimer le nombre total d'images dans une catégorie, même approximativement. Dans l'avenir, nous allons développer une base de données de test partageable à grande échelle pour évaluer le rappel. Rappelons que la précision est le nombre d'images pertinentes retrouvées rapporté au nombre d'images total proposé pour une requête donnée. quand un utilisateur interroge le corpus, il souhaite que les images proposées en réponse à son interrogation correspondent à son attente. Toutes les images retournées superflus ou non pertinents constituent du bruit. Si elle est élevée, cela signifie que peu d'images inutiles sont proposés par le système et que ce dernier peut être considéré comme "précis"⁵.

Pour comparer les approches, la moyenne des valeurs de précision est calculée. Nous présentons dans 6.21 quelques résultats des clusters obtenus. Dans les approches de CLUE et de SIMPLiCITY, une requête image est fournie. Ce sont les images en haut à gauche de chaque cluster. Dans *Emergsem*, l'image obtenue par la sémantique représentative est également placée en haut et à gauche des clusters. Sous chaque cluster est indiqué les résultats des

5. http://fr.wikipedia.org/wiki/Précision_et_rappel

deux recherches effectuées. Le premier résultat (8 images sur 11 si nous considérons le premier groupe de clusters de CLUE) est le résultat obtenu après la proposition de la requête image et le deuxième résultat (19 sur 25) est le nombre d'images obtenus après le retour de pertinence. Avec *Emergsem*, nous avons un cluster définitif. Nous n'y appliquons plus une recherche basée sur le retour de pertinence. Dans la suite des expérimentations, seuls les résultats obtenus après le retour de pertinence sur CLUE et SIMPLcity seront comparés aux résultats des clusters d'*Emergsem*. Les recherches sont menées sur plusieurs requêtes mais nous consignons seulement deux catégories d'images : une image de nourriture et une image de monuments historiques. Pour chacune des clusters d'images, la moyenne de la précision est calculée. Dans le tableau de la figure 6.21, nous allons tenir compte des images des derniers résultats de clusters de CLUE et de SIMPLcity.

Le graphe 6.22 présente la moyenne des précisions des trois approches. Les résultats de la comparaison des approches CLUE et SIMPLcity sont consignés dans le travail mené dans [183]. Les précisions moyennes globales sont déterminées, soit 0,538 pour CLUE et 0,477 pour SIMPLcity, alors que nous avons calculé la précision moyenne globale d'*Emergsem*, nous avons constaté qu'elle est égale à 0,82. Ceci est dû au fait que dans les approches de CLUE et de SIMPLcity, les classifications se font sur les caractéristiques extraites des images qui ont au préalable été grisées. Qui dit griser dit qu'il y a perte d'informations. Ceci a fait que certains résultats obtenus n'ont aucune relation avec l'image de la requête, tandis qu'avec *Emergsem*, la classification se fait sur les sémantiques attribuées aux images et non sur les caractéristiques physiques des images. C'est la raison pour laquelle *Emergsem* est mieux que les deux autres approches.

6.4/ CONCLUSION

Les résultats que nous avons obtenus ont montré d'une part que l'annotation sémantique collaborative répond aux besoins de description complète des images, ce qui améliore par la suite leur recherche. Nos résultats de recherche d'images sont très encourageants parce que la technique de recherche que nous proposons tient compte des relations sémantiques qui existent entre les images pour proposer des clusters pertinents aux utilisateurs. Bien qu'il existe un compromis entre la complexité et la performance, l'annotation sémantique collaborative est un choix viables quand la performance est recherchée.

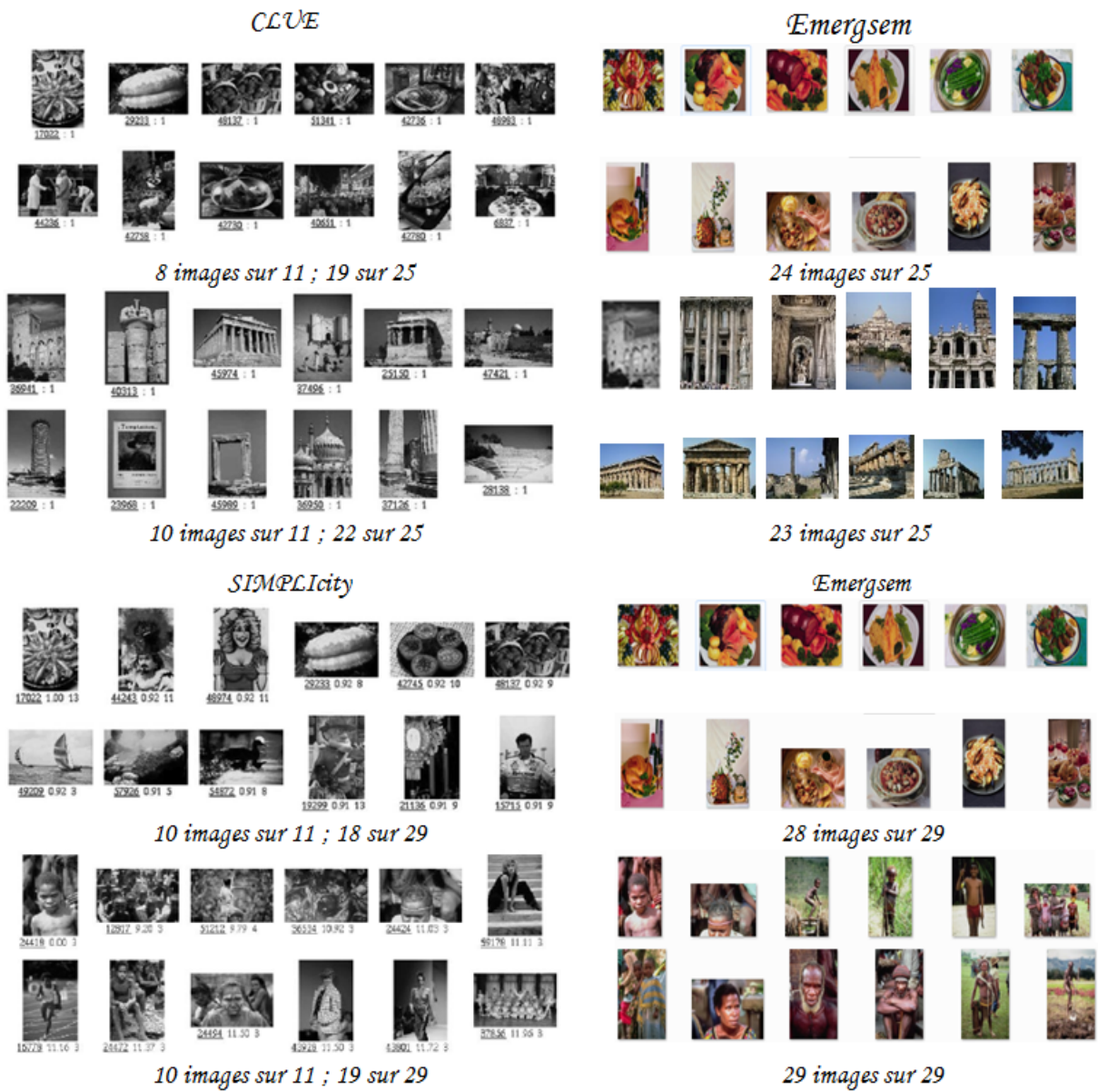


FIGURE 6.21 – Comparaison des résultats entre *Emergsem* avec Clue et SIMPLIcity

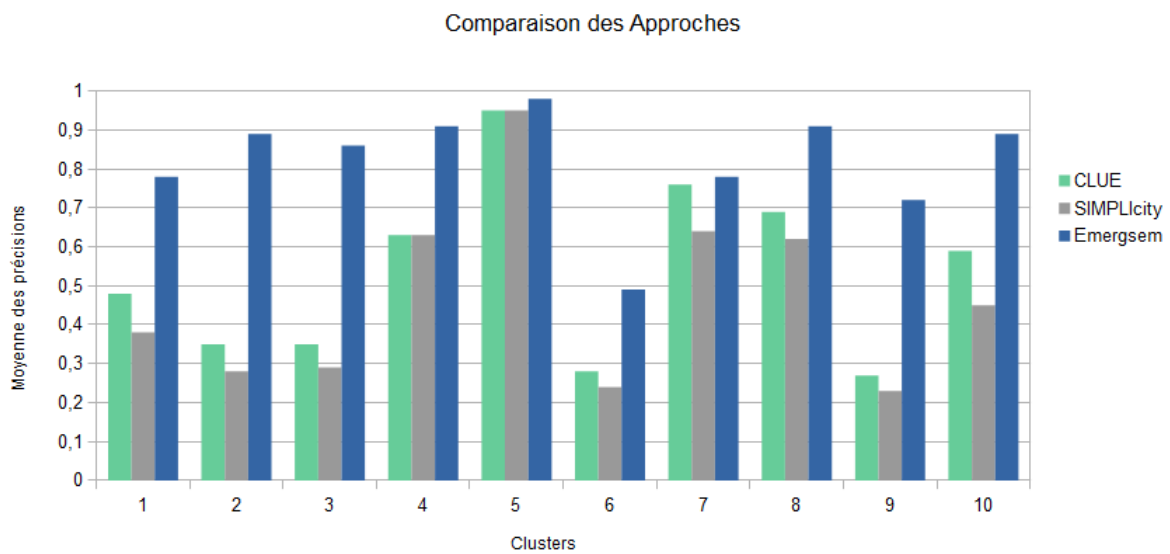


FIGURE 6.22 – Graphe de comparaison des approches

CONCLUSION ET PERSPECTIVES

Dans ce chapitre, nous résumons dans un premier temps la problématique et les propositions faites dans cette thèse, et dans un deuxième temps les principales perspectives de recherche et de développement identifiées à partir des résultats de cette thèse.

Sommaire

7.1 Conclusion	144
7.2 Perspectives	147

7.1/ CONCLUSION

Cette thèse a pour but la constitution d'un système d'annotation et de recherche d'images basé sur la sémantique : le système *Emergsem*. Les travaux qui y sont menés visent à résoudre les problèmes auxquels les approches existantes sont confrontées : la capture de la sémantique des images du point de vue humain afin de réduire le fossé sémantique.

Pour réduire ce fossé, nous sommes amenés dans cette thèse à fournir une définition précise et rigoureuse de la sémantique des images du point de vue humain. A travers *Emergsem*, nous avons focalisé nos efforts sur la mise en place d'un système capable de capturer et de faire émerger la sémantique des images. Ce système a permis de (1) représenter les images du point de vue humain par des outils capables de représenter non seulement les objets présents dans une image, mais également les relations qui existent entre ces objets dans un langage compris par la machine et par les humains et de (2) rechercher les images via à leur sémantique. *Emergsem* se base sur deux approches fondamentales :

- L'émergence de la sémantique des images est rendue possible grâce à une technique d'annotation sémantique collaborative. L'annotation collaborative consiste à faire annoter une images par un groupe d'annotateurs où chaque membre du groupe propose une description sémantique de la même image grâce à un descripteur sémantique et un vocabulaire lexical. L'annotation sémantique collaborative est composée de trois étapes principales :
 - La représentation sémantique des images est la première étape. Dans cette thèse, nous avons proposé une ontologie à deux niveaux pour capturer la sémantique d'une image. Le premier niveau est une ontologie générique constituée d'un méta concept abstrait *Image* et d'un descripteur sémantique. Le méta concept abstrait est un concept de très haut niveau qui contient tous les concepts pouvant décrire les images quelque soit le domaine auquel elles appartiennent. Le descripteur sémantique quant à lui contient un ou plusieurs concepts (abstrait et concrets) et des relations. Le deuxième niveau est une ontologie d'application dérivée de l'ontologie générique pour décrire la sémantique des images d'un domaine bien précis. C'est une ontologie personnalisée dans laquelle on peut créer autant de concepts abstraits de base qui peuvent à leur tour contenir un ou plusieurs concepts (abstrait ou concrets) et des relations reliant les concepts.
 - L'annotation collaborative est la deuxième étape au cours de laquelle une image est décrite par une communauté d'annotateurs où chaque annotateur de la communauté propose une sémantique de l'image grâce à une ontologie d'application et un dictionnaire lexical (WordNet). Cette étape est basée sur l'instanciation des concepts de l'ontologie d'application par les termes du dictionnaire lexical. Chaque annotateur de la communauté propose (ou attache) un terme ou mot-clé décrivant une région d'image aux concepts concrets de l'ontologie d'application. Les annotateurs proposent également des relations entre les concepts concrets. Le nombre de sémantiques obtenues pour une image est égal au nombre d'annotateurs ayant décrit l'image.
 - La génération de la sémantique représentative est la troisième étape qui consiste à faire converger les sémantiques proposées par les annotateur vers une seule sémantique : la

sémantique émergente. Le principe est le suivant. Une fois les sémantiques proposées par les annotateurs, nous procédons à la comparaison des instances et des relations proposées dans le but de déterminer pour chaque concept et chaque relation de l'ontologie, les mots-clés les mieux utilisées par les annotateurs. La comparaison est réalisée grâce à la fonction f_{emerg} qui combine la fréquence et le poids des instances des sémantiques. Les termes pertinents sont mappés dans l'ontologie d'application pour donner la sémantique émergente qui servira à annoter définitivement l'image.

- La recherche sémantique d'images est réalisée grâce aux descripteurs sémantiques associés à celles-ci. Elle est composée de quatre étapes principales :
 - L'extraction des sémantiques représentatives est la première étape. Le principe est le suivant. Lorsque la requête de mots-clés est proposée par l'utilisateur, nous procédons à l'extraction des sémantiques qui répondent mieux aux mots-clés de la requête. L'extraction des sémantiques représentative se base sur deux méthodes : l'appariement et les mesures de similarité. L'appariement permet de comparer les mots-clés aux instances des sémantiques du corpus. Le but est de calculer le poids (la fréquence) de ces mots-clés dans chaque sémantique du corpus. Pour cela nous avons proposé un algorithme qui calcule les poids. Par ailleurs, nous avons utilisé plusieurs mesures pour calculer la similarité entre les mots-clés et les instances des sémantiques. L'utilisation des mesures nous a amenés à considérer deux vecteurs de caractéristiques sémantiques : un vecteur de caractéristiques sémantiques des mots-clés et un vecteur de caractéristiques sémantiques des instances. Certaines mesures effectuent une recherche exacte sur l'ensemble des mots-clés fournis dans la requête, ce qui permet d'afficher les sémantiques qui contiennent tous les mots-clés de la requête. D'autres mesures par contre ne tiennent pas compte de la comparaison exacte mais de tous les éléments des vecteurs de caractéristiques sémantiques. Dans les deux cas, un seuil est défini. Toutes les sémantiques ayant une similarité en dessous de ce seuil ne sont pas considérées comme des sémantiques représentatives de la requête de l'utilisateur.
 - L'extraction des sémantiques voisines est la deuxième étape. Elle permet de déterminer les sémantiques voisines de chaque sémantique représentative. L'extraction des sémantiques voisines est proposée parce que nous avons décidé d'étendre la recherche d'images à toute la collection d'images. Le principe d'extraction du voisinage des sémantiques représentatives est le suivant. Une fois que les sémantiques représentatives sont obtenues, nous procédons à la comparaison du vecteur de caractéristiques sémantiques de chaque sémantique représentative aux vecteurs de caractéristiques sémantiques des autres sémantiques du corpus. La comparaison est basée sur la similarité cosinus des vecteurs de caractéristiques sémantiques des instances. Chaque voisinage des sémantiques représentatives est constitué grâce à un seuil.
 - Le clustering d'images est la troisième étape de recherche sémantique d'images. Il permet de proposer des groupes de sémantiques pertinentes aux utilisateurs. Pour former les clusters d'images, nous avons dans un premier temps procédé à la formation des groupes

de sémantiques au sein des sémantiques voisines. La formation des groupes de sémantiques est basée sur la granularité, c'est-à-dire, la comparaison des structures de sémantiques ayant servi à annoter les images. Pour cela, nous avons proposé une technique basée sur le parcours en profondeur des structures (graphes) et qui compare les arcs (relation entre deux nœuds). La technique permet d'obtenir la signature des structures de sémantiques du corpus. Les structures ayant la même même signature sont groupées ensemble. Dans un deuxième temps, nous formons les clusters des sémantiques pertinentes en calculant la matrice de confusion des groupes obtenus grâce à la signature des structures. La matrice de confusion permet de rechercher dans les différents ensembles de sémantiques voisines, les groupes de sémantiques qui sont similaires.

- La recommandation des sémantiques est la quatrième étape. Elle permet de proposer des sémantiques aux utilisateurs en se basant sur leurs préférences. Elle est composée de trois niveaux. Le premier niveau est l'acquisition des profils d'utilisateurs. Deux profils se distinguent : les profils statiques qui représentent les informations personnelles des utilisateurs et les profils dynamiques qui représentent les préférences des utilisateurs, c'est-à-dire les sémantiques qu'ils utilisent pour rechercher les images. Le deuxième niveau est la classification des profils. Elle est basée sur l'algèbre de Galois. L'algèbre de Galois est une approche qui permet de regrouper les profils selon sur similarités. Il utilise sur la table des contextes où chaque élément est un couple de profils (statique, dynamique). Chaque couple doit être un couple complet selon les relations liant les profils statiques aux profils dynamiques. Cela signifie que les deux propriétés : $f(X)$ qui associe aux profils statiques, les profils dynamiques Y communs à tous les profils statiques et $g(Y)$ qui associe aux profils dynamiques, les profils statiques X communs à tous les profils dynamiques sont tel que $f(X) = Y$ et $g(Y) = X$. Le diagramme de Hasse associé à l'algèbre de Galois présente des groupes de profils similaires. Il représente une relation de généralisation / spécialisation entre groupes de profils. Le troisième niveau est la génération des liste de recommandations. Une fois les groupes de profils similaires constitués, la génération permet d'éliminer les doublons et de ne pas recommander plusieurs fois la même sémantique aux utilisateurs au même moment.

La phase de l'expérimentation nous a permis d'implémenter l'approche *Emergsem*. Nous avons comparé les résultats que nous avons obtenus d'une part avec les résultats obtenus via deux autres approches classiques que nous avons implémentées, et d'autre part avec les résultats de deux autres approches de recherches d'images. Nous avons constaté que nos résultats sont très encourageants comme le montre la table 6.22. Comme nous pouvons également le constater dans les résultats, la recommandation des sémantiques a joué un rôle très important dans la performance d'*Emergsem* parce qu'elle a permis de diminuer de façon considérable le temps mis pour rechercher les images.

7.2/ PERSPECTIVES

Selon les résultats obtenus, nos conditions de faisabilité intégrant la collaboration dans le processus d'annotation des images se révèle bénéfique aux systèmes récupération des images. Elle a permis d'avoir des annotations plus précises. Par ailleurs, malgré l'apport du système recommandation qui accélère la recherche des images, nous remarquons que l'annotation collaborative requiert beaucoup de ressources humaines et par conséquent beaucoup de temps. A l'issue de cette thèse, nous envisageons que :

- l'intégration d'une technique de propagation des annotations rendrait le processus d'annotation moins complexe. Ceci aura pour but de propager les annotations d'images à d'autres images ayant des objets en commun. Au lieu d'annoter toutes les images d'une collection, certaines images *témoins* pourront être répertoriées et annotées de façon collaborative.
- la prise en compte de l'avis des utilisateurs qui utilisent les descripteurs sémantiques pour rechercher les images serait indispensable pour améliorer la sémantique des images grâce au retour de pertinence. Jusqu'ici, seul le point de vue de la communauté d'annotateurs est pris en compte dans la description des images.
- la proposition d'un traitement automatique des ontologies qui prend en compte les relations spatiales pour aider à la formation des clusters de sémantiques pourrait permettre d'affiner le résultat des recherches.

Dans ce chapitre nous présentons les annexes relatives à notre travail. Dans un premier temps, nous présentons les ontologies développées, dans un deuxième temps quelques outils de construction annexe présente quelques codes d'interface et la troisième annexe compare l'utilisation d'une ontologie à l'utilisation d'un modèle relationnel pour stocker la description des images.

Sommaire

8.1	Les Ontologies	151
8.1.1	Ontologie d'application pour annoter les images	151
8.1.2	Ontologie des profils	157
8.2	Quelques outils de construction d'ontologie	161
8.3	Extrait du code d'enregistrement des annotations	163
8.4	Extrait du dictionnaire lexical	169
8.5	Publications	173

ANNEXE A

8.1/ LES ONTOLOGIES

Dans ce rapport, nous avons choisi de décrire les concepts et les relations avec le langage OWL - Ontology Web Language. Les concepts sont appelés classes et les relations propriétés dans Langage OWL. Une classe fournit un mécanisme d'abstraction pour le regroupement des ressources ayant des caractéristiques similaires tandis que les propriétés relient les classes. Ici nous utilisons OWL qui est un des langages utilisés par le Web Sémantique pour interpréter les ontologies.

8.1.1/ ONTOLOGIE D'APPLICATION POUR ANNOTER LES IMAGES

```
<?xml version="1.0" ?>

<rdf :RDF

xmlns="http://www.owl-ontologies.com/Ontology1414226260.owl"

xmlns :rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns"

xmlns :protege="http://protege.stanford.edu/plugins/owl/protege"

xmlns :xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl"

xmlns :owl="http://www.w3.org/2002/07/owl"

xmlns :xsd="http://www.w3.org/2001/XMLSchema"

xmlns :swrl="http://www.w3.org/2003/11/swrl"

xmlns :swrlb="http://www.w3.org/2003/11/swrlb"

xmlns :rdfs="http://www.w3.org/2000/01/rdf-schema"

xml :base="http://www.owl-ontologies.com/Ontology1414226260.owl">

<owl :Ontology rdf :about=""/>

<rdfs :Class rdf :ID="Acteur"/>

<rdfs :Class rdf :ID="Vegetation">
```

```
<rdfs :subClassOf>
<rdfs :Class rdf :ID="Environnement"/>
</rdfs :subClassOf>
</rdfs :Class>
<rdfs :Class rdf :ID="Construction">
<rdfs :subClassOf rdf :resource="Acteur"/>
</rdfs :Class>
<rdfs :Class rdf :ID="Animal">
<rdfs :subClassOf>
<rdfs :Class rdf :ID="Etre"/>
</rdfs :subClassOf>
</rdfs :Class>
<rdfs :Class rdf :ID="Inattendu">
<rdfs :subClassOf>
<rdfs :Class rdf :ID="Evenement"/>
</rdfs :subClassOf>
</rdfs :Class>
<rdfs :Class rdf :ID="Temps"/>
<rdfs :Class rdf :ID="Place">
<rdfs :subClassOf rdf :resource="Environnement"/>
</rdfs :Class>
<rdfs :Class rdf :ID="Organise">
<rdfs :subClassOf rdf :resource="Evenement"/>
</rdfs :Class>
<rdfs :Class rdf :ID="Insecte">
<rdfs :subClassOf>
```

```
<rdfs :Class rdf :about="Etre"/>
</rdfs :subClassOf>
</rdfs :Class>
<rdfs :Class rdf :ID="Objet">
<rdfs :subClassOf rdf :resource="Acteur"/>
</rdfs :Class>
<rdfs :Class rdf :ID="Oiseau">
<rdfs :subClassOf>
<rdfs :Class rdf :about="Etre"/>
</rdfs :subClassOf>
</rdfs :Class>
<rdfs :Class rdf :ID="Humain">
<rdfs :subClassOf>
<rdfs :Class rdf :about="Etre"/>
</rdfs :subClassOf>
</rdfs :Class>
<rdfs :Class rdf :ID="Espace">
<rdfs :subClassOf rdf :resource="Environnement"/>
</rdfs :Class>
<rdfs :Class rdf :about="Etre">
<rdfs :subClassOf rdf :resource="Acteur"/>
</rdfs :Class>
<owl :ObjectProperty rdf :ID="Decrit">
<rdfs :range rdf :resource="Etre"/>
<rdfs :domain rdf :resource="Acteur"/>
<owl :inverseOf>
```

```
<owl :ObjectProperty rdf :ID="EstDecritDans"/>
</owl :inverseOf>
</owl :ObjectProperty>
<owl :ObjectProperty rdf :ID="Contient">
<rdfs :domain>
<owl :Class>
<owl :unionOf rdf :parseType="Collection">
<rdfs :Class rdf :about="Acteur"/>
<rdfs :Class rdf :about="Environnement"/>
<rdfs :Class rdf :about="Evenement"/>
<rdfs :Class rdf :about="Temps"/>
</owl :unionOf>
</owl :Class>
</rdfs :domain>
<owl :inverseOf>
<owl :ObjectProperty rdf :ID="EstContenuDans"/>
</owl :inverseOf>
</owl :ObjectProperty>
<owl :ObjectProperty rdf :about="EstContenuDans">
<owl :inverseOf rdf :resource="Contient"/>
</owl :ObjectProperty>
<owl :ObjectProperty rdf :about="EstDecritDans">
<rdfs :range rdf :resource="Acteur"/>
<rdfs :domain rdf :resource="Etre"/>
<owl :inverseOf rdf :resource="Decrit"/>
</owl :ObjectProperty>
```

```
<owl :SymmetricProperty rdf :ID="EstUn1">
  <rdfs :domain rdf :resource="Environnement"/>
  <owl :inverseOf rdf :resource="EstUn1"/>
  <rdfs :range>
    <owl :Class>
      <owl :unionOf rdf :parseType="Collection">
        <rdfs :Class rdf :about="Espace"/>
        <rdfs :Class rdf :about="Place"/>
        <rdfs :Class rdf :about="Vegetation"/>
      </owl :unionOf>
    </owl :Class>
  </rdfs :range>
  <rdf :type rdf :resource="http ://www.w3.org/2002/07/owlObjectProperty"/>
</owl :SymmetricProperty>

<owl :SymmetricProperty rdf :ID="EstUn2">
  <owl :inverseOf rdf :resource="EstUn2"/>
  <rdfs :range>
    <owl :Class>
      <owl :unionOf rdf :parseType="Collection">
        <rdfs :Class rdf :about="Inattendu"/>
        <rdfs :Class rdf :about="Organise"/>
      </owl :unionOf>
    </owl :Class>
  </rdfs :range>
  <rdfs :domain rdf :resource="Evenement"/>
  <rdf :type rdf :resource="http ://www.w3.org/2002/07/owlObjectProperty"/>
```



```
</owl :SymmetricProperty>
<owl :SymmetricProperty rdf :ID="EstUn3">
<rdfs :range>
<owl :Class>
<owl :unionOf rdf :parseType="Collection">
<rdfs :Class rdf :about="Animal"/>
<rdfs :Class rdf :about="Humain"/>
<rdfs :Class rdf :about="Insecte"/>
<rdfs :Class rdf :about="Oiseau"/>
</owl :unionOf>
</owl :Class>
</rdfs :range>
<rdf :type rdf :resource="http ://www.w3.org/2002/07/owlObjectProperty"/>
<owl :inverseOf rdf :resource="EstUn3"/>
<rdfs :domain rdf :resource="Etre"/>
</owl :SymmetricProperty>
<Espace rdf :ID="Nuages"/>
<Humain rdf :ID="Enfant"/>
<Temps rdf :ID="Matin"/>
<Animal rdf :ID="Chien"/>
<Vegetation rdf :ID="Arbres"/>
<Place rdf :ID="Parking"/>
<Organise rdf :ID="Promenade"/>
<Objet rdf :ID="Voiture"/>
<Construction rdf :ID="Appartement"/>
<Humain rdf :ID="Homme"/>
```

```
</rdf :RDF>
```

8.1.2/ ONTOLOGIE DES PROFILES

```
<!DOCTYPE rdf :RDF [
  <!ENTITY owl "http ://www.w3.org/2002/07/owl" >
  <!ENTITY swrl "http ://www.w3.org/2003/11/swrl" >
  <!ENTITY swrlb "http ://www.w3.org/2003/11/swrlb" >
  <!ENTITY xsd "http ://www.w3.org/2001/XMLSchema" >
  <!ENTITY rdfs "http ://www.w3.org/2000/01/rdf-schema" >
  <!ENTITY rdf "http ://www.w3.org/1999/02/22-rdf-syntax-ns" >
  <!ENTITY protege "http ://protege.stanford.edu/plugins/owl/protege" >
  <!ENTITY xsp "http ://www.owl-ontologies.com/2005/08/07/xsp.owl" >
]>
<rdf :RDF xmlns="http ://www.owl-ontologies.com/Ontology1426086641.owl"
xml :base="http ://www.owl-ontologies.com/Ontology1426086641.owl"
xmlns :rdfs="http ://www.w3.org/2000/01/rdf-schema"
xmlns :swrl="http ://www.w3.org/2003/11/swrl"
xmlns :protege="http ://protege.stanford.edu/plugins/owl/protege"
xmlns :xsp="http ://www.owl-ontologies.com/2005/08/07/xsp.owl"
xmlns :owl="http ://www.w3.org/2002/07/owl"
xmlns :xsd="http ://www.w3.org/2001/XMLSchema"
xmlns :swrlb="http ://www.w3.org/2003/11/swrlb"
xmlns :rdf="http ://www.w3.org/1999/02/22-rdf-syntax-ns">
  <owl :Ontology rdf :about="" />
  <rdfs :Class rdf :ID="Adresse">
    <rdfs :subClassOf rdf :resource="Statiques" />
  </rdfs :Class>
```

```

<IdInternaute rdf :ID="Annot1" / >< owl : ObjectProperty rdf : ID = "Contient" >
<owl :inverseOf rdf :resource="Est-contenu-dans"/>
<rdfs :range rdf :resource="Profiles"/>
</owl :ObjectProperty>
<Nom rdf :ID="Durand"/>
<rdfs :Class rdf :ID="Dynamiques">
<rdfs :subClassOf rdf :resource="Profiles"/>
</rdfs :Class>
<owl :ObjectProperty rdf :ID="Est">
<rdfs :domain rdf :resource="Profiles"/>
<owl :inverseOf rdf :resource="Est-dans"/>
<rdfs :range>
<owl :Class>
<owl :unionOf rdf :parseType="Collection">
<rdfs :Class rdf :about="Dynamiques"/>
<rdfs :Class rdf :about="Statiques"/>
</owl :unionOf>
</owl :Class>
</rdfs :range>
</owl :ObjectProperty>
<owl :ObjectProperty rdf :ID="Est-contenu-dans">
<owl :inverseOf rdf :resource="Contient"/>
<rdfs :range rdf :resource="Profiles"/>
</owl :ObjectProperty>
<owl :ObjectProperty rdf :ID="Est-dans">
<rdfs :domain>

```

```

<owl :Class>
<owl :unionOf rdf :parseType="Collection">
<rdfs :Class rdf :about="Dynamiques"/>
<rdfs :Class rdf :about="Statiques"/>
</owl :unionOf>
</owl :Class>
</rdfs :domain>
<owl :inverseOf rdf :resource="Est"/>
<rdfs :range rdf :resource="Profils"/>
</owl :ObjectProperty>
<owl :ObjectProperty rdf :ID="Est-possede">
<rdfs :domain>
<owl :Class>
<owl :unionOf rdf :parseType="Collection">
<rdfs :Class rdf :about="Adresse"/>
<rdfs :Class rdf :about="IdInternaute"/>
<rdfs :Class rdf :about="Mot_de_passe"/ >< rdfs : Classrdf : about = "Nom" / >
<rdfs :Class rdf :about="Prenom"/>
<rdfs :Class rdf :about="Pseudo"/>
</owl :unionOf>
</owl :Class>
</rdfs :domain>
<owl :inverseOf rdf :resource="Poss232 ;de"/>
<rdfs :range rdf :resource="Statiques"/>
</owl :ObjectProperty>
<rdfs :Class rdf :ID="IdInternaute">

```

```

<rdfs :subClassOf rdf :resource="Statiques"/>

</rdfs :Class>

<rdfs :Class rdf :ID="Mot_de_passe" >< rdfs : subClassOf rdfs : resource = "Statiques" / >

</rdfs :Class>

<rdfs :Class rdf :ID="Nom">

<rdfs :subClassOf rdf :resource="Statiques"/>

</rdfs :Class>

<Mot_de_passe rdfs : ID = "Passeword12" / >< Pseudordf : ID = "Peter12" / >

<Prenom rdf :ID="Pierre"/>

<owl :ObjectProperty rdf :ID="Poss232;de">

<rdfs :domain rdf :resource="Statiques"/>

<owl :inverseOf rdf :resource="Est-possede"/>

<rdfs :range>

<owl :Class>

<owl :unionOf rdf :parseType="Collection">

<rdfs :Class rdf :about="Adresse"/>

<rdfs :Class rdf :about="Mot_de_passe" / >< rdfs : Class rdfs : about = "Nom" / >

<rdfs :Class rdf :about="Prenom"/>

<rdfs :Class rdf :about="Pseudo"/>

</owl :unionOf>

</owl :Class>

</rdfs :range>

</owl :ObjectProperty>

<rdfs :Class rdf :ID="Prenom">

<rdfs :subClassOf rdf :resource="Statiques"/>

```

```

</rdfs :Class>

<rdfs :Class rdf :ID="Profiles"/>

<rdfs :Class rdf :ID="Pseudo">

<rdfs :subClassOf rdf :resource="Statiques"/>

</rdfs :Class>

<Adresse rdf :ID="Rue_des_Moutons_21000_Dijon"/>

<Dynamiques rdf :ID="Semantique1"/ >< rdfs : Classrdf : ID = "Statiques" >

<rdfs :subClassOf rdf :resource="Profiles"/>

</rdfs :Class>

</rdf :RDF>

```

8.2/ QUELQUES OUTILS DE CONSTRUCTION D'ONTOLOGIE

Pour construire les ontologies, plusieurs outils sont proposés dans la littérature :

- **Ontolingua** : Ontolingua est un serveur d'édition d'ontologies. Il utilise des classes, des relations, des fonctions, des instances et des axiomes pour décrire une ontologie. Une relation peut contenir des propriétés nécessaires (contraintes) ou nécessaires et suffisantes qui définissent la relation. En plus le serveur Ontolingua offre la possibilité d'intégrer les ontologies Ontolingua, ce qui permet la construction modulaire des ontologies.
- **OntoSaurus** : OntoSaurus de l'Information Science Institute de l'Université de Southern California est composé de deux modules : un serveur utilisant LOOM comme langage de représentation des connaissances, et un serveur de navigation créant dynamiquement des pages HTML qui affichent la hiérarchie de l'ontologie ; le serveur utilise des formulaires HTML pour permettre à l'utilisateur d'éditer l'ontologie.
- **WebOnto** : WebOnto du Knowledge Media Institute de l'Open University, est une application Web pour naviguer et développer collaborativement les ontologies. Il supporte la navigation collaborative, la création et l'édition d'ontologies sur le Web. Les ontologies WebOnto sont implémentées dans le langage OCML. WebOnto distingue quatre types d'ontologies : ontologie de domaine, ontologie de tâche, ontologie de méthode, et ontologie d'application.
- **OilEd** : OilEd (Oil Editor) est un éditeur d'ontologies utilisant le formalisme OIL. Il est essentiellement dédié à la construction de petites ontologies dont on peut ensuite tester la cohérence à l'aide de FACT, un moteur d'inférences bâti sur OIL.
- **Protégé** : est une interface modulaire permettant l'édition, la visualisation, le contrôle d'ontologie, l'extraction d'ontologies à partir de sources textuelles, et la fusion semi-automatique d'ontologies. Le modèle de connaissances sous-jacent à protégé 2000 est issu du modèle des frames et contient des classes, des propriétés et des valeurs des propriétés et contraintes, ainsi

que des instances des classes et des propriétés. Il autorise la définition de méta-classes, dont les instances sont des classes, ce qui permet de créer son propre modèle de connaissances avant de bâtir une ontologie.

- ODE et WebOde : L'outil ODE (Ontology Design Environment) permet de construire des ontologies au niveau connaissance, comme le préconise la méthodologie METHONTOLOGY. L'utilisateur construit son ontologie dans un modèle de type frame, en spécifiant les concepts du domaine, les termes associés, les attributs et leurs valeurs, les relations de subsomption.
- OntoEdit : OntoEdit (Ontology Editor) est également un environnement de construction d'ontologies indépendant de tout formalisme. Il permet l'édition des hiérarchies de concepts et de relations et l'expression d'axiomes algébriques portant sur les relations, et de propriétés telles que la généralité d'un concept. Des outils graphiques dédiés à la visualisation d'ontologies sont inclus dans l'environnement. OntoEdit intègre un serveur destiné à l'édition d'une ontologie par plusieurs utilisateurs. Un contrôle de la cohérence de l'ontologie est assuré à travers la gestion des ordres d'édition.

ANNEXE B

8.3/ EXTRAIT DU CODE D'ENREGISTREMENT DES ANNOTATIONS

Cette sous-section présente quelques codes de l'application que nous avons développée. L'application est développée sous le langage c-sharp.

```
public void charger_grid()
{
    int p = dataGridView4.RowCount;

    if (p == 0)
    {
        return;
    }
    if (p == 1)
    {
        DataSet ds = new DataSet("1");
        DataTable dt = new DataTable("Mots_cles");
        ds.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 1 + ".xml");
        dt = ds.Tables["Mots_cles"];
        dataGridView1.DataSource = dt;
    }
    if (p == 2)
    {
        DataSet ds1 = new DataSet("1");
        DataTable dt1 = new DataTable("Mots_cles");
        ds1.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 1 + ".xml");
        dt1 = ds1.Tables["Mots_cles"];
        dataGridView1.DataSource = dt1;

        DataSet ds2 = new DataSet("2");
        DataTable dt2 = new DataTable("Mots_cles");
        ds2.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 2 + ".xml");
        dt2 = ds2.Tables["Mots_cles"];
        dataGridView2.DataSource = dt2;
    }
}
```



```
if (p == 3)
{
    DataSet ds1 = new DataSet("1");
    DataTable dt1 = new DataTable("Mots_cles");
    ds1.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 1 + ".xml");
    dt1 = ds1.Tables["Mots_cles"];
    dataGridView1.DataSource = dt1;

    DataSet ds2 = new DataSet("2");
    DataTable dt2 = new DataTable("Mots_cles");
    ds2.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 2 + ".xml");
    dt2 = ds2.Tables["Mots_cles"];
    dataGridView2.DataSource = dt2;

    DataSet ds3 = new DataSet("3");
    DataTable dt3 = new DataTable("Mots_cles");
    ds3.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 3 + ".xml");
    dt3 = ds3.Tables["Mots_cles"];
    dataGridView5.DataSource = dt3;

    int i = dataGridView1.ColumnCount;
    int ii = 0;
    int lign = 0;
    while (ii < i)
    {
        ;
        dataGridView3.Rows[lign].Cells[ii].Value = dataGridView1.Rows[0].Cells[ii].Value;
        ii = ii + 1;
    }
    dataGridView3.Rows.Add();
    int j = dataGridView2.ColumnCount;
    int jj = 0;

    while (jj < j)
    {
        dataGridView3.Rows[lign].Cells[jj].Value = dataGridView2.Rows[0].Cells[jj].Value;
        jj = jj + 1;
    }
    dataGridView3.Rows.Add();
    int k = dataGridView2.ColumnCount;
    int kk = 0;

    while (k < kk)
    {
        dataGridView3.Rows[lign].Cells[jj].Value = dataGridView5.Rows[0].Cells[kk].Value;
        kk = kk + 1;
    }
}
```

```
if (p == 6)
{
    DataSet ds1 = new DataSet("1");
    DataTable dt1 = new DataTable("Mots_cles");
    ds1.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 1 + ".xml");
    dt1 = ds1.Tables["Mots_cles"]; /
    dataGridView1.DataSource = dt1;

    DataSet ds2 = new DataSet("2");
    DataTable dt2 = new DataTable("Mots_cles");
    ds2.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 2 + ".xml");
    dt2 = ds2.Tables["Mots_cles"];
    dataGridView2.DataSource = dt2;

    DataSet ds3 = new DataSet("3");
    DataTable dt3 = new DataTable("Mots_cles");
    ds3.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 3 + ".xml");
    dt3 = ds3.Tables["Mots_cles"];
    dataGridView5.DataSource = dt3;

    DataSet ds4 = new DataSet("4");
    DataTable dt4 = new DataTable("Mots_cles");
    ds4.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 4 + ".xml");
    dt4 = ds4.Tables["Mots_cles"];
    dataGridView6.DataSource = dt4;

    DataSet ds5 = new DataSet("5");
    DataTable dt5 = new DataTable("Mots_cles");
    ds5.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 5 + ".xml");
    dt5 = ds5.Tables["Mots_cles"];
    dataGridView7.DataSource = dt5;

    DataSet ds6 = new DataSet("6");
    DataTable dt6 = new DataTable("Mots_cles");
    ds6.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 6 + ".xml");
    dt6 = ds6.Tables["Mots_cles"];
    dataGridView8.DataSource = dt6;
}
```

```
if (p == 7)
{
    DataSet ds1 = new DataSet("1");
    DataTable dt1 = new DataTable("Mots_cles");
    ds1.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 1 + ".xml");
    dt1 = ds1.Tables["Mots_cles"];
    dataGridView1.DataSource = dt1;

    DataSet ds2 = new DataSet("2");
    DataTable dt2 = new DataTable("Mots_cles");
    ds2.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 2 + ".xml");
    dt2 = ds2.Tables["Mots_cles"];
    dataGridView2.DataSource = dt2;

    DataSet ds3 = new DataSet("3");
    DataTable dt3 = new DataTable("Mots_cles");
    ds3.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 3 + ".xml");
    dt3 = ds3.Tables["Mots_cles"];
    dataGridView5.DataSource = dt3;

    DataSet ds4 = new DataSet("4");
    DataTable dt4 = new DataTable("Mots_cles");
    ds4.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 4 + ".xml");
    dt4 = ds4.Tables["Mots_cles"];
    dataGridView6.DataSource = dt4;

    DataSet ds5 = new DataSet("5");
    DataTable dt5 = new DataTable("Mots_cles");
    ds5.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 5 + ".xml");
    dt5 = ds5.Tables["Mots_cles"];
    dataGridView7.DataSource = dt5;

    DataSet ds6 = new DataSet("6");
    DataTable dt6 = new DataTable("Mots_cles");
    ds6.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 6 + ".xml");
    dt6 = ds6.Tables["Mots_cles"];
    dataGridView8.DataSource = dt6;

    DataSet ds7 = new DataSet("7");
    DataTable dt7 = new DataTable("Mots_cles");
    ds7.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 7 + ".xml");
    dt7 = ds7.Tables["Mots_cles"];
    dataGridView9.DataSource = dt7;
}
```

```
if (p == 8)
{
    DataSet ds1 = new DataSet("1");
    DataTable dt1 = new DataTable("Mots_cles");
    ds1.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 1 + ".xml");
    dt1 = ds1.Tables["Mots_cles"];
    dataGridView1.DataSource = dt1;

    DataSet ds2 = new DataSet("2");
    DataTable dt2 = new DataTable("Mots_cles");
    ds2.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 2 + ".xml");
    dt2 = ds2.Tables["Mots_cles"];
    dataGridView2.DataSource = dt2;

    DataSet ds3 = new DataSet("3");
    DataTable dt3 = new DataTable("Mots_cles");
    ds3.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 3 + ".xml");
    dt3 = ds3.Tables["Mots_cles"]; /
    dataGridView5.DataSource = dt3;

    DataSet ds4 = new DataSet("4");
    DataTable dt4 = new DataTable("Mots_cles");
    ds4.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 4 + ".xml");
    dt4 = ds4.Tables["Mots_cles"];
    dataGridView6.DataSource = dt4;

    DataSet ds5 = new DataSet("5");
    DataTable dt5 = new DataTable("Mots_cles");
    ds5.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 5 + ".xml");
    dt5 = ds5.Tables["Mots_cles"];
    dataGridView7.DataSource = dt5;

    DataSet ds6 = new DataSet("6");
    DataTable dt6 = new DataTable("Mots_cles");
    ds6.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 6 + ".xml");
    dt6 = ds6.Tables["Mots_cles"]; /
    dataGridView8.DataSource = dt6;

    DataSet ds7 = new DataSet("7");
    DataTable dt7 = new DataTable("Mots_cles");
    ds7.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 7 + ".xml");
    dt7 = ds7.Tables["Mots_cles"];
    dataGridView9.DataSource = dt7;

    DataSet ds8 = new DataSet("8");
    DataTable dt8 = new DataTable("Mots_cles");
    ds8.ReadXml(@"\Annot_Sem\Annot_Sem\bin\Debug\Annotation\" + 8 + ".xml");
    dt8 = ds8.Tables["Mots_cles"];
    dataGridView10.DataSource = dt8;
}
```

```

private void Form2_Load(object sender, EventArgs e)
{
    dataGridView3.Visible = false;
    string direct;
    direct = (Directory.GetCurrentDirectory());
    direct = "";
    DirectoryInfo dir = new DirectoryInfo(direct + @"\Annot_Sem\Annot_Sem\bin\Debug\Annotation");
    //DirectoryInfo dir = (Directory.GetCurrentDirectory() + @"\images\Bd");
    dataGridView4.DataSource = dir.GetFiles("*.xml");

    // Affichage des annotations
    if (dataGridView3.Rows[0].Cells[0].Value == null)
    {
        charger_grid();
    }
}

public void remplir_grid_des_poids(DataGridView a, int i, int j)
{
    string nom_colonne="";
    while (j < 17)
    {
        dataGridView14.Rows[i].Cells[0].Value = a.Rows[0].Cells[j].Value;
        nom_colonne = a.Columns[j].HeaderText;
        dataGridView14.Rows[i].Cells[2].Value=nom_colonne;
        j = j + 1;
        i = i + 1;
    }
}

public void remplir_grid_des_poids_evaluation(int v)
{
    /** première colonne
    // Afficher les données dans la grid
    int j = 0;
    // string colonne = "";
    string concept1 = "";
    while (j < (dataGridView3.RowCount - 1))
    {
        concept1 = dataGridView3.Rows[j].Cells[v].Value.ToString();

        j = j + 1;
        if (concept1 != dataGridView3.Rows[j].Cells[1].Value.ToString() )
        {
            dataGridView15.Rows.Add(concept1);
            //dataGridView15.Rows[j].Cells[1].
        }
    }
}

```

ANNEXE C

8.4/ EXTRAIT DU DICTIONNAIRE LEXICAL

Cette annexe présente quelques parties du dictionnaire lexical que nous avons mis à la disposition de la communauté d'annotateurs pour l'annotation sémantique collaborative des images.

```
00013236-r fra:lemma mal
00013328-v fra:lemma jouer
00013328-v fra:lemma jouet
00013328-v fra:lemma jeu
00013429-r fra:lemma bien
00013429-r fra:lemma puits
00013442-a fra:lemma objectif
00013615-v fra:lemma représenter
00013615-v fra:lemma jeu
00013615-v fra:lemma jouer
00013615-v fra:lemma sembler
00013615-v fra:lemma agir
00013615-v fra:lemma présenter
00013626-r fra:lemma largement
00013626-r fra:lemma bien
00013626-r fra:lemma puits
00013626-r fra:lemma richement
00013662-a fra:lemma réel
00013662-a fra:lemma tangible
```

```
00005567-r fra:lemma automatiquement
00005599-a fra:lemma absolu
00005599-a fra:lemma implicite
00005674-r fra:lemma horriblement
00005674-r fra:lemma terriblement
00005718-a fra:lemma infini
00005779-r fra:lemma largement
00005779-r fra:lemma immensément
00005779-r fra:lemma vastement
00005779-r fra:lemma beaucoup
00005779-r fra:lemma abondamment
00005815-v fra:lemma tousser
00005815-v fra:lemma toux
00006024-n fra:lemma Hétérotrophie
00006024-n fra:lemma hétérotrophie
00006032-a fra:lemma relatif
00006032-a fra:lemma relative
```

```
00007015-r fra:lemma environ
00007015-r fra:lemma à peu près
00007015-r fra:lemma peu ou prou
00007096-a fra:lemma hygrosopique
00007328-v fra:lemma bâillement
00007328-v fra:lemma bâiller
00007347-n fra:lemma cause
00007347-n fra:lemma raison
00007488-r fra:lemma vraiment
00007488-r fra:lemma absolument
00007516-a fra:lemma spongieux
00007549-v fra:lemma renifler
00007697-a fra:lemma assoiffé
00007703-r fra:lemma partiellement
00007703-r fra:lemma partie
00007703-r fra:lemma en partie
00007739-v fra:lemma ciller
```

13157595-n fra:lemma épingle
 13157595-n fra:lemma aiguille
 13161506-n fra:lemma lobe
 13161998-n fra:lemma rosette
 13162164-n fra:lemma Ligule
 13162164-n fra:lemma ligule
 13162297-n fra:lemma écorce
 13162297-n fra:lemma aboyer
 13162297-n fra:lemma cortex
 13163093-n fra:lemma angostura
 13163250-n fra:lemma branche
 13163471-n fra:lemma herbe
 13163649-n fra:lemma chaume
 13163991-n fra:lemma brindille
 13163991-n fra:lemma rameau
 13164583-n fra:lemma pousse
 13164881-n fra:lemma pigeonner
 13165086-n fra:lemma bourgeon
 13165727-n fra:lemma branchage
 13165727-n fra:lemma rameau
 13165727-n fra:lemma branche
 13165815-n fra:lemma tribu
 13165815-n fra:lemma tronc
 13165815-n fra:lemma coffre

00014201-v fra:lemma frissonner
 00014285-r fra:lemma amplement
 00014285-r fra:lemma puits
 00014285-r fra:lemma terriblement
 00014285-r fra:lemma bien
 00014285-r fra:lemma sérieusement
 00014285-r fra:lemma copieusement
 00014285-r fra:lemma largement
 00014285-r fra:lemma considérablement
 00014405-v fra:lemma se reposer
 00014490-a fra:lemma ample
 00014490-a fra:lemma large
 00014490-a fra:lemma profus
 00014490-a fra:lemma abondant
 00014490-a fra:lemma plantureux
 00014490-a fra:lemma grand
 00014616-r fra:lemma alors
 00014616-r fra:lemma puits
 00014616-r fra:lemma bien
 00014738-r fra:lemma mal
 00014742-v fra:lemma dormir
 00014858-a fra:lemma étendu

11479816-n fra:lemma diamagnétisme
 11480091-n fra:lemma ferrimagnétisme
 11480284-n fra:lemma ferromagnétisme
 11480521-n fra:lemma paramagnétisme
 11480930-n fra:lemma son
 11481209-n fra:lemma ultrason
 11481334-n fra:lemma vol
 11481334-n fra:lemma trajectoire
 11481334-n fra:lemma trajet
 11481487-n fra:lemma balistique
 11482013-n fra:lemma énergie mécanique
 11482140-n fra:lemma réchauffement
 11482140-n fra:lemma dégel
 11482312-n fra:lemma micro-onde
 11482579-n fra:lemma soleil de minuit
 11482706-n fra:lemma brume
 11482706-n fra:lemma brouillard
 11482873-n fra:lemma mistral
 11482985-n fra:lemma moment
 11483472-n fra:lemma dipôle électrostatique
 11483829-n fra:lemma moment magnétique
 11483990-n fra:lemma météore
 11483990-n fra:lemma étoile filante
 11484260-n fra:lemma bolide
 11484375-n fra:lemma mirage
 11484570-n fra:lemma mousson
 11484741-n fra:lemma mousson
 11484975-n fra:lemma lune

03106846-a fra:lemma pyrotechnique
 03107152-n fra:lemma cordite
 03107596-a fra:lemma qatari
 03107596-a fra:lemma Qatarien
 03107596-a fra:lemma qatarien
 03107801-a fra:lemma quadratique
 03107904-n fra:lemma noyau
 03107904-n fra:lemma central
 03107904-n fra:lemma cœur
 03108069-n fra:lemma noyau
 03108077-a fra:lemma quadratique
 03108853-n fra:lemma bouchon
 03108853-n fra:lemma liège
 03108861-a fra:lemma rabelaisien
 03109141-a fra:lemma récessif
 03109150-n fra:lemma tire-bouchon
 03109350-n fra:lemma Université Cornell
 03109486-n fra:lemma tournant
 03109486-n fra:lemma corne
 03109486-n fra:lemma coin
 03109486-n fra:lemma cor
 03109693-n fra:lemma coin
 03109795-a fra:lemma résidentiel

06095022-n fra:lemma astronomie
 06096451-n fra:lemma mécanique spatiale
 06096600-n fra:lemma astrométrie
 06096759-n fra:lemma radioastronomie
 06096913-n fra:lemma astronautique
 06096913-n fra:lemma aéronautique
 06097231-n fra:lemma avionique
 06097231-n fra:lemma Avionique
 06097478-n fra:lemma biophysique
 06097594-n fra:lemma mécanique céleste
 06097775-n fra:lemma astrophysique
 06098195-n fra:lemma cosmogonie
 06098195-n fra:lemma cosmologie
 06098195-n fra:lemma Cosmogonie
 06098687-n fra:lemma cryogénie
 06098876-n fra:lemma cristallographie
 06099107-n fra:lemma électromagnétisme
 06099269-n fra:lemma électronique
 06100101-n fra:lemma électrostatique
 06100236-n fra:lemma mécanique
 06100555-n fra:lemma physique nucléaire
 06100555-n fra:lemma physique atomique

06545137-n fra:lemma geste
 06545137-n fra:lemma action
 06545137-n fra:lemma titre
 06545137-n fra:lemma document
 06545137-n fra:lemma acte
 06545137-n fra:lemma pièce
 06547059-n fra:lemma mandat
 06548110-n fra:lemma lettre de cachet
 06548671-n fra:lemma retour
 06548671-n fra:lemma déclaration d'impôt
 06549661-n fra:lemma autorisation
 06549661-n fra:lemma permettre
 06549661-n fra:lemma permis
 06549661-n fra:lemma certificat
 06549661-n fra:lemma licence
 06549661-n fra:lemma débit
 06550046-n fra:lemma permis de construire
 06550206-n fra:lemma permis de conduire
 06550381-n fra:lemma permis de pêche
 06550891-n fra:lemma lettre de course
 06550891-n fra:lemma lettre de marque
 06551627-n fra:lemma brevet
 06551627-n fra:lemma lettre patente
 06551784-n fra:lemma arrêt

00004722-r fra:lemma sauf
 00004722-r fra:lemma cependant
 00004722-r fra:lemma seulement
 00004819-v fra:lemma éternuement
 00004819-v fra:lemma éternuer
 00004967-r fra:lemma simplement
 00004967-r fra:lemma seulement
 00005041-v fra:lemma inhaler
 00005041-v fra:lemma inspirer
 00005041-v fra:lemma respirer
 00005055-r fra:lemma simplement
 00005055-r fra:lemma seulement
 00005055-r fra:lemma strictement

ANNEXE D

8.5/ PUBLICATIONS

1- Exploiting Semantic Indexing Images for Emergence Recommendation Semantics System in International Conference on Systems and Networks Communications (ICSNC), Italie, October 2013.

2- Collaborative Semantic Annotation Images : Ontology-Based Model in Signal Image Processing : An International Journal (SIPIJ) Vol.4, No.6, December 2013.

3- Contributions of Semantic Web Tools to Relational Models in International Journal of Computer Theory And Engineering (IJCTE), Vol.6, No.3, June 2014.

4- Emergsem : Semantic Emergent And Recommendation System For Image Retrieval, International Conference on Signal-Image Technology Internet-Based Systems, December 2014.

BIBLIOGRAPHIE

- [1] Jérôme Champavère. De la représentation des connaissances au web sémantique : Un survol. Tiré de : <http://www.grappa.univ-lille3.fr/champavere/Enseignement/0809/13miashs/ia/rc-ws.pdf>, Decembre 2013.
- [2] Henri F. and Lundgren-Cayrol K. Apprentissage collaboratif à distance. *Presses de l'Université du Québec*, 2001.
- [3] Lynne Dunckley. Multimedia database : An object-relational approach. *Book, Published in Great Britain*, pages 337–338, 2003.
- [4] Staab S. Petridis K. Anastasopoulos D. Saathoff C., Timmermann N. and Kompatsiaris Y. M-ontomat-annotizer : Linking ontologies with multimedia low-level features for automatic image annotation. *In : Demos and Posters of the 3rd European SemanticWeb Conference (ESWC)*, 2006.
- [5] I. L. Coman I. K. Sethi. Mining association rules between low-level image features and high-level concepts. *Proceedings of the SPIE Data Mining And Knowledge Discovery*, III :279–290, 2001.
- [6] A. Vailaya A. K Jain. Image retrieval using colour and shape. *Pattern Recognition*, 29 :1233–1244, 1993.
- [7] A. D. Bimbo G. Serra A. D. Bagdanov, M. Bertini and C. Torniai. Semantic annotation and retrieval of video events using multimedia ontologies. *In IEEE Computer Society Conference*, pages 713–720, 2007.
- [8] D. Forsyth N. de Freitas D. M. Blei K. Barnard, P. Duygulu and M. I. Jordan. Matching words and pictures. *JMLR*, 3 :1107–1135, 2003.
- [9] S. Dasiopoulou and I. Kompatsiaris. Trends and issues in description logics frameworks for image interpretation. *In Artificial Intelligence : Theories, Models and Applications*, 6040 :611–770, November 2010.
- [10] M. Dong C. Yang and J. Hua. Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning. *in Conference on Computer Vision and Pattern Recognition (CVPR) 06*, pages 2057–2063, 2006.
- [11] L. Zhang C. Wang, F. Jing and H.J. Zhang. Image annotation refinement using random walk with restarts. *in ACM MULTIMEDIA 06*, pages 647–650, 2006.
- [12] Z. Li W.Y. Ma X. Rui, M. Li and N. Yu. Bipartite graph reinforcement model for web image annotation. *in MULTIMEDIA 07*, pages 585–594, 2007.

- [13] Sylvie Cazalens Anthony Ventresque, Philippe Lamarre. Echange d'information grâce à des caractéristiques sémantiques. *Modèles Formels de l'interaction - MFI 05*, Caen :France 2005.
- [14] Simone Marchi Simonetta Montemagni Claudio Andreatta Rodolfo Stecher Roberto Bartolini, Emiliano Giovannetti and Paolo Bouquet. Multimedia information extraction in ontology-based semantic annotation of product catalogues. *Semantic Web Applications and Perspectives - SWAP*, October 27 2006.
- [15] I. Terziev D. Manov A. Kiryakov, B. Popov and D. Ognyanoff. Semantic annotation, indexing, and retrieval. *Journal of Web Semantics*, 2(1), November 2004.
- [16] G. Adomavicius and A. Tuzhilin. Towards the next generation of recommender systems : A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Engin.*, 17(6) :734–749, 2005.
- [17] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. *In CHI' 04*, 2004.
- [18] Kyle Heath Zixuan Wang, Omprakash Gnawali and Leonidas J. Guibas. Collaborative image annotation using image webs. *Army Science Conference*, 2010.
- [19] Apostol (Paul) Natsev Timo Volkmer, John R. Smith. A web-based system for collaborative annotation of large image and video collections. *MM'05*, November 6–12, 2005, Singapore.
- [20] Michal Hradis Roman Jurànek Jiri Havel Radovan Josth Adam Herout, Pavel Zemcik and Lukas Polok. Image features for real-time object detection. *ICCVG 2008 Proceedings of the International Conference on Computer Vision and Graphics : Revised Papers*, pages 239–248, 2008.
- [21] Carsten Saathoff Thomas Franz Miroslav Vacura, Vojtech Svatek and Raphael Troncy. Describing low-level image features using the comm ontology. *in Proc. 15th International Conference on Image Processing (ICIP) Workshop on Multimedia Information Retrieval*, 2008.
- [22] Wolfgang Förstner. A framework for level feature extraction. *ECCV-European conference on Computer Vision*, 2 :383–394, 1994.
- [23] H. Soliman M.M. El-gayar and N. Meki. A comparative study of image low level feature extraction algorithms. *Egyptian Informatics journal*, 4 :175–181, July 2013.
- [24] W.C. Karl H. Feng, D.A. Castanon. A curve evolution approach for image segmentation using adaptive flows. *Proceedings of the International Conference on Computer Vision (ICCV)*, 01 :494–499, 2001.
- [25] B.S. Majunath W.Y. Ma. Edge flow : a framework of boundary detection and image segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 744–749, 1997.

- [26] Guojun Lu Wei-Ying Ma Ying Liu, Dengsheng Zhang. A survey of content-based image retrieval with high-level semantics. *Journal Pattern Recognition Archive*, 40 :262–282, Issue 1, January 2007.
- [27] Nicolas ZLATOFF. Indexation d'images 2d vers une reconnaissance d'objets multi-critères. *Thèse*, 12 Juillet 2006.
- [28] Zhang T. Tan. Bref review of invariant texture analysis methods. *Pattern Recognition*, 35 :735–747, 2002.
- [29] J. Weszka A. Rosenfeld. Picture recognition. *In Digital Pattern Recognition*, pages 135–166, K. Fu(Ed.) Springer-Verlag 1980.
- [30] G. Lu D. Zhang. Study and evaluation of different fourier methodes for image retrieval. *Image Vision Computing*, 23, 2005.
- [31] J. Daugman. Uncertainty relation for resolution in space, saptial frequency and orientation optimised by two-dimensionnal visual cortical filters. *Journal of the Optical Society of America*, 2 :1160–1169, 1985.
- [32] W. Giesler A. Bovik, M. Clark. Multichannel texture analysis using localised spatial filters. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12 :53–73, 1990.
- [33] S. Mallat. Multifrequency channel decomposition of image and wavelet models. *IEEE Trans. Acoustic, Speech and Signal Processing*, 37 (12) :2091–2110, 1989.
- [34] J. Fan A. Laine. Texture classification by wavelet packet signatures. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15 (11) :1186–1191, 1993.
- [35] V. Hlavac M. Sonka and R. Boyle. Image processing, analysis and machine vision. *PWS Publishing, seconde edition*, 1999.
- [36] M.G. Strintzis V. Mezaris, I. Kompatsiaris. An ontology approach to object-based image retrieval. *Proceedings of the ICIP*, 02 :511–514, 2003.
- [37] D. Chan G. Wiederhold J.Z. Wang, J. Li. Semantics-sensitive retrieval for digital picture libraries. *Digital Library Magazine*, 40, 1999.
- [38] B. Manjunath W.Y. Ma. Netra : a toolbox for navigating large image databases. *Proceedings of the IEEE International Conference on Image Processing*, 40 :568–571, 1997.
- [39] A. Zhang Y. Song, W. Wang. Automatic annotation and retrieval of images. *J. World Wide Web*, 6 (2) :209–231, 2003.
- [40] B. Rogowitz A. Mojsilovic. Isee : perceptual features for image library navigation. *Proceedings of the SPIE, Human Vision and Electronic Imaging*, 4662 :266–277, 2002.
- [41] Nesrine Chehata Franck Taillandier Laurent Guigues, Roger Trias-Sanz and Matthieu Deveau. Segmentation multi-échelles d'images : théorie et applications. *Bulletin d'Information Scientifique Technique de l'IGN*, 75(1), January 2006.
- [42] P. Meer D. Comaniciu. Robust analysis of feature spaces : color image segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 75(1) :750–755, 1997.

- [43] B. Dimitrov P.L. Stanchev, D. Green Jr. High level color similarity retrieval. *Int. J. Inf. Theories Appl.*, 10(3) :363–369, 2003.
- [44] J.-H. Oh K.A. Hua, K. Vu. Sammatch : a flexible and efficient sampling-based image retrieval technique for large image databases. *Proceedings of the Seventh ACM International Multimedia Conference*, 10(3) :225–234, November 1999.
- [45] Jean-Yves Audibert Anne-Marie Touscha, Stéphane Herbin. Semantic hierarchies for image annotation : A survey. *Pattern Recognition*, 45 (1) :333–345, January 2012.
- [46] Marcel Worring IEEE-Simone Santani Amarnath Gupta Ramesh Jain Arnold Smeulders, IEEE. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 22, N 12, 2000.
- [47] Philippe Laublet Jean Charlet and Chantal Reynaud. Action spécifique 32 cnrs / stic web sémantique :. *Rapport final*, December 2003.
- [48] Collections SWRL. <http://protege.cim3.net/cgi-bin/wiki.pl?collectionsswrl>. October 2011.
- [49] Xavier Lacot. Introduction à owl, un langage xml d'ontologies web. Juin 2005.
- [50] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6) :907–928, 1995.
- [51] Guus Schreiber-Jan Wielemaker Bob Wielinga Laura Hollink, Giang Nguyen and Marcel Worring. Adding spatial semantics to image annotations. *4th International Workshop On Knowledge Markup And Semantic Annotation At Iswc'04*, 2004.
- [52] Nicola Guarino. Formal ontology in information systems. *Proceedings of FOIS'98, IOS Press*, pages 3–15, June 1998.
- [53] R. M. Smullyan. First-order logic. *Springer-Verlag*, Berlin,1968.
- [54] S. Decker-D. Fensel F. van Harmelen J. Broekstra, M. Klein and I. Horrocks. Enabling knowledge representation on the web by extending rdf schema. *Computer Networks*, 39(5) :609–634, 2002.
- [55] Thi Anh Le PHAM. De l'optimisation à la décomposition de l'ontologie dans la logique de description. *Thèse*, 2008.
- [56] Raphaël Troncy Jean Charlet, Bruno Bachimont. Ontologies pour le web sémantique. *Rapport*, 2003.
- [57] Jérôme Champavère. De la représentation des connaissances au web sémantique : Un survol. *Rapport*, Mars 2010.
- [58] P. Lepain P. Aigrain, P. Joly and V. Longueville. Content-based representation and retrieval of visual media : A-state-of-the-art review. *Journal of Multimedia Tools and Applications*, 3 :179–202, 1996 1996.
- [59] J.J. Fan and K.Y. Su. An efficient algorithm for matching multiple patterns. *IEEE Trans. Knowl. Data Eng.*, 5(2) :339–351, April 1993.

- [60] G. Salton and M. J. McGill. An introduction to modern information retrieval. *McGraw-Hill*, 1983.
- [61] G. Bonfiglio-Dosio M. Agosti and N. Ferro. A historical and contemporary study on annotations to derive key features for systems design. *in Int. Digital Libraries Journal*, 8(1) :1–19, 2007.
- [62] J. Bradley. Pliny : A model for digital support of scholarship. *Journal of Digital Information*, 9(1), 2008.
- [63] J.M. Jolion S. Bres and F. Lebourgeois. Traitement et analyse des images numériques. *Hermes Lavoisier*, 2003.
- [64] D. Lingrand. Introduction au traitement d'images. *Vuibert*, 2004.
- [65] Fabio Crestani Sasa NESIC, Mehdi Jazayeri and Dragan GASEVIC. Concept-based semantic annotation, indexing and retrieval of office-like document units. *RIAO 10 Adaptivity, Personalization and Fusion of Heterogeneous Information*, pages 134–135, 2010.
- [66] Yanfeng Sun-Hongjiang Zhang Mary Cz-erwinski Brent Field Liu Wenyin, Susan Dumais and One Microsoft Way. Semi-automatic image annotation. strategy. *Microsoft Research China*, 1999.
- [67] A. Moffat and J. Zobel. Self-indexing inverted files for fast text retrieval. *ACM Trans. Inf. Syst.*, 14(4) :349–379, 1996.
- [68] R. Sacks Davis A. J. Kent and K. Ramamohanarao. A signature file scheme based on multiple organizations for indexing very large text databases. *JASIS*, 41(7) :508–534, 1990.
- [69] R. Manmatha Lavrenko V. and J. Jeon. A model for learning the semantics of pictures. *In : Advances in Neural Information Processing Systems*, 16, 2004.
- [70] J. F. G. de Freitas Duygulu P., K. Barnard and D. A. Forsyth. Object recognition as machine translation : Learning a lexicon for a fixed image vocabulary. *In : ECCV-European conference on Computer Vision*, pages 97–112, 2002.
- [71] P. J. Moreno Carneiro G., A. B. Chan and N. Vasconcelos. Supervised learning of semantic classes for image annotation and retrieval. *IEEE TPAMI*, 29(3), 2007.
- [72] V. Lavrenko Jeon, J. and R. Manmatha. Automatic image annotation and retrieval using cross – media relevance models. *In : Proc. ACM SIGIR Conf. Research and Development in Information Retrieval*, page 119–126, 2003.
- [73] Chai J.Y. Jin R. and L. Si. Effective automatic image annotation via a coherent language model and active learning. *In : Proc. ACM Multimedia Conference*, page 892–899, 2004.
- [74] Gao Y. and J. Fan. Incorporating concept ontology to enable probabilistic concept reasoning for multi-level image annotation. *In : Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, page 79–88, 2006.
- [75] Vladimir Pavlovi Ameesh Makadi and Sanjiv Kuma. Baselines for image annotation. *International Journal on Computer Vision (IJCV)*, 2010.

- [76] Christophe Millet. Annotation automatique d'images : annotation cohérente et création automatique d'une base d'apprentissage. *Thèse*, 2008.
- [77] Yanfeng Sun HongJiang Zhang Mary Czerwinski Liu Wenyin, Susan Dumais and Brent Field. Semi-automatic image annotation. *In Proc. of Interact : Conference on Human-Computer Interaction*, 2001.
- [78] R. Wong and C. Leung. Automatic semantic annotation of real-world web images. *IEEE Trans. Pattern Anal. Mach. Intell*, 30(11) :1933–1944, 2008.
- [79] Y. Gao J. P. Fan and H. Z. Luo. Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation. *IEEE Trans. Image Processing*, 17(3) :407–426, 2008.
- [80] X. Xue Y. Gao, J. Fan and R. Jain. Automatic image annotation by incorporating feature hierarchy and boosting to scale up svm classifier. *in ACM MULTIMEDIA 06*, pages 901–910, 2006.
- [81] Sabine Barat. Modèles graphiques probabilistes pour la reconnaissance de formes. *Thèse*, 2009.
- [82] Hakim Hacid. Annotation semi-automatique de grandes bd images : Approche par graphes de voisinage. *CORIA*, pages 205–211, 2006.
- [83] Z. Lin G. Liu and Y. Yu. Radon representation-based feature descriptor for texture classification. *IEEE Trans. Image Processing*, 18(3) :921–928, 2009.
- [84] Urban J. and Jose J.M. Adaptive image retrieval using a graph model for semantic feature integration. *ACM Press, New York*, pages 117–126, 2006.
- [85] Faloutsos C. Pan J.Y., Yang H.J. and Duygulu P. Gcap : Graph-based automatic image captioning. *IEEE Computer Society Press, Los Alamitos*, 9 :146, 2004.
- [86] Nhu-Van Nguyen. Représentations visuelles de concepts textuels pour la recherche et l'annotation interactives d'images. *Thèse*, 2011.
- [87] Jana Urban Vassilios Stathopoulos and Joemon Jose. Semantic relationships in multi-modal graphs for automatic image annotation. *ECIR'08 Proceedings of the IR research, 30th European conference on Advances in information retrieval*, pages 490–497, 2008.
- [88] Pham Minh Hai. Apprentissage automatique. travail d'intérêt personnel encadré. *Institut de la Francophonie pour l'Informatique*, 2004.
- [89] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2) :121–167, 1998.
- [90] L. Breiman. Random forests. *in Machine Learning*, pages 5–32, 2001.
- [91] Sabine Barrat and Salvatore Tabbone. Classification and automatic annotation extension of images using a bayesian network. *Traitement du signal*, 26, 2009.
- [92] C.P. Town and D. Sinclair. Content-based image retrieval using semantic visual categories. *Society for Manufacturing Engineers, Technical Report MV01-211*, 2001.

- [93] Alain Boucher and Thi-Lan Le. Comment extraire la sémantique d'une image. *International Conference on Sciences of Electronic, Technologies of Information and Telecommunications - SETIT*.
- [94] Bruno Bouzy. Apprentissage par renforcement. *Cours de d'apprentissage automatique*, September 2005.
- [95] R. Manmatha S. Feng and V. Lavrenko. Multiple bernoulli relevance models for image and video annotation. *in Conference on Computer Vision and Pattern Recognition (CVPR) '04*, 2 :1002–1009, 2004.
- [96] David M. Blei and Michael I. Jordan. Modeling annotated data. *In SIGIR'03*, pages 127–134, 2003.
- [97] M. Li W.-Y. Ma R. Zhang, Z. M. Zhang and H.-J. Zhang. A probabilistic semantic model for image annotation and multi-modal image retrieval. *in ICCV '05*, 1 :846–851, 2005.
- [98] Sabine Barrat and Salvatore Tabbone. Classification et extension automatique d'annotations d'images en utilisant un réseau bayésien. *Colloque International Francophone sur l'Écrit et le Document - CIFED*, 2008.
- [99] Z. Li W. Y. Ma J. R. Wen D. Cai, X. He. Hierarchical clustering of www image search results using visual, textual and link information. *Proceedings of the ACM International Conference on Multimedia*, 2004.
- [100] T. S. Chua H. Feng, R. Shi. A bootstrapping framework for annotating and retrieving www images. *Proceedings of the ACM International Conference on Multimedia*, 2004.
- [101] Z. Li W. Y. Ma J.-R. Wen D. Cai, X. He. Hierarchical clustering of www image search results using visual, textual and link information. *Proceedings of the ACM International Conference on Multimedia*, 2004.
- [102] W. Y. Ma J.-R. Wen-H. Zhang D. Cai, X. He. Organizing www images based on the analysis of page layout and web link structure. *Proceedings of the International Conference on Multimedia and Expo(ICME)*, 2004.
- [103] Jan Wielemaker Laura Hollink, Guus Schreiber and Bob Wielinga. Semantic annotation of image collections. *In Workshop on Knowledge Markup and Semantic Annotation*, 2003.
- [104] Johanna Vompras and Stefan Conrad. A semi-automated framework for supporting semantic image annotation. *International Conference on Management of Data*, pages 944–194, 1998.
- [105] Oria V. Lin S., Tamer Ozsu M. and Ng R. An extensible hash for multi-precision similarity querying of image databases. *Proc. of the 27th Int. Conf. on Very Large DataBase, Roma (Italy)*, pages 221–230, 2003.
- [106] Roberto Di Salvo Daniela Giordano Isaak Kavasidis, Simone Palazzo and Concetto Spampinato. An innovative web-based collaborative platform for video annotation. *Multimedia Tools and Applications*, 2013.

- [107] Michael Lottko Dominik Renzel, Yiwei Cao and Ralf Klamma. Collaborative video annotation for multimedia sharing between experts and amateurs. *Workshop on Interoperable Social Multimedia Applications*, 2010.
- [108] Yi Hong and Stephan Reiff-Marganiec. Towards a collaborative framework for image annotation and search. *CAiSE Workshops-Lecture Notes in Business Information Processing*, 83 :564–574, 2011.
- [109] Xiao Huang David McGee Paulo Barthelmeß, Edward Kaiser and Philip Cohen. Collaborative multimodal photo annotation over digital paper. *Proceeding ICMI '06 Proceedings of the 8th international conference on Multimodal interfaces*, pages 4–11, 2006.
- [110] M. Marszalek and C. Schmid. Semantic hierarchies for visual object recognition. *In Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [111] R. Socher L.-J. Li K. Li J. Deng, W. Dong and L. Fei-Fei. Imagenet : A large-scale hierarchical image database. *In Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.
- [112] J. Tesic-S.-F. Chang W. Hsu L. Kennedy A. Hauptmann M. Naphade, J. R. Smith and J. Curtis. Large-scale concept ontology for multimedia. *In IEEE MultiMedia*, July 2006.
- [113] A. Zisserman-W. T. Freeman J. Sivic, B. C. Russell and A. A. Efros. Unsupervised discovery of visual object class hierarchies. *In Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [114] Y. Lim-D. Blei L.-J. Li, C. Wang and L. Fei-Fei. Building and using a semantic visual image hierarchy. *In Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [115] Chung-Cheng Li Shu Lei Chen Von-Wun Soo, Chen-Yu Lee and Ching chih Chen. Automatic metadata creation : Automated semantic annotation and retrieval based on shareable ontology and case-based learning techniques. *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries*, May 2003.
- [116] N. Magesh and P. Thangaraj. Semantic image retrieval based on ontology and sparql query. *International Conference on Advanced Communication Technology (ICACT)*, 2011.
- [117] Mezzanotte R.J. Biederman I. and Rabinowitz J.C. Scene perception : Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, 14(2) :143–177, June 1982.
- [118] Y. Avrithis G. Stamou N. Simou, V. Tzouvaras and S. Kollias. A visual descriptor ontology for multimedia reasoning. *In Workshop on Image Analysis for Multimedia Interactive Services – WIAMIS*, 153 :255–262, April 2005.
- [119] I. Kompatsiaris S. Dasiopoulou, V. Tzouvaras and M. G. Strintzis. Enquiring mpeg-7 based multimedia ontologies. *Multimedia Tools Appl.*, 46, January 2010.
- [120] J. Hunter. Adding multimedia to the semantic web - building an mpeg-7 ontology. *In Semantic Web Working Symposium-SWWS*, pages 261–281, 2001.

- [121] Beatriz Lopez Miquel Montaner and Josep Lluís de la Rosa. Evaluation of recommender systems through simulated users. *AAMAS Workshop on Coordinating Agents' Plans and Schedules (CAPS)*, pages 1–8, May 2007.
- [122] Suzanne Little Roberto Garcia Raphael Troncy, Oscar Celma and Chrisa Tsinaraki. Mpeg-7 based multimedia ontologies : Interoperability support or interoperability issue ? *International Semantic Web Conference, ISWC 2007*, 46, November 2007.
- [123] Ghislain Auguste Ateazing Mari Carmen Suarez-Figueroa and Oscar Corcho. The landscape of multimedia ontologies in the last decade. *Multimedia Tools Applications*, 62, December 2013.
- [124] Thanos Athanasiadis Yannis Avrithis Kosmas Petridis, Frederic Precioso and Yiannis Kompatsiaris. Combined domain specific and multimedia ontologies for image understanding. *Mixed-reality as a challenge to image understanding and artificial intelligence*, September 2005.
- [125] Monique Thonnat Nicolas Maillot and Sophia Antipolis. Ontology based object learning and recognition : Application to. image retrieval. *In Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*, page 620–625, November 2004.
- [126] Handschuh S. Vargas-Vera M. Motta E. Uren V., Cimiano P. and Ciravegna F. Semantic annotation for knowledge management : requirements and a survey of the state of the art. *In Journal of Web Semantics, Science, Services and Agents on the World Wide Web*, 4(1).
- [127] S. Melzer-R. MOLLER S. E. Peraldi, A. Kaya and M.Wessel. Multimedia interpretation as abduction. *In Digital Libraries Journal*, 2007.
- [128] B. Neumann and R. MOLLER. On scene interpretation with description logics. *In Image Vision Computing*, 26(1) :82–101, 2008.
- [129] Avril Styrman. Ontology-based image annotation and retrieval. *Master of Science Thesis*, May 2008.
- [130] Chia L. T. Wang H. and Liu S. Semantic retrieval with enhanced matchmaking and multimodality ontology. *IEEE International Conference on Multimedia and Expo.*, 26(1) :516–519, 2007.
- [131] Bowden M. Srikanth M., Varner J. and Moldovan D. Exploiting ontologies for automatic image annotation. *In : Proceedings of the 28th International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 26(1) :552–558, 2005.
- [132] Jesús A. González Aurelio López Manuel Montes Eduardo Morales Enrique Sucar H. Jair Escalante, Carlos Hernández-Gracidas and Luis VILLASENOR. Segmenting and annotating the iapr-tc12 benchmark. *Reporte Técnico No. CCC-08-005*, November 2008.
- [133] Ning HUANG Ning Ruan and Wen HONG. Semantic-based image retrieval in remote sensing archive : An ontology approach. *IEEE INTERNATIONAL CONFERENCE*, 4, 2006.

- [134] Strintzis M. Mezaris V., Kompatsiaris I. An ontology approach to object-based image retrieval. *Proceedings of the IEEE International Conference on Image Processing, ICIP03*, 2 :511–514, 2003.
- [135] G. Lu Y. Liu, D. Zhang and W.-Y. Ma. A survey of contentbased image retrieval with high-level semantics. *Journal of Pattern Recognition*, 41(1) :262–282, 2007.
- [136] W. I. Grosky and R. Zhao. Negotiating the semantic gap : From feature maps to semantic landscapes. in *SOFSEM '01*, pages 33–52, 2001.
- [137] A. Benitez and C. Shih-Fu. Perceptual knowledge construction from annotated image collections. in *ICME '02*, 1 :189–192, 2002.
- [138] Vincent Oria Marta Rukoz. Geneviève Jomier, Maude Manouvrier. Indexation multi-niveau pour la recherche globale et partielle par le contenu. *20èmes Journées Bases de Données Avancées, BDA, Actes (Informal Proceedings)*, Montpellier, octobre 2004.
- [139] Nathalie Hernandez. Ontologies de domaine pour la modélisation du contexte en recherche d'information. *thèse*, 2005.
- [140] L. Lin M. W. Lee B. Yao, X. Yang and S. C. Zhu. Image parsing to text description. *In Proc. of IEEE*, 2009.
- [141] Benoit Favre. Indexation multimedia caractérisation du déséquilibre entre les modalités texte et parole. *Mémoire - DEA Informatique*, Juin 2003.
- [142] Sabrina Tollari and Hervé Glotin. Recherche visuo-textuelle d'images sur le web améliorée par sélection de la dimension. *CORIA*, pages 7–22, 2008.
- [143] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Sciences*, 27(3) :129–146, 1976.
- [144] W Woods. Conceptual indexing : A better way to organize knowledge. *Technical report SMLI TR-97-61, Sun Microsystems Laboratories, Mountain view, CA*, 1997.
- [145] R. McCool R.V. Guha and E. Miller. Semantic search. *In Proceedings of the 12th International World Wide Web Conference*, pages 700–709, 2003.
- [146] Khan L. and F. Luo. Ontology construction for information selection. *In Proc. of 14th IEEE International Conference on Tools with Artificial Intelligence*, pages 122–127, November 2002.
- [147] R. Mihalcea and D.I. Moldovan. Semantic indexing using wordnet senses. *In Proceedings of ACL Workshop on IR and NLP*, 2000.
- [148] F. Neri R. Cucchiarelli, R. Navigli and P. Velardi. Extending and enriching wordnet with ontolearn. *In Proceedings of the 2nd Global WordNet Conference*, 2004.
- [149] N. Aussenac-Gilles M. Baziz, M. Boughanem and C. Chrisment. Semantic cores for representing documents in ir. *In Proceedings of the 20th ACM Symposium on Applied*, 57 :1011–1017, 2005.
- [150] M.A. Hearst and C. Karadi. Cat-a-cone : an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. *In Proceedings of the 20th*

- International conference on Research and Development in Information Retrieval, SIGIR*, pages 246–257, 1997.
- [151] Pierre Alain MOELLIC Adrian Popescu and Christophe Millet. Ontology driven content based image retrieval. *Conference on Image and Video Retrieval - CIVR*, pages 113–116, 2007.
- [152] R. Burke. Hybrid recommender systems : Survey and experiments. *User Model. User Adapt. Inter.*, 12 :331–370, 2002.
- [153] G. Adomavicius and A. Tuzhilin. Towards the next generation of recommender systems : A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Engin.*, 17(6) :734–749, 2005.
- [154] J. A. Konstan B. N. Miller and J. Riedl. Pocketlens : Toward a personal recommender system. *ACM Trans. Inf. Syst.*, 22(3) :437–476, 2005.
- [155] D.-R. Liu and Y.-Y. Shih. Integrating ahp and data mining for product recommendation based on customer lifetime value. *Inf. Manag.*, 42 :387–400, 2005.
- [156] J. L. De La Rosa M. Montaner, B. López. A taxonomy of recommender agents on the internet. *AI Review, Kluwer Publishers*, 19, 2003.
- [157] M. Deshpande and G. Karypis. *Acm trans. Inf. Manag. Inf. Syst.*, 22(1) :143–177, 2004.
- [158] H. Nguyen and P. Haddawy. Diva : Applying decision theory to collaborative filtering. *AAAI Worksh. Recomm. Syst., Madison, WI*, July 1998.
- [159] H. Nguyen and P. Haddawy. The decision-theoretic video advisor. *15th Conf. Uncert. Artif. Intell., Stockholm, Sweden*, pages 494–501, 1999.
- [160] J. G. Jr. Lynch D. Ariely and M. Aparicio. Learning by collaborative and individual-based recommendation agents. *J. Consum. Psych*, 14(1/2) :81–94, 2004.
- [161] J. Youll G. Tewari and P. Maes. Personalized location-based brokering using an agent-based intermediary architecture. *Dec. Supp. Syst.*, 34 :127–137, 2002.
- [162] S. H. Choi and Y. H. Cho. An utility range-based similar product recommendation algorithm for collaborative companies. *Exp. Syst. Appl.*, 27 :549–557, 2004.
- [163] R. Baeza-Yates and B. Ribeiro-Neto. Modern information retrieval. *Addison-Wesley*, 1999.
- [164] Romain Picot-Clément. Une architecture générique de systèmes de recommandation de combinaison d'items. application au domaine du tourisme. *Mémoire de thèse*, 2011.
- [165] Adonis Dimas. Dogi : An automatic image annotation tool for images of dog breeds. *Dissertation thesis*, 2011.
- [166] Amardeilh F. Web sémantique et informatique linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle. *Thèse de doctorat*, Mai 2007.
- [167] Alexandre Piquet. Guide pratique du travail collaboratif : Théories, méthodes et outils au service de la collaboration. *Document destiné au « Groupe Communication » du réseau Isolement Social*, 17(2) :7–9, Août 2009.

- [168] Rana EL CHARIF. Analyse des paramètres de pondération dans le cadre de collections volumineuses. *Rapport de DEA Informatique*, page 23, 2006.
- [169] L. T. Chia H. Wang and S. Liu. Image retrieval web image retrieval with an enhanced multi-modality ontology. *Multimedia Tools Appl.*, pages 189–215, Mars 2008.
- [170] S. Liu. H. Wang and L. T. Chia. Does ontology help in image retrieval? *MULTIMEDIA '06 : Proceedings of the 14th annual ACM international conference on Multimedia, New York, NY, USA*, 39(2) :109–112, Mars 2006.
- [171] Hirst G. Budanitsky A. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32 :13–47, 2006.
- [172] Fellbaum C. Wordnet : An electronic lexical database. *Cambridge, MA, MIT Press.*, 1998.
- [173] C. Wemmert. Classification hybride distribué par collaboration des méthodes non supervisées. *Thèse de doctorat, Université de Strasbourg*, 2000.
- [174] Germain Forestier. Connaissance et clustering d'objets complexes multisources. *Thèse de doctorat, Université de Strasbourg, ISBN 0201788993*, 2010.
- [175] Pablo Castells Iván Cantador, Alejandro Bellogín. Ontology-based personalised and context-aware recommendations of news items. *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE Computer Society Washington, DC, USA*, pages 562–565, 2008.
- [176] E. A. Fox S. Perugini, M. A. Gonçalves. A connection-centric survey of recommender systems research. *Journal of Intelligent Information Systems*, 23 (1), 2003.
- [177] C. Berrut A-T. Nguyen, N. Denos. Cartes de communautés pour l'adaptation interactive de profils dans un système de filtrage d'information. *Actes de la 32ème Congrès INFORSID, France*, 2005.
- [178] R. Wille. Line diagrams of hierarchical concept systems. *Int. Classif.* 11, pages 77–86, 1984.
- [179] K. E. Wolff. A first course in formal concept analysis - how to understand line diagrams. *In F. Faulbaum (Ed.), SoftStatt'93, Advances in Statistical Software 4*, pages 429–438, 1993.
- [180] M. et B. Monjardet Barbut. A first course in formal concept analysis - how to understand line diagrams. *Ordre et classification, Algèbre et combinatoire, Tome 2. Hachette*, 1970.
- [181] G. Birkhoff. Lattice theory. *New York : American Mathematical Society*, 25, 1940.
- [182] M. Crampes J. Villerd, S. Ranwez. Navigation sur des cartes de connaissances supportées par un treillis de galois. *Colloque Carto 2.0, Noisy-le-Grand, France, April 3*, 2008.
- [183] IEEE James Z. Wang Member IEEE Yixin Chen, Member and Robert Krovetz. Clue : Cluster-based retrieval of images by unsupervised learning. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 14, No. 8, AUGUST 2005.

Résumé :

L'extraction de la sémantique d'une image est un processus qui nécessite une analyse profonde du contenu de l'image. Elle se réfère à leur interprétation à partir d'un point de vue humain. Dans ce dernier cas, la sémantique d'une image pourrait être générique (par exemple un véhicule) ou spécifique (par exemple une bicyclette). Elle consiste à extraire une sémantique simple ou multiple de l'image afin de faciliter sa récupération. Ces objectifs indiquent clairement que l'extraction de la sémantique n'est pas un nouveau domaine de recherche.

Cette thèse traite de l'annotation sémantique collaborative des images et de leur récupération. Il aborde d'une part, la façon dont les annotateurs pourraient décrire et représenter le contenu des images en se basant sur les informations visuelles, et d'autre part comment la récupération des images pourrait être considérablement améliorée grâce aux récentes technologies, notamment les systèmes de recommandation. Pour atteindre ces objectifs, l'exploitation des outils de description implicite du contenu des images, des interactions des annotateurs qui décrivent la sémantique des images et celles des utilisateurs qui utilisent la sémantique produite pour récupérer les images seraient indispensables.

Dans cette thèse, nous nous sommes penchés vers les outils du Web Sémantique, notamment les ontologies pour décrire les images de façon structurée. L'ontologie permet de représenter les objets présents dans une image ainsi que les relations entre ces objets (les scènes d'image). Autrement dit, elle permet de représenter de façon formelle les différents types d'objets et leurs relations. Les ontologies codent la structure relationnelle des concepts que l'on peut utiliser pour décrire et raisonner. Cela les rend éminemment adaptées à de nombreux problèmes comme la description sémantique des images qui nécessite une connaissance préalable et une capacité descriptive et normative.

La contribution de cette thèse est focalisée sur trois points essentiels: La Représentation Sémantique d'Images, l'Annotation Sémantique Collaborative d'Images et la Recherche Sémantique d'Images.

- La Représentation Sémantique permet de proposer un outils capable de représenter la sémantique des images. Pour capturer la sémantique des images, nous avons proposer une ontologie d'application dérivée d'une ontologie générique.
- L'Annotation Sémantique Collaborative d'Images que nous proposons consiste à faire émerger la sémantique des images à partir des sémantiques proposées par une communauté d'annotateurs.
- La Recherche Sémantique d'Images permet de rechercher les images avec les sémantiques fournies par l'annotation sémantique collaborative. Elle est basée sur deux techniques: le clustering et la recommandation. Le clustering permet de regrouper les images similaires à la requête d'utilisateur et la recommandation a pour objectif de proposer des sémantiques aux utilisateurs en se basant sur leurs profils statiques et dynamiques. Elle est composée de trois étapes à savoir: la formation de la communauté des utilisateurs, l'acquisition des profils d'utilisateurs et la classification des profils d'utilisateurs avec l'algèbre de Galois.

Des expérimentations ont été menées pour valider les différentes approches proposées dans ce travail.

Mots-clés : Annotation, collaboration, Images, Recommandations, Annotateurs, Sémantique, Émergence, Ontologie

The logo for the SPIM (École doctorale SPIM) features a stylized 'S' followed by the letters 'PIM' in a clean, sans-serif font. The 'S' is white and partially overlaps an orange horizontal bar on the left.