

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Mohamed Khider Biskra
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre :.....

Série :.....



Thèse

Présentée en vue de l'obtention du diplôme de docteur en sciences

***Sélection des concepts et calcul de proximité
sémantique pour réduire le silence dans la recherche
d'images par le texte : vers des moteurs qui se
configurent automatiquement***

Option : Informatique

Présentée par :

Debbagh Farah

Soutenue le : 28/10/2017

Devant le jury :

KamalEddine MELKEMI	Professeur	Université de Biskra	Président
Med Chaouki BABAHENINI	Professeur	Université de Biskra	Rapporteur
Med Lamine KHERFI	MCA	Université de Ouargla	Co-rapporteur
Med Chaouki BATOUCHE	Professeur	Université de Constantine 2	Examineur
Abdelmalik TALEB-AHMED	Professeur	Université de Valenciennes	Examineur
Abdelhamid DJEFFAL	MCA	Université de Biskra	Examineur

Remerciements

Tout d'abord, je remercie le puissant DIEU, qui m'a éclairé le chemin et m'a aidé de mener à bien ce travail.

Je tiens à exprimer ma gratitude et ma reconnaissance à mon rapporteur Prof. Med Chaouki BABAHENINI et à mon co-rapporteur Dr. Med Lamine KHERFI pour leurs orientations, leurs remarques pertinentes, leurs discussions, leurs lectures attentives, leur soutien moral encouragement tout en long de ce travail. Je les adresse mes remerciements les plus distingués pour leur gentillesse et leur simplicité. Ils me donnent toujours le bon exemple.

Mes remerciements vont également aux honorables membres du jury qui ont accepté d'examiner et d'évaluer cette étude, à savoir: Prof. KamalEddine MELKEMI de l'Université de Biskra, Prof. Med Chaouki BATOUCHE de Université de Constantine, Prof. Abdelmalik TALEB-AHMED de l'Université de Valenciennes et Dr. Abdelhamid DJEFFAL de l'Université de Biskra.

Mes remerciements vont également aux professeurs Ms et Mme Cheriet de l'École de technologie supérieure (ÉTS) et au polytechnique à Montréal pour leur discussion et conseils importantes sur mon travail. Je remercie également le professeur Ismail Biskri de l'université de trois rivières, le professeur Jian-Yun Nie et le professeur Guy Lapalme de l'université de Montréal pour ma réception dans leurs bureaux et discussion autour du mon travail. Je remercie le Prof. Kobus Barnard de L'université d'Arizona (UA) pour les liens de téléchargement de la base d'images annotées Korel 5K.

J'aimerais remercier tout particulièrement ma collègue et sœur Mme Bouanenene qu'elle était avec moi tout en long de ces années de travail, avec ses discussions, ses conseils et ses propositions, ainsi que son soutien morale. Je la remercie également pour sa lecture attentive de ce manuscrit.

Je tiens à exprimer ma gratitude à mes collègues et frères Bilel et Oussama pour leur aide profond et leur révision attentive de mon papier. Je les remercie profondément pour être prêt de moi dans les moments difficiles. Je remercie ma collègue et sœur Olaya pour la mise en forme de ce manuscrit et pour son encouragement. Je remercie également mon amie Leila pour son encouragement.

Mes remerciements vont également aux professeurs de la langue française, Ms Bourenene, Ms Zouzou, Ms Mouad, Ms Benslimane med seghir et Mme Benzid Najia pour leurs révisions attentives de ce manuscrit. Finalement, mes remerciements à tous ceux qui m'ont encouragé.

Dédicaces

Je dédie ce travail

A mes chers Parents,

A ma sœur, mes frères, mes nièces et mes neveux...

A mon mari et mes enfants Sara et Hacem, qui m'ont aidé beaucoup par leur patience et leurs prières...

A mon beau père et sa famille

Mes dédicaces vont tendrement à mes chères éducatrices de l'école coranique et à mes chères amies Dr Benaouda, Mme Bouanane, Leïla, Olaya, Hanna, Nassima, Sounia et sabah.

A l'esprit des mes chères amies Mounira et Sana.. !

A mes collègues de l'équipe imagerie, Meriem, Bilel, Oussama, Abdelmadjid, ramla, zineb et lamis.

A tous mes collègues du département d'informatique.

A tous qui m'aiment et je les aime

Résumé

Dans la recherche d'images par le texte, la comparaison binaire est une technique qui retrouve pour une requête basée-concept Q , les images annotées avec Q . Par conséquent, la performance du système de recherche est très influencée par la qualité d'annotation. Comme il est extrêmement difficile d'avoir une base d'images bien annotée, la recherche ignore de nombreuses images pertinentes parce que tout simplement elles ne sont pas annotées avec les concepts de la requête. Ce problème est communément appelé le silence. Il dégrade considérablement les performances de TBIR.

L'objectif principal de cette thèse est de minimiser un tel problème, et par conséquent obtenir une performance supérieure dans les systèmes TBIR. Nous proposons un système TBIR qui intègre la proximité sémantique entre les concepts au sein de la recherche. Pour calculer la proximité sémantique entre les concepts, nous avons utilisé un ensemble d'articles Wikipedia comme source de connaissances externes, le schéma de pondération statistique TF_ICTF (Term Frequency_ Inverse Collection Term Frequency en anglais) et la mesure de similarité cosinus. Par la suite, les résultats de proximité sémantique sont intégrés au sein du mécanisme de recherche. Autrement dit, nous avons calculé la pertinence de chaque image en fonction des valeurs de proximité sémantique entre ses concepts d'annotation et la requête utilisateur.

La méthode proposée présente de nombreux avantages. Elle est entièrement automatique, car aucune intervention humaine n'est requise. En outre, elle peut être appliquée soit dans la phase de recherche pour détecter plus d'images pertinentes, ou bien dans la phase d'annotation pour détecter et compléter des concepts pertinents manquants. De plus, la méthode statistique que nous avons adoptée pour calculer la proximité sémantique entre les concepts se caractérise par sa simplicité et sa performance. Elle est plus simple par rapport aux méthodes graphiques probabilistes, et plus performante que d'autres méthodes statistiques.

Les résultats expérimentaux ont montré que la méthode proposée a atteint une grande précision dans la détection d'images pertinentes. En outre, la méthode proposée a montré sa performance par rapport à la comparaison binaire et d'autres travaux existants.

Mots clés : Recherche d'images par le texte, annotation, silence, TF_ICTF, similarité cosinus, proximité sémantique, articles Wikipedia, coefficient de corrélation de Pearson, WordSimilarity-353.

Abstract

In Text-Based Image Retrieval (TBIR), matching is a technique that retrieves for a concept-based query Q , images annotated with Q . As consequent, the performance of the retrieval engine is very influenced by the annotation quality. Since it is extremely difficult to have a well annotated data-set, the retrieval neglects many relevant images simply because they are not annotated with the query concepts. This problem is commonly known as the missing and it degrades considerably the performance of TBIR.

The main aim of this thesis is to minimize such a problem, and thus to guarantee a superior performance in TBIR systems. We propose a TBIR system that integrates the semantic relatedness between concepts within retrieval. To compute the semantic relatedness between concepts, we have used a set of Wikipedia articles as an external source of knowledge, the statistical weighting scheme TF_ICTF (Term Frequency_Inverse Collection Term Frequency) and the cosine similarity measurement. Afterwards, we have incorporated the semantic relatedness results into the retrieval mechanism. That is to say, we have computed the relevance of each image, based on the semantic relatedness values between its concept annotation and the user query.

The proposed method has many advantages. It is fully automatic, as no human intervention is required. In addition, it can be applied either in the retrieval stage to detect missing relevant images, or in the annotation stage to detect and complete missing relevant concepts. Furthermore, the statistical method used to compute the semantic relatedness between concepts is characterized by its simplicity and its performance. It is simpler compared to probabilistic graphical methods, and more performant than other statistical methods.

The experimental results show that the proposed method has achieved a high precision in detecting missing relevant images. In addition, the proposed method has proven its strength comparing with matching and other works.

Keywords: Text-based image retrieval (TBIR), annotation, missing, TF_ICTF, cosine similarity, semantic relatedness, Wikipedia articles, Pearson correlation, WordSimilarity-353.

ملخص

في أنظمة البحث عن الصور بواسطة النص (TBIR)، تعتبر المقارنة الثنائية طريقة تسمح بإيجاد من أجل طلب المستعمل المقدم على شكل مفردات، الصور المشروحة بتلك المفردات. نتيجة لذلك، يتأثر أداء محرك البحث كثيرا بجودة الشروحات المسندة للصور. بما أنه من الصعب للغاية الحصول على قاعدة صور مشروحة شرحا كاملا ودون أية أخطاء، فإن محرك البحث يمكنه أن يهمل العديد من الصور الموافقة للمطلوب لأنها وببساطة ليست مشروحة بالمفردات المقدمة من طرف المستعمل. تعرف هذه المشكلة عادة بمصطلح الصمت وهو يؤثر سلبيا وبشكل كبير على أداء هذه الأنظمة.

الهدف الرئيسي لهذه الأطروحة هو التقليل من هذه المشكلة، وبالتالي التحصل على أداء أحسن لأنظمة TBIR. لهذا نقتراح نظام TBIR والذي يأخذ بعين الاعتبار القرابة السيميائية الدلالية بين المفردات في عملية البحث.

من أجل حساب القرابة الدلالية بين المفردات، استعملنا مجموعة من مقالات ويكيبيديا كمصدر خارجي للمعرفة، كما استخدمنا مخطط التريج الإحصائي TF_ICTF ومقياس تشابه جيب التمام. بعد ذلك، تم إدراج نتائج القرابة الدلالية في آلية البحث. بعبارة أخرى، قمنا بحساب مدى موافقة كل صورة للمطلوب اعتمادا على قيم القرابة الدلالية بين مفردات شروحاتها ومفردات طلب المستعمل.

تقدم الطريقة المقترحة العديد من المزايا. فهي تعمل بشكل آلي تام بدون أي تدخل بشري. بالإضافة إلى ذلك، يمكن تطبيقها، إما في مرحلة البحث لرصد عدد أكبر من الصور التي تجيب على طلب المستعمل، أو في مرحلة الشرح التوضيحي للصور لإيجاد واستكمال المفردات الناقصة ذات الصلة بالصور. علاوة على ذلك، فإن الطريقة الإحصائية التي اعتمدها لحساب القرابة الدلالية بين المفردات تتميز ببساطتها وكفاءتها، فهي أبسط مقارنة بالأساليب الاحتمالية البيانية، وأكثر كفاءة من طرق إحصائية أخرى.

أظهرت النتائج التجريبية أن الطريقة المقترحة حققت درجة عالية من الدقة في البحث عن الصور الموافقة للمطلوب، كما أظهرت الطريقة تفوقها على المقارنة الثنائية وكذا على بعض الطرق الأخرى الموجودة.

الكلمات المفتاحية: البحث عن الصور بواسطة النص، الشروحات، الصمت، مخطط التريج الإحصائي، تشابه جيب التمام، القرابة السيميائية الدلالية، مقالات ويكيبيديا، معامل ترابط بيرسن، تشابه الكلمات-353.

Table des matières

Remerciements.....	ii
Dédicaces	iii
Résumé	iv
Table des matières	vii
Liste des figures	xi
Liste des tables	xii
Chapitre I. Introduction générale.....	13
I.1 Introduction	13
I.2 Problématique.....	15
I.3 Vue d'ensemble des travaux de la littérature	17
I.4 Motivations.....	19
I.5 Contributions	21
I.6 Structure de la thèse	22
Chapitre II. Recherche d'images par le texte (TBIR) : Définitions, Architecture, Composants, Applications et Problèmes majeurs	24
II.1 Introduction.....	24
II.2 Définition d'un système TBIR.....	25
II.3 Architecture d'un système TBIR	25
II.4 Les principales composantes d'un système de recherche d'images TBIR	28
II.4.1 Annotation des images	28
II.4.2 Formulation de requêtes	29
II.4.3 Calcul de pertinence (recherche).....	30
II.5 Exemples de systèmes TBIR	31
II.5.1 Google Images.....	31
II.5.2 Atlas WISE.....	31
II.5.3 PicHunter.....	32
II.5.4 Bing Images.....	32
II.5.5 ALIPR	32
II.6 Problèmes majeurs d'un système TBIR.....	33
II.6.1 Problème de Silence	34
II.6.2 Problème de Bruit.....	34

II.6.3 Défi de l'annotation des images	35
II.7 Conclusion	35
Chapitre III. Minimisation du silence dans un TBIR : état de l'art	36
III.1 Introduction	36
III.2 Familles des méthodes permettant de minimiser le silence.....	37
III.3 Méthodes visuelles	38
III.3.1 Méthodes basées-modèles.....	38
III.3.2 Méthodes basées-données	39
III.4 Limites des méthodes visuelles	41
III.5 Méthodes sémantiques.....	42
III.5.1 Méthodes sémantiques basées-corpus local.....	43
III.5.1.1 Un apprentissage semi-supervisé multi-labels basé graphe	43
III.5.1.2 Classification multi-labels en utilisant un réseau de dépendance conditionnelle	46
III.5.1.3 Accomplissement des tags pour la recherche d'images	48
III.5.1.4 Un modèle de langage de similarité sémantique pour améliorer l'annotation automatique d'images	51
III.5.1.5 Recherche d'images sociales basée-tag: vers des résultats pertinentes et diverses	53
III.5.1.6 Similarité sémantique des images basée-contexte	54
III.5.2 Méthodes sémantiques basées-corpus global.....	56
III.5.2.1 Recherche d'images sur le Web améliorée par une ontologie : aidée par Wikipedia et la théorie de l'activation de propagation	56
III.5.2.2 Thesaurus de concepts assisté-Wikipedia pour une meilleure compréhension du Media Web	59
III.5.2.3 Indexation basée-ontologies multiples des documents multimédias dans Internet	61
III.6 Limites des méthodes sémantiques.....	63
III.7 Etat de l'art des méthodes de calcul de proximité sémantique entre les concepts	65
III.7.1 Méthodes topologiques	65
III.7.1.1 Wikirelate! Calcul de proximité sémantique en utilisant Wikipedia	66
III.7.1.2 Mesurer la proximité sémantique des entités en utilisant Wikipedia.....	66
III.7.1.3 Une mesure efficace, peu coûteuse de proximité sémantique obtenue à partir des liens Wikipedia.....	68
III.7.1.4 Wikiwalk: Marche aléatoire dans Wikipedia pour la proximité sémantique	69
III.7.1.5 Utilisation de la sémantique Wikipedia pour calculer la proximité contextuelle	70

III.7.2 Méthodes statistiques	71
III.7.2.1 Calcul de proximité sémantique en utilisant une analyse sémantique explicite basée-Wikipedia	71
III.7.2.2 Proximité sémantique en utilisant l'analyse sémantique saillante	74
III.8 Conclusion	77
Chapitre IV. Méthode pour réduire le silence en recherche d'images basée proximité sémantique	79
IV.1 Introduction	79
IV.2 Analyse des relations sémantiques entre les concepts.....	79
IV.3 Similarité sémantique et Proximité sémantique	81
IV.4 Aperçu sur la méthode proposée	82
IV.5 Etapes de la solution pour réduire le silence dans un TBIR.....	85
IV.5.1 Calcul automatique de proximité sémantique entre les concepts	85
IV.5.1.1 Construction du corpus de connaissances	85
IV.5.1.2 Calcul des poids de pondération des concepts	87
IV.5.1.3 Calcul de la similarité cosinus entre les concepts	91
IV.5.2 Exploitation de la proximité sémantique pour réduire le silence en recherche d'images	93
IV.5.2.1 Traitement de requête atomique.....	93
IV.5.2.2 Traitement de requête visuelle	95
IV.5.3 Exploitation de la proximité sémantique pour réduire le silence dans l'annotation...	95
IV.6 Conclusion.....	96
Chapitre V. Résultats, Evaluation et Validation de la méthode basée proximité sémantique	97
V.1 Introduction.....	97
V.2 La configuration expérimentale	98
V.2.1 La base d'images COREL 5K.....	98
V.2.2 Le benchmark de proximité sémantique WordSimilarity-353.....	99
V.2.3 Mesures de performances.....	100
V.2.3.1 Métrique de performance de calcul de proximité sémantique entre les concepts	100
V.2.3.2 Métrique de performance de la recherche d'images	101
V.3 Résultats expérimentaux	101
V.3.1 Proximité sémantique entre les concepts	101
V.3.1.1 Calcul de proximité sémantique entre les concepts.....	102

V.3.1.2 Evaluation des résultats de proximité sémantique par rapport à WordSimilarity-353	103
V.3.1.3 Evaluation des résultats de proximité sémantique par rapport à quelques travaux connexes	104
V.3.2 Recherche et annotation d'images	105
V.3.2.1 Résultats de recherche	105
V.3.2.2 Résultats d'accomplissement d'annotation	107
V.3.2.3 Comparaison avec quelques méthodes de l'état de l'art	109
V.4 Conclusion	113
Conclusion générale et perspectives	115
Bibliographie.....	121

Liste des figures

Figure 1. Exemple d'une collection d'images avec annotations	17
Figure 2. Images retournées suite à la comparaison binaire.....	17
Figure 3. Architecture typique d'un système TBIR	27
Figure 4. Premières images récupérées par 'Google Images' pour la requête 'cascade'	30
Figure 5. Exemple d'un réseau de dépendance conditionnelle (Guo and Gu, 2011)	47
Figure 6. Accomplissement des tags pour la recherche d'images(Wu et al., 2013).....	49
Figure 7. Calcul de proximité sémantique en utilisant une analyse sémantique explicite basée-Wikipedia (Gabrilovich and Markovitch, 2007).....	72
Figure 8. L'organigramme de la méthode proposée.....	84
Figure 9. Un exemple d'un article Wikipedia.	86
Figure 10. Quelques images représentatives de COREL 5K.....	98
Figure 11. Exemples d'images retournées depuis la base d'images Corel 5K avec leurs annotation correspondantes, suite aux requêtes : (a) City, (b) Forest, (c) Sunset et (d) Desert.	107
Figure 12. Exemples d'enrichissement d'annotation pour des images de Corel 5K. Les nouvelles annotations sont en gras.	109
Figure 13. Une comparaison des résultats de recherche obtenus par les différentes méthodes, en utilisant les requêtes : 'city', 'forest', 'sunset' et 'desert' respectivement.....	112

Liste des tables

Table 1. Valeurs de proximité sémantiques, pour quelques paires de concepts dans le benchmark WordSimilarity-353.....	100
Table 2. Valeurs de proximité sémantique pour quelques paires de concepts du COREL 5K.	103
Table 3. La corrélation linéaire entre WordSimilarity-353 et nos valeurs de proximité sémantique, calculée par le coefficient ρ	103

Chapitre I. Introduction générale

I.1 Introduction

La popularité et la simplicité de manipulation des appareils numériques, la montée en puissance des capacités de stockage, la grande disponibilité sur le marché avec des prix bas ainsi que la démocratisation de l'Internet, ont conduit à une explosion extraordinaire du nombre d'images, que ce soit pour des collections personnelles, professionnelles ou dans le Web. Par conséquent, la recherche manuelle des images est devenue laborieuse et très insatisfaisante pour les utilisateurs du point de vue précision et temps de réponse et il est devenu donc indispensable de développer des systèmes automatisant cette tâche(Kherfi et al., 2004).

La littérature a connu une concurrence très active dans le domaine du développement des systèmes et des moteurs de recherches d'images. La majorité des travaux peuvent être classés en deux catégories : La recherche d'images par le contenu visuel, dite Content-Based Image Retrieval ou CBIR en anglais(Agarwal and Maheshwari, 2015; Ayech and Amiri, 2016; Bakar et al., 2013; Syam and Srinivasa Rao, 2012; Vipparthi and Nagar, 2015; Yang et al., 2010; Yue et al., 2011). Elle est appelée aussi la recherche de bas niveau. La deuxième catégorie est la recherche d'images par le texte, dite Text-Based Image Retrieval ou TBIR en anglais(Chen et al., 2010; Liu et al., 2011). Elle est appelée aussi la recherche de haut niveau.

Tout système de recherche d'images se résume en deux phases principales : une phase offline nommée généralement indexation, et une phase online de recherche qui commence dès que l'utilisateur introduit sa requête. La différence entre les systèmes de recherche d'images réside dans l'une de ces deux phases. C'est-à-dire soit dans la méthode utilisée pour indexer (représenter) les images d'une base, ou bien dans le mécanisme développé pour la recherche.

Dans un CBIR, la phase offline consiste à représenter chaque image de la base avec une ou plusieurs caractéristiques visuelles de bas niveau telle que la couleur, la texture, la forme et les points d'intérêts (Bakar et al., 2013; Liu and Yang, 2013; Neelima and Reddy, 2015; Yue et al., 2011). Ces caractéristiques sont organisées en vecteurs dits vecteurs de caractéristiques visuelles. Le résultat de cette phase est une base d'images accompagnées de leurs vecteurs visuels. La phase online se lance quand l'utilisateur introduit sa requête généralement sous forme d'une (ou de plusieurs) image exemple. Ainsi, s'il s'agit d'une requête en dehors de la base, le système calcule son vecteur visuel de la même façon que dans la phase offline. Ensuite, il identifie des images pertinentes en se basant sur une mesure de similarité visuelle entre le vecteur de caractéristiques de la requête et ceux des images de la base. Plusieurs distances ont été utilisées pour mesurer la similarité visuelle, telle que la distance Euclidienne et la distance de Minkowski (Kherfi et al., 2004). Enfin, le système retourne les images à l'utilisateur par ordre croissant de distance.

Bien que les considérables efforts pour le développement des CBIR aient mené à des résultats encourageants, notamment si l'utilisateur est intéressé par l'apparence visuelle de l'image seulement, la recherche par le contenu visuel reste confrontée à deux grands défis. Le premier est la représentation efficace des images. Autrement dit, trouver de bonnes caractéristiques visuelles qui représentent précisément des images parmi un énorme nombre de caractéristiques existant dans la littérature, surtout que l'efficacité d'une caractéristique peut différer d'un ensemble d'images à un autre. Le deuxième défi est le gap sémantique entre les caractéristiques visuelles de bas niveau de description des images, et la sémantique interprétant ces images et qui peut être de haut niveau de description des images (Kherfi et al., 2004). Le gap sémantique est introduit du fait que le CBIR ne s'intéresse qu'aux caractéristiques visuelles des images, et ignore complètement la sémantique qui leur sont associée (Gorisse, 2010).

Dans la recherche d'images par le texte (TBIR), la phase offline appelée annotation, consiste à assigner à chaque image de la base un ou plusieurs mots clés (ou concepts) décrivant son contenu. Ces concepts peuvent ainsi décrire le bas niveau d'abstraction d'une image (tel que sa couleur et sa forme), le niveau intermédiaire (tel que les objets contenus dans l'image) ou bien le haut niveau d'abstraction (tel que les scènes et les sensations). Le résultat de la phase d'annotation est une base d'images avec annotations. La phase online se lance dès que l'utilisateur introduit sa requête textuelle sous forme de mots clés ou bien du texte libre et consiste à localiser des images pertinentes. Ainsi, un système de recherche

classique identifie des images pertinentes en se basant sur une comparaison binaire, dite *matching* en anglais, entre la requête et les annotations des images de la base. Autrement dit, une image annotée explicitement par la requête est considérée comme pertinente, et une image qui ne contient pas cette requête parmi ses annotations est écartée. Enfin, le système retourne les images à l'utilisateur par ordre décroissant de pertinence.

Bien que l'avantage majeur d'un TBIR par rapport à un CBIR c'est qu'il tient en compte de la sémantique associée aux images grâce aux annotations qui peuvent exprimer non seulement le bas niveau d'abstraction (c'est le cas du CBIR), mais aussi le niveau intermédiaire (tel que les objets contenus dans l'image) et le haut niveau (tel que les scènes, les sensations)(Yang et al., 2010), il reste confronté à deux problèmes majeurs, le bruit et le silence. Le bruit signifie que le moteur retourne des images erronées, ce qui est dû à la comparaison binaire avec des annotations (mots clés) anonymes ou bien erronées. Le silence signifie que le système rate des images pertinentes, ce qui est dû à la comparaison binaire avec des annotations incomplètes.

I.2 Problématique

Du fait qu'en réalité, l'être humain préfère exprimer ses besoins oralement ou par écrit plutôt que d'utiliser une image, la recherche par le texte est une recherche plus appropriée et plus naturelle pour l'utilisateur que la recherche par une image exemple. De plus, grâce aux annotations (mots clés), un TBIR peut capturer la sémantique associée aux images dans les différents niveaux d'abstraction de l'utilisateur. Mais malheureusement, avec un TBIR classique, cette capture de la sémantique que ce soit pour une image elle-même (la sémantique interne de l'image) ou bien pour un ensemble d'images (la sémantique extraite lorsque nous regardons un ensemble d'images) reste insuffisante. Ceci s'explique par le fait qu'en réalité ces mots clés ne sont pas indépendants les uns des autres, mais il y a, plutôt, une certaine sémantique qui les relie. Autrement dit, il existe une certaine interdépendance sémantique qui relie les concepts du monde réel. Par conséquent, l'interdépendance entre les concepts doit être prise en considération par un TBIR afin qu'il puisse refléter cette sémantique d'une façon correcte, adéquate et suffisante.

En fait, le calcul de l'interdépendance sémantique entre les concepts dans un système TBIR, présente une très grande difficulté et un défi réel aux chercheurs, comparé à un CBIR où

ce problème n'existe même pas. Ceci est dû au fait qu'il est très difficile pour une machine de capturer la sémantique comme font les êtres humains. D'ailleurs, indépendamment des domaines d'application, ce problème a attiré l'attention d'une importante communauté des chercheurs et toute une panoplie des travaux ont focalisé sur la manière de mesurer la sémantique entre les concepts (Gabrilovich and Markovitch, 2007), (Hassan and Mihalcea, 2011; Jabeen et al., 2012), (Ni et al., 2016; Taieb et al., 2014; Pakhomov et al., 2010; Aouicha and Taieb, 2016).

Ainsi, et en raison de cette difficulté, un TBIR n'est pas encore parvenu à capturer la sémantique de façon complète et surtout suffisante, ce qui fait qu'un tel système de recherche d'images, reste insatisfaisant vis-à-vis des exigences des utilisateurs. En effet, une des limitations les plus courantes est le silence. Le silence veut dire qu'un TBIR peut écarter plusieurs images bien qu'elles soient pertinentes à la requête de l'utilisateur, parce que tout simplement elles ne sont pas annotées d'une façon explicite par cette requête (comme le cas d'un TBIR classique) ou bien parce qu'elles sont annotées avec des concepts que le système les considère non pertinents à cette requête. Ce problème dégrade la qualité du résultat d'une façon considérable. Les principales causes de ce problème est le manque dans les annotations, surtout que les annotations manuelles sont ambiguës, bruitées et incomplètes (Chen et al., 2010; Wu et al., 2013), conjointement avec un mécanisme de recherche qui est classique dans la plupart des cas (c.à.d. la comparaison binaire).

Pour bien clarifier le problème du silence, prenons un exemple d'un TBIR classique : supposons qu'une collection d'images est composée des images de la Figure. 1, où chaque image est annotée avec les concepts qui sont justes au-dessous, et qu'un utilisateur formule sa requête avec le concept « désert ». Un système de recherche d'images par le texte (TBIR) recherche toutes les images qui sont annotées explicitement avec le concept « désert » (comparaison binaire). Les images retournées sont présentées dans la figure. 2

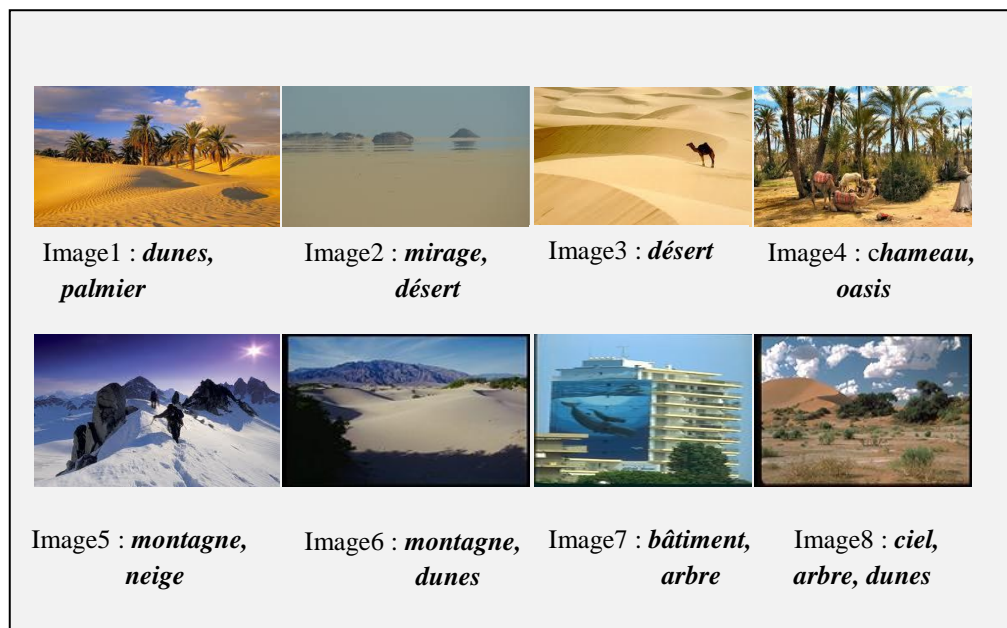


Figure 1. Exemple d'une collection d'images avec annotations



Figure 2. Images retournées suite à la comparaison binaire

D'après cette figure, il est clair que les images image1, image4, image6 et image8 sont omises et considérées comme non pertinentes, parce que tout simplement elles ne sont pas annotées par le concept « désert ». Or, d'après les concepts qui les annotent, elles représentent bel et bien le désert. Ceci est le problème de silence.

I.3 Vue d'ensemble des travaux de la littérature

Dans la littérature, plusieurs travaux ont été faits pour réduire le problème de silence, et donc augmenter de la précision des systèmes de recherche d'images. Parmi ces travaux, il ya ceux qui ont concentré sur la phase offline d'un TBIR. Autrement dit, la phase

d'annotation. Du fait que la comparaison binaire ne conduit pas au problème de silence quand la base d'images est très bien annotée. Autrement dit, le silence dans le résultat est dû, de leur point de vue, au manque existant dans les annotations. Par conséquent, leur objectif (Bao et al., 2011; Cui et al., 2015; Guo and Gu, 2011; Xu et al., 2016; Zha et al., 2009) était d'assigner aux images d'une base des concepts décrivant leur contenu sémantique, de telle sorte que ces annotations soient correctes et complètes. Ainsi, ils ont mis l'effort afin d'arriver à créer des annotations correctes et complètes pour des bases d'images non annotées, à compléter des annotations manquantes ainsi qu'à raffiner des annotations existantes mais bruitées par l'élimination de celles incorrectes dans des bases d'images légèrement annotées (Wu et al., 2013). Pour ce faire, ils ont tenté de déduire automatiquement la relation entre les caractéristiques visuelles représentant le bas niveau d'abstraction d'un ensemble d'images d'apprentissage annotées, et la sémantique (les concepts) associée à ces images et qui est de haut niveau. Par la suite, l'annotation des images non annotées se fait automatiquement en se basant sur la relation déduite.

Tandis que, d'autres travaux ont focalisé sur la phase online d'un TBIR (Fan and Li, 2006; Pesquita et al., 2009; Popescu et al., 2007; Rodrigues et al., 2014; Wang et al., 2008; Wang et al., 2010; Maree et al., 2016), c'est-à-dire la recherche d'images. En effet, leur but est de ne pas se limiter à la comparaison binaire entre des concepts, mais d'essayer de tirer profit de la sémantique associée aux images de telle sorte qu'un système de recherche soit capable de détecter des images pertinentes même si elles ne sont pas annotées explicitement par la requête de l'utilisateur. Pour atteindre cet objectif, ils ont commencé par modéliser cette sémantique sous forme de thésaurus ou ontologies composés de concepts et de relations sémantiques entre eux, puis ils l'ont exploité au moment de la recherche. Par exemple, en modélisant la taxinomie des concepts qui représente la relation *Est-un*, Popescu et al. (Popescu et al., 2007) ont développé un système qui retourne, en plus des images annotées explicitement par le concept requête, les images annotées par les concepts liés à cette requête par la relation *est-un*.

Il convient de noter que, dans cette section, nous donnons seulement un aperçu global des travaux concernés par le problème de silence en recherche d'images. Alors que nous consacrons un chapitre entier (chapitre 3) pour présenter les différentes méthodes proposées pour cette fin.

I.4 Motivations

Bien que les grands efforts déployés par les chercheurs pour résoudre le problème de silence ont abouti à la proposition de plusieurs méthodes performantes (Xu et al., 2016), ces dernières ont de sérieuses limitations qui devraient être prises en considération. Dans ce qui suit, nous en présentons les plus importantes ;

1. La majorité des méthodes d'annotation automatique nécessitent un ensemble important d'images annotées pour la phase d'apprentissage. La condition principale pour réussir l'apprentissage est que cet ensemble doit être sérieusement annoté (annotations précises et complètes). Ce qui signifie, pas de manque, pas de bruit et pas d'ambiguïté dans les annotations. Toutefois, il est très difficile de satisfaire cette condition idéale pour la plupart des images du monde réel (Wu et al., 2013) où se sont des utilisateurs qui fournissent manuellement ces annotations. Ce qui veut dire que les annotations peuvent être, contrairement à la condition décrite plus haut, incohérentes, générales, ambiguës, bruitées, incomplètes et parfois inappropriées (Chen et al., 2010; Wu et al., 2013). De plus, l'annotation manuelle est laborieuse, requiert beaucoup de temps et subjective, du fait que chaque annotateur peut interpréter une image selon son point de vue et ses intérêts.

2. La plupart des méthodes d'annotation considèrent uniquement la relation caractéristique visuelle-concept, et elles négligent totalement la relation concept-concept. En réalité, les concepts ne sont pas indépendants les uns des autres mais il existe une certaine sémantique qui les relie. Donc, ignorer une telle relation, influence certainement le processus d'annotation.

3. Il est très difficile de faire l'apprentissage des concepts de haut niveau d'abstraction, qui ont des apparences visuelles variées. Par exemple, des images annotées par le concept 'joie' peut avoir des vecteurs visuels différents.

4. La majorité des travaux ont utilisé des corpus locaux (des bases d'images locales qui sont annotées) pour déduire la relation caractéristique visuelle-concepts, ou bien la relation concept-concept (Franzoni et al., 2015; Guo and Gu, 2011; Xu et al., 2016; Yang et al., 2011; Zha et al., 2009). En effet, se limiter à l'information de cooccurrence est inappropriée, car la

fréquence d'apparition des concepts dans une base d'images est souvent très déséquilibrée(Xu et al., 2016). Ainsi, l'interdépendance extraite peut ne pas être généralisée pour des cas réels.

Tous les problèmes mentionnés ci-dessus peuvent influencer considérablement le processus d'apprentissage, et par conséquent dégrader la qualité des résultats d'annotation.

5. La majorité des travaux qui ont focalisé sur la phase de recherche afin de réduire le silence dans le résultat retourné, n'ont modélisé que quelques relations sémantiques(Ambika and Samath, 2012; Manzoor et al., 2012).Par exemple les auteurs de(Pesquita et al., 2009) se sont limités à la relation hiérarchique *est-un*, le travail(Fan and Li, 2006) a modélisé les relations *sous-classe-de*, *super-classe-de* et *classe équivalente*, les travaux présentés dans(Wang et al., 2008) et (Wang et al., 2010) ont présenté les relations de *Synonymie*, *polysémie*, *est-un*, *partie-de*, *concept-associé*. Par conséquent, les modèles obtenus ne reflètent pas la richesse en relations sémantiques qui existent en réalité. Ceci, peut conduire à une estimation biaisée de la mesure de similarité sémantique. Alors, il faut considérer beaucoup plus de relations pour refléter la sémantique qui se trouve dans la réalité.

6. Dans la majorité des méthodes sémantiques, le mécanisme de recherche infère généralement sur les relations directes et explicites entre une requête et les autres concepts. Cependant, il peut exister des relations implicites et indirectes entre les concepts. Par exemple, les images de '*forêt*' peuvent être retournées pour une requête '*arbre*' en se basant sur la relation sémantique explicite '*arbre est-partie-de forêt*', alors que les images de '*cascade*' sont complètement écartées malgré l'existence d'une relation implicite entre '*cascade*' et '*arbre*'.Par conséquent, l'ignorance de telles relations va nécessairement conduire à la mise en écart de plusieurs images pertinentes et entrainera donc un problème de silence

Compte tenu des problèmes suscités, nous sommes motivés à proposer une nouvelle solution pour réduire le problème de silence. La section suivante résume notre contribution et montre comment nous avons abordé le problème du silence.

I.5 Contributions

À cause de la comparaison binaire entre des annotations incomplètes et une requête utilisateur, un TBIR souffre du problème du silence. Dans cette thèse, nous proposons une solution automatique pour réduire ce problème. La solution proposée peut être classifiée dans les deux catégories de solutions présentées dans la section précédente. Autrement dit, elle peut être classifiée dans la famille des travaux sur l'annotation puisqu'elle est capable de détecter et compléter des concepts manquants dans des bases d'images légèrement annotées, comme elle peut être classée dans la famille des travaux focalisant sur la recherche, puisque le mécanisme de recherche proposé est capable de détecter des images pertinentes et qui ne sont pas annotées explicitement par la requête. L'idée principale dans notre méthode est d'intégrer des degrés de proximité sémantique entre les concepts, dite *semantic relatedness* en anglais, dans le mécanisme de recherche. Toutefois, il faut noter que les différents modèles de corrélation entre les concepts présentés dans les approches d'annotations sémantiques, nous ont inspirés pour modéliser ce degré de proximité sémantique entre les concepts.

Les contributions majeures de notre travail par rapport aux travaux de la littérature résident principalement dans les aspects suivants :

1. Contrairement aux méthodes de la littérature focalisant sur la recherche et qui considèrent des types limités de relations sémantiques et appliquent des mécanismes de recherches valables pour ces types de relations uniquement, la solution proposée n'est pas limitée à aucune relation prédéfinie, mais elle calcule plutôt le degré de proximité sémantique entre chaque paire de concepts. Ce degré de proximité reflète la sémantique de n'importe quelle relation qui peut exister entre les concepts.
2. A notre connaissance, tous les travaux qui ont proposé des méthodes de calcul de proximité sémantique entre les concepts pour réduire le silence dans un système de recherche d'images, se sont limités à l'information de cooccurrence entre les concepts dans des bases d'images locales. Cependant, la solution proposée calcule cette proximité depuis une source de connaissance externe indépendante de toute base d'images. Cette source externe est créée par une communauté humaine à travers le monde (ex. Wikipedia). Ce qui rend notre solution plus proche de la façon dont les êtres humains estiment la proximité sémantique entre les concepts du monde réel.
3. Différemment du mécanisme de recherche par la comparaison binaire entre une requête utilisateur et les annotations des images, la méthode de recherche proposée utilise les

valeurs de proximité sémantique entre la requête et les concepts d'annotation pour calculer la valeur de pertinence de chaque image dans un intervalle entre zéro et un (au lieu de zéro ou un comme dans la comparaison binaire).

Avec la solution proposée, un système de recherche d'images par le texte est capable de retourner des images pertinentes, malgré des manques dans ses annotations. Autrement dit, la performance de recherche n'est plus influencée s'il ya des manques dans les annotations.

La méthode proposée a prouvé sa performance par rapport à la méthode traditionnelle de recherche, et par rapport à différentes méthodes récentes de la littérature. De plus, le mécanisme de recherche proposé est capable d'une part, de détecter des annotations manquantes, et d'autre part, de retourner plus d'images pertinentes.

I.6 Structure de la thèse

En plus du premier chapitre qu'était consacré à une introduction générale, cette thèse est organisée comme suit :

- Au chapitre 2, nous introduisons le contexte général d'un système de recherche d'images par le texte(TBIR). Nous expliquons des notions importantes pour comprendre un tel système. Ces notions concernent l'annotation des images, la formulation de la requête et la comparaison binaire. Ensuite, nous présentons quelques exemples de systèmes de recherche d'images par le texte ainsi que les problèmes communs rencontrés par les systèmes TBIR.
- Au chapitre 3, nous présentons un état de l'art des travaux permettant de minimiser le silence en recherche d'images. Parmi ces travaux, il y a ceux qui ont focalisé sur l'annotation, et d'autres qui ont focalisé sur la recherche. En fait, afin de mieux positionner notre travail et de mieux faciliter la compréhension aux lecteurs, nous présentons l'ensemble de ces travaux selon la méthode adoptée elle-même pour réduire le silence. Par conséquent, nous présentons deux familles : famille des méthodes visuelles et famille des méthodes sémantiques. De plus, et puisque la phase de calcul de proximité sémantique entre les concepts est une phase clé dans notre travail, nous faisons un tour d'horizon de quelques méthodes de la littérature autour de ce calcul.

- Dans le chapitre 4, nous détaillons notre solution pour réduire le problème du silence. Premièrement, nous analysons et discutons les différentes relations sémantiques qui peuvent exister entre les concepts. C'est à partir de cette analyse que nous avons dégagé l'idée principale de notre solution. Ensuite, nous présentons une description détaillée des différentes étapes de la solution proposée. Dans ce chapitre, nous expliquons comment nous avons construit notre corpus de connaissances externes à partir de Wikipedia, comment nous l'avons exploité pour calculer la proximité sémantique entre les concepts et comment nous avons intégré cette information sémantique pour minimiser le silence que ce soit dans la recherche ou bien dans l'annotation.

- Au chapitre 5, nous commençons par une description détaillée de la configuration expérimentale sur laquelle nous avons effectué nos expériences, impliquant la base d'images, le benchmark (ou vérité terrain) de proximité sémantique et les mesures de performance utilisées. Ensuite, nous rapportons et évaluons nos résultats de calcul de proximité sémantique entre les concepts par rapport au jugement humain, ainsi que nous analysons et discutons la performance de la méthode adoptée par rapport à quelques travaux connexes. De plus, nous présentons nos résultats de recherche d'images, leur évaluation en termes de précision ainsi que leur comparaison avec quelques travaux de la littérature.

A la fin de cette thèse, nous mettons en relief les principales conclusions du travail et nous suggérons quelques perspectives et travaux futurs.

Chapitre II. Recherche d'images par le texte (TBIR) : Définitions, Architecture, Composants, Applications et Problèmes majeurs

II.1 Introduction

Le domaine de recherche d'images est un domaine en plein essor et très actif de nos jours. L'objectif principal de ce domaine de recherche est de développer des systèmes permettant de libérer l'utilisateur d'une recherche d'images manuelle laborieuse et fastidieuse. Ainsi, un utilisateur est demandé d'exprimer ses besoins sous forme de requêtes visuelles ou textuelles, et c'est le système de recherche qui prend la charge de localiser un maximum d'images pertinentes à ces requêtes en un temps raisonnable.

La littérature a connu deux grandes familles de systèmes de recherche d'images. La famille des systèmes de recherche d'images par le contenu visuel, dite Content-Based Image Retrieval ou CBIR en anglais (Bakar et al., 2013; Yang et al., 2010; Yue et al., 2011; Agarwal and Maheshwari, 2015; Aych and Amiri, 2016; Syam and Srinivasa Rao, 2012; Vipparthi and Nagar, 2015), et la famille des systèmes de recherche d'images par le texte, dite Text-Based Image Retrieval ou TBIR en anglais (Chen et al., 2010; Liu et al., 2011).

Dans un CBIR, le mécanisme de recherche se base sur l'aspect visuel des images. Ainsi, il consiste à faire une comparaison purement visuelle entre le vecteur de caractéristiques visuelles d'une requête utilisateur, sous forme d'image exemple, et ceux qui correspondent aux images de la base, en utilisant certaines mesures (ou distances). Par conséquent, l'ignorance de l'aspect sémantique des images a rapidement conduit les systèmes CBIR à se confronter avec un problème très sérieux, c'est le gap sémantique entre les caractéristiques visuelles de bas niveau

de description des images, et l'interprétation ou la sémantique associée à ces images, et qui peut être de haut niveau de description.

Contrairement aux systèmes CBIR, les Systèmes TBIR tiennent en compte la sémantique associée aux images. Ceci grâce aux annotations qui peuvent exprimer non seulement le bas niveau d'abstraction (c'est le cas du CBIR), mais aussi le niveau intermédiaire (tel que les objets contenu dans l'image) et le haut niveau (tel que les scènes, les sensations)(Yang et al., 2010). Un autre avantage des systèmes TBIR est qu'ils permettent aux utilisateurs d'exprimer facilement leurs besoins sous forme de requêtes en langage naturel.

Comme notre travail s'inscrit dans le cadre des systèmes TBIR, l'objectif de ce chapitre est de présenter les éléments essentiels pour assurer une bonne intelligibilité de ce manuscrit. Ainsi, après une définition du système TBIR, nous présentons son architecture et ses composants principaux, y compris annotation des images, formulation des requêtes et le calcul de pertinence. Nous présentons également quelques exemples de systèmes de recherche d'images par le texte en plus des problèmes communs rencontrés par les systèmes TBIR.

II.2 Définition d'un système TBIR

Un système de recherche d'images par le texte TBIR (Text-Based Image Retrieval), comme son nom l'indique, vise à localiser automatiquement des images pertinentes à une requête utilisateur textuelle, en exploitant l'aspect sémantique associé aux images sous forme d'annotations. Ainsi, un tel système identifie des images pertinentes en se basant sur une comparaison binaire entre la requête et les annotations des images de la base. Autrement dit, une image annotée explicitement par la requête est considérée comme pertinente, et une image qui ne contient pas la requête parmi ses annotations est écartée. Comme tout système de recherche d'images, un haut degré de précision et une recherche à grande vitesse (c'est-à-dire un temps de réponse court) sont deux critères principaux souhaitables pour la performance du système.

II.3 Architecture d'un système TBIR

Tous les systèmes de recherche d'images possèdent une même architecture générale qui se résume en deux phases principales. Une première phase nommée généralement

indexation. Elle consiste à assigner à chaque image de la base un code qui permet de l'identifier. Elle se fait en offline afin de minimiser le temps de réponse pour les utilisateurs. Alors qu'une deuxième phase qui se fait en online, est appelée la recherche. Elle commence dès qu'un utilisateur introduit sa requête, et consiste à rechercher des images qui répondent à cette requête à travers une comparaison entre l'identifiant de la requête et ceux des images de la base, en utilisant une certaine mesure de similarité.

Les différences qui peuvent exister entre les systèmes de recherche d'images résident dans la façon de réaliser ces deux phases. C'est-à-dire, dans la méthode utilisée pour indexer (représenter) les images d'une base, ou bien dans le mécanisme développé pour la recherche.

Dans la recherche d'images par le texte (TBIR), la phase offline est appelée l'annotation. Elle consiste à assigner à chaque image de la base un ou plusieurs mots clés (ou concepts) décrivant son contenu. Ces concepts peuvent ainsi décrire le bas niveau d'abstraction d'une image (tel que sa couleur ou sa forme), le niveau intermédiaire (tel que les objets contenus dans l'image) ou bien le haut niveau d'abstraction (tel que les scènes et les sensations). Alors le résultat de la phase d'annotation est une base d'images avec annotations.

Après que toutes les images de la base soient annotées, la phase online (la recherche) se lance dès qu'un utilisateur introduit sa requête textuelle sous forme de mots clés (ou de concepts) ou bien du texte libre. Ainsi, le système identifie des images pertinentes en se basant sur une comparaison binaire entre la requête et les annotations des images de la base.

Supposons que nous avons une base d'images annotée, de telle sorte que chaque image est annotée par un ou plusieurs concepts en offline, et que nous voulons rechercher des images à partir de cette base d'images. Un système TBIR typique est illustré dans la figure. 3

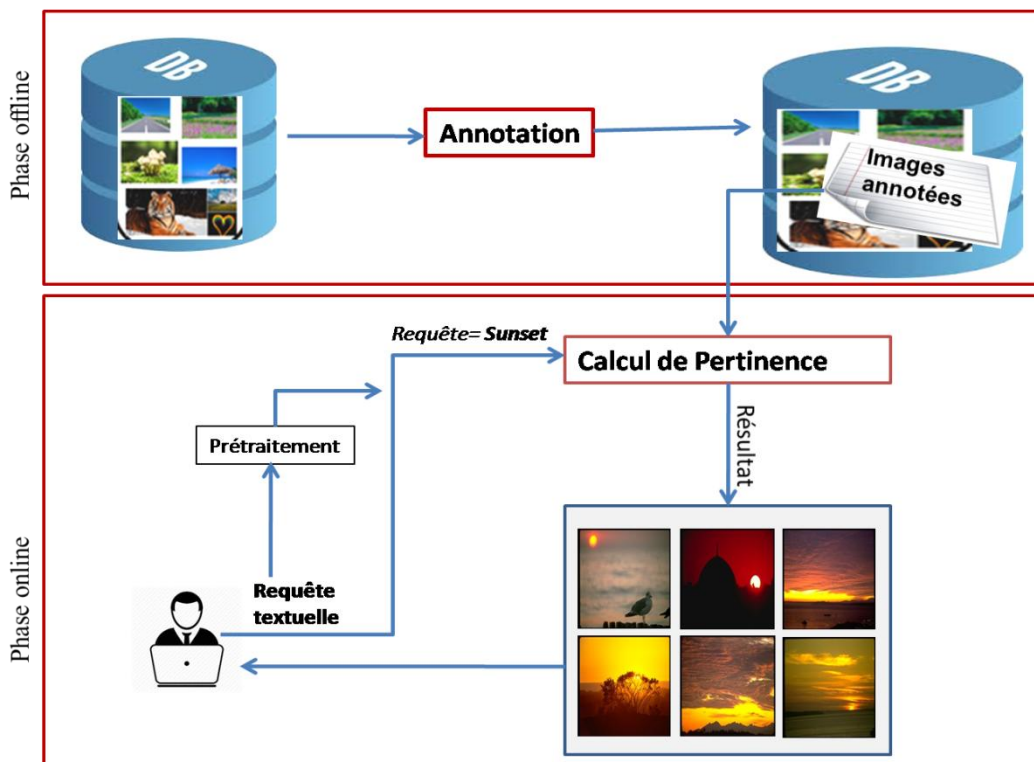


Figure 3. Architecture typique d'un système TBIR

Les étapes décrites dans la figure. 3 peuvent être expliquées comme suit :

1. L'utilisateur introduit sa requête textuelle sous forme de concepts ou bien du texte libre.
2. Si la requête est un texte libre, le système lance un prétraitement de cette requête pour la transformer en un ou plusieurs concepts.
3. Le système mesure la pertinence sémantique de chaque image de la base par rapport à la requête, grâce à une comparaison binaire entre la requête et les annotations des images.
4. Le système trie les images par ordre décroissant de pertinence. Ensuite, il affiche les N premières à l'utilisateur.

Dans la section suivante, nous détaillons les composantes principales d'un système TBIR.

II.4 Les principales composantes d'un système de recherche d'images TBIR

Un système TBIR s'articule sur trois composantes fonctionnelles principales qui sont : l'annotation des images, la formulation des requêtes et le calcul de pertinence.

II.4.1 Annotation des images

Le processus d'annotation consiste à assigner à chaque image de la base un ou plusieurs mots clés (ou concepts) décrivant son contenu. Ces concepts peuvent ainsi décrire le bas niveau d'abstraction d'une image (tel que sa couleur sa forme), le niveau intermédiaire (tel que les objets contenus dans l'image) ou bien le haut niveau d'abstraction (tel que les scènes et les sensations). En fait, l'annotation est primordiale dans un TBIR classique, car les résultats de recherches sont profondément liés à la complétude et à la qualité des annotations. Autrement dit, une annotation de qualité permet au TBIR de récupérer avec succès le maximum d'images pertinentes et donc avoir un degré élevé de précision.

Le processus d'annotation d'images peut être manuel, semi-automatique ou bien automatique.

- **L'annotation manuelle :** Le processus d'annotation s'effectue manuellement par des opérateurs humains. Ainsi, un annotateur assigne des mots clés (ou bien des concepts) aux images en se basant sur les connaissances qu'il possède à propos de ces images. Par conséquent, un des inconvénients du processus manuel est qu'il est subjectif, du fait que chacun peut interpréter (décrire) une image selon ses propres intérêts et son point de vue personnel. De plus, ce processus est fastidieux, laborieux et consomme beaucoup de temps surtout pour de grandes collections. En outre, les annotations fournies peuvent être incohérentes, générales, ambiguës, bruitées, incomplètes et parfois inappropriées(Chen et al., 2010; Wu et al., 2013), du fait que rien n'assure le sérieux durant tout le processus. Par conséquent, l'annotation manuelle peut influencer négativement les performances d'un TBIR. C'est la raison pour laquelle les annotations d'images semi-automatiques et automatiques ont été introduites pour répondre à ces défis, et ont reçu un intérêt de recherche croissant.
- **L'annotation semi-automatique :** L'annotation semi-automatique, comme son nom l'indique, se divise en une première phase qui s'effectue manuellement et une deuxième qui s'effectue automatiquement. La phase manuelle consiste à choisir un échantillon représentatif de

la collection d'images à annoter, et de l'annoter manuellement de telle sorte que les annotations doivent être correctes et complètes. Cet échantillon est généralement appelé ensemble d'apprentissage. Ensuite, dans la deuxième phase, les images restantes, appelées ensemble de test, sont annotées automatiquement en exploitant l'ensemble d'apprentissage. Pour ce faire, il existe plusieurs approches d'annotation dans la littérature, qui peuvent être classées en approches visuelles et d'autres sémantiques. Les approches visuelles se basent sur l'exploitation de la relation entre les caractéristiques visuelles et les concepts. Tandis que, les approches sémantiques constituent une extension des approches visuelles, et considèrent en plus les relations entre les concepts eux même. Plus de détail sur les approches d'annotation est présenté dans le chapitre de l'état de l'art.

- **L'annotation automatique :** C'est un processus complètement automatique sans aucune intervention de l'être humain. Ainsi, le système d'annotation se charge d'extraire les concepts caractéristiques d'un échantillon d'images, en exploitant des informations existantes sur ces images dans le Web, telles que le nom, le titre, l'adresse URL, le texte alternatif (affiché si l'image ne peut être visualisée), les métadonnées, ou bien par l'analyse du texte des pages Web contenant ces images. Ensuite, pour le reste des images, le processus d'annotation se poursuit soit avec le même principe que l'échantillon, ou bien comme dans la deuxième phase de l'annotation semi-automatique.

II.4.2 Formulation de requêtes

Après que toutes les images d'une base soient annotées en offline, un TBIR peut répondre aux besoins des utilisateurs. En effet, un des avantages d'un TBIR c'est en plus du fait qu'il donne la possibilité à l'utilisateur d'introduire des requêtes visuelles, il lui permet aussi d'exprimer facilement ses besoins en langage naturel. Autrement dit, un utilisateur utilise une image ou un texte pour introduire sa requête. Ainsi une requête peut être formulée comme suit :

Requête en texte libre : Dans ce type, l'utilisateur a la liberté totale d'écrire le texte qu'il veut, pour exprimer ses besoins. Ensuite, le système déclenche un processus de prétraitement de cette requête, en utilisant des techniques de traitement et d'analyse du langage naturel (TAL), afin de transformer ce texte en un ensemble de concepts clés bien définis(Claveau and Nie, 2016; Jiang and Tan, 2006; Contreras et al., 2004). Par conséquent, une telle requête sera transformée en l'un des deux types suivants.

- **Requête par concept atomique :** Dans ce type l'utilisateur introduit une requête sous forme d'un seul concept qu'il choisit parmi un ensemble bien défini de concepts par le système TBIR. Notons qu'un traitement de concepts synonymes et multilingues est possible grâce à l'utilisation des structures sémantiques telles que les thésaurus et les ontologies (Malone et al., 2016).
- **Requête composée de plusieurs concepts :** Ce type est une extension du type précédent. Ainsi, au lieu d'introduire un seul concept, l'utilisateur introduit plusieurs concepts avec connecteurs logiques tels que ET, OU et NON. Alors, dans ce type, la requête est une expression logique.
- **Requête visuelle :** Dans ce type l'utilisateur choisit une ou plusieurs images de la base comme requête. Par conséquent, le système TBIR prend les annotations correspondantes comme requête. Ensuite, il lance un processus de raffinement de requête pour comprendre ce que l'utilisateur veut exactement. Après le processus de raffinement, la nouvelle requête sera soit une requête atomique ou bien composée de plusieurs concepts.

La figure 4. montre les premières images retrouvées par 'Google Images' en utilisant une requête textuelle par le concept atomique 'cascade'.



Figure 4. Premières images récupérées par 'Google Images' pour la requête 'cascade'

II.4.3 Calcul de pertinence (recherche)

Après qu'un utilisateur ait introduit sa requête, le système TBIR prend la charge de localiser des images pertinentes à cette requête. Pour calculer la pertinence d'une image, un

TBIR suit généralement un mécanisme très utilisé en recherche d'information, c'est la comparaison booléenne ou binaire entre les concepts composants la requête et ceux annotant chaque image de la base. Autrement dit, s'il s'agit d'une requête atomique, une image qui est annotée explicitement avec le concept requête est considérée comme pertinente (valeur de pertinence = 1), et donc retournée à l'utilisateur, alors qu'une image qui ne l'est pas est considérée comme non-pertinente (valeur de pertinence = 0), et donc ignorée. Dans le cas d'une requête composée, une image est considérée comme pertinente si son annotation satisfait l'expression booléenne de la requête.

II.5 Exemples de systèmes TBIR

Il existe de nombreux systèmes de recherche d'images par le texte qui sont généralement destinés pour rechercher des images sur le Web. Dans ce qui suit nous allons citer quelques-uns.

II.5.1 Google Images

Le moteur Google est l'un des moteurs les plus utilisés à travers le monde pour la recherche sur le Web, créé par Google en 2001. Il permet de rechercher tout type d'information, textuelle, images, audio et vidéo. Pour rechercher une image, l'utilisateur a la possibilité d'introduire soit une requête textuelle sous forme de mots clés ou bien une requête visuelle par image exemple. Ainsi, pour déterminer si une image dans le Web est pertinente, la recherche textuelle de Google se base sur certaines informations tel que le nom de l'image, son URL, le texte alternatif qui s'affiche dans le cas où l'image ne peut être visualisée et qui permet de décrire le contenu de l'image, l'analyse du texte de la page Web contenant cette image, ainsi que le contenu situé à proximité de l'image y compris toute légende descriptive ou titre.

II.5.2 Atlas WISE

Atlas WISE(Kherfi et al., 2003)est un moteur de recherche d'images sur le Web, qui utilise des caractéristiques visuelles et textuelles pour décrire et rechercher des images du Web. Pour ce faire, il collecte des images depuis des pages Web populaires telles que celles du Google et Yahoo. Ensuite, les images collectées sont indexées visuellement en utilisant l'histogramme de couleur et l'histogramme de gradient orienté (HOG). De plus, ces images sont annotées par des concepts caractéristiques extraites depuis leurs noms, tags et titres des pages correspondantes. Les utilisateurs d'Atlas WISE peuvent exprimer leurs besoins sous forme de

requêtes visuelles ou bien textuelles par concepts sémantiques. La recherche textuelle d'Atlas WISE suit un même principe que le modèle vectoriel utilisé dans les approches de recherche d'information(Haddad et al., 1996).

II.5.3 PicHunter

PicHunter(Cox et al., 2000)est développé par l'institut de recherche NEC. C'est un système CBIR qui a été amélioré par l'introduction de l'aspect sémantique des images. Ainsi, PicHunter décrit les images en utilisant l'histogramme des couleurs et la répartition spatiale des couleurs avec des annotations. Pour la comparaison visuelle, il utilise la distance de Minkowski. Pour l'introduction de l'aspect sémantique, PicHunter a utilisé un ensemble de concepts d'annotation qui contient 147 concepts sémantiques, dont 138 concepts sont identifiés par l'un des auteurs qui avait une exploration large de la base d'images expérimentale, et les 9 concepts restants sont des concepts de catégories, telle que la catégorie '*animal*'. La base d'images expérimentale se constitue de 1500 images depuis la base d'images COREL. Ces images sont examinées visuellement et annotées manuellement par des concepts pertinents parmi les 138 concepts. Pour les 9 concepts de catégories, ils sont assignés aux images d'une façon automatique. Ainsi, chaque image est représentée par un vecteur booléen de taille égale à 147, où une valeur égale à 1 signifie que le concept correspond est présent dans l'image, et une valeur égale à 0 signifie le contraire. Pour la recherche textuelle, PicHunter, utilise la distance de Hamming normalisée comme mesure de similarité sémantique entre les vecteurs d'annotation des images.

II.5.4 Bing Images

C'est un moteur de recherche rival de Google, créé par Microsoft en 2009. Bing Images propose quasiment les mêmes options de recherche que Google. Pour la recherche textuelle des images, il suit ainsi le même principe que Google Images.

II.5.5 ALIPR

Automatic Linguistic Indexing of Pictures- Real Time ALIPR(Li and Wang, 2008) est un système d'annotation automatique et de recherche d'images en temps réel. Autrement dit, il

permet d'identifier des mots d'annotation pour n'importe quelle image dans le web présentée par son URL, ou bien pour des images uploadées par des utilisateurs. Pour ce faire, les développeurs de ce système ont créé un nouvel algorithme de regroupement (c.à.d. clustering en anglais) inspiré depuis l'algorithme K-Means, appelé 'D2-clustering', où D2 signifie que des objets sont représentés par des distributions discrètes. Cet algorithme minimise la distance totale dans un cluster. De plus, ils ont développé une nouvelle méthode de modélisation de mixture (the hypothetical local mapping (HLM) method) pour obtenir une mesure de probabilité efficace sur l'espace de distribution discrète. ALIPR est également un système de recherche d'images. Il offre une recherche textuelle basées-mots clés, ainsi qu'une recherche visuelle par image exemple. Sa recherche basées-mots clés consiste à faire une comparaison binaire entre les mots clés d'une requête utilisateur et les mots d'annotations des images de la base. Une version de démonstration est disponible au site Web <http://alipr.com>, qu'est a été rendu public le 1er novembre 2006. Afin de rendre ce site plus utile pour la recherche d'images, ses développeurs ont ajouté plus d'un million d'images depuis terragalleria.com et Flickr.com, et ils ont vérifié les tags fournis par ces sites en utilisant ALIPR.

II.6 Problèmes majeurs d'un système TBIR

L'avantage majeur d'un système TBIR par rapport à un CBIR c'est qu'il considère l'aspect sémantique associé aux images, grâce aux annotations. Ces dernières peuvent exprimer la sémantique dans les différents niveaux d'abstraction. Autrement dit, les concepts associés aux images peuvent être de bas niveau d'abstraction comme le fait un CBIR (par exemple un concept couleur exprime la couleur d'une image), de niveau intermédiaire (par exemple une image peut être annotée par les noms des objets qu'elle contient) et de haut niveau d'abstraction (par exemple une image peut être annotée par des concepts de scènes ou de sensation tel que le concept *joie* ou *fête*). De plus, un TBIR répond efficacement par rapport à son rival CBIR dans le cas où il y a un gap sémantique entre la représentation visuelle et l'interprétation des images. Autrement dit, dans le cas où l'utilisateur introduit une requête visuelle qui exprime un certain concept sémantique (généralement de haut niveau d'abstraction) alors que ce concept peut avoir différentes apparences visuelles, ou dans le cas où il existe plusieurs images similaires à sa requête visuelle, mais qui expriment des concepts sémantiques différents dans des contextes variés. Pour ce genre de requête, un CBIR peut conduire à de mauvais résultats, puisqu'il s'appuie uniquement sur les caractéristiques visuelles. Autrement dit, la comparaison visuelle

peut ignorer plusieurs images pertinentes à la requête utilisateur, du point de vue sémantique, du fait que leur ressemblance visuelle avec l'image requête est légère. Comme il peut retourner plusieurs images bruits, du fait qu'elles se ressemblent visuellement à la requête, alors qu'elles sont complètement différentes de ce que veut l'utilisateur. Par conséquent, un utilisateur trouve le système TBIR plus intéressant, surtout dans le cas où sa requête est orientée sémantique plutôt que visuelle. A titre d'exemple, si un utilisateur veut chercher des images de *joie* c'est plus intéressant d'introduire une requête textuelle que donner une image exemple.

Cependant, le système TBIR est confronté à un certain nombre de limitations et problèmes communs qui devraient être abordés pour améliorer ses performances. Dans ce qui suit, nous résumons les problèmes les plus courants, ainsi que le grand défi d'un système TBIR.

II.6.1 Problème de Silence

Le silence veut dire qu'un TBIR peut écarter plusieurs images pertinentes à la requête de l'utilisateur parce que tout simplement elles ne sont pas annotées explicitement par cette requête. Ce problème dégrade la qualité et la précision des résultats d'une façon considérable.

Les causes principales de ce problème, sont le manque dans les annotations conjointement avec un mécanisme de recherche binaire. Une explication détaillée de ce problème est déjà faite dans le chapitre précédent.

L'objectif principal du présent travail est de donner une solution pour minimiser le problème de silence d'une manière entièrement automatique. Dans le prochain chapitre, nous présentons un état de l'art de quelques méthodes permettant de minimiser le silence en recherche d'images.

II.6.2 Problème de Bruit

Le bruit signifie que le moteur retourne des images erronées par rapport à ce qui a été demandé par l'utilisateur dans sa requête. Ce problème est dû à la comparaison binaire de la requête utilisateur avec des concepts d'annotations anonymes ou bruités. Des concepts d'annotation anonymes (ou polysémies) signifient que les images peuvent être annotées par un même concept que la requête mais dans un autre contexte sémantique. Des concepts d'annotation bruités ou erronés, signifient que les images sont annotées par des concepts qui ne les expriment pas correctement. Le bruit dans les annotations est dû principalement au processus d'annotation manuel qui est subjectif, laborieux et fastidieux.

II.6.3 Défi de l'annotation des images

D'après le principe d'un système TBIR classique décrit dans les sections précédentes, il est clair que son succès pour rechercher des images pertinentes est hypothéqué par la qualité de processus d'annotation. L'annotation semi-automatique et automatique des images est devenue un domaine de recherche très actif et qui a attiré l'attention de beaucoup de chercheurs en ces dernières années. Par conséquent, plusieurs travaux ont été développés afin d'annoter des nouvelles bases d'images, de compléter des annotations des bases d'images légèrement annotées, ou bien d'éliminer des annotations erronées. Le but ultime de ces travaux est d'avoir des bases d'images avec annotations complètes et correctes, et par conséquent assurer une recherche d'images avec un haut degré de précision. Dans le chapitre suivant, nous présentons quelques travaux autour de ce défi.

II.7 Conclusion

La recherche d'images par le texte TBIR est plus intéressante que son alternative, la recherche d'images par le visuel CBIR, dans le cas où la requête utilisateur est orientée sémantique plus que visuelle. Ceci se justifie par le fait qu'un TBIR tient en compte l'aspect sémantique associé aux images, et ce , grâce aux annotations qui peuvent décrire les différents niveaux d'abstraction de l'être humain.

Ce chapitre a été principalement consacré à donner l'essentiel pour faire comprendre un système TBIR ainsi que ses problèmes majeurs. En particulier, nous avons expliqué le principe d'un TBIR typique, ainsi que ses composants principaux, y compris l'annotation des images, la formulation des requêtes et le calcul de pertinence. Nous avons terminé ce chapitre par l'explication des problèmes majeurs des systèmes TBIR qui influencent considérablement la précision et la performance des résultats de recherche. Nous avons expliqué et illustré l'un des problèmes qui reste encore ouvert, le problème de silence, qui est notre préoccupation principale dans cette thèse.

Chapitre III. Minimisation du silence dans un TBIR : état de l'art

III.1 Introduction

Le silence est l'un des problèmes majeurs qui influence la précision et la qualité d'un système de recherche d'images en général. Dans un TBIR classique, ce problème est dû principalement au mécanisme de recherche basé sur une comparaison binaire entre une requête utilisateur et des annotations incomplètes des images. Dans la littérature, des efforts importants ont été consacrés afin de minimiser le silence, et donc améliorer la précision des systèmes de recherche d'images. Néanmoins et malgré ces efforts, le silence reste un problème ouvert.

Dans ce chapitre, nous présentons un état de l'art des travaux qui ont contribué à la résolution d'un tel problème. Nous commençons par une proposition d'une catégorisation de ces travaux. Ensuite, nous détaillons les méthodes de chaque catégorie. De plus, vu que l'étape de calcul de proximité sémantique entre les concepts est une étape principale dans notre solution, et parce que cette tâche constitue un réel défi pour la communauté scientifique spécialisée, et que toute une panoplie des travaux ont focalisé sur la manière de mesurer la sémantique entre les concepts (Gabrilovich and Markovitch, 2007), (Hassan and Mihalcea, 2011; Jabeen et al., 2012), (Ni et al., 2016; Taieb et al., 2014; Pakhomov et al., 2010; Aouicha and Taieb, 2016) et ce, indépendamment du domaine d'application, nous allons consacrer également une section pour présenter quelques-uns de ces travaux.

III.2 Familles des méthodes permettant de minimiser le silence

Dans la littérature, de nombreux travaux ont contribué à la réduction du problème de silence, et donc à l'augmentation de la précision des systèmes de recherche d'images. Parmi ces travaux, il y a ceux qui ont été concentrés sur l'annotation des images, comme il y a d'autres qui se sont focalisés sur la recherche. L'objectif des premiers est d'assigner aux images d'une base des concepts décrivant leur contenu sémantique, de telle sorte que ces annotations soient correctes et complètes. Ainsi, ils ont mis l'effort afin d'arriver à créer des annotations correctes et complètes pour des bases d'images non annotées, à compléter des annotations manquantes ainsi qu'à raffiner des annotations existantes mais bruitées par l'élimination de celles incorrectes dans des bases d'images légèrement annotées(Wu et al., 2013). Par conséquent, quand les annotations sont complètes et correctes, la comparaison binaire ne conduit pas au problème de silence. Tandis que l'objectif des travaux focalisant sur la recherche était de développer des mécanismes de recherche mieux que la comparaison binaire entre une requête utilisateur et les annotations des images d'une base. Autrement dit, développer des mécanismes de recherche qui soient capables de détecter des images pertinentes même si la requête ne figure pas explicitement dans leurs annotations.

Bien que notre solution, pour réduire le silence dans un TBIR, peut être exploitée dans la recherche pour détecter plus d'images pertinentes, comme dans l'annotation pour détecter et compléter des annotations pertinentes manquantes dans des bases d'images légèrement annotées, nous préférons présenter l'ensemble des travaux de l'état de l'art selon une autre projection qui permet de mieux positionner notre travail et de mieux faciliter la compréhension aux lecteurs. En fait, c'est une projection selon la méthode adoptée elle-même pour réduire le silence. Par conséquent, nous pouvons distinguer deux familles de méthodes : la famille des méthodes visuelles et la famille des méthodes sémantiques. Les méthodes visuelles n'ont pris en considération que la relation image-concepts. Ce sont généralement des méthodes destinées à l'annotation, tandis que les méthodes sémantiques ont considéré en plus de la relation image-concepts la relation concept-concept afin d'annoter ou rechercher des images. Autrement dit, les méthodes sémantiques sont des extensions des méthodes visuelles, elles contiennent une partie visuelle où la relation image-concepts est prise en considération, et une autre partie sémantique (c'est l'extension) où la relation concept- concept est prise en considération.

Nous allons commencer par un aperçu global et rapide des méthodes visuelles, pour comprendre leurs lacunes et justifier l'importance et la nécessité de considérer cette dernière relation afin de minimiser le silence à un degré satisfaisant.

III.3 Méthodes visuelles

Sous cette famille se trouve un nombre important des travaux destinés beaucoup plus à l'annotation. Ainsi, ils ont développé des techniques d'annotation qui exploitent la relation entre les caractéristiques visuelles de bas niveau et les concepts d'annotation d'un nombre suffisant d'images annotées manuellement, dit ensemble d'apprentissage, et ce afin de prédire des concepts pour des images non annotées, dites images de test. Ainsi, pour annoter les images de test, plusieurs méthodes y ont été développées. Elles partent toutes d'une même hypothèse que les annotations de l'ensemble d'apprentissage sont parfaites. C'est-à-dire, qu'elles ne contiennent ni des concepts bruits, ni des concepts manquants, ni des concepts ambigus. Elles peuvent être classées en deux familles : les méthodes basées-modèles, et les méthodes basées-données(Cui et al., 2015). Dans ce qui suit, nous présentons brièvement le principe de chaque famille.

III.3.1 Méthodes basées-modèles

Ces méthodes ont proposé des techniques d'apprentissage automatique qui visent à rendre la machine capable d'apprendre les concepts sémantiques en fonction des caractéristiques visuelles des images ou des régions correspondantes, et ceci dans l'ensemble d'apprentissage. Autrement dit, la machine soit capable de faire sortir les modèles visuels des concepts en fonction des images d'apprentissage et ses annotations correspondantes. La condition principale de l'ensemble d'apprentissage c'est que ses annotations doivent être complètes et correctes. Ensuite, les modèles visuels obtenus sont utilisés pour prédire des concepts (annotations) pour l'ensemble du test. Parmi les méthodes basées-modèles, nous pouvons trouver les méthodes probabilistes et les méthodes de classification(Cui et al., 2015). Le principe des méthodes probabilistes consiste à inférer la probabilité de distribution jointe des caractéristiques visuelles et des concepts d'annotation de l'ensemble d'apprentissage. Ainsi, le modèle obtenu est une probabilité de distribution utilisée pour prédire des concepts pour les images de test(Cui et al., 2015). Plus du détail sur les méthodes probabilistes est présenté dans(Cui et al., 2015). Dans les

méthodes de classification, la phase d'apprentissage a pour objectif de modéliser la relation entre les vecteurs visuels des images (ou régions) et les concepts. Autrement dit, elle consiste à utiliser les vecteurs des caractéristiques visuelles des images (Guo and Gu, 2011; Zha et al., 2009) ou des régions (Bao et al., 2011) d'apprentissage et les annotations correspondantes afin de sortir les modèles visuels des concepts. Ensuite, dans la phase test, les modèles obtenus sont utilisés afin de prédire des concepts pour les images de test qui sont représentées par leurs vecteurs visuels. Certaines méthodes de classification traitent les concepts séparément en utilisant des classifieurs binaires, un pour chaque concept, pour prédire la présence ou bien l'absence d'un certain concept dans une image de test. (Xu et al., 2016; Cui et al., 2015), tandis que d'autres méthodes de multi-classification utilisent un apprentissage multi-label (multi-label learning en anglais), pour prédire plus d'un seul concept pour une image de test (Bao et al., 2011; Guo and Gu, 2011; Zha et al., 2009). Les travaux dans (Guo and Gu, 2011; Zha et al., 2009) se sont basés sur l'hypothèse que les images qui sont visuellement similaires sont susceptibles de partager des mêmes annotations et considèrent les caractéristiques globales des images. Par contre, des autres travaux comme (Bao et al., 2011; Parashar, 2009; Manipoonchelvi and Muneeswaran, 2011; Zhang et al., 2012a; Belloulata et al., 2014; Yang et al., 2014; Gallas et al., 2015; Manipoonchelvi and Muneeswaran, 2015) ont montré qu'habituellement, chaque concept ne correspond mieux et bien qu'à une région locale au sein d'une image. Par conséquent, ils ont exploité les caractéristiques locales extraites depuis des régions au lieu des caractéristiques globales.

III.3.2 Méthodes basées-données

Une méthode de base très connue sous cette catégorie, c'est celle des plus proches voisins (en anglais nearest-neighbor NN). Elle s'appuie sur l'hypothèse que des images visuellement similaires sont plus susceptibles de partager des annotations. Ainsi, pour annoter une nouvelle image il suffit de propager les concepts des images voisines les plus proches. Cette méthode se déroule en deux phases. La première phase consiste à trouver parmi les images d'apprentissage, celles qui sont les K-voisines les plus proches à la nouvelle image dans l'espace visuel. Ensuite, la deuxième phase consiste à sélectionner, parmi les annotations de ces voisins, les concepts les plus pertinents, et les faire propager à cette image (Cui et al., 2015).

Plusieurs travaux de la littérature ont utilisé cette méthode afin d'annoter une base d'images. Les différences entre ceux-ci résident dans les mécanismes utilisés dans ces deux phases. Comme instance, dans la première phase de recherche des images voisines les plus proches à une image de test, en exploitant leurs caractéristiques visuelles, les auteurs de (Makadia et al., 2008) ont représenté les images d'une base par des caractéristiques visuelles de bas niveau qui sont la couleur et la texture. Pour la couleur, ils ont utilisé 3 espaces de couleurs, à savoir : RGB, HSV et LAB. Pour la texture, ils ont utilisé les Ondelettes de Gabor et Haar. Ensuite, ils ont utilisé les distances de base : *KL*-divergence, *Chi2*-statistic, la distance-*L1* et la distance-*L2* pour mesurer la similarité visuelle entre l'image à annoter et les images d'apprentissage, de telles sortes que pour chaque caractéristique, ils ont utilisé la distance la plus appropriée. Par exemple, suite à une évaluation des distances sur l'ensemble d'apprentissage, ils ont trouvé que la distance *KL*-divergence donne de bons résultats pour la caractéristique de couleur LAB. De plus, ces auteurs et d'autres (Wu et al., 2009; Wu et al., 2011; Zhang et al., 2012b) ont tenté de définir une mesure de distance entre images via une combinaison linéaire des distances dans les différentes dimensions de l'espace des caractéristiques visuelles. Donc, les *K*-voisins les plus proches sont les *K*-premières images similaires à l'image de test. Cependant, les auteurs d'un travail plus récent (Cui et al., 2015) ont proposé un mécanisme de tri qui permet d'optimiser l'ordre relatif des images d'apprentissage, au lieu de leurs distances obsolètes avec l'image de test. Ainsi, ils ont développé un algorithme de tri qui utilise les techniques de learning-to-rank (LTR) (Liu, 2009) et qui exploite l'information de préférence implicite cachée dans les images d'apprentissage.

Dans la deuxième phase de sélection des concepts à propager, la plupart des études ont utilisé des règles heuristiques pour sélectionner des concepts les plus pertinents (Cui et al., 2015). Comme instance, les auteurs de (Torralba et al., 2008) ont trié les concepts en fonction de leurs fréquences d'apparition dans les voisins. Ceci a présenté un handicap, du fait que des concepts généraux et fréquents dans l'ensemble d'apprentissage dominent les résultats. Pour cela, les auteurs de (Li and Wang, 2008) ont multiplié cette fréquence par la fréquence inverse du concept dans cet ensemble, et donc ils ont proposé une règle heuristique semblable à la méthode de pondération statistique très connue en recherche d'information, qui est la méthode TF-IDF (Term Frequency-Inverse-Term-Frequency). Les auteurs de (Makadia et al., 2008) ont utilisé un mécanisme de propagation des concepts des *K*-voisins les plus proches à l'image de test et qui exploite l'information de fréquence d'apparition ainsi que l'information de co-occurrence entre les concepts. Leur mécanisme de propagation peut être décrit comme suit :

Partant d'une image de test I à annoter avec n concepts et ses K -voisins les plus proches $k_{proches} = \{I_1, I_2, \dots, I_k\}$. La propagation des concepts se fait comme suit :

-Trier les concepts d'annotation de I_l , l'image la plus similaire à I , selon leurs fréquences d'apparition dans l'ensemble d'apprentissage.

-Propager les n premiers concepts vers l'image de test I .

-Si le nombre de concepts propagés $< n$ alors

-Trier les concepts d'annotation des images K -voisins les plus proches restantes $\{I_2, \dots, I_k\}$ en se basant sur leur co-occurrence avec les concepts de I_l ainsi que leurs fréquences d'apparition dans l'ensemble $\{I_2, \dots, I_k\}$.

-Propager les $n - |I_l|$ premiers concepts vers l'image de test I .

Cependant, les auteurs de (Cui et al., 2015), ont déclaré qu'il n'y a aucune garantie de la qualité de sélection des concepts à propager en utilisant ces règles heuristiques. Par conséquent, ils ont choisi d'utiliser des méthodes d'apprentissage supervisées, dont l'objectif est d'apprendre une fonction de score fiable pour évaluer la pertinence des concepts d'apprentissage pour l'image de test. Pour ce faire, ils ont utilisé la méthode structurelle de machine à vecteurs de supports (structural SVM ou structural support vector machine en anglais) (Tsochantaridis et al., 2005) avec un algorithme des plans sécants (cutting plane algorithm) (Joachims et al., 2009).

III.4 Limites des méthodes visuelles

Bien que les méthodes précédentes aient fourni des efforts considérables, et qui ont mené à des résultats encourageants et importants, elles restent confrontées avec quelques limitations importantes, qui doivent être abordées, à savoir :

1. La majorité des méthodes d'annotation proposées exigent un nombre important d'images d'apprentissage et supposent qu'une condition très critique pour la réussite du processus d'annotation soit vérifiée. Ils supposent que les annotations manuelles de l'ensemble d'apprentissage soient précises. Ce qui signifie qu'elles ne contiennent ni de manque, ni de bruit ni d'ambiguïté. En pratique, il est très difficile de satisfaire cette condition idéale pour la plupart des images du monde réel (Wu et al., 2013). Par conséquent, cela peut conduire à une estimation

biaisée de la prédiction, ce qui dégrade la performance du processus d'annotation, et par conséquent, les résultats de recherche !

2. Elles considèrent la relation entre les caractéristiques visuelles des images ou régions et les concepts d'annotation. Cependant, elles ignorent complètement une relation très importante, qui est la relation entre les concepts eux même. En réalité, les concepts ne se trouvent pas indépendants, mais il y a une certaine corrélation entre eux. Par exemple, le concept '*forêt*' habituellement co-occure avec le concept '*arbre*'. L'ignorance de cette sémantique entre les concepts peut influencer négativement les résultats d'annotation, et par conséquent, les résultats de recherche.

3. Il est très difficile pour les méthodes visuelles d'annotation de modéliser des concepts ayant plusieurs apparences visuelles, comme le concept '*paysage*' ou '*fête*'. Ce genre de concepts correspond généralement à un haut niveau d'abstraction. Egalement pour les méthodes de recherche visuelle, il est difficile de chercher ce genre de concepts en introduisant une image exemple. Au fait, donner une image exemple peut amener plusieurs images pertinentes à rester en silence en raison de leur grande distance visuelle à la requête.

4. Les approches visuelles permettent d'annoter les images par des concepts perceptuels locaux de premier niveau d'abstraction, décrivant le contenu local de ces images. Cependant, il existe d'autres concepts sémantiques globaux de niveau intermédiaire ou bien de haut niveau d'abstraction qui ne sont pas pris en considération.

III.5 Méthodes sémantiques

Sous cette famille, s'inscrit un nombre important des travaux récents destinés soit à l'annotation ou bien à la recherche. L'avantage majeur de ces travaux par rapport aux travaux précédents, c'est qu'ils ont exploité, en plus des relations entre les caractéristiques visuelles et les concepts, les relations d'interdépendance entre les concepts eux-mêmes. Cette dernière relation est très importante du fait qu'en réalité, les concepts ne s'y trouvent pas séparément, mais il y a une certaine sémantique qui les relie. Néanmoins, pour calculer l'interdépendance entre les concepts d'une façon automatique, un ordinateur doit être disposé d'une source de connaissance externe qui simule la connaissance humaine. Ainsi, en fonction de cette source de connaissance, nous pouvons distinguer deux familles de méthodes sémantiques : méthodes basées-corpus local et méthodes basées-corpus global. Les méthodes basées-corpus local ont utilisé comme source de connaissance l'information sémantique de cooccurrence entre les

concepts présentée dans l'annotation d'une base d'images locale, c'est la base expérimentale d'apprentissage, tandis que les méthodes basées-corpus global ont extrait la sémantique entre les concepts en utilisant des sources de connaissance externes indépendamment de toute base d'image expérimentale, tel que Wikipedia.

Après le calcul (modélisation) de l'interdépendance entre les concepts, les modèles obtenus sont incorporés dans la méthode de prédiction des concepts pour les images de test (c.à.d. dans le processus d'annotation), ou bien au sein du moteur de recherche d'images. Dans ce qui suit, nous présentons quelques travaux de la littérature de chacune de ces deux familles, dont l'objectif est de savoir comment ils ont modélisé l'interdépendance entre les concepts, ainsi que comment ils ont intégré les modèles obtenus au sein du processus d'annotation ou de recherche des images de test.

III.5.1 Méthodes sémantiques basées-corpus local

Les méthodes de cette famille peuvent être considérées comme une extension des méthodes visuelles. En fait, elles exploitent la relation entre les caractéristiques visuelles et les concepts d'annotation, et ce conjointement avec l'information sémantique d'interdépendance entre les concepts d'annotation en fonction de leur cooccurrence, et ceci dans l'ensemble d'images d'apprentissage. Ainsi, l'interdépendance entre les concepts est modélisée sous forme d'un réseau, un graphe ou bien une matrice (Guo and Gu, 2011; Zha et al., 2009; Xu et al., 2016; Yang et al., 2011; Franzoni et al., 2015; Feng and Bhanu, 2011; Metzler and Manmatha, 2004). En fait, nous pouvons distinguer deux classes de méthodes de modélisation de l'interdépendance entre les concepts. Sous la première classe s'inscrivent les méthodes de calcul statistique. Elles fournissent le résultat sous forme d'une matrice dite de similarité sémantique ou de corrélation. Tandis que les méthodes de la deuxième classe utilisent des graphes ou bien des réseaux pour modéliser l'ensemble des données (images et concepts), et utilisent des calculs probabilistes approximatifs.

III.5.1.1 Un apprentissage semi-supervisé multi-labels basé graphe

(Graph based semi-supervised learning with multiple labels)(Zha et al., 2009)

Dans ce travail, les auteurs ont présenté une méthode d'apprentissage multi-labels basée graphe pour la détection des concepts dans des clips vidéo qui sont segmentés en plusieurs images. Ainsi, leur méthode visuelle s'inscrit dans la classe des méthodes de multi-classification

utilisant un apprentissage multi-label, où une nouvelle image peut être annotée par plusieurs concepts à la fois. Par exemple, une image peut être annotées à la fois par ‘*nuage*’, ‘*montagne*’ et ‘*eau*’. En fait, l’avantage majeur de cette méthode contrairement aux méthodes directes d’apprentissage multi-label qui traitent les concepts séparément l’un des autres (par l’utilisation d’un ensemble de classifieurs binaires où les concepts d’une nouvelle image sont déterminés en fonction de leurs résultats) alors qu’en réalité ils sont corrélés, c’est qu’elle modélise la corrélation entre les concepts et qu’elle l’intègre au sein du processus d’apprentissage basé-graphe. Dans ce qui suit, nous présentons comment les auteurs ont modélisé la corrélation entre les concepts, et comment par la suite, ils l’ont intégré dans la méthode de prédiction.

1. Modélisation de la corrélation entre les concepts

Pour modéliser la corrélation entre les concepts, les auteurs ont adopté des calculs statistiques sur les concepts d’annotation de l’ensemble d’apprentissage. Ainsi, partant d’un ensemble d’images annotées d’apprentissage $\{x_1, x_2, \dots, x_n\}$, et d’un ensemble de concepts d’annotation $\{c_1, c_2, \dots, c_k\}$, la corrélation entre les concepts est modélisé sous forme d’une matrice C , dite matrice de corrélation. Elle est de taille $(k \times k)$, où k est le nombre de concepts. La matrice C s’obtient par la soustraction de deux autres matrices C' et D_c .

a) Calcul de la matrice C'

La matrice C' a pour objectif de capturer la distance entre chaque pair de concepts, en fonction de leurs occurrences dans les annotations de l’ensemble d’images d’apprentissage. Pour ce faire, l’ensemble d’apprentissage est représenté par une matrice Y dite d’annotation. Elle est de taille $(n \times k)$, où chaque ligne i de cette matrice correspond au vecteur d’annotation f_i de l’image x_i , et chaque colonne j correspond au vecteur d’occurrence f_j' du concept c_j dans l’ensemble d’apprentissage. Un élément $Y[i, j]$ a une valeur égale à 1 si l’image x_i est annotée par le concept c_j , et une valeur égale à 0 autrement.

La matrice C' est une matrice symétrique carrée de taille $(k \times k)$. Une valeur de cette matrice $C'[i, j]$ représente la distance euclidienne moyenne entre les vecteurs des deux concepts c_i et c_j . Elle est calculée par la formule suivante:

$$C'[i, j] = \exp\left(-\frac{\|f'_i - f'_j\|^2}{2\sigma_c^2}\right) \quad (1)$$

Où f'_i est le vecteur du concept c_i représenté par la colonne i de la matrice Y . De même pour f'_j . $\|f'_i - f'_j\|$ signifie la distance euclidienne entre ces deux vecteurs, et σ_c c'est leur distance moyenne, elle est définie comme suit :

$$\sigma_c = E(\|f'_i - f'_j\|) \quad (2)$$

b) Calcul de la matrice D_c

Après le calcul de la matrice C' , la matrice D_c est une matrice diagonale. Un élément de cette matrice $D_c[i, i]$ représente la somme des distances du concept c_i avec tous les autres concepts.

$$D_c[i, i] = \sum_{j=1}^k C'[i, j] \quad (3)$$

Après le calcul des deux matrices C' et D_c , la matrice de corrélation entre les concepts C est le résultat de soustraction de ces deux matrices.

$$C = C' - D_c \quad (4)$$

Ainsi la matrice C est une matrice carrée symétrique. Elle est de taille $(k \times k)$, où k est le nombre de concepts. Un élément de cette matrice $C[i, j]$ contient la valeur de corrélation de deux concepts c_i et c_j .

2. Intégration de la corrélation entre les concepts dans la méthode d'annotation

La méthode de détection des concepts adoptée par les auteurs est une méthode d'apprentissage multi-labels basée sur les graphes. Les stratégies d'apprentissage basées-graphe modélisent l'ensemble de toutes les données sous forme d'un graphe, dont les nœuds de ce graphe correspondent aux images d'apprentissage et de test, et les arcs reflètent la similarité entre les nœuds. Elles visent à estimer une fonction optimale sur le graphe qui doit vérifier certaines propriétés. Ainsi, les auteurs ont tenté d'optimiser une fonction de prédiction des concepts suivante :

$$E(F) = E_l(F) + E_s(F) + E_c(F) \quad (5)$$

Les deux premiers termes sont définis de la même façon que celle adoptée par les autres méthodes basée graphe. Ainsi, $E_l(F)$ est une fonction de perte (loss function en anglais). Elle

permet de vérifier la propriété que la fonction $E(F)$ soit proche (close) à une annotation donnée dans les nœuds d'apprentissage. Cette propriété pénalise la déviation d'une annotation donnée. $E_s(F)$ et $E_c(F)$ sont des régulateurs. $E_s(F)$ est un régulateur pour redresser la lisibilité des annotations (label smoothness en anglais) dans le graphe au complet.

La nouveauté introduite par les auteurs concerne l'intégration du deuxième régulateur, qui permet d'assurer que la fonction $E(F)$ soit consistante avec la corrélation entre les concepts. Autrement dit, il permet d'assurer que la prédiction des concepts pour chaque image satisfait la corrélation entre les concepts. Pour ce faire, ils ont défini une mesure e_i qui reflète la cohérence du vecteur d'annotation f_i de l'image x_i avec la matrice de corrélation. e_i qui est défini par la formule suivante :

$$e_i = f_i^T \times C \times f_i \quad (6)$$

Depuis cette formule, il est clair que plus que e_i est grande, plus que l'annotation f_i de l'image x_i est plus cohérente avec la corrélation entre les concepts. Par Conséquent, ils ont généralisé cette mesure pour spécifier $E_c(F)$ comme suit :

$$E_c(F) = -tr(Y \times C \times Y^T) \quad (7)$$

Où Y est la matrice d'annotation de l'ensemble d'apprentissage, C est la matrice de corrélation entre les concepts qui est calculée dans l'étape précédente. $tr(M)$ signifie la trace d'une matrice M .

III.5.1.2 Classification multi-labels en utilisant un réseau de dépendance conditionnelle

(Multi-label classification Using Conditional Dependency Network)(Guo and Gu, 2011)
 Comme le travail présenté dans (Zha et al., 2009), ce travail s'inscrit dans la famille des méthodes de classification multi-labels, dont l'objectif commun(ultime) est d'apprendre une fonction de prédiction multi-labels, et de l'utiliser pour classifier les images de test . La contribution des auteurs était de capturer les dépendances entre les concepts au sein du processus de classification afin d'améliorer sa performance. Pour ce faire, ils ont choisi une modélisation graphique avec des calculs probabilistes, sur l'ensemble d'apprentissage. Ainsi, ils ont modélisé les dépendances entre les concepts par un réseau de dépendance conditionnelle RDC combiné avec des classifieurs binaires. Nous présentons plus de détail dans ce qui suit.

Partant d'un ensemble d'images d'apprentissage annotées $= \{x_1, x_2, \dots, x_n\}$, un ensemble d'images de test à annoter $D' = \{x^1, x^2, \dots, x^m\}$ et un ensemble de k concepts d'annotation $Y = \{y_1, y_2, \dots, y_k\}$. A chaque image x_i est associé un vecteur d'annotation (y_1^i, \dots, y_k^i) , de telle sorte qu'une valeur y_j^i de ce vecteur est égale à 1 si l'image x_i est annotée avec le concept y_j et -1 autrement.

1. Phase d'apprentissage

Dans la phase d'apprentissage, les auteurs ont modélisé les dépendances entre les concepts par un réseau de dépendance conditionnelle RDC, qui exploite l'information de co-occurrence entre les concepts. Les nœuds du réseau représentent les concepts ainsi que les vecteurs visuels des images, et les arcs représentent les dépendances directes entre les concepts. Ce réseau est un graphe bidirectionnel, cyclique, conditionnel et complètement connecté. Ainsi, chaque concept nœud est interdépendant à tous les autres concepts, et conditionné par les caractéristiques visuelles.

Le choix d'utiliser un réseau de dépendance conditionnelle et non pas d'autres modèles graphiques probabilistes comme les réseaux bayésiens (Bielza et al., 2011) ou les champs aléatoires conditionnels (conditional random fields ou CRFs en anglais) (Ghamrawi and McCallum, 2005), est justifié par le fait que l'identification de la structure optimale d'un RB, ainsi que l'apprentissage de la structure générale des champs aléatoires conditionnels, sont considérés comme des problèmes NP-difficile (Guo and Gu, 2011), alors qu'un réseau de dépendance conditionnelle permet un apprentissage et une prédiction beaucoup plus simple.

La figure.5 présente un exemple d'un RDC avec 4 concepts :

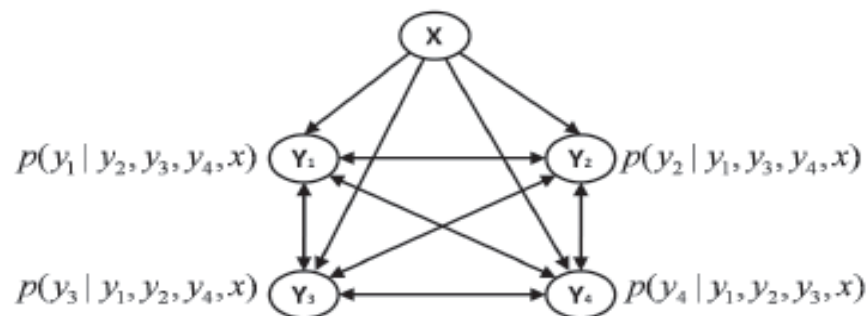


Figure 5. Exemple d'un réseau de dépendance conditionnelle (Guo and Gu, 2011)

L'apprentissage consiste à entraîner ce réseau RDC, afin de faire sortir ses paramètres notés $\{\theta_1, \dots, \dots, \theta_k\}$, où chaque paramètre est associé à un concept, et encode la distribution de probabilité conditionnelle (CPD) associée à ce concept sachant tous les autres concepts et une image x .

Pour définir les distributions de probabilité conditionnelle des concepts nœuds, les auteurs ont entraîné des classifieurs binaires probabilistes et d'autres non probabilistes. Pour la classification probabiliste, ils ont utilisé la régression logistique binaire régularisée.(Bröcker, 2010), pour entraîner K classifieurs binaires probabilistes, un pour chaque concept nœud, de telle sorte que chaque classifieur définit une distribution de probabilité conditionnelle sur le concept correspond sachant les autres concepts et le vecteur visuel d'une image d'apprentissage. Tandis que pour la classification non probabiliste, ils ont entraîné K classifieurs binaires en utilisant SVM (support vector machine). Les auteurs ont utilisé les classifieurs SVM afin de profiter de leur performance supérieure de classification, et d'étendre leur modèle afin qu'il permet une distribution de probabilité conditionnelle discrète sur les concepts (soit 1 ou 0)

Après l'entraînement du réseau de dépendance conditionnelle via l'entraînement des classifieurs binaires, le résultat obtenu est un réseau RDC avec un ensemble de paramètres noté $\{\theta_1, \dots, \dots, \theta_k\}$.

2. Phase de test

La phase de test consiste à utiliser le réseau RDC afin de prédire le vecteur d'annotation $Y = \{y_1, y_2, \dots, y_k\}$ pour une image de test x . Pour ce faire, les auteurs ont utilisé la méthode Gibbs Sampling (Neal, 1993), qui est une méthode d'inférence approximative appropriée dans les modèles graphiques(Guo and Gu, 2011). L'algorithme de cette méthode est présenté dans (Guo and Gu, 2011).

III.5.1.3 Accomplissement des tags pour la recherche d'images

(Tag completion for image retrieval)(Wu et al., 2013)

Les auteurs ont proposé une méthode semi-supervisée pour réduire le problème de silence et de bruit dans les systèmes de recherche d'images, et donc augmenter leur précision. En fait, les auteurs ont focalisé sur les annotations, et ont commencé depuis la problématique que les annotations manuelles qui sont bruitées, inconsistantes et souffrent de silence ou incomplètes (des concepts pertinents mais manquants), influencent négativement le résultat d'une recherche

d'images. Ainsi, l'amélioration de la qualité des annotations manuelles entraînera certainement l'amélioration des résultats de recherche. C'est pour cela, la solution proposée cible principalement deux objectifs, l'accomplissement des annotations manquantes et rectifications de celles bruitées. En fait, les auteurs ont formalisé leur problématique en un problème d'optimisation d'une matrice d'annotation qui représente la relation image-concept (ou tag), de telle sorte qu'elle soit consistante avec les annotations et les caractéristiques visuelles. Pour ce faire, ils ont proposé une méthode schématisée par la figure suivante :

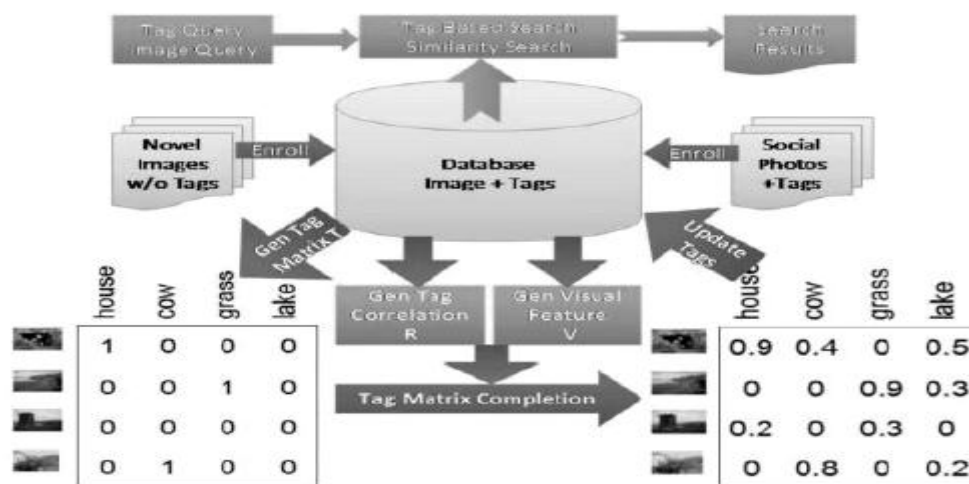


Figure 6. Accomplissement des tags pour la recherche d'images(Wu et al., 2013)

Les étapes de leur solution sont :

- Partons de n images et m concepts d'annotation, l'ensemble des annotations manuelles correspondant à la relation image-concept est représenté par une matrice \hat{T} de taille $n \times m$, dont $\hat{T}_{i,j}$ à une valeur 1 si l'image I_i est annotée avec le concept c_j , et 0 sinon. Alors l'objectif était de compléter cette matrice, et donc obtenir une nouvelle matrice $T \in R^{n \times m}$, dont $T_{i,j}$ est un nombre réel indiquant la probabilité d'assigner le concepts c_j à l'image I_i . Cette matrice est à la base de trois autres matrices : la matrice d'annotation \hat{T} , la matrice des caractéristiques visuelles et la matrice de corrélation entre les concepts. Ainsi, pour calculer la matrice T les auteurs ont procédé comme suit :
- Représenter l'ensemble des caractéristiques visuelles des images sous forme d'une matrice $\in R^{n \times d}$, où d est le nombre de caractéristiques visuelles. Chaque ligne i de cette matrice correspond au vecteur visuel de l'image I_i .

- Représenter la corrélation entre les concepts (ou l'interdépendance entre eux) par une matrice de corrélation $R \in R^{m \times m}$, où une valeur $R_{i,j}$ de cette matrice représente la valeur de corrélation entre le concept c_i et le concept c_j , et elle se calcule par la formule suivante :

$$R_{i,j} = \frac{f_{i,j}}{(f_i + f_j) - f_{i,j}} \quad (8)$$

Où f_i est le nombre d'occurrence du concept c_i dans l'ensemble des annotations des images (il est égale à la somme de la colonne i de la matrice \hat{T}), et $f_{i,j}$ est le nombre d'occurrence simultané de deux concepts c_i et c_j dans cet même ensemble.

- Considérer trois critères pour construire la matrice T :
 1. La matrice complétée T doit être similaire à la matrice \hat{T} . Pour cela, ils ont pénalisé la différence entre T et \hat{T} avec une norme de Frobenius comme suit :

$$\|T - \hat{T}\|_F^2 \quad (9)$$

Où la norme de Frobenius d'une matrice A est donné par la formule suivante :

$$\|A\|_F = (\text{tr } A^*A)^{1/2} \quad (10)$$

Où A^* désigne la matrice adjointe de A , et tr désigne la trace.

2. La matrice complétée T doit refléter le contenu visuel des images, qui est représenté par la matrice V . Pour cela, ils ont proposé de comparer les similarités visuelles des images avec les similarités entre les annotations correspondantes. Plus spécifiquement, ils ont calculé la similarité visuelle entre une image I_i représentée par son vecteur V_i et une autre image I_j représentée par son vecteur V_j par la formule suivante :

$$\text{Sim}_{visuelle}(I_i, I_j) = V_i^T \times V_j \quad (11)$$

Comme ils ont comparé ces deux images, en se basant sur leurs annotations complètes T_i et T_j depuis la matrice d'annotation complète T par la formule suivante :

$$\text{Sim}_{sim}(I_i, I_j) = T_i^T \times T_j \quad (12)$$

Alors, la matrice complète T reflétant le contenu visuel des images signifie que pour n'importe quelles deux images I_i et I_j , la valeur de $|V_i^T V_j - T_i^T T_j|^2$ est petite. Par conséquent, la norme de Frobenius soit petite pour l'ensemble de toutes les images, et elle est calculée par la formule suivante :

$$\|TT^T - VV^T\|_F^2 = \sum_{i,j=1}^n |V_i^T V_j - T_i^T T_j|^2 \quad (13)$$

3. La matrice complétée T doit être consistante avec la matrice de corrélation entre les concepts R . Ceci implique que la valeur de la norme de Frobenius suivante soit petite :

$$\|TT^T - R\|_F^2 \quad (14)$$

Donc, par la combinaison de trois critères précédents, ils ont obtenu le problème d'optimisation suivant afin de trouver la matrice d'annotation complète T :

$$\min_{T \in \mathbb{R}^{n \times m}} \|TT^T - VV^T\|_F^2 + \gamma \|TT^T - R\|_F^2 + \mu \|T - \hat{T}\|_F^2 \quad (15)$$

De plus, les auteurs ont amélioré cette dernière formule afin de prendre en considération les différents poids des caractéristiques visuelles, ainsi que pour régler la densité de la matrice T et la rendre creuse.

Après la complétion de la matrice d'annotation \hat{T} et l'obtention de la matrice T , les auteurs l'ont appliqué pour la recherche d'images. Ainsi, pour répondre à une requête atomique composée d'un seul concept c_j , une image est considérée comme pertinente si elle contient ce concept requête dans son annotation (comparaison binaire). Ainsi, ils ont trié les images par leur ordre décroissant des valeurs de pertinence, et qui correspondent à la colonne j de la matrice T . Pour une requête composée de plusieurs concepts $=(q_1, \dots, q_m) \in \{0, 1\}$, les valeurs de pertinence des images sont obtenues par le produit : TWq , où W est la matrice de corrélation entre les concepts. Elle est estimée en se basant sur la matrice d'annotation complète T .

$$W = \pi_{[0,1]}(T^T T) \quad (16)$$

Où $\pi_{[0,1]}$ projette chaque entrée de A en une valeur entre 0 et 1.

III.5.1.4 Un modèle de langage de similarité sémantique pour améliorer l'annotation automatique d'images

A semantic similarity language model to improve automatic image annotation) (Gong et al., 2010)

Les auteurs ont présenté une nouvelle méthode d'annotation qui prend en considération la corrélation entre les concepts, en utilisant un modèle probabiliste. En effet, ils ont proposé un

modèle de langage de similarité sémantique où les annotations qui sont sémantiquement plus cohérentes, ont une probabilité plus élevée d'être choisies pour annoter une image. Dans ce modèle de langage, les auteurs ont présenté chaque concept d'annotation par un vecteur sémantique qui présente la distribution du concept en termes de la force de co-occurrence avec les autres concepts dans une même annotation. Ainsi, le système génère des annotations sémantiquement cohérentes en utilisant le modèle de langage de similarité sémantique proposé avec le modèle probabiliste de corrélation image-concept pour estimer la probabilité antérieure d'un ensemble de mots au lieu de concepts individuels

Les auteurs ont utilisé l'information de co-occurrence entre les concepts dans les annotations de l'ensemble d'apprentissage pour modéliser la probabilité d'un concept donné sachant d'autres concepts dans les annotations et la probabilité de l'annotation. Ainsi, un concept w est représenté par un vecteur sémantique $V_w = \langle \langle v_1, v_2, \dots, v_i, \dots, v_m \rangle \rangle$, où m est le nombre de concepts. Chaque valeur v_i de ce vecteur est calculée en utilisant la probabilité conditionnelle comme suit :

$$v_i = \frac{p(\text{context}_i | w)}{p(\text{context}_i)} \quad (17)$$

Où $p(\text{context}_i)$ est le nombre d'occurrence du concept context_i sur le nombre total d'occurrence de tous les concepts, et la probabilité conditionnelle $p(\text{context}_i | w)$ est bien la fréquence relative du nombre de co-occurrence des deux concepts context_i et w dans toutes les annotations sur le nombre total d'occurrence de w dans les annotations :

$$p(\text{context}_i | w) = \frac{\text{count}(\text{context}_i, w)}{\text{count}(w)} \quad (18)$$

Alors la probabilité d'un ensemble de concepts d'annotations $A = \{w_1, \dots, w_n\}$ est mesurée par la similarité entre chacun de ces concepts et le restant des autres concepts. Elle est définie par la formule suivante :

$$p(A) \propto \frac{1}{n(n-1)} \sum_{w_i \in A} \sum_{w_j \in A, j \neq i} \text{sim}(w_i, w_j) \quad (19)$$

Où chaque concept w_i depuis A est représenté par son vecteur sémantique. La similarité $\text{sim}(w_1, w_2)$ est la similarité cosinus de Salton, elle est définie comme suit :

$$\text{sim}(w_1, w_2) = \frac{V_{w_1} \times V_{w_2}}{\|V_{w_1}\| \times \|V_{w_2}\|} \quad (20)$$

Où le produit de deux vecteurs V_{w_1} et V_{w_2} est calculé comme suit :

$$V_{w_1} \times V_{w_2} = \sum_{i=1}^m v_{w_1,i} \times v_{w_2,i} = \sum_{i=1}^m \frac{p(c_i|w_1)}{p(c_i)} \times \frac{p(c_i|w_2)}{p(c_i)} \quad (21)$$

III.5.1.5 Recherche d'images sociales basée-tag: vers des résultats pertinentes et diverses

Tag-Based Social Image Search: Toward Relevant and Diverse Results(Yang et al., 2011)

Dans ce travail (Yang et al. 2011), les auteurs ont présenté une méthode de recherche d'images qui permet de localiser des images qui doivent être non seulement pertinentes à la requête utilisateur mais aussi elles soient diverses. 'Des images diverses' signifient des images avec des apparences visuelles variées. Par exemple, des images d'une fête peuvent avoir différentes apparences visuelles. Ainsi, après qu'un utilisateur introduit un concept requête, leur méthode de recherche des images pertinentes et diverses se déroule en deux grandes étapes :

- L'estimation de la valeur de pertinence d'une image par rapport à ce concept requête, en fonction de ses caractéristiques visuelles ainsi que la proximité sémantique entre ses concepts d'annotation (c.à.d.ses tags) et le concept requête.
- Le tri des images en fonction de leurs valeurs de pertinence estimées, et en optimisant une nouvelle métrique de performance qui prend en compte la diversité des images. Cette métrique est une généralisation (ou extension) de la métrique conventionnelle : la précision moyenne (average Precision AP), ils l'ont appelé la précision diverse moyenne (Average Diverse Precisoins ADP).

Ce qui nous intéresse dans ce travail, ce n'est pas la diversité, mais plutôt de savoir comment les auteurs ont calculé la proximité sémantique entre les concepts et comment ils l'ont exploité pour calculer la pertinence d'une image.

1. Le calcul de la proximité sémantique entre les concepts

Pour calculer la proximité sémantique entre les concepts, les auteurs ont utilisé une formule inspirée de la distance de Google(Cilibrasi and Vitanyi, 2007). En fait, la distance de Google mesure la proximité entre deux concepts en se basant sur leur occurrence dans les pages web, tandis que, leur mesure de similarité de deux concepts c_i et c_j est basée sur leur occurrence dans les annotations des images, et elle est définie par l'équation suivante :

$$sim(c_i, c_j) = \exp\left(-\frac{\max(\log x(c_i), \log x(c_j)) - \log(x(c_i, c_j))}{\log M - \min(\log x(c_i), \log x(c_j))}\right) \quad (22)$$

Où $x(c_i)$ et $x(c_j)$ correspondent aux nombre d'images Flickr annotées avec le concept c_i et c_j respectivement, $x(c_i, c_j)$ est le nombre d'images annotées avec les deux concepts c_i et c_j simultanément, et M correspond au nombre total des images dans la base Flickr.

Le résultat de calcul de proximité sémantique est présenté sous forme d'une matrice dite matrice de similarité des concepts.

2. Le calcul de pertinence des images

Ils ont calculé la pertinence d'une image x_i à un concept requête c_q en fonction de la proximité sémantique entre ses concepts d'annotation (tags) et ce concept requête, et elle est égale à la moyenne des valeurs de proximité sémantique entre ses concepts et le concept requête. Elle est calculée selon la formule suivante :

$$sim(c_q, Annot_{x_i}) = \frac{1}{|Annot_{x_i}|} \sum_{c_k \in Annot_{x_i}} sim(c_q, c_k) \quad (23)$$

Où $Annot_{x_i}$ c'est l'ensemble de concepts d'annotation de l'image x_i .

III.5.1.6 Similarité sémantique des images basée-contexte

(Context-based Image Semantic Similarity CISS)(Franzoni et al., 2015)

Les auteurs ont développé un système de recherche d'images qui mesure la similarité entre les images par la comparaison sémantique entre leurs tags ou concepts d'annotation. En fait, ils ont choisi 520 images Flickr, de telle sorte que les tags associés par l'auteur de chacune de ces images constituent son annotation (son groupe de concept) qui la caractérise. Ensuite, pour comparer les annotations $(Ti_1, Ti_2, \dots, Ti_m)$ et $(Tj_1, Tj_2, \dots, Tj_n)$ de deux images I_i et I_j , ils ont défini une mesure de similarité de groupe DI_{ij} appelée 'Context-based Image Semantic Similarity (CISS)', et qui est donnée par la formule suivante :

$$DI_{ij} = AVG2\{AVG1[SEL(dT_{im \rightarrow jn})], AVG1[SEL(dT_{jn \rightarrow im})]\} \quad (24)$$

Où d est une mesure élémentaire entre deux concepts, et qui est l'une des mesures sémantiques existantes dans la littérature, telle que : la Confiance mutuelle ponctuelle PMI (en anglais c'est

Pointwise Mutual Confidence)(Lin, 1998; Turney, 2001), La distance de Google normalisée NGD(en anglais c'est Normalized Google Distance)(Enser et al., 2005; Franzoni and Milani, 2012; Pallottelli et al., 2015)ou bien la confiance mutuelle CM (en anglais c'est Mutual Confidence)(Franzoni et al., 2015)

$$dT_{im} \rightarrow dT_{jn} = \begin{pmatrix} dT_{i1 \rightarrow j1}, & dT_{i1 \rightarrow j2}, & dT_{i1 \rightarrow j3}, & \dots & dT_{i1 \rightarrow jn} \\ dT_{i2 \rightarrow j1}, & dT_{i2 \rightarrow j2}, & dT_{i2 \rightarrow j3}, & \dots & dT_{i2 \rightarrow jn} \\ \dots, & \dots, & \dots, & \dots & \dots \\ dT_{in \rightarrow j1}, & dT_{in \rightarrow j2}, & dT_{in \rightarrow j3}, & \dots & dT_{in \rightarrow jn} \end{pmatrix}$$

Les auteurs ont essayé 9 combinaisons différentes des mesures sémantiques par variation de $SEL \in \{MAX, AVG, MIN\}$ et $d \in \{PMI, NGD, CM\}$. La mesure la plus précise était la mesure CISS avec MAX-CM. CM (Mutual Confidence) est une mesure symétrique ($DI_{ij} = DI_{ji}$) fondée sur la mesure de confiance (confidence measurement en anglais). CM est la moyenne de la mesure de confiance dans les deux sens $x \rightarrow y$ et $(y \rightarrow x)$, et elle est définie par la formule suivante :

$$CM = \frac{\text{confidence}(x \rightarrow y) + \text{confidence}(y \rightarrow x)}{2} \quad (25)$$

Où $\text{confidence}(x \rightarrow y)$ permet de mesurer la probabilité conditionnelle qu'un concept x provoque un autre concept, et elle est définie comme suit :

$$\text{confidence}(x \rightarrow y) = \frac{P(x \wedge y)}{P(x)} = P(x|y) \quad (26)$$

Par le remplacement de $SEL = MAX$ et $d = CM$, la formule (15) est réécrite comme suit :

$$DI_{ij} = AVG2\{AVG1[MAX(CM(T_{im \rightarrow jn}))], AVG1[MAX(CM(T_{jn \rightarrow im}))]\} \quad (27)$$

Où

$$\begin{aligned} & AVG1[MAX(CM(T_{im \rightarrow jn}))] \\ & = avg[MAX(CM(T_{i1 \rightarrow jn}), MAX(CM(T_{i2 \rightarrow jn}), \dots, MAX(CM(T_{im \rightarrow jn}))) \quad (28) \end{aligned}$$

$$\begin{aligned}
 &AVG1[MAX(CM(T_{jn \rightarrow im}))] \\
 &= avg[MAX(CM(T_{j1 \rightarrow im}), MAX(CM(T_{j2 \rightarrow im}), \dots, MAX(CM(T_{jn \rightarrow im})))] \quad (29)
 \end{aligned}$$

Où $MAX(CM(T_{i1 \rightarrow jn}))$ est la valeur de confiance mutuelle CM maximale entre le concept T_{i1} de l'image I_i et tous les autres concepts de l'image I_j . De même pour $MAX(CM(T_{j1 \rightarrow im}))$

III.5.2 Méthodes sémantiques basées-corpus global

Les méthodes de cette famille ont choisi d'utiliser une source de connaissance externe indépendante de toute base d'images locales, et qui est partagée par la communauté humaine à travers le monde, c'est le cas de Wikipedia. En fait, l'ensemble de ces méthodes sont des méthodes topologiques utilisant certaines propriétés de Wikipedia, telles que les titres des sections des pages, les liens de redirection, sa structure hiérarchique, et ceci pour construire automatiquement des modèles sémantiques comme les thésaurus et les ontologies contenant des concepts qui sont liés par des relations sémantiques.

A nos connaissances, toutes les méthodes de la littérature utilisant Wikipedia comme une source de connaissance pour construire des thésaurus ou des ontologies, ont focalisé sur la recherche d'images comme domaine d'application de leurs modèles sémantiques. Dans ce qui suit, nous présentons quelques méthodes de la littérature sur la recherche d'images- basées ontologie ou thésaurus construits automatiquement depuis Wikipedia. La dernière méthode(Maree et al., 2016) que nous allons présenter, a exploité directement deux ontologies déjà existantes pour indexer des pages web multimédia et, par conséquent, les rechercher d'une façon plus performante. Ces deux ontologies sont WordNet et YAGO, dont YAGO est construite à partir de Wikipedia.

III.5.2.1 Recherche d'images sur le Web améliorée par une ontologie : aidée par Wikipedia et la théorie de l'activation de propagation

(Ontology Enhanced Web Image Retrieval: Aided by Wikipedia & Spreading Activation Theory)(Wang et al., 2008)

Les auteurs ont développé un système 'OntoEnhanced' de recherche d'images à base d'ontologie. Ce système contient deux parties, une partie offline de construction d'une ontologie,

et une partie online de recherche d'images. Le schéma de ce système est présenté par la figure suivante :

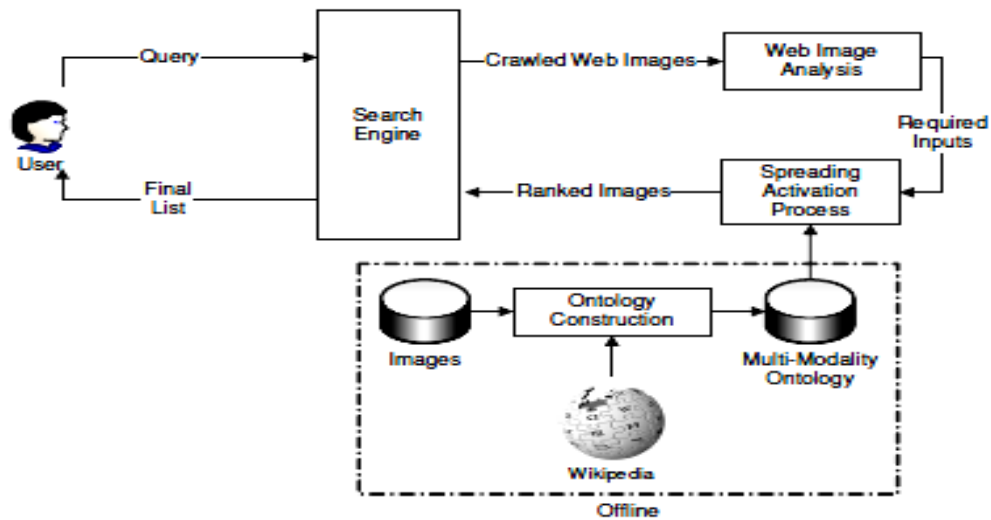


Figure 7. Système pour la recherche d'images à base d'ontologie (Wang et al., 2008)

1. Partie offline de construction d'une ontologie : Ils ont construit d'une façon automatique une ontologie multimodale qui modélise des concepts d'un point de vue sémantique et visuel. En fait, le processus de construction est divisé en deux phases séparées, à savoir : la construction des concepts et des relations sémantiques entre eux en utilisant Wikipedia comme une source de connaissance, et construction des modèles visuels des concepts à partir des vecteurs visuels des images.

- **Extraction des concepts et relations sémantiques depuis Wikipedia :** Les auteurs ont utilisé Wikipedia pour construire une ontologie qui comprend des domaines variés. Cette ontologie contient trois types de relations, à savoir : *taxinomie* (*est-un* ou *partie-de*), *Synonymie* et *concept-liés*. Ainsi, pour construire la taxinomie des concepts, ils ont utilisé la structure hiérarchique des catégories Wikipedia dont les liens entre les concepts expriment la relation *est-un* ou bien la relation *partie-de*. Pour trouver des concepts synonymes à un concept requête, ils ont exploité les liens de redirection existants dans Wikipedia, et qui permettent de diriger d'une façon automatique un concept requête vers ses synonymes. Ils ont extrait les concepts liés à un concept requête (qui sont des concepts possédant des hyperliens vers leurs propres pages de définition Wikipedia) depuis les paragraphes de l'article qui lui correspond. Ces concepts sont ainsi connectés avec le concept requête par des relations qui sont explicitement définies dans les titres des sections. Par exemple : concepts 'Canada', 'Alaska' sont extraits depuis la section

'population and distribution' et ils sont liés avec le concept requête 'Arcticfox' par la relation *has-distribution*.

- **Construction des concepts visuels à partir des images :** les auteurs ont utilisé les techniques de reconnaissance pour construire un vocabulaire visuel et entraîner des classifieurs en utilisant une machine à support de vecteur(SVM). En fait , ils ont utilisé le détecteur Harris-Laplace(Mikolajczyk and Schmid, 2004) pour détecter les points d'intérêts dans une image, le descripteur SIFT(Lowe, 2004) pour représenter l'information de forme autour des points d'intérêts et le descripteur 'opponent angle' pour décrire l'information de couleur autour des points d'intérêts. Par conséquent, un vecteur visuel final d'une image était de dimension 164 (128 pour SIFT et 36 pour opponent angle).Ensuite, ils ont construit 1000 mots visuels résultat d'un clustering K-means des vecteurs visuels de toutes les images. Par conséquent, chacune des images est représentée par un vecteur de mots visuels de dimension égale à 1000.

Finalement, les concepts et leurs relations extraits dans les deux phases sont combinés pour obtenir une ontologie multimodale.

2. Partie online de recherche d'images : Pour rechercher des images qui répondent à une requête utilisateur, introduite sous forme d'un ou de plusieurs concepts sélectionnés depuis l'ontologie, les auteurs ont procédé comme suit :

- Considérer leur ontologie comme un réseau sémantique, dont les nœuds sont les concepts de l'ontologie, et les arcs représentant les relations entre les concepts.

- Pour calculer les valeurs de pertinence des images d'une base, ils ont lancé un processus d'inférence sur ce réseau sémantique pour chacune de ces images en utilisant la méthode d'inférence 'activation de la propagation' (en anglais c'est Spreading Activation Theory SAT)(Anderson, 1983).

En fait, pour calculer la valeur de pertinence d'une image par rapport à un concept requête c_q , le processus d'activation de la propagation est formulé comme suit :

$$O = [\varepsilon - (1 - \alpha)w^T]^{-1} I \quad (30)$$

$O = [O_1, O_2, \dots, O_q, \dots, O_n]$: est le vecteur de l'image résultat du processus d'activation de propagation, où chaque élément de ce vecteur O_q représente la valeur de pertinence de l'image par rapport au concept requête c_q . α est un facteur représentant la perte d'énergie dans le

processus d'activation de la propagation. ε est une matrice d'identité d'ordre n , et de taille $n \times n$, dont n est le nombre de concepts de l'ontologie (nombre de nœuds du réseau). $I = [I_1, I_2, \dots, I_n]$: est le vecteur initial de l'image, utilisé comme entrée au processus d'activation de la propagation. Autrement dit, ses valeurs sont les valeurs d'activation initiales des nœuds du réseau. Une valeur I_j de ce vecteur est calculée par la formule suivante :

$$I_j = \frac{freq(c_j)}{\sum_{i=1}^n freq(c_i)} \quad (31)$$

Où $freq(c_j)$ est la fréquence d'apparition du concept c_j dans le texte qui entoure l'image.

W est la matrice de l'ontologie. Autrement dit, elle contient les valeurs des arcs du réseau sémantique. Un élément de cette matrice $w_{i,j}$ représente la valeur de la relation entre les deux concepts c_i et c_j , et il est calculé par la formule suivante :

$$w_{i,j} = \frac{freq(r_{i,j})}{\sum_{all j} freq(r_{i,j})} \quad (32)$$

Où $freq(r_{i,j})$ est la fréquence d'apparition de la relation $r_{i,j}$ dans un ensemble de données.

Le processus d'activation de la propagation permet de propager la valeur d'activation de chacun des nœuds vers ses nœuds voisins. Après stabilisation du réseau sémantique, les valeurs $[O_1, O_2, \dots, O_q, \dots, O_n]$ du vecteur O sont ainsi obtenues, dont la valeur O_q représente la valeur de pertinence de l'image par rapport au concept requête c_q .

Après le lancement de processus d'activation pour chacune des images de la base, les images sont retournées à l'utilisateur par leur ordre décroissant de pertinence (c.à.d. ordre décroissant des valeurs O_q) par rapport au concept requête c_q .

III.5.2.2 Thésaurus de concepts assisté-Wikipedia pour une meilleure compréhension du Media Web

Wikipedia-Assisted Concept thesaurus for Better Web Media understanding (Wang et al., 2010)

Les auteurs ont présenté une méthode de construction automatique d'un thésaurus de concepts depuis Wikipedia, et qui encode quatre relations sémantiques, à savoir : *Synonymie*, *polysémie*, *concept-parent/catégorie*, *concept-associé* entre ces concepts, en exploitant les

propriétés internes (contenu des pages) et externes (structure hiérarchique des catégories) de Wikipedia. Ce thésaurus est appliqué pour la recherche d'images sur le Web. Les étapes de leur travail sont comme suit :

1. Construction du thésaurus : la construction du thésaurus consiste à extraire les concepts et les relations sémantiques entre eux à partir de Wikipedia.

- **Extraction des concepts :** les auteurs ont utilisé la décharge Wikipedia (la version anglaise), de telle sorte que chaque article correspond à un concept, où le titre constitue le nom du concept. Ceci (utilisation de la décharge au complet) offre l'avantage que le thésaurus ne soit pas limité à un domaine spécifique. Ils ont obtenu 5.836.166 concepts.

- **Détection des relations :** ils ont considéré les quatre types de relations sémantiques, à savoir : *Synonymie*, *concept-parent/catégorie*, *concept-liés* et *polysémie*. Pour les trois premiers types, ils ont procédé de la même façon que dans (Wang et al., 2008). Pour trouver des polysémies (un concept avec différentes significations dans différents contextes), ils ont exploité la page de désambiguïsation, qui est fournie par Wikipedia comme une référence supplémentaire.

Après la construction du thésaurus, un utilisateur peut l'interroger via un concept requête, afin d'avoir ses synonymes, ses polysémies, son père, ses fils et ses concepts associés (liés).

2. Application du thésaurus pour la recherche d'images sur le Web : Pour construire une base expérimentale, les auteurs ont téléchargé 13.856 images et leurs web pages associées depuis le moteur de recherche Yahoo ! image.

En introduisant un concept requête, les auteurs ont calculé la distance sémantique entre cette requête et les images de la base (représentées par leurs pages web associées). Pour ce faire, ils ont procédé comme suit :

- Représenter chaque page web associée à une image de la base par son vecteur TF_IDF (Term Frequency- Inverse Document Frequency), qui représente les poids des concepts dans cette page. Ainsi, pour l'ensemble de n pages web, ils ont obtenu une matrice CF des poids TF-IDF

$$CF = \begin{bmatrix} cf_{1,1} & cf_{1,2} & \dots & cf_{1,d-1} & cf_{1,d} \\ cf_{2,1} & cf_{2,2} & \dots & cf_{2,d-1} & cf_{2,d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ cf_{n-1,1} & cf_{n-1,2} & \dots & cf_{n-1,d-1} & cf_{n-1,d} \\ cf_{n,1} & cf_{n,2} & \dots & cf_{n,d-1} & cf_{n,d} \end{bmatrix}$$

Où $cf_{i,j} = Tf_IDF$ du concept c_j dans la page web dc_i . Cette valeur est calculée par la formule suivante :

$$TF_IDF(c_j, dc_i) = TF(c_j, dc_i) \times IDF(c_j) \quad (33)$$

$$TF(c_j, dc_i) = \frac{\text{Frèquence_de_}c_j\text{_dans la page_}dc_i}{\text{Total_des_Frequences_des_concepts_dans_}dc_i} \quad (34)$$

Où $TF(c_j, dc_i)$ signifie la fréquence du terme (en anglais Term Frequency). Elle mesure la fréquence d'apparition normalisée du concept c_j dans la page Web dc_i .

$$IDF(c_j) = \log\left(\frac{n}{|d_k : c_j \in d_k|}\right) \quad (35)$$

Où $IDF(c_j)$ signifie la fréquence inverse du document (en anglais Inverse Document Frequency). n est le nombre total des pages, $|d_k : c_j \in d_k|$ est le nombre des pages où le concept c_j apparaît.

- Transmettre le concept requête au thesaurus et récupérer un ensemble T contenant ses concepts synonymes, ses polysémies, son père, ses fils et ses concepts liés. Ensuite, à partir de T construire un vecteur thesaurus $V = [v_1, v_2, \dots, v_n]^T$ de taille d (c.à.d. le nombre de concepts dans le thesaurus), de telle sorte qu'une entrée v_i a une valeur égale à 1 si le concept i est présent dans T , et 0 autrement.
- Construire une matrice diagonale W de taille $d \times d$. Elle encode les poids des concepts en fonction de leurs relations avec la requête. Par exemple, un concept lié avec la relation synonyme reçoit un poids le plus élevé, et c'est le contraire pour un concept lié.
- Construire une matrice S de saillance sémantique $S = W \times V$
- Transformer la matrice des poids TF_IDF en une nouvelle matrice $F = CF \times S$. Ainsi, à partir de cette matrice, les pages web sont triées, et par conséquent, leurs images correspondantes sont affichées par ordre décroissant de pertinence.

III.5.2.3 Indexation basée-ontologies multiples des documents multimédias dans Internet

Multiple Ontology-Based Indexing of Multimedia Documents on the WorldWide Web (Maree et al., 2016)

Les auteurs ont choisi d'utiliser deux ontologies déjà existantes pour l'indexation sémantique des pages web multimédia (c.à.d. des pages contenant en plus du texte des images, des audios et des vidéos). Ces ontologies sont WordNet (Miller, 1995) et YAGO (Fabian et al.,

2007), dont YAGO est construite à partir de Wikipedia. En fait, leur système d'indexation sémantique est complètement automatique, et l'index résultat est un réseau sémantique dont les nœuds sont des termes et les arrêtes sont les liens sémantiques entre les termes. Pour construire ce réseau sémantique, ils ont procédé comme suit :

- Sélectionner d'une façon manuelle 300 pages web, où chaque lot de 100 pages correspondent à un type de document multimédia (images, audio, vidéo). Les pages sélectionnées sont celles qui respectent une certaine structure traitable correctement par leur algorithme de segmentation de pages web 'Document Object Model Tree' (DOM)(Fauzi et al., 2009).

- Segmenter ces 300 pages en utilisant l'algorithme DOM Tree qu'ils ont proposé dans (Fauzi et al., 2009). Cet Algorithme permet d'extraire le texte qui entoure les images, l'audio et les vidéos.

- Raffiner les segments de texte extraits par l'élimination des mots d'arrêts, l'extraction des unités lexicales(en anglais c'est n-gram text tokenization) et le marquage des parties de la parole (en anglais c'est part-of-speech(Pos) tagging). L'algorithme d'extraction des unités lexicales à partir des segments de texte extrait des unités lexicales n-gram de longueur de 1 à 4 mots. Par exemple '*Java Idonesian*' est une unité lexicale de longueur 2. Le tagger POS est utilisé pour classer les unités lexicales dans les catégories lexicales à qui elles correspondent.

- Construire d'une façon incrémentale des réseaux sémantiques en se basant sur les deux ontologies WordNet et YAGO (une ontologie peut produire un ou plusieurs réseaux sémantiques). Pour ce faire, ils ont transmis chaque unité lexicale n-gramme à chacune des ontologies pour savoir si elle y existe ou non. Par conséquent, les unités lexicales reconnues sont représentées comme des nœuds dans les réseaux sémantiques. De plus, s'il existe une relation sémantique entre deux unités lexicales dans une ontologie, alors les nœuds correspondants sont liés par une arrête dans ces réseaux sémantiques.

- Combiner tous les réseaux sémantiques obtenus en un seul réseau cohérent, en utilisant leur algorithme de fusion, présenté dans(Maree and Belkhatir, 2010). Ainsi, ce réseau constitue l'index sémantique résultat du système proposé.

- Pour les unités lexicales non reconnues par aucune des deux ontologies, les auteurs ont utilisé un algorithme d'enrichissement du réseau sémantique par ces unités lexicales, dont le principe général est le suivant

- o mesurer pour chacune des unités lexicales non reconnues sa proximité sémantique avec les nœuds du réseau sémantique résultat d'indexation. Pour ce faire, ils ont utilisé la

Distance de Recherche Normalisée (en anglais c'est Normalized Retrieval Distance NRD), qu'ils ont proposé dans (Maree and Belkhatir, 2015). Cette distance est une forme adaptée de la distance de Google normalisée (Cilibrasi and Vitanyi, 2007). Elle mesure une probabilité conditionnelle symétrique de cooccurrence entre des entités. Ainsi, la distance NRD pour deux entités c_i et c_j est donnée par la formule suivante :

$$NRD(c_i, c_j) = \exp\left(-\frac{\max(\log f(c_i), \log f(c_j)) \log(f(c_i, c_j))}{\log M - \min(\log f(c_i), \log f(c_j))}\right) \quad (36)$$

Où c_i est une unité lexicale non reconnue, c_j est une unité lexicale existante dans le réseau sémantique (et dans les ontologies), $f(c_i)$ et $f(c_j)$ sont le nombre d'occurrence de c_i et c_j respectivement dans les pages web indexées. $f(c_i, c_j)$ est le nombre d'occurrence de c_i et c_j simultanément dans les pages web. M est le nombre total des pages web indexées (dans leur cas c'est 300). Une distance égale à zéro signifie que les deux entités apparaîtraient toujours ensemble.

- Sur la base des valeurs de proximité sémantiques NRD entre les unités lexicales non reconnues et les nœuds du réseau sémantique résultat d'indexation, ces nouvelles unités sont ainsi ajoutées au réseau sémantique (processus d'enrichissement du réseau) de telle sorte qu'elles soient liées avec les nœuds les plus proches (valeur de proximité sémantique maximale ou valeur de NRD minimale).

Après la construction de l'index sémantique de 300 pages, les auteurs ont évalué la précision de leur système d'indexation sémantique sur des bases d'images, des bases audio et des bases de vidéos respectivement. En fait, ils ont introduit 6 requêtes par type multimédia, et ils ont comparé le résultat obtenu avec une vérité terrain qui s'agissait de valeurs de pertinences assignées d'une façon manuelle.

III.6 Limites des méthodes sémantiques

Les deux familles des méthodes sémantiques (basées-corpus local et basées-corpus global) possèdent un avantage majeur par rapport aux méthodes visuelles, c'est qu'elles considèrent en plus de la relation image-concepts, la relation concept-concept afin d'annoter ou de rechercher des images. Cette dernière relation est très importante du fait qu'en réalité les concepts ne se trouvent pas séparément, mais il y a une certaine sémantique qui les relie.

Les méthodes sémantiques utilisant un ensemble d'images d'apprentissage locales pour calculer l'interdépendance entre les concepts, ont obtenu des performances compétitives. Néanmoins, leur inconvénient majeur réside dans cet ensemble. En fait, comme les approches visuelles, les annotations de l'ensemble d'apprentissage peuvent être incomplètes, bruitées et ambiguës. Par conséquent, cela peut influencer la précision du calcul de l'interdépendance entre les concepts. De plus, le calcul de l'interdépendance entre les concepts en se basant sur leur cooccurrence au sein d'une base d'images locale peut être inapproprié, car les statistiques sur l'occurrence des concepts (Xu et al., 2016) ont montré que la distribution des concepts est souvent très déséquilibrée. Par conséquent, les résultats de calcul de l'interdépendance entre les concepts peuvent être loin de jugement humain et donc ne peuvent pas refléter des cas réels.

Dans un autre côté, les méthodes sémantiques basées-corpus global ont focalisé sur la recherche d'images, et ont fourni des efforts considérables qui ont mené à des systèmes de recherche d'images plus performants qu'un système TBIR traditionnel. En fait, ils ont construit des modèles sémantiques (c.à.d. ontologie et thésaurus) permettant de modéliser des concepts et des relations sémantiques entre eux. Ensuite, ils ont exploité ces modèles pour rechercher des images. Cependant, leur inconvénient majeur réside dans les modèles sémantiques construits. En fait, les modèles sémantiques n'ont modélisé que quelques relations sémantiques. Ce sont les relations qui peuvent être extraites automatiquement à partir du corpus externe (dont Wikipedia est le corpus le plus utilisé), telles que les relations de *Synonymie*, *polysémie*, *est-un*, *partie-de*, *concept-lié*. Par conséquent, ces modèles ne reflètent pas la richesse en relations sémantiques qui existent en réalité. Ceci, peut conduire à une estimation biaisée de la mesure de similarité sémantique des images.

Nous avons exposé un état de l'art des différentes méthodes permettant de minimiser le silence et augmenter la précision dans un système de recherche d'images. Nous avons remarqué qu'il existe des méthodes visuelles exploitant la relation caractéristique visuelle-concept, ou bien des méthodes qui calculent en plus l'interdépendance entre les concepts. En fait, nous devons mentionner que ce dernier calcul constitue un réel défi pour la communauté scientifique spécialisée, et toute une panoplie de travaux ont focalisé sur la manière de mesurer la sémantique entre les concepts (Gabrilovich and Markovitch, 2007), (Hassan and Mihalcea, 2011; Jabeen et al., 2012), (Ni et al., 2016; Taieb et al., 2014; Pakhomov et al., 2010; Aouicha and Taieb, 2016) indépendamment du domaine d'application. Pour cela, nous consacrons la section suivante pour présenter quelques-uns de ces travaux.

III.7 Etat de l'art des méthodes de calcul de proximité sémantique entre les concepts

Il existe une panoplie des travaux qui ont focalisé sur le calcul de l'interdépendance entre les concepts (Gabrilovich and Markovitch, 2007), (Hassan and Mihalcea, 2011; Jabeen et al., 2012; Medina et al., 2012), (Ni et al., 2016; Taieb et al., 2014; Pakhomov et al., 2010; Aouicha and Taieb, 2016) indépendamment du domaine d'application. En fait, l'ensemble des méthodes proposées (à nos connaissances) sont des méthodes basées-Wikipedia. Autrement dit, ils ont choisi d'utiliser les propriétés de Wikipedia comme une source d'information sémantique, du fait qu'il s'agit d'une source de connaissance encyclopédique hautement organisée, indépendante de toute base d'images et qui est partagée par la communauté humaine à travers le monde. Nous pouvons ainsi classer l'ensemble de ces méthodes, en fonction des propriétés du Wikipedia exploitées, en deux familles : des méthodes topologiques et des méthodes statistiques. Les méthodes topologiques exploitent des propriétés externes de Wikipedia comme la structure hiérarchique des catégories Wikipedia, et les liens Wikipedia. Tandis que les méthodes statistiques exploitent les propriétés internes des articles Wikipedia. C'est-à-dire qu'elles font des statistiques sur le texte contenu dans les articles Wikipedia. Dans ce qui suit, nous présentons quelques méthodes de chaque classe. Néanmoins, il faut signaler que notre méthode de calcul de proximité sémantique entre les concepts est une méthode statistique.

III.7.1 Méthodes topologiques

Les méthodes topologiques exploitent des propriétés externes de Wikipedia qui peuvent être déduites ou être inférées à partir des articles. Comme instance, il y a la structure hiérarchique des catégories Wikipedia (Strube and Ponzetto, 2006; Taieb et al., 2013; Medina et al., 2012), et la structure d'hyperlien qui lie les articles Wikipedia entre eux (Yeh et al., 2009; Witten and Milne, 2008b).

Les méthodes utilisant la structure hiérarchique des catégories Wikipedia partent du principe que tous les articles Wikipedia doivent avoir au moins une catégorie commune, et donc sont considérés comme liés. Ainsi, la proximité sémantique entre n'importe lequel de deux articles (concepts) de la hiérarchie des catégories, s'obtient comme suit :

- Construction de la hiérarchie des catégories entre ces deux articles. Elle est sous la forme d'un graphe qui est parcouru à partir de ces deux articles vers le haut jusqu'à trouver un article parent commun.

- La proximité sémantique entre n'importe lequel de deux articles dans cette hiérarchie, est égale à la distance entre les deux articles, calculée en fonction du nombre d'arêtes ou bien du nombre de nœuds.

Pour les méthodes utilisant la structure d'hyperlien entre les articles Wikipedia, le principe général de calcul de proximité sémantique entre deux concepts (représentés par leurs articles) consiste à exploiter le plus court chemin entre ses articles (par exemple la somme des poids des arêtes).

Dans ce qui suit nous présentons quelques travaux de ces deux familles.

III.7.1.1 Wikirelate! Calcul de proximité sémantique en utilisant Wikipedia

Wikirelate! Computing semantic relatedness using Wikipedia (Strube and Ponzetto, 2006)

Selon (Franzoni et al., 2015), Wikirelate est le premier modèle dans la littérature de calcul de proximité sémantique utilisant Wikipedia. Ce travail utilise la structure hiérarchique des catégories Wikipedia pour calculer la proximité sémantique entre deux mots. Pour le faire, les auteurs ont adopté la mesure présentée dans (Leacock and Chodorow, 1998) qui définit la proximité sémantique entre deux mots comme le nombre de nœuds du plus court chemin entre les nœuds correspondant dans la taxonomie, divisé par la profondeur de la taxonomie dans laquelle se trouvent les concepts, comme moyen de normalisation. La formule de la proximité sémantique entre deux mots c_1 et c_2 adoptée est la suivante :

$$sem(c_1, c_2) = -\log \frac{length(c_1, c_2)}{2D} \quad (37)$$

Où $length(c_1, c_2)$ est le nombre de nœuds du plus court chemin entre les deux nœuds, et D est la profondeur maximale de la taxonomie.

III.7.1.2 Mesurer la proximité sémantique des entités en utilisant Wikipedia

Measuring Entity Semantic Relatedness using Wikipedia (Medina et al., 2012)

Les auteurs ont proposé une mesure de proximité sémantique entre des concepts scientifiques, en utilisant la taxonomie (la structure hiérarchique) des catégories Wikipedia. Cette

mesure est en fonction de la somme des poids des arrêtes du chemin des catégories Wikipedia entre deux concepts. La figure suivante montre un exemple d'une hiérarchie entre les deux concepts 'Boosting' et 'Feature Learning'.

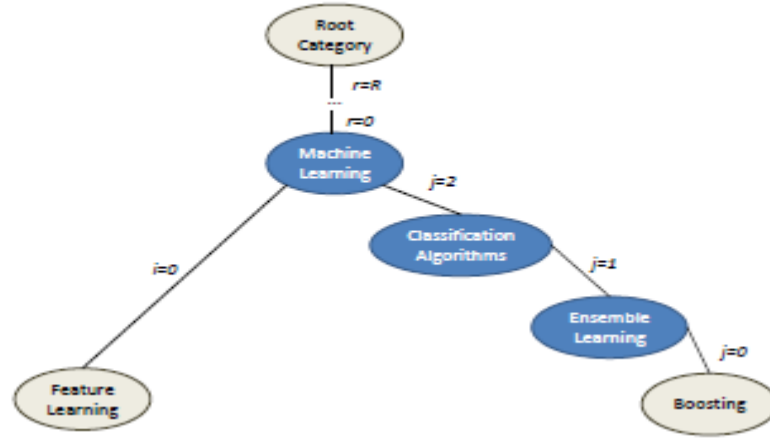


Figure 8. Chemin des catégories entre les deux concepts 'Featurelearning' et 'Boosting'(Medina et al., 2012)

En fait, pour calculer la proximité sémantique entre deux concepts c_1 et c_2 , les auteurs ont calculé la proximité sémantique entre c_1 et le premier concept commun dans la hiérarchie des catégories (par exemple c'est le concept 'machine learning' dans la figure précédente), et la proximité sémantique entre c_1 et le concept racine de cette hiérarchie('RootCategory' mentionné dans la figure précédente). Elle est égale à la somme des poids des arêtes. De même pour c_2 . Ainsi la proximité sémantique globale entre c_1 et c_2 est donnée par la formule suivante :

$$d(c_1, c_2) = \frac{\sum_{i=0}^I w_i^1 + \sum_{i=0}^J w_i^2}{\sum_{i=0}^R w_i^1 + \sum_{i=0}^R w_i^2} \quad (38)$$

Où w_i^1 est le poids de l'arête d'indice i dans le chemin des catégories entre c_1 et le premier concept commun. De même pour w_i^2 . I et J sont les indices des deux dernières arêtes des deux chemins respectivement. R est l'indice de la dernière arête du chemin entre le premier concept commun et le concept racine (comme le montre la figure précédente). Le poids d'une i^{eme} arête w_i est calculé comme suit :

$$w_i = \beta^{\alpha i} \quad (39)$$

Où α et β sont des paramètres prédéfinis (ils n'ont pas été expliqués par les auteurs).

III.7.1.3 Une mesure efficace, peu coûteuse de proximité sémantique obtenue à partir des liens Wikipedia

An effective, low-cost measure of semantic relatedness obtained from Wikipedia links (Witten and Milne, 2008b)

Les auteurs ont proposé une mesure basée-liens Wikipedia, nommée WLM (c.à.d. Wikipedia Linke-based Measure) qui permet de calculer la proximité sémantique entre les concepts en utilisant les liens trouvés dans leurs articles Wikipedia correspondants (autrement dit, le réseau de liens inter-articles Wikipedia. Pour ce faire, ils ont combiné deux mesures, la première est inspirée de la mesure statistique TF-IDF, tandis que la deuxième est inspirée de la Distance de Google Normalisée NGD. En fait, pour mesurer la proximité entre deux articles a et b , les auteurs ont procédé comme suit :

- Calculer une première proximité sémantique entre eux en utilisant une mesure inspirée de la mesure statistique TF-IDF suivi par la mesure de cosinus.

- o Pour chacun des deux concepts a et b , calculer son modèle vectoriel de liens Wikipedia (en anglais c'est Wikipedia Link Vector Model (WLVM), dont chaque élément de ce vecteur contient le poids du lien correspond vers l'article cible. Ainsi, pour calculer un élément de ce vecteur, les auteurs ont utilisé une mesure inspirée de la mesure statistique TF-IDF. En fait, TF-IDF d'un concept donné compte le nombre de concepts pondéré par la probabilité que ce concept se produit. Tandis que leur mesure est définie par le nombre total de liens vers l'article cible sur le nombre total d'articles. Ainsi, si s et t sont les articles source et cible, alors le poids w du lien $s \rightarrow t$, noté $w(s \rightarrow t)$ est calculé par la formule suivante :

$$w(s \rightarrow t) = \log\left(\frac{|W|}{|T|}\right) \text{ si } s \in T, 0 \text{ sinon} \quad (40)$$

Où $|T|$ est le nombre de tous les articles qui sont liés à t (autrement dit c'est le nombre total de liens vers l'article cible), et $|W|$ est le nombre total des articles Wikipedia. $w(s \rightarrow t)$ prend une valeur égale à zéro si ce lien n'existe pas.

Après la génération des vecteurs de poids des liens décrivant chacun des deux articles a et b , la proximité sémantique entre eux est obtenue par la similarité cosinus entre les deux vecteurs. Elle est égale à 0° si les articles contiennent des listes de liens identiques, à 90° s'il n'y

a pas de chevauchement entre eux. La similarité cosinus θ entre deux vecteurs V_{c_i} et V_{c_j} est donnée par la formule suivante :

$$\cos \theta = \frac{V_{c_i} \times V_{c_j}}{\|V_{c_i}\| \times \|V_{c_j}\|} \quad (41)$$

$$\text{Où } V_{c_i} \times V_{c_j} = (vc_{i1} \times vc_{j1}) + (vc_{i2} \times vc_{j2}) + \dots + (vc_{in} \times vc_{jn}) \quad (42)$$

$$\|V_{c_i}\| = \sqrt{(vc_{i1}^2 + vc_{i2}^2 + \dots + vc_{in}^2)} \quad (43)$$

- Calculer une deuxième mesure inspirée de la distance de Google normalisée (Normalized Google Distance). En fait, la distance de Google est basée sur l'occurrence des concepts dans les pages web. Tandis que, leur mesure inspirée est basée sur les liens Wikipedia. Formellement, la mesure de proximité entre deux articles a et b est donnée comme suit :

$$sr(a, b) = \left(\frac{\log(\max(|A|, |B|)) - \log(|A \cap B|)}{\log|W| - \log(\min(|A|, |B|))} \right) \quad (44)$$

Où $|A|$ et $|B|$ sont les ensembles de tous les articles qui sont liés à a et à b respectivement, et W est l'ensemble de tous les articles Wikipedia.

- La proximité sémantique entre les deux concepts est la moyenne des deux valeurs précédentes (la mesure cosinus et la distance de Google).

III.7.1.4 Wikiwalk: Marche aléatoire dans Wikipedia pour la proximité sémantique

Wikiwalk: random walks on Wikipedia for semantic relatedness (Yeh et al., 2009)

Les auteurs ont présenté une méthode de calcul de la proximité sémantique en utilisant une marche aléatoire (Personalized Page Rank) (Haveliwala, 2003) dans un graphe dérivé depuis la structure d'hyperlien entre les articles Wikipedia, de telle sorte que les nœuds soient les articles Wikipedia et les arêtes soient les liens entre ces articles. Ainsi, leur calcul de proximité sémantique d'une paire de concept se résume en trois étapes :

1. Mapper chaque concept à ses nœuds correspondants dans le graphe.
2. Exécuter l'algorithme 'Personalized PageRank' (Haveliwala, 2003) pour calculer la distribution stationnaire d'une chaîne de Markov de chaque concept.
3. Mesurer la similarité entre les deux distributions stationnaires des deux concepts avec la mesure de similarité cosinus.

Après les différentes expérimentations faites par les auteurs, ces derniers ont déclaré que l'utilisation du texte de Wikipedia fournit des résultats de proximité sémantiques plus efficaces que la structure des liens. Ceci est dû au fait que certains liens sont informatifs pour la proximité sémantique tandis que d'autres ne le sont pas, et que Wikipedia est une source bruyante d'informations de lien.

III.7.1.5 Utilisation de la sémantique Wikipedia pour calculer la proximité contextuelle

Harnessing Wikipedia Semantics for Computing Contextual Relatedness (Jabeen et al., 2012)

Les auteurs ont proposé une méthode 'WikiSim' de calcul de proximité sémantique entre les concepts basée sur les sens de Wikipedia et la structure d'hyperlien. Ainsi, pour calculer la proximité sémantique entre deux concepts, les auteurs ont procédé comme suit :

1. Extraction des sens des concepts : Un concept peut avoir des polysémies (un concept avec différentes significations selon le contexte où il est trouvé), comme il peut avoir des concepts synonymes (concepts qui signifient la même chose). A cet effet, pour identifier le sens approprié de chacun des concepts, les auteurs ont exploité des informations existantes dans Wikipedia, tel qu'un contexte entre parenthèses qui suit le concept, les concepts alternatives qui sont habituellement des synonymes, des hyponymes (c.à.d. fait partie-du concept) ou des hypernym (c.à.d. que c'est un parent du concept).

2. Calcul de la proximité sémantique

Pour calculer la proximité sémantique entre deux concepts w_a et w_b , les auteurs ont extrait, pour chacun des sens candidats des concepts w_a et w_b , tous les liens entrants (tous les articles faisant référence à l'article du concept requête) et tous les liens sortants (tous les articles renvoyés par l'article du concept requête). Ensuite ils ont comparé chaque sens du concept w_a avec chacun des sens du concept w_b . Ainsi, chaque paire de sens est assignée une valeur de proximité basée sur les liens partagés. C'est la probabilité des liens partagés. Elle est calculée par la formule suivante :

$$w(s_i, s_j) = \frac{|S|}{|T|}, \text{ si } S \neq \emptyset \quad (45)$$

Où s_i est un sens de w_a et s_j est un sens de w_b , $|S|$ est le nombre de liens partagés par cette paire de sens, et $|T|$ est le nombre total de liens entrants et sortants des deux sens. S'il n'y a pas de liens partagés entre les deux sens, alors leur proximité est nulle.

Après l'évaluation de leurs résultats de proximité sémantique par rapport à l'ensemble de données WordSim-353, les auteurs ont obtenu une valeur de corrélation mieux que les travaux de l'état de l'art (Strube and Ponzetto, 2006; Witten and Milne, 2008b).

III.7.2 Méthodes statistiques

Ces méthodes exploitent les propriétés internes des articles Wikipedia. Autrement dit, ils font des statistiques dans le texte des articles. Dans ce qui suit, nous présentons quelques méthodes de cette famille.

III.7.2.1 Calcul de proximité sémantique en utilisant une analyse sémantique explicite basée-Wikipedia

Computing semantic relatedness using Wikipedia-based Explicit Semantic Analysis (Gabrilovich and Markovitch, 2007)

Ce travail est une référence de base (Baseline) par la majorité des travaux sur le calcul de proximité sémantique entre des concepts. En fait, ses auteurs ont proposé une méthode d'analyse sémantique explicite ESA (c.à.d. Explicit Semantic Analysis) afin de calculer la proximité sémantique entre deux fragments de texte ou bien deux mots. Une analyse sémantique est explicite dans le sens où ils ont représenté les significations des fragments de texte en manipulant des concepts clairs et naturels définis par les êtres humains. En fait, ils ont utilisé la totalité de la version anglaise des articles Wikipedia, appelée décharge Wikipedia (en anglais c'est English Wikipedia dump), comme une source de connaissance qui reflète le monde réel. La figure .7 montre les étapes de cette méthode.

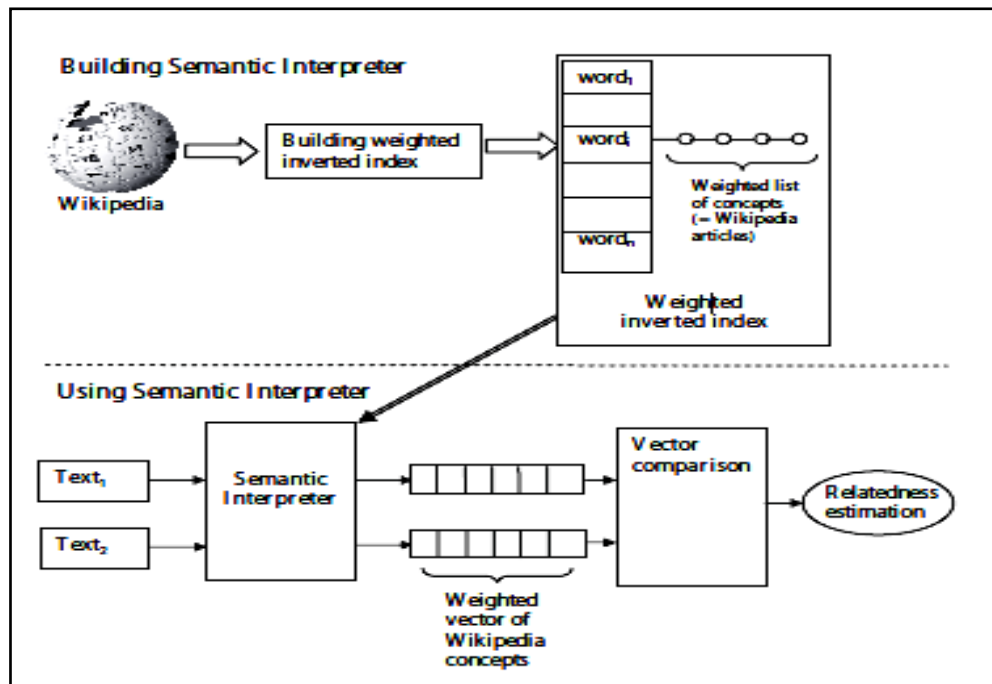


Figure 7. Calcul de proximité sémantique en utilisant une analyse sémantique explicite basée-Wikipedia (Gabrilovich and Markovitch, 2007)

Les étapes de leur travail sont :

1. Construction d'une source de connaissances :

Pour construire un corpus de connaissance, les auteurs ont téléchargé la version anglaise de la totalité des pages des articles Wikipedia en SQL et XML, c.à.d. la décharge Wikipedia (en anglais c'est English Wikipedia dump). Ce corpus est accessible au lien : https://en.Wikipedia.org/wiki/wikipedia:Database_download. Ensuite, ils ont fait un prétraitement du corpus téléchargé et qui consiste à éliminer les mots d'arrêt (stopwords) et les mots rares, ainsi qu'une lemmatisation du texte. La lemmatisation désigne l'analyse lexicale du contenu d'un texte regroupant les mots d'une même famille. Chacun des mots d'un contenu se trouve ainsi réduit en une entité appelée lemme (forme canonique). Prenant comme exemple l'adjectif *petit*, il existe sous quatre formes : *petit*, *petite*, *petits* et *petites*. La forme canonique de tous ces mots est le lemme « *petit* ». Ce prétraitement est implémenté par un script perl qu'ils ont appelé *Wikiprep* et qui est accessible au lien :

abrilivitch.com/ressources/codes/wikiprep/wikiprep.html). Ainsi, le résultat de cette étape, est un corpus de n articles prétraités représentant des concepts du monde réel.

2. Calcul de proximité sémantique entre les concepts :

Du fait que chaque article Wikipedia s'intéresse à un seul sujet décrit par son titre, alors les auteurs ont considéré que chaque article $Wikipediadc_i$ correspond à un seul concept c_i , et que

chaque concept prend comme nom le titre de l'article correspond. Alors l'ensemble de n articles prétraités correspond à n concepts. Ainsi pour calculer la proximité sémantique entre les paires de concepts de l'ensemble de n concepts, les auteurs ont procédé comme suit :

a. Calcul des poids des concepts

Cette étape consiste à calculer pour chaque concept c_i parmi les n concepts et qui est présenté par son article dc_i , son vecteur d'interprétation $Vc_i=(vc_{i1},vc_{i2},\dots,vc_{ij},\dots,vc_{in})$, où chaque élément vc_{ij} quantifie la force de l'association entre les deux concepts c_i et c_j et représente le poids du concept c_j dans l'article dc_i en utilisant la méthode de pondération statistique TF_IDF (Term Frequency- Inverse Document Frequency). La formule TF-IDF est déjà présentée dans le travail (Wang et al., 2010).

b. Elimination des concepts non significatifs

L'objectif de cette étape est d'éliminer tout concept non significatif. Un concept est non significatif si ses poids dans les différents articles sont inférieurs à un certain seuil, c.à.d. des valeurs très petites. Ainsi, pour faire la correspondance d'un concept c_j avec ses poids dans les différents articles, les auteurs ont utilisé l'index inversé du concept c_j .

c. la similarité cosinus

La dernière étape consiste à calculer la proximité sémantique entre deux concepts ou bien entre deux fragments de texte, par la comparaison de leurs vecteurs d'interprétation dans l'espace défini par les concepts. Pour ce faire, ils ont utilisé la mesure de cosinus, qui est donnée par la formule suivante :

$$prox_sémant(Vc_i, Vc_j) = \cos \theta = \frac{Vc_i \times Vc_j}{\|Vc_i\| \times \|Vc_j\|} \quad (46)$$

$$Vc_i \times Vc_j = (vc_{i1} \times vc_{j1}) + (vc_{i2} \times vc_{j2}) + \dots + (vc_{in} \times vc_{jn}) \quad (47)$$

$$\|Vc_i\| = \sqrt{(vc_{i1}^2 + vc_{i2}^2 + \dots + vc_{in}^2)} \quad (48)$$

Les auteurs ont évalué leur résultat de calcul de proximité sémantique avec des benchmarks de jugement humain, ils ont prouvé l'efficacité de leur méthode de capturer la proximité sémantique entre des concepts mieux que les méthodes présentées dans (Strube and Ponzetto, 2006), (Witten and Milne, 2008b), (Yeh et al., 2009) et (Jabeen et al., 2012).

III.7.2.2 Proximité sémantique en utilisant l'analyse sémantique saillante

Semantic Relatedness Using Salient Semantic Analysis (Hassan and Mihalcea, 2011)

Les auteurs ont déclaré que l'être humain juge la proximité sémantique, en se basant non seulement sur la large connaissance qu'il acquit suite aux expériences accumulées, mais en plus, il utilise ses capacités de pensée conceptuelle, d'abstraction et de généralisation. Par conséquent, pour qu'un système puisse calculer la proximité sémantique efficacement, il doit non seulement être en mesure d'acquérir et d'utiliser une large source de connaissances externes, mais il devrait aussi pouvoir l'abstraire et la généraliser. Pour cette raison, les auteurs ont développé une méthode de mesure de proximité sémantique non supervisée qui repose sur l'idée que le sens d'un mot peut être caractérisé par les concepts saillants trouvés dans son contexte immédiat, ce qui signifie qu'ils sont les plus conceptuellement pertinents pour ce concept. Les concepts saillants à un concept c_i qui est représenté par l'article Wikipedia dc_i , sont les concepts liés dans le texte de dc_i , c.à.d. les concepts qui possèdent des liens permettant de passer automatiquement de l'article dc_i aux articles Wikipedia trouvés dans son contexte immédiat et qui correspondent à ces concepts liés. Par conséquent, chaque concept c_i est représenté par un vecteur Vc_i qui contient les poids de ses concepts saillants, et la proximité sémantique entre deux concepts est la distance entre leurs vecteurs de concepts saillants.

Pour calculer les vecteurs $Vc_{i,i:1..n}$ où n est le nombre de concepts (c.à.d le nombre d'articles Wikipedia), et la proximité entre les concepts, ils ont procédé comme suit :

1. Construction d'un corpus d'articles Wikipedia avec annotation explicite des concepts saillants :

En se basant sur les indications des directives Wikipedia, que seuls les mots ou expressions importants pour la compréhension d'un certain texte devraient être liés, alors les auteurs ont utilisé les liens manuels fournis par les utilisateurs de Wikipedia, et ils ont considéré donc que les concepts liés sont des concepts saillants, et de même pour les expressions liées. Ainsi, ils ont calculé le nombre d'occurrence d'un concept saillant dans un article Wikipedia, en tenant compte même les occurrences non liées, et chaque nouvelle occurrence non liée est donc liée à l'article correspond. Ensuite, ils ont appliqué une méthode de désambiguïsation qui permet d'assigner les articles Wikipedia aux concepts (ou expressions) les plus liés. En fait, cette méthode de désambiguïsation calcule pour chaque concept saillant (et expression saillante) une probabilité qui est égale au nombre de fois qu'il apparaît dans un lien affecté manuellement divisé par le nombre total de son occurrence dans Wikipedia (lié ou non). Ainsi, les concepts (ou

les expressions) qui ont une probabilité ≥ 0.5 de pointer vers un seul article sont annotés avec l'article correspondant.

2. Construction des vecteurs des poids des concepts :

Les auteurs ont utilisé le corpus construit dans l'étape précédente, noté C pour calculer pour chaque concept c_i (qui correspond à un article Wikipedia) son vecteur V_{c_i} des poids de ses concepts saillants. Pour ce faire, ils ont procédé d'une façon formelle comme suit:

Partant du corpus C avec m unité lexicale et un ensemble de concepts de taille n ,

- Construire une matrice E de cooccurrence de taille $n \times n$, qui représente les fréquences accumulées de cooccurrence de chacun des concepts par rapport aux autres concepts. Un élément de cette matrice E_{ij} est calculé par l'équation suivante :

$$E_{ij} = f(c_i, c_j) \quad (49)$$

Où $f(c_i, c_j)$ est le nombre de fois où les deux concepts c_i et c_j co-occurrent ensemble dans le corpus entier.

A partir de la matrice de cooccurrence E , générer une autre matrice P de taille $n \times n$, dont un élément de cette matrice mesure l'information mutuelle ponctuelle (Pointwise Mutual Information (PMI) en anglais) entre deux concepts. PMI est une mesure très utilisée en statistique et la théorie de l'information pour mesurer l'association entre deux événements (Franzoni et al., 2015). Cette mesure est donnée par la formule suivante :

$$P_{ij} = PMI(c_i, c_j) = \log_2 \frac{f(c_i, c_j) \times m}{f^c(c_i) \times f^c(c_j)} \quad (50)$$

Où $f^c(c_i)$ et $f^c(c_j)$ sont les fréquences d'apparition des concepts c_i et c_j respectivement, dans tout le corpus.

Ensuite chaque ligne P_i de cette matrice est filtrée afin d'éliminer les associations non pertinentes, en gardant uniquement les premières β_i cellules les plus grandes, et mettre le reste à zéro. La valeur de sélection β_i de ligne i est calculée comme suit :

$$\beta_i = (\log_{10}(f^c(c_i)))^2 \times \frac{\log_2(n)}{\delta}, \delta \geq 1 \quad (51)$$

Où δ est un constant qui est ajusté selon la taille du corpus.

3. Calcul de proximité sémantique entre les concepts :

Les auteurs ont utilisé la matrice P après filtrage pour calculer la proximité sémantique entre deux concepts A et B , et qui exprime le chevauchement entre leurs profils sémantiques. Pour ce faire, ils ont adopté deux mesures avec des petites modifications : la similarité cosinus, et la mesure SOCPMI (Second Order Co-Occurrence Pointwise Mutual Information) dont il a été démontrée qu'elle est plus forte que la similarité cosinus. Les équations de ces mesures sont respectivement :

$$Score_{cos}(A, B) = \frac{\sum_{y=1}^n (P_{iy} \times P_{jy})^\gamma}{\sqrt{\sum_{y=1}^n P_{iy}^{2\gamma}} \times \sqrt{\sum_{y=1}^n P_{jy}^{2\gamma}}} \quad (52)$$

$$Score_{soc}(A, B) = \ln \left(\frac{\sum_{y=1}^n (P_{iy})^\gamma}{\beta_i} + \frac{\sum_{y=1}^n (P_{jy})^\gamma}{\beta_j} + 1 \right) \quad (53)$$

Où le paramètre γ permet de contrôler le biais de poids. En outre, étant donné que le cosinus est une métrique normalisée qui donne 1 pour des concepts identiques, elle est affectée négativement par un espace dispersé car elle tend à fournir de faibles scores pour des synonymes proches. Cela crée un large fossé sémantique entre les concepts identiques et les concepts fortement liés. Pour combler ce fossé et fournir des résultats plus significatifs, ils ont utilisé également un facteur normalisation λ , comme montre l'équation suivante

$$Sim(A, B) = \begin{cases} 1 & Score_{cos}(A, B) > \lambda \\ \frac{Score_{cos}(A, B)}{\lambda} & Score_{cos}(A, B) \leq \lambda \end{cases} \quad (54)$$

Les deux travaux sus-cités sont les travaux avec lesquels nous allons comparer pour évaluer nos calculs de proximité sémantique. En fait, nous allons les référencier sous l'abréviation ESA et SSA respectivement. Ce choix se justifie par les points suivants :

- Notre méthode s'inscrit dans la même famille que ESA et SSA, elles sont toutes des méthodes statistiques.
- Selon les différents travaux de l'état de l'art (topologiques ou bien statistiques), la méthode ESA présente les meilleurs résultats à ce jour sur l'ensemble de données WordSim-353 (Yeh et al., 2009; Witten and Milne, 2008b; Jabeen et al., 2012; Strube and Ponzetto, 2006).

III.8 Conclusion

Dans ce chapitre, nous avons concentré notre attention sur les travaux qui ont contribué à réduire le silence dans un TBIR. Nous avons présenté des différentes méthodes existantes, à savoir des méthodes visuelles et des méthodes sémantiques.

Les méthodes visuelles sont destinées beaucoup plus à l'annotation. Ainsi, elles ont développé des techniques qui exploitent la relation entre les caractéristiques visuelles et les concepts d'annotation d'un ensemble d'apprentissage (images ou régions) annoté manuellement, afin de prédire des concepts pour des images de test non annotées. En fait, ces méthodes aient mené à des résultats encourageants et importants. Cependant, elles restent confrontées avec quelques limitations importantes, dont les plus importantes sont: la difficulté d'assurer un ensemble d'apprentissage avec des annotations complètes et précises, la difficulté de modéliser des concepts ayant plusieurs apparences visuelles, se limiter généralement à des concepts d'annotation de premier niveau d'abstraction, et qui décrivent le contenu local perceptuel des images. De plus, leur limitation majeure c'est qu'elles ont ignoré complètement la sémantique d'interdépendance qui existe entre les concepts du monde réel. Par conséquent, ces limitations dégradent certainement la performance du processus d'annotation, et donc, les résultats de la recherche d'images.

Contrairement aux méthodes visuelles, les méthodes sémantiques, comme leur nom l'indique, elles ont considéré la sémantique qui relie des concepts afin d'annoter ou rechercher des images. Ainsi, elles ont calculé (modélisé) l'interdépendance entre les concepts d'une façon automatique. Ensuite, elles ont incorporé les résultats obtenus dans la méthode de prédiction des concepts pour les images de test (c.à.d. dans le processus d'annotation), ou bien au sein du moteur de recherche d'images. Ces méthodes sémantiques peuvent être classées, en fonction de la source de connaissance utilisée pour calculer l'interdépendance entre les concepts, en deux familles: méthodes basées-corpus local et méthodes basées-corpus global. La première famille a utilisé comme source de connaissance l'information sémantique de cooccurrence entre les concepts présentée dans l'annotation d'une base d'images locale, c'est la base expérimentale d'apprentissage, tandis que la deuxième famille a extrait cette sémantique en utilisant des sources de connaissance externes indépendamment de toute base d'image expérimentale, dont la plus utilisée c'est le Wikipedia.

Les méthodes sémantiques basées-corpus local ont obtenu des performances compétitives. Néanmoins, les résultats de calcul de l'interdépendance entre les concepts peuvent être loin de jugement humain et donc ne peuvent pas refléter des cas réels. Ceci est due à l'information sémantique de cooccurrence entre les concepts qui est extraites à partir des annotations d'apprentissage qui peuvent être incomplètes, bruitées et ambiguës. De plus, les résultats obtenus peuvent être inapproprié, car les statistiques sur l'occurrence des concepts (Xu et al., 2016) ont montré que la distribution des concepts est souvent très déséquilibrée.

Les méthodes sémantiques basées-corpus global ont construit des modèles sémantiques (c.à.d. ontologie et thésaurus) permettant de modéliser des concepts et leurs relations sémantiques, afin d'être exploités, généralement, en recherche d'images. En fait, leur inconvénient majeur c'est qu'ils n'ont modélisé que les relations sémantiques extractibles automatiquement depuis Wikipedia, telles que les relations de *Synonymie*, *polysémie*, *est-un*, *partie-de*, *concept-lié*. Par conséquent, ces modèles ne reflètent pas la richesse en relations sémantiques qui existent en réalité. Ceci, peut conduire à une estimation biaisée de la mesure de similarité sémantique des images.

En outre, le calcul de proximité sémantique est une étape clé dans notre travail, et il constitue un réel défi pour une grande communauté des chercheurs, indépendamment du domaine d'application. En fait, l'ensemble des méthodes pour cette fin (à nos connaissances) sont des méthodes basées-Wikipedia. Nous pouvons ainsi les classer, en fonction des propriétés du Wikipedia exploitées, en deux familles: des méthodes topologiques et des méthodes statistiques. Les méthodes topologiques exploitent des propriétés externes de Wikipedia, tandis que les méthodes statistiques exploitent les propriétés internes des articles Wikipedia. Les méthodes statistiques ont prouvé leur corrélation plus performante avec du jugement humain par rapport aux méthodes topologiques. Pour cela, nous allons choisi de développer une méthode sémantique statistique de calcul de proximité sémantique entre les concepts.

Chapitre IV. Méthode pour réduire le silence en recherche d'images basée proximité sémantique

IV.1 Introduction

Comme il a été déjà mentionné, le problème de silence en recherche d'images par le texte est dû à la comparaison binaire entre une requête utilisateur et des annotations incomplètes.

Dans la littérature, de nombreuses approches ont été proposées pour atténuer ce problème, telles que celles présentées dans le chapitre précédent. Mais malheureusement, les solutions proposées restent confrontées à certaines limitations importantes et sensibles. Par conséquent, le problème de silence reste posé, et la concurrence reste encore ouverte afin d'atteindre une performance convaincante.

Dans ce chapitre, nous présentons notre solution, qui constitue notre principale contribution pour réduire le silence dans un système TBIR. Nous commençons par montrer comment nous avons dégagé l'idée principale de la solution proposée, à travers une analyse et discussion des différentes relations sémantiques qui peuvent exister entre les concepts. Ensuite, nous présentons un aperçu global sur la solution proposée, ainsi que nos principales contributions. Enfin, nous décrivons en détail chaque étape de cette solution.

IV.2 Analyse des relations sémantiques entre les concepts

Comme nous avons présenté dans le chapitre de l'état de l'art, la majorité des travaux focalisant sur l'exploitation des relations sémantiques entre des concepts, afin de réduire le

silence et améliorer le résultat de recherche, ont utilisé principalement les relations '*est-un*' et '*équivalent-à*'. Ainsi, deux concepts c_1 et c_2 reliés, à titre d'exemple, par la relation '*est-un*' notée, c_1 '*est-un*' c_2 , signifie que c_1 est très pertinent à c_2 . Par conséquent, lorsqu'un utilisateur formule sa requête par c_1 , le système de recherche considère les images annotées par c_1 comme pertinentes, ainsi que les images annotées par c_2 . Nous pouvons alors, appliquer ce même principe pour n'importe quel type de relation sémantique, telle que la relation de spécialisation '*sous-classe-de*', la relation de composition '*se-compose-de*', la relation de signe '*signe-de*', la relation de *cooccurrence*, la relation d'*exclusion*, etc. Autrement dit, n'importe quelle relation sémantique $R(c_1, c_2)$ qui pourrait exister entre deux concepts c_1 et c_2 apporte un certain degré de pertinence sémantique entre ces deux concepts. Et c'est ce degré de pertinence qui compte au moment de la recherche. Par conséquent, nous pouvons remplacer toute relation par le degré de pertinence qu'elle apporte. Ce degré peut être vu comme une valeur de probabilité. Elle peut être maximale égale à un, comme pour les relations '*est-un*' et '*équivalent-à*', ce qui signifie que les deux concepts sont sémantiquement les mêmes, zéro indiquant l'indépendance et entre ces deux valeurs indiquant des relations intermédiaires.

À partir de cette analyse, les défis d'un système de recherche d'images deviennent : d'abord, calculer des degrés de pertinence entre les différents concepts. Deuxièmement, développer des mécanismes de recherche qui infèrent la pertinence des images par rapport à une requête utilisateur en exploitant ces degrés de pertinence.

Le premier défi est très important et très difficile à atteindre. Il est très important, du fait que ses résultats influencent directement la performance du mécanisme de recherche. Cependant, il est très difficile à atteindre, par ce que tout simplement, il est très difficile pour une machine de capturer la sémantique comme les êtres humains le font ! En fait, indépendamment des domaines d'application, ce défi a attiré l'attention d'une communauté importante de chercheurs. et de plusieurs travaux qui se sont concentrés sur la manière de mesurer la sémantique entre les concepts, existant dans la littérature.(Gabrilovich and Markovitch, 2007) et (Hassan and Mihalcea, 2011; Jabeen et al., 2012), (Ni et al., 2016; Taieb et al., 2014; Pakhomov et al., 2010; Aouicha and Taieb, 2016).

IV.3 Similarité sémantique et Proximité sémantique

Après une recherche approfondie dans plusieurs travaux de la littérature afin de trouver le terme approprié de degré de pertinence décrit ci-dessus, nous avons constaté qu'il y a deux termes très utilisés, qui sont : la similarité sémantique, dite *semantic similarity* en anglais, et la proximité sémantique, dite *semantic relatedness* en anglais. Ces deux termes sont très proches, ce qui a conduit à une certaine confusion et à leur utilisation de manière indifférente dans de nombreux travaux de la littérature. Cependant, dans leur *Survey* extensif sur les mesures de proximité, Budanitsky et Hirst (Budanitsky and Hirst, 2006) ont soutenu que la notion de proximité est plus générale que celle de la similarité. Ils ont en outre veillé à ce que les applications de la linguistique informatique nécessitent souvent des mesures de proximité plutôt que des mesures de similarité. En fait, la similarité sémantique est plus spécifique que la proximité sémantique puisqu'elle est limitée aux structures hiérarchiques entre des concepts. Autrement dit, la similarité sémantique consiste à estimer les valeurs de la relation sémantique '*est-un*' entre des concepts dans une hiérarchie (ou taxinomie) de concepts. Elle est d'ailleurs appelée 'similarité topologique'. Comme instance, plusieurs travaux ont utilisé la hiérarchie catégorique de Wikipedia (Strube and Ponzetto, 2006; Taieb et al., 2013), ou bien sa structure d'hyperlien (Yeh et al., 2009; Witten and Milne, 2008a), pour calculer la similarité sémantique entre les concepts. Tandis que, la proximité sémantique consiste à estimer la force de connexion sémantique entre deux concepts quelle que soit la relation sémantique entre eux. Selon (Hassan and Mihalcea, 2011), la proximité sémantique quantifie la force des connexions sémantiques qui existent entre les unités textuelles, qu'elles soient des paires de mots, des paires de phrases ou des paires de documents. Autrement dit, elle répond à la question « combien une unité textuelle A est liée sémantiquement à l'unité textuelle B ». La réponse à cette question est généralement une valeur comprise entre -1 et 1, ou entre 0 et 1, où 1 signifie que les deux unités sont identiques.

Dans notre travail, nous sommes intéressés à la proximité sémantique entre les concepts pour deux raisons: premièrement, la similarité sémantique se limite aux structures hiérarchiques, et ceci n'est pas notre objectif. Deuxièmement, la proximité sémantique est très appropriée au degré de pertinence que nous avons expliqué dans la section précédente.

IV.4 Aperçu sur la méthode proposée

Du fait que le silence est dû à la comparaison binaire entre une requête utilisateur et des annotations incomplètes, et puisque c'est très difficile d'assurer la complétude des annotations quelle que soit la méthode d'annotation adoptée, alors nous proposons un autre mécanisme de recherche autre que la méthode traditionnelle (comparaison binaire) afin de minimiser le silence dans un TBIR. Ce mécanisme est basé sur la proximité sémantique entre les concepts. Pour ce faire, dans la phase offline de notre TBIR, nous calculons la proximité sémantique entre des concepts. Ensuite, dans la phase online, nous proposons un mécanisme de recherche d'images qui exploite les valeurs de proximité sémantiques obtenues afin de détecter des images pertinentes même si le concept requête ne figure pas dans leurs annotations.

Nos principales contributions sont :

1. Relations sémantiques

Au lieu de se limiter à certains types de relations sémantiques, l'utilisation des degrés de proximité sémantique nous permettent de refléter le degré de pertinence sémantique pour n'importe quelle relation sémantique. Ceci rend notre solution plus complète du point de vue richesse sémantique, plus souple du fait qu'elle n'est limitée à aucune relation prédéfinie et plus consistante du fait que le mécanisme de recherche sera indépendant de toute relation.

2. Corpus pour le calcul de proximité sémantique

Au lieu d'utiliser l'information de cooccurrence dans un corpus local (base d'images annotées), l'utilisation d'une source de connaissances externes créée par une communauté humaine à travers le monde, nous permet de calculer la proximité sémantique entre des concepts d'une façon plus performante et d'obtenir des résultats plus proches du jugement humain.

3. Méthode de recherche

Au lieu d'une comparaison binaire (zéro ou un) entre une requête utilisateur et les annotations des images, l'utilisation des valeurs de proximité sémantique entre la requête et les concepts d'annotation, nous permet de calculer une valeur de pertinence pour chaque image de la base, et ceci dans un intervalle entre zéro et un. Par conséquent, nous pouvons localiser plus d'images pertinentes, et donc réduire le silence.

La solution proposée se compose de deux phases principales : calcul de proximité sémantique entre des concepts et la recherche. Pour calculer la proximité sémantique entre des concepts, nous avons utilisé la méthode de pondération statistique TF-ICTF (en anglais c'est Term Frequency-Inverse Collection Term Frequency) ainsi que la similarité cosinus. La méthode TF-ICTF permet d'estimer l'importance (ou bien le poids) d'un concept contenu dans un document. Nous avons rassemblé un ensemble de documents (c.à.d. articles) qui correspondent à nos concepts depuis une source de connaissance externe construite par la communauté humaine à travers le monde, c'est le Wikipedia. Ensuite, après le calcul des poids de pondération de chaque concept, le résultat obtenu est utilisé comme une entrée pour calculer la similarité cosinus entre ces concepts. Les valeurs obtenues reflètent la proximité sémantique. Durant la recherche, l'utilisateur peut naturellement exprimer ses besoins en introduisant une requête textuelle sous forme d'un concept. Le système de recherche estime la pertinence de chaque image de la base en fonction des valeurs de proximité sémantique entre ses concepts d'annotation et la requête. Ensuite, il retourne les images par ordre décroissant de pertinence. La Figure. 6 présente l'organigramme de la méthode proposée.

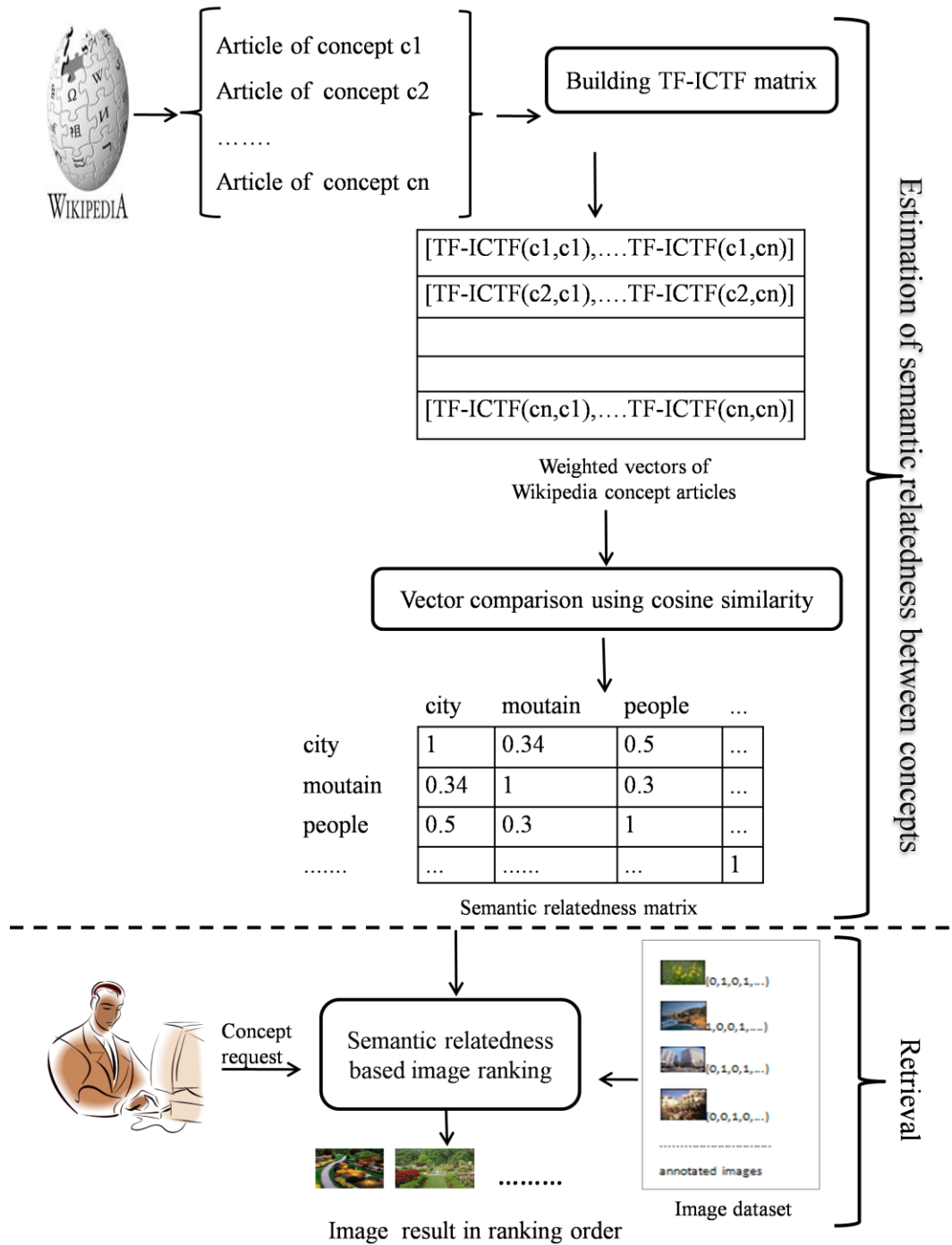


Figure 8. L'organigramme de la méthode proposée

Nous allons présenter ci-après une description détaillée de chaque étape de notre solution

IV.5 Etapes de la solution pour réduire le silence dans un TBIR

La méthode proposée s'articule sur deux phases principales, une phase offline de calcul de proximité sémantique entre les concepts, et une phase online de recherche. Dans ce qui suit une description détaillée de ces deux phases.

IV.5.1 Calcul automatique de proximité sémantique entre les concepts

Étant donné deux concepts, il est très facile pour l'être humain de juger leur proximité sémantique. Et ceci, en se basant sur la large connaissance qu'il possède à propos de ces deux concepts. Mais pour une machine cette tâche est très difficile. Par conséquent, pour qu'un ordinateur soit capable de l'accomplir, il doit disposer d'une source de connaissance externe qui simule la connaissance humaine. En fait, la majorité des travaux de la littérature a utilisé des bases de données avec annotation (c.à.d. corpus local) comme source de connaissance. Mais malheureusement, se limiter à l'information de cooccurrence est inapproprié, car la fréquence d'apparition des concepts dans une base d'images est souvent très déséquilibrée (Xu et al., 2016). Ainsi, les valeurs de proximité sémantique peuvent être loin du jugement humain.

À cet effet, nous avons choisi d'utiliser une source de connaissance indépendante de toute base d'images, et qui est partagée par la communauté humaine à travers le monde. C'est le Wikipedia. En fait, Wikipedia stocke une grande quantité d'informations non seulement sur les concepts eux-mêmes, mais aussi sur divers aspects des relations entre les concepts. Ainsi, pour calculer la proximité sémantique entre des paires de concepts d'une collection $C = \{c_1, c_2, \dots, c_i, \dots, c_n\}$, nous allons procéder comme suit :

IV.5.1.1 Construction du corpus de connaissances

La première étape consiste à construire d'une façon automatique une source de connaissance. Elle s'agit d'une collection de documents $D = \{dc_1, dc_2, \dots, dc_i, \dots, dc_n\}$ qui correspond à nos concepts. Pour ce faire, une requête Wikipedia se lance pour chacun des concepts et le document retourné est automatiquement téléchargé et stocké. Ainsi, chaque document retourné dc_i se réfère à un article Wikipedia ayant le concept c_i comme titre. Par exemple, le concept '*desert*' correspond à l'article dans Wikipedia sous le titre '*desert*'. La figure. 3 montre une partie de ce document.

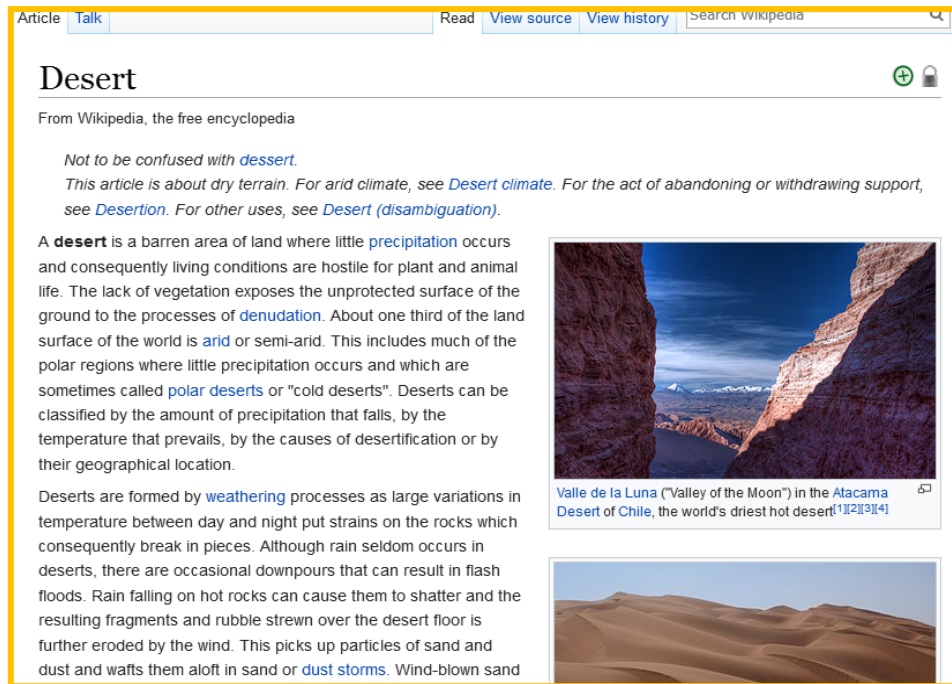


Figure 9. Un exemple d'un article Wikipedia.

L'algorithme 1 montre les étapes de construction de la source de connaissance externe

Algorithme 1 **source_connaissance**

Input: $C = \{c_1, c_2, \dots, c_i, \dots, c_n\}$ un ensemble de concepts.

Output: $D = \{dc_1, dc_2, \dots, dc_i, \dots, dc_n\}$: une collection de document

Début

Pour chaque concept c_i depuis la collection C **faire**

Début

Lancer une requête Wikipedia avec le concept c_i ;

Enregistrer le code HTML de la page correspondante comme document texte dc_i ;

Fin pour ;

Fin.

IV.5.1.2 Calcul des poids de pondération des concepts

À partir de la collection de documents construite dans l'étape précédente $D = \{dc_1, dc_2, \dots, dc_i, \dots, dc_n\}$, nous calculons, d'une façon automatique, les poids (ou bien l'importance) des concepts dans chacun des documents de cette collection. Par exemple, prenons un document dc_i qui correspond au concept c_i , nous calculons les poids de tous les concepts dans ce document. Pour ce faire, nous utilisons la méthode de pondération statistique TF-ICTF (TermFrequency Inverse Collection Term Frequency) (Kwok, 1995). Le résultat obtenu est le vecteur TF-ICTF du concept c_i , noté $V_{c_i} = (vc_{i1}, vc_{i2}, \dots, vc_{ij}, \dots, vc_{in})$, où chaque élément vc_{ij} représente $TF_ICTF(c_j, dc_i)$, le poids du concept c_j dans le document dc_i .

La méthode TF-ICTF est une variante de la méthode originale TF-IDF (Term Frequency-Inverse Document Frequency). Ces méthodes de pondération statistique étant souvent utilisées en recherche d'information et en particulier dans la fouille de textes pour établir la description des documents dans un modèle vectoriel, et par conséquent, la mesure de similarité entre une requête utilisateur textuelle et un document, deviennent une comparaison entre deux vecteurs.

Afin de distinguer entre les deux méthodes TF-IDF et TF-ICTF, ainsi que justifier le choix de la deuxième méthode, nous présentons chacune d'elles :

TF-IDF et TF-ICTF permettent, toutes les deux, de calculer le poids d'un concept c_j dans un document dc_i . La différence entre elles, réside dans la formule de calcul où TF-IDF utilise l'équation (1) alors que, TF-ICTF utilise l'équation (2).

$$TF_IDF(c_j, dc_i) = TF(c_j, dc_i) \times IDF(c_j) \quad (1)$$

$$TF_ICTF(c_j, dc_i) = TF(c_j, dc_i) \times ICTF(c_j) \quad (2)$$

Notons que dans les deux formules, le même premier terme de la multiplication, $TF(c_j, dc_i)$, est utilisé alors que le second terme est différent.

$TF(c_j, dc_i)$ est défini comme suit :

$$TF(c_j, dc_i) = \frac{\text{Fréquence_de_}c_j\text{_dans_le_document_}dc_i}{\text{Total_des_Fréquences_des_concepts_dans_}dc_i} \quad (3)$$

Où $TF(c_j, dc_i)$ signifie la fréquence du terme (en anglais Term Frequency). Elle mesure la fréquence d'apparition normalisée du concept c_j dans le document dc_i .

La fréquence d'apparition d'un concept c_j dans un document dc_i est simplement le nombre d'occurrence de ce concept dans le document considéré. Mais puisque les documents différent en longueur, il est possible qu'un concept apparaît beaucoup plus souvent dans les longs documents que dans les plus courts. Ce qui peut faire dominer le résultat. Par exemple, un concept c_j apparaît 6 fois dans un document de 30 mots, et 20 fois dans un autre document de 1000 mots. C'est-à-dire que c_j représente 20% du premier document, et 2% du deuxième document. Si nous ne considérons que cette fréquence d'apparition brute, le concept c_j sera plus important dans le deuxième document que dans le premier. Mais en réalité c'est le contraire. C'est pour cette raison que la fréquence est souvent divisée par les fréquences totales des concepts considérés dans le document comme moyen de normalisation.

Le deuxième terme de la multiplication pour la méthode TF_IDF est défini comme suit :

$$IDF(c_j) = \log\left(\frac{|D|}{|d_k : c_j \in d_k|}\right) \quad (4)$$

Où $IDF(c_j)$ signifie la fréquence inverse de document (en anglais Inverse Document Frequency). Elle est égale au logarithme (en base 10) de l'inverse de la proportion de documents de la collection contenant le concept c_j . Ainsi, $|D|$ est le nombre total de documents dans la collection D , $|d_k : c_j \in d_k|$ est le nombre de documents où le concept c_j apparaît. La mesure $IDF(c_j)$ calcule l'importance d'un concept c_j dans une collection de documents D en fonction de sa présence dans cette collection. Cette mesure vise à donner un poids plus important aux termes les moins fréquents, et considérés comme plus discriminants. L'IDF est généralement utilisé pour écarter des concepts non discriminants, du fait qu'ils sont très fréquents dans de nombreux documents d'une collection (comme les articles : un, une, la, le, les).

Le deuxième terme de la multiplication pour la méthode TF ICTF est défini par l'équation. (5).

$$ICTF(c_j) = \log\left(\frac{tf(D)}{tf(c_j)}\right) \quad (5)$$

Où $tf(D)$ est la somme des fréquences d'apparition de tous les concepts dans la collection de documents D , et $tf(c_j)$ est la fréquence d'apparition du concept c_j dans la collection D .

Le choix de la mesure ICTF au lieu de la mesure IDF est justifié par les deux raisons suivantes :

- Nous calculons les poids de pondération sur un ensemble de concepts de notre choix. Autrement dit, nous n'avons pas de concepts à écarter, pour utiliser IDF.
- La mesure ICTF capte mieux l'importance d'un concept dans une collection de documents que la mesure IDF, car l'ICTF tient compte de l'information de la fréquence d'occurrence d'un concept dans une collection de documents, tandis que IDF se limite à la présence ou l'absence d'un concept dans cette collection. Pour mieux comprendre cela, prenons un exemple d'une collection de trois documents dc_1, dc_2, dc_3 , où les fréquences d'apparition d'un concept c_i sont de 5 dans dc_1 , 0 dans dc_2 et 9 dans dc_3 . Et pour un concept c_j elles sont de 1 dans dc_1 , 0 dans dc_2 et 2 dans dc_3 . Dans cette collection les deux concepts c_i et c_j vont avoir la même valeur IDF (égale à $3/2$) du fait qu'ils ont une même présence dans les trois documents. Par contre, et puisque l'ICTF prend en considération les fréquences d'apparition, les deux concepts c_i et c_j vont avoir deux valeurs ICTF différentes.

Le résultat de cette étape est un ensemble de vecteurs, $Vc_i (i: 1..n)$, où chaque vecteur $Vc_i = (vc_{i1}, vc_{i2}, \dots, vc_{ij}, \dots, vc_{in})$ correspond au concept c_i , et chaque élément vc_{ij} représente $TF_ICTF(c_j, dc_i)$, le poids du concept c_j dans le document dc_i . L'ensemble de tous les vecteurs est organisé dans une matrice de pondération M de taille $n \times n$ (où n est le nombre de concepts), de telle sorte que chaque ligne i de cette matrice correspond au vecteur Vc_i du concept c_i , et chaque élément $M[i,j]$ représente l'élément vc_{ij} de ce vecteur.

L'algorithme 1 montre les étapes de calcul des poids de pondération TF ICTF des concepts

Algorithme 1 Procédure Calcul poids concepts

Input: $C = \{c_1, c_2, \dots, c_i, \dots, c_n\}$: un ensemble de concepts, $D = \{dc_1, dc_2, \dots, dc_i, \dots, dc_n\}$: une collection de document

Output: M: une matrice de pondération TF-ICTF de taille $n \times n$

Début

Pour chaque concept c_j depuis la collection C **faire**

Début

Calculer la fréquence de c_j dans D ;

$$tf(c_j) = \sum_{i=1}^n \text{fréquence de } c_j \text{ dans } dc_i$$

Fin pour ;

Calculer le total des fréquences de tous les concepts de C dans D ;

$$tf(D) = \sum_{j=1}^n tf(c_j)$$

Pour chaque concept c_i depuis la collection C **faire**

Début

Calculer le total des fréquences de tous les concepts de C dans dc_i ;

$$\text{fréquence_totale_}dc_i = \sum_{j=1}^n \text{fréquence de } c_j \text{ dans } dc_i$$

Pour chaque concept c_j depuis la collection C **faire**

Début

Calculer $M[i,j]$ le poids TF-ICTF de c_j dans le document dc_i ;

Calculer la fréquence normalisée de c_j dans dc_i ;

$$TF(c_j, dc_i) = \frac{\text{Fréquence de } c_j \text{ dans } dc_i}{\text{fréquence_totale_}dc_i}$$

Calculer ICTF de c_j dans dc_i ;

$$ICTF(c_j) = \log\left(\frac{tf(D)}{tf(c_j)}\right)$$

$$M[i, j] \leftarrow TF(c_j, dc_i) \times ICTF(c_j) ;$$

Fin pour ;

Fin pour ;

Fin.

IV.5.1.3 Calcul de la similarité cosinus entre les concepts

Après l'établissement des modèles vectoriels des différents concepts de la collection, l'étape courante consiste à calculer la proximité sémantique entre eux. Pour ce faire, nous choisissons une mesure très utilisée dans la fouille de textes (Claveau and Nie, 2016), c'est la similarité cosinus, dite aussi cosinus de Salton (Gérard Salton qui a proposé le modèle vectoriel en recherche d'information en 1970). Elle consiste à mesurer le cosinus de l'angle θ entre deux vecteurs $V_{c_i} = (vc_{i1}, vc_{i2}, \dots, vc_{in})$ et $V_{c_j} = (vc_{j1}, vc_{j2}, \dots, vc_{jn})$. Dans notre travail, les éléments d'un vecteur V_{c_i} correspondent aux éléments de ligne i de la matrice de pondération M . La similarité cosinus est calculée par l'équation. (6)

$$\begin{aligned} S(i, j) &= \text{proximité_sémantique}(V_{c_i}, V_{c_j}) \\ &= \cos \theta = \frac{V_{c_i} \times V_{c_j}}{\|V_{c_i}\| \times \|V_{c_j}\|} \end{aligned} \quad (6)$$

Où $S(i, j)$ est la valeur de proximité sémantique entre les deux concepts c_i et c_j (représentés par leurs vecteurs V_{c_i} et V_{c_j} respectivement). Le numérateur de la fraction consiste au produit scalaire de deux vecteurs V_{c_i} et V_{c_j} . Il est donné par l'équation (7)

$$V_{c_i} \times V_{c_j} = (vc_{i1} \times vc_{j1}) + (vc_{i2} \times vc_{j2}) + \dots + (vc_{in} \times vc_{jn}) \quad (7)$$

Le dénominateur représente le produit de deux normes de vecteurs V_{c_i} et V_{c_j} . La norme d'un vecteur V_{c_i} est donnée par l'équation. (8)

$$\|V_{c_i}\| = \sqrt{(vc_{i1}^2 + vc_{i2}^2 + \dots + vc_{in}^2)} \quad (8)$$

Comme la valeur $\cos \theta$ est comprise dans l'intervalle $[-1, 1]$, la valeur -1 signifie que les deux concepts sont contradictoires, 0 signifie l'indépendance, 1 signifie les mêmes et les valeurs intermédiaires permettent d'évaluer le degré de proximité.

La mesure cosinus est largement utilisée dans plusieurs travaux de la littérature pour représenter la proximité sémantique entre des concepts (Hassan and Mihalcea, 2011; Zhiqiang et al., 2009; Gabrilovich and Markovitch, 2007). Notons que le travail présenté dans (Hassan and Mihalcea, 2011), a utilisé une version modifiée de cette mesure.

La raison de choisir la mesure cosinus au lieu des autres distances très utilisées dans littérature telle que la distance de Google qui est utilisée par exemple dans (Yang et al., 2011), et la distance euclidienne qui est utilisée à titre d'exemple dans (Zha et al., 2009) se justifie par le fait que la distance Google est appropriée pour les structures topologiques. Autrement dit, elle convient bien et mieux pour le calcul de similarité sémantique. D'un autre côté, la distance Euclidienne ne reflète pas la distance sémantique correctement. Autrement dit, deux concepts qui sont proches sémantiquement peuvent être considérer le contraire, du fait qu'ils ont une grande distance Euclidienne entre eux. Par contre la distance cosinus reflète cette sémantique adéquatement.

Le résultat de cette étape est une matrice symétrique S , où chaque élément $S(i, j)$ représente la proximité sémantique entre deux concepts c_i et c_j .

Après avoir calculé la proximité sémantique entre les différents concepts, nous allons exploiter cette information sémantique (c.à.d. matrice de proximité sémantique) pour réduire le silence. Un avantage majeur de la méthode proposée, est qu'elle peut réduire le silence dans l'annotation comme dans la recherche. D'une part, elle peut être utilisée pour détecter et accomplir des annotations manquantes dans des bases d'images, et donc enrichir chacune des annotations des images avec plus de concepts pertinents. Elle peut aussi être exploitée pour rechercher plus d'images pertinentes à une requête utilisateur comme nous allons l'expliquer dans le paragraphe qui suit.

IV.5.2 Exploitation de la proximité sémantique pour réduire le silence en recherche d'images

Le but d'un système de recherche d'images par le texte (TBIR) est de localiser le maximum possible d'images pertinentes à une requête utilisateur. Le mécanisme de recherche traditionnelle ne considère une image comme pertinente, sauf si elle est annotée explicitement par les concepts de la requête. Mais comme il est impossible de trouver des bases d'images avec des annotations complètes, ce mécanisme de comparaison binaire conduit au problème de silence. Alors, pour réduire ce problème, nous proposons dans cette section, un autre mécanisme de recherche plus performant. L'avantage majeur de ce mécanisme est qu'il intègre la proximité sémantique entre les concepts.

La phase de recherche est une phase online. Elle commence dès qu'un utilisateur introduit sa requête textuelle. Une requête peut être un texte libre, atomique contenant un seul concept, composée de plusieurs concepts ou bien visuelle contenant une ou plusieurs images annotées. Pour une requête en texte libre il y a toute une panoplie de méthodes de traitement de texte qui ont réussi à faire sortir les concepts clés d'un texte libre (Claveau and Nie, 2016), et donc rendre la requête en texte libre en une requête composée de concepts.

Dans ce présent travail, nous nous sommes limités aux requêtes atomiques textuelles (sous formes d'un concept) ou bien visuelles (une image annotée par un seul concept). Les autres types de requêtes restent comme des perspectives pour un travail futur.

IV.5.2.1 Traitement de requête atomique

Étant donné une requête utilisateur Q qui s'agit d'un seul concept, et une base d'images où chacune des images I est annotée par un ou plusieurs concept (c_1, c_2, \dots, c_k) . Notre moteur de recherche d'images calcule la pertinence de chaque image I par rapport à ce concept requête, en exploitant la proximité sémantique entre les concepts (c_1, c_2, \dots, c_k) annotant cette image, et le concept requête. Le défi, est de trouver la bonne formule qui donne un bon tri des images pertinentes. En fait, après l'expérimentation de plusieurs formules, nous avons obtenu de meilleurs résultats de tri en utilisant la formule suivante :

$$Relevance(I, Q) = \frac{maxRelatedness + meanRelatedness}{2} \quad (9)$$

Tel que, $maxRelatedness$ est la valeur de proximité maximale entre les concepts (c_1, c_2, \dots, c_k) de l'image I et la requête Q , $meanRelatedness$ est leur valeur moyenne. $maxRelatedness$ et $meanRelatedness$ sont définies par les formules (10) et (11) respectivement.

$$MaxRelatedness = \max_{1 \leq i \leq k} S(c_i, Q) \quad (10)$$

$$MeanRelatedness = \frac{\sum_{i=1}^k S(c_i, Q)}{k} \quad (11)$$

Tel que, $S(c_i, Q)$ est la valeur de proximité sémantique entre un concept c_i et la requête Q récupérée depuis la matrice de proximité sémantique S . k est le nombre de concepts annotant l'image I .

Après le calcul des valeurs de pertinence de toutes les images de la base. Le système trie les images par ordre décroissant de pertinence, et affiche les N premières.

Le processus de recherche est effectué selon l'Algorithme 2.

Algorithme 2 Procédure Recherche d'images par requête atomique

Input: S : la matrice de proximité sémantique, Bd : une base d'images annotées, Q : un concept requête

Output: des images pertinentes à Q

Début

Pour chaque image I depuis Bd , annotée par les concepts (c_1, c_2, \dots, c_k) **faire**

Début

Calculer $Relevance(I, Q)$ la valeur de pertinence de l'image I par rapport au concept requête Q

Depuis la matrice S

Trouver la valeur de proximité maximale **MaxRelatedness** entre les concepts ($c_1,$

c_2, \dots, c_k) de I et le concept requête Q ;

$$\mathbf{MaxRelatedness} = \max_{1 \leq i \leq k} S(c_i, Q)$$

Calculer la valeur de proximité moyenne **MeanRelatedness** entre les concepts

(c_1, c_2, \dots, c_k) et le concept Q ;

$$\mathbf{MeanRelatedness} = \frac{\sum_{i=1}^k S(c_i, Q)}{k}$$

$$\mathbf{Relevance}(I, Q) = \frac{\mathbf{maxRelatedness} + \mathbf{meanRelatedness}}{2}$$

Fin pour ;

Trier les images par ordre décroissant de pertinence ;

Afficher à l'utilisateur les N premières images qui ont les valeurs de pertinence les plus élevées

Fin.

IV.5.2.2 Traitement de requête visuelle

Nous nous sommes limités à une requête visuelle qui contient une seule image annotée, choisie parmi les images de la base. En fait, le traitement de cette requête n'est qu'un traitement d'une requête textuelle composée d'un ou de plusieurs concepts. C'est par ce qu'un TBIR ne considère pas l'image mais plutôt les annotations qui lui sont associées.

IV.5.3 Exploitation de la proximité sémantique pour réduire le silence dans l'annotation

Étant donné une base d'images légèrement annotée, nous pouvons exploiter la proximité sémantique entre les concepts afin de détecter le silence dans les annotations des images. Autrement dit, détecter des concepts pertinents pour certaines images, mais qui sont absents dans leurs annotations. Par conséquent, nous allons réduire ce silence, par l'ajout des concepts manquants dans les annotations appropriées. Par exemple, pour détecter si un concept c_i est pertinent pour certaines images, mais qu'il est absent dans leurs annotations,

nous exploitons notre mécanisme de recherche. C'est-à-dire, nous recherchons les N premières images pertinentes à c_i . Parmi ces images, il y a celles qui ne contiennent pas c_i dans leurs annotations, alors elles seront enrichies par ce concept c_i

Le processus d'enrichissement des annotations des images d'une base se fait selon l'Algorithme 3.

Algorithme 3 Procédure Enrichissement annotation

Input: S : La matrice de proximité sémantique, Bd : une base d'images annotées, $C = \{c_1, c_2, \dots, c_i, \dots, c_n\}$: une collection de concepts

Output: La base d'images Bd avec plus d'annotations

Début

Pour chaque concept c_i depuis C **faire**

Début

 Lancer la recherche avec c_i

 Appel à la procédure Recherche d'images par requête atomique

Pour chaque image I des N premières images retournées **faire**

 Ajouter le concept c_i à l'annotation de I

Fin pour ;

Fin.

IV.6 Conclusion

Dans ce chapitre, nous avons explicité la méthode que nous proposons pour réduire le silence dans un système de recherche d'images par le texte (TBIR), et ceci dans l'annotation tout comme dans la recherche. L'idée principale de cette méthode est de calculer la proximité sémantique entre les concepts et de l'exploiter au moment de la recherche. Nous avons montré que notre méthode de calcul de proximité sémantique entre les concepts est basée sur des statistiques faites sur le contenu des articles Wikipedia, une source de connaissances partagée à travers le monde. Ainsi, elle est indépendante de toute base d'images locales. En outre, pour inférer la pertinence d'une image, notre mécanisme de recherche ne s'est pas limité à une relation sémantique prédéfinie, mais plutôt il exploite les valeurs de proximité sémantique entre la requête et l'annotation de l'image.

Chapitre V. Résultats, Evaluation et Validation de la méthode basée proximité sémantique

V.1 Introduction

L'objectif de ce chapitre est d'évaluer la méthode proposée, afin de prouver son efficacité pour réduire le silence dans un système TBIR. Ainsi, Nous créons un système TBIR basé sur notre conception du calcul de proximité sémantique. Il prend comme entrée une requête textuelle atomique, composée de plusieurs concepts, ou bien une requête visuelle qui contient une image annotée choisie parmi les images de la base. Ensuite, il donne comme résultat un ensemble d'images triées par ordre décroissant de pertinence.

Ce chapitre est divisé en deux parties, la configuration expérimentale et les résultats expérimentaux. Dans la première partie, nous présentons les conditions expérimentales avec lesquelles nous avons réalisé nos expériences, impliquant la base d'images, le benchmark (ou vérité terrain) de proximité sémantique et les mesures de performance utilisées. Dans la deuxième partie, nous rapportons nos résultats, accompagnés d'une analyse et une discussion des constatations obtenues. Cette deuxième partie est divisée, à son tour, en deux sections. En première section, nous reportons nos calculs de proximité sémantique ainsi que l'évaluation des valeurs obtenus par rapport au jugement humain et par rapport à quelques travaux de la littérature. Alors que la deuxième section est consacrée à la démonstration de nos résultats de

recherche d'images et d'annotation ainsi que l'analyse et la discussion de performance de la méthode proposée par rapport à quelques travaux récents de la littérature.

V.2 La configuration expérimentale

Nous explicitons dans cette partie les conditions expérimentales avec lesquelles nous avons réalisé nos expériences.

V.2.1 La base d'images COREL 5K

La base d'images COREL 5K (Duygulu et al., 2002; Gong et al., 2010) est une base d'images très utilisée partout dans le monde. Elle est souvent utilisée dans le contexte de l'annotation ainsi que la recherche d'images. Elle est composée de 5000 images décrivant plusieurs aspects de la vie quotidienne comme les gens, les paysages, les sports,...Etc. Ces images sont annotées avec un nombre total de 374 concepts sémantiques dans des différents contextes, de tel sorte que le nombre de concepts d'annotation par image varie de un à cinq concepts (Gong et al., 2010). La base d'images COREL 5K est partitionnée en un ensemble d'apprentissage composé de 4500 images et un ensemble de tests avec 500 images. La Figure 10 montre quelques images représentatives de COREL 5K.

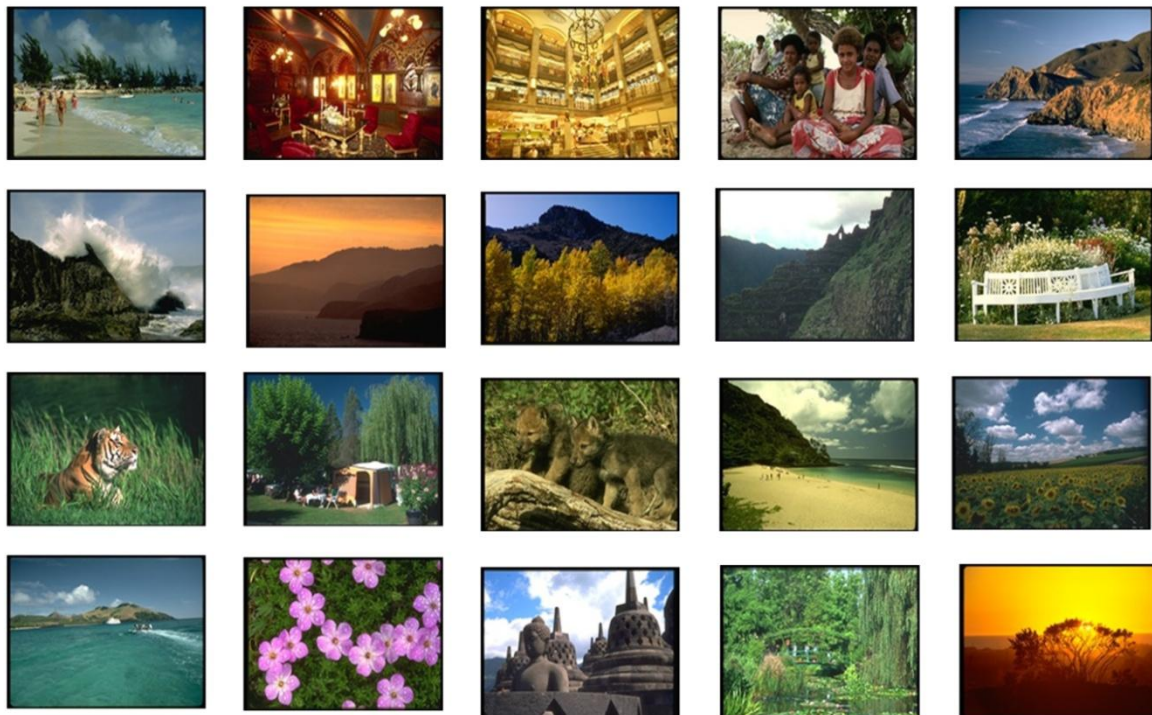


Figure 10. Quelques images représentatives de COREL 5K

Pour évaluer notre méthode, nous avons utilisé l'ensemble d'images du COREL 5K destiné à l'apprentissage. Cet ensemble a été utilisé par plusieurs travaux de la littérature qui ont focalisé soit sur l'annotation ou bien sur la recherche (Chen et al., 2010; Wu et al., 2013; Gong et al., 2010; Duygulu et al., 2002; Cui et al., 2015; Xu et al., 2016; Bao et al., 2011). En fait, ils ont supposé que les annotations de cet ensemble d'apprentissage sont complètes, alors qu'elles ne le sont pas.

V.2.2 Le benchmark de proximité sémantique WordSimilarity-353

Pour évaluer la performance de notre méthode de calcul de proximité sémantique entre les concepts, nous devons comparer les valeurs obtenues avec une vérité terrain. Pour cela, nous avons utilisé un benchmark de jugement humain disponible sur le web, c'est le WordSimilarity-353 (Finkelstein et al., 2001). C'est une collection de 353 paires de concepts avec valeurs de proximité sémantique basées sur le jugement humain, publiée en 2001. En effet, pour chaque paire de concepts, 13 à 16 personnes ont été demandées de donner des valeurs de proximité sémantiques entre 0 (pour indiquer deux concepts non liés) et 10 (pour indiquer deux concepts très proches ou identiques). Ensuite, la valeur de proximité pour une paire de concepts s'obtient par la moyenne des valeurs données. La table 1. donne un aperçu de quelques paires de ce benchmark

Paire de concepts	Moyenne des jugements humains
<i>(tiger, cat)</i>	7.35
<i>(book, paper)</i>	7.46
<i>(computer, internet)</i>	7.58
<i>(student, professor)</i>	6.81
<i>(baby, mother)</i>	7.85
<i>(car, automobile)</i>	8.94
<i>(coast, shore)</i>	9.10
<i>(bird, cock)</i>	7.10
<i>(coast, forest)</i>	3.15
<i>(tiger, carnivore)</i>	7.08
<i>(jaguar, cat)</i>	7.42
<i>(news, report)</i>	8.16
<i>(hospital, infrastructure)</i>	4.63

<i>(life, death)</i>	7.88
<i>(summer, nature)</i>	5.63

Table 1. Valeurs de proximité sémantiques, pour quelques paires de concepts dans le benchmark WordSimilarity-353.

Toutefois, il ya d'autres benchmarks disponibles tel que le benchmark Mturk-771(Halawi et al., 2012), qui est une collection de 771 paires de concepts publiée en 1 février 2012. Comme il ya d'autres qui sont non accessibles, tels que le benchmark de Rubenstein et Goodenough R&G(Rubenstein and Goodenough, 1965) qui contient 65 paires de concepts, M&C(Miller and Charles, 1991)qui est un sous ensemble du benchmark R&G qui comprend 30 paires de concepts.

V.2.3 Mesures de performances

V.2.3.1 Métrique de performance de calcul de proximité sémantique entre les concepts

Pour mesurer la performance de notre méthode de calcul de proximité sémantique par rapport au benchmark de jugement humain WordSimilarity-353, nous avons calculé le coefficient de corrélation de Pearson ρ (en anglais Pearson's correlation coefficient) entre nos valeurs de proximité sémantique, et celle du benchmark.

Le coefficient de corrélation de Pearson ρ permet de mesurer la corrélation linéaire entre deux variables X et Y . Il varie entre -1 et 1. La valeur 1 signifie une corrélation positive parfaite entre X et Y , une valeur de -1 implique une corrélation négative parfaite et une valeur de 0 signifie qu'il n'y a pas de corrélation linéaire entre les deux variables. Sa formule est donnée comme suit:

$$\rho = \frac{1}{n} \times \sum_{i=1}^n \frac{(X_i - \mu_X)}{\sigma_X} \times \frac{(Y_i - \mu_Y)}{\sigma_Y} \quad (12)$$

Où les valeurs X_i et Y_i correspondent aux valeurs des variables X et Y respectivement. μ_X, σ_X et μ_Y, σ_Y dénotent la moyenne et l'écart type de X et Y respectivement.

Dans notre cas, les valeurs de la variable X correspondent à nos valeurs de proximité sémantiques, et celles de Y correspondent aux valeurs de jugement humain WordSimilarity-353. Il est évident que plus le coefficient de corrélation ρ est élevé par rapport à WordSimilarity353, plus la méthode de calcul de proximité sémantique entre les concepts est précise (ou appropriée).

V.2.3.2 Métrique de performance de la recherche d'images

La métrique que nous avons utilisée pour mesurer la performance de notre méthode de recherche d'images est la précision. Elle est définie par:

$$\text{Précision} = \frac{\text{nombre d'images pertinentes retournées}}{\text{nombre total d'images retournées}} \quad (13)$$

V.3 Résultats expérimentaux

Dans cette partie nous allons commencer par rapporter nos calculs de proximité sémantique entre les concepts, l'évaluation des résultats obtenus par rapport au benchmark de jugement humain WordSimilarity-353, ainsi que l'analyse et la discussion de performance de la méthode adoptée par rapport à quelques travaux connexes. Ensuite, Nous allons présenter nos résultats pour la recherche ainsi que pour l'accomplissement d'annotation d'images, suivi par leurs évaluations, en termes de précision, par rapport à quelques travaux de la littérature.

V.3.1 Proximité sémantique entre les concepts

Dans cette section, nous présentons nos résultats de calcul de proximité sémantique, leur évaluation par rapport à WordSimilarity-353, ainsi que leur comparaison avec les résultats de quelques travaux connexes.

V.3.1.1 Calcul de proximité sémantique entre les concepts

Pour calculer automatiquement la proximité sémantique entre des concepts, nous avons utilisé des articles Wikipedia comme une source de connaissance externe. Nous avons sélectionné la version anglaise de Wikipedia. Ce choix se justifie par le fait que tous les benchmarks de proximité sémantiques, que nous avons trouvé, sont en anglais. Pour la collection des concepts C , nous avons utilisé l'ensemble de 374 concepts utilisés dans l'annotation de la base d'images COREL 5K. Notons que notre méthode de calcul peut mesurer la proximité sémantique pour n'importe quelles paires de concepts du monde réel. Le calcul de la proximité sémantique entre les concepts du COREL 5K se fait selon la méthode décrite dans le chapitre précédent qui est constituée de trois étapes :

- Dans la première étape, nous avons construit une collection de documents de 374 articles Wikipedia correspondants aux concepts du COREL 5K.
- Dans la deuxième étape, nous avons calculé la matrice M des poids de pondération des concepts. Elle est de taille 374×374 .
- Dans la troisième étape, nous avons calculé une matrice S de proximité sémantique entre les concepts de COREL 5K. La matrice S est une matrice symétrique de taille 374×374 .

La table. 2montre les valeurs de proximité sémantique obtenues pour quelques paires de concepts.

Concepts	<i>City</i>	<i>Mountain</i>	<i>Sky</i>	<i>Sun</i>	<i>Water</i>	<i>Clouds</i>	<i>Tree</i>	<i>Lake</i>	<i>Sea</i>	.
<i>City</i>	1	0.335	0.29	0.288	0.232	0.364		0.176	0.386	.
<i>Mountain</i>	0.335	1	0.309	0.287	0.232	0.362	0.292	0.238	0.384	.
<i>Sky</i>	0.29	0.309	1	0.527	0.203	0.514	0.228	0.151	0.31	.
<i>Sun</i>	0.288	0.287	0.527	1	0.185	0.408	0.23	0.132	0.335	.
<i>Water</i>	0.232	0.232	0.203	0.185	1	0.283	0.186	0.27	0.419	.
<i>Clouds</i>	0.364	0.362	0.514	0.408	0.283	1	0.292	0.189	0.387	.
<i>Tree</i>	0.254	0.292	0.228	0.23	0.186	0.292	1	0.131	0.243	.
<i>Lake</i>	0.176	0.238	0.151	0.132	0.27	0.189	0.131	1	0.251	.
<i>Sea</i>	0.386	0.384	0.31	0.335	0.419	0.387	0.243	0.251	1	.

.....
-------	---	---	---	---	---	---	---	---	---	---

Table 2. Valeurs de proximité sémantique pour quelques paires de concepts du COREL 5K.

V.3.1.2 Evaluation des résultats de proximité sémantique par rapport à WordSimilarity-353

L'évaluation de la performance de notre méthode de calcul de proximité sémantique entre les concepts se fait par le calcul du coefficient de corrélation de Pearson ρ entre nos valeurs résultats de proximité sémantiques et celles du benchmark WordSimilarity-353 (Finkelstein et al., 2001).

Dans la table 3., nous présentons pour quelques paires de concepts, les valeurs de proximité sémantiques dans WordSimilarity-353 et celles obtenues par notre méthode ainsi que la valeur obtenue du coefficient ρ .

Pairs de concepts	Valeurs de WS-353 (/10)	Nos résultats
<i>(wood, forest)</i>	0.773	0.463
<i>(coast, forest)</i>	0.315	0.143
<i>(coast, shore)</i>	0.91	0.785
<i>(tiger, cat)</i>	0.735	0.335
<i>(tiger, tiger)</i>	1	1
<i>(food, fruit)</i>	0.752	0.326
<i>(canyon, landscape)</i>	0.753	0.238
<i>(man, woman)</i>	0.83	0.867
Coefficient de corrélation de Pearson ρ	1	0.79

Table 3. La corrélation linéaire entre WordSimilarity-353 et nos valeurs de proximité sémantique, calculée par le coefficient ρ .

Depuis cette table, il est clair que la corrélation linéaire entre nos valeurs de proximité sémantiques et celles de WordSimilarity353 est bonne ($\rho = \mathbf{0.79}$ c'est presque une corrélation à

80%). Ce qui signifie que notre méthode de calcul de proximité sémantique est proche de jugement humain.

V.3.1.3 Evaluation des résultats de proximité sémantique par rapport à quelques travaux connexes

Comme nous l'avons mentionné dans le chapitre de l'état de l'art, les méthodes de calcul de proximité sémantique entre les concepts, basées Wikipedia, sont de deux familles, des méthodes topologiques exploitant l'architecture hiérarchique des catégories des Wikipedia ou bien les liens Wikipedia, et des méthodes statistiques utilisant des calculs statistiques sur le contenu du Wikipedia. Ainsi, les valeurs de proximité sémantique obtenues par les méthodes de la première famille ne peuvent pas être comparées avec ceux obtenues par les méthodes de la deuxième famille. C'est par ce que le principe de calcul est complètement différent. En fait, les premières calculent une instance, tandis que les deuxièmes calculent une similarité. Par conséquent, pour la première famille, plus la distance est faible, plus la similarité sémantique est grande. En revanche, pour la deuxième famille, plus la mesure est grande, plus il y a de similarité sémantique.

Comme notre méthode s'inscrit dans la deuxième famille, nous l'avons comparé avec des méthodes de cette famille. Ainsi, nous avons choisi les méthodes ESA et SSA, qui sont présentées dans (Gabrilovich and Markovitch, 2007) et (Hassan and Mihalcea, 2011) respectivement. Ces deux travaux ont utilisé le même benchmark que nous, c.à.d. WordSimilarity-353, pour évaluer leurs résultats de proximité sémantique. Ainsi, pour évaluer la performance de nos calculs par rapport à ces méthodes, nous comparons les valeurs du coefficients ρ obtenus par chacune de méthodes. Les valeurs de ρ : sont **0.5** pour la méthode ESA(Gabrilovich and Markovitch, 2007), **0.59** pour la méthode SSA(Hassan and Mihalcea, 2011) et **0.79** pour nous.

Depuis ces valeurs, il est clair que la valeur de ρ obtenue par notre méthode dépasse significativement celles des deux autres méthodes. Ceci confirme la performance (puissance) de notre méthode pour capturer la proximité sémantique d'une façon plus appropriée et plus précise que les autres.

En fait, la performance de notre méthode par rapport aux deux méthodes ESA et SSA, qui sont étroitement liées l'une à l'autre, est due à la mesure de pondération des concepts utilisée. Autrement dit, nous avons utilisé la méthode TF_ICTF, alors qu'ESA et SSA ont utilisé la méthode TF_IDF. Ceci confirme notre justification du choix de TF_ICTF que nous avons présenté dans le chapitre précédent.

V.3.2 Recherche et annotation d'images





Dans cette section, nous montrons nos résultats pour la recherche ainsi que pour l'accomplissement d'annotation d'images, suivi par leurs évaluation, en terme de précision, par rapport à quelques travaux de l'état de l'art.

V.3.2.1 Résultats de recherche

Une fois que l'utilisateur introduit sa requête textuelle depuis la collection des concepts C , notre système de recherche d'images calcule la pertinence de chacune des images I de la base d'images COREL 5K, comme décrit dans le chapitre précédent. Ensuite, il retourne les 320 premières images après un tri par ordre décroissant de pertinence.

La vérité terrain de pertinence d'une image est obtenue par la collection des votes de 100 étudiants qui ont participé à cette expérience, et qui sont familiers à notre système de recherche d'images, de tel sorte que chaque participant vote chaque image comme pertinente ou non, en se basant sur son jugement personnel. Nous considérons alors une image comme pertinente, si elle a des votes de pertinence plus que ceux de non pertinence, et vice versa. Les auteurs de (Franzoni et al., 2015) ont utilisé une méthode semblable pour obtenir une vérité terrain de similarité sémantique pour chaque paire d'images (ils ont l'appelé Human Similarity Evaluation HSE en anglais) afin d'évaluer leur méthode de calcul de similarité sémantique entre les images.








La figure 11. montre un échantillon d'images que notre système a pu retourner suite aux requêtes atomiques '*City*', '*Forest*', '*Sunset*' et '*Desert*' respectivement.

Images							
Annotation courante dans Corel 5K	buildings	people	people	People	people, buildings	sky, buildings	tree, buildings








(a)

Images							
Annotation courante dans Corel 5K	tree	tree, people	tree, people	tree, grass	tree, sculpture	tree, road	tree, maui

(b)

Images							
Annotation courante dans Corel 5K	sun, sea	sun, horizon, sunrise	sky, sun, clouds	sky, sun, tree	sun, clouds, boats	sun, clouds, sea	Mount-ain, sun, clouds

(c)

Images							
Annotation courante dans Corel 5K	sky, sand dunes	sand, dunes	sky,hills, sand,dunes	sky, tree,sand, dunes	moutain, sand, valley, dunes	sky, hills, dunes	rocks, sand

(d)

Figure 11. Exemples d'images retournées depuis la base d'images Corel 5K avec leurs annotation correspondantes, suite aux requêtes : (a) City, (b) Forest, (c) Sunset et (d) Desert.






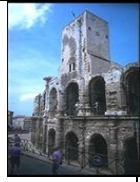

Cette figure montre clairement la capacité de notre système de détecter des images pertinentes qui ne sont pas annotées explicitement par une requête, mais annotées par des concepts proches sémantiquement à cette requête. Par exemple, si on introduit le concept '*desert*' comme requête atomique, nous pouvons constater qu'il ya des images pertinentes retournées qui ne sont pas annotées explicitement par '*desert*', mais annotées avec des concepts proches sémantiquement de ce concept tel que '*dunes*', '*rocks*', '*sand*'. Ces images ont été ignorées lorsque nous avons appliqué la comparaison binaire. De même, nous remarquons que les annotations associées aux images retournées sont très pertinentes à leurs requêtes. Cela confirme la force de la méthode proposée et sa capacité de réduire le silence. De même, à partir des résultats, il semble que la base d'images Corel 5K contient des manques dans ses annotations, ce qui contredit l'hypothèse sur laquelle se basent de nombreuses approches d'annotation utilisant cette base d'images, que ses annotations sont complètes. En fait, il est extrêmement difficile d'avoir des bases d'images avec des annotations complètes.

V.3.2.2 Résultats d'accomplissement d'annotation








Comme nous avons montré dans la section précédente, notre système est capable de détecter des images pertinentes qui ne sont pas annotées explicitement avec un concept requête, mais annotées par des concepts proches sémantiquement à cette requête. Alors, si nous exploitons cette capacité du point de vue annotation, nous pouvons confirmer que via le mécanisme de recherche proposé, nous pouvons détecter des concepts manquants dans les annotations des images (ce sont ceux introduits comme des requêtes). Autrement dit, détecter

des concepts pertinents pour certaines images, mais qui sont absents dans leurs annotations. Par conséquent, nous pouvons réduire ce silence, par l’ajout des concepts manquants dans les annotations appropriées. Pour ce faire, nous avons lancé la recherche avec chacun des 374 concepts de la base d’image Corel 5K afin de trouver toutes les images (et donc toutes les annotations) où le concept requête est pertinent mais absent. Par conséquent, nous accomplissons les annotations de ces images par ce nouveau concept.

Prenons le même exemple précédent, nous introduisons les concepts ‘City’, ‘Forest’, ‘Sunset’ et ‘Desert’ comme requêtes afin de détecter et donc compléter des annotations manquantes. La figure 12. montre les annotations ajoutées aux images, elles sont en gras.

Images							
Annotation courante dans Corel 5K	building, city	People, city	People, city	people, city	people, buildings, city	sky, buildings, city	tree, buildings, city

(a)








Images							
Annotation courante dans Corel 5K	Tree, forest	tree, people, forest	tree, people, forest	tree, grass, forest	tree, sculpture, forest	tree, road, forest	tree,maui, forest

(b)

Images							
--------	---	---	---	---	--	---	---

Annotation courante dans Corel 5K	sun, sea, sunset	sun, horizon, sunrise, sunset	sky, sun, clouds, sunset	sky, sun, tree, sunset	sun, clouds, boats, sunset	sun, clouds, sea, sunset	mountain, sun, clouds, sunset
--	----------------------------	---	---------------------------------------	-------------------------------------	--	--	---

(c)

Images							
Annotation courante dans Corel 5K	sky, sand dunes, desert	sand, dunes, desert	sky,hills,s and, dunes, desert	sky, tree, sand, dunes, desert	moutain, sand, valley, dunes, desert	sky, hills, dunes, desert	rocks, sand, desert

(d)

Figure 12. Exemples d'enrichissement d'annotation pour des images de Corel 5K. Les nouvelles annotations sont en gras.

V.3.2.3 Comparaison avec quelques méthodes de l'état de l'art

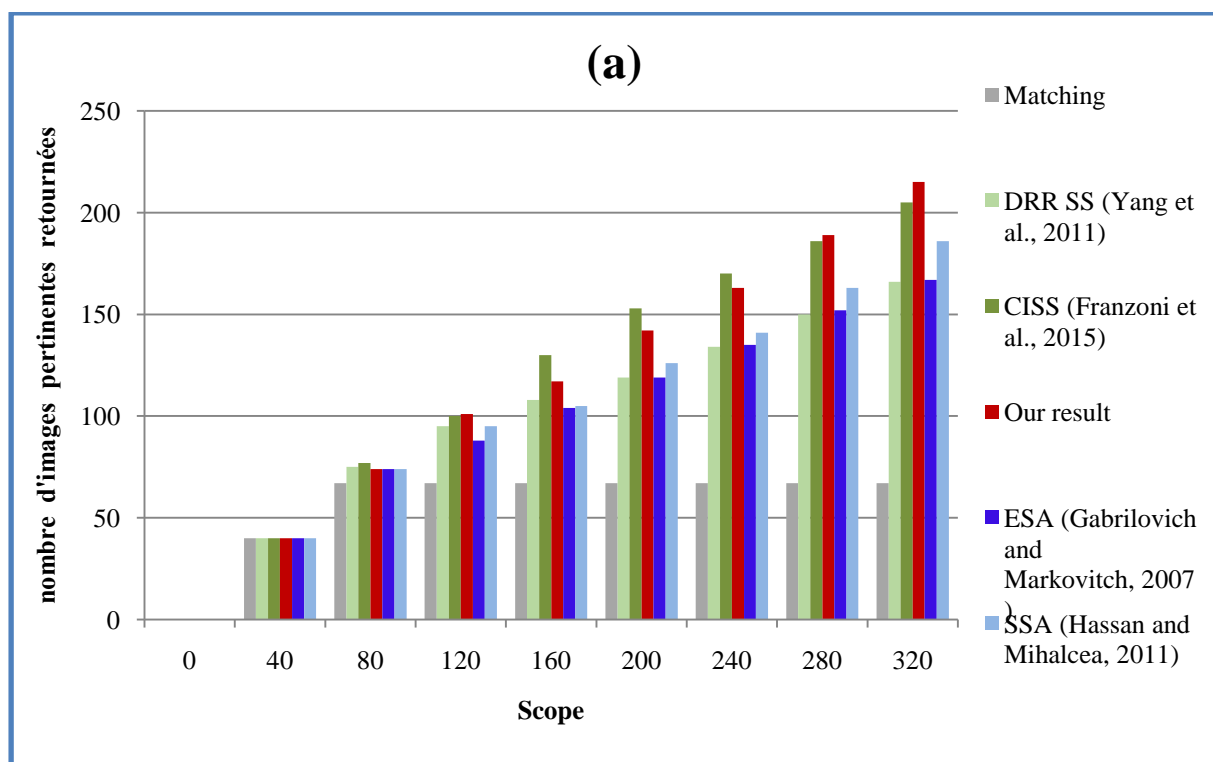
L'objectif de cette section est de comparer la performance de la méthode proposée avec quelques méthodes de l'état de l'art. Ainsi, nous avons comparé avec :

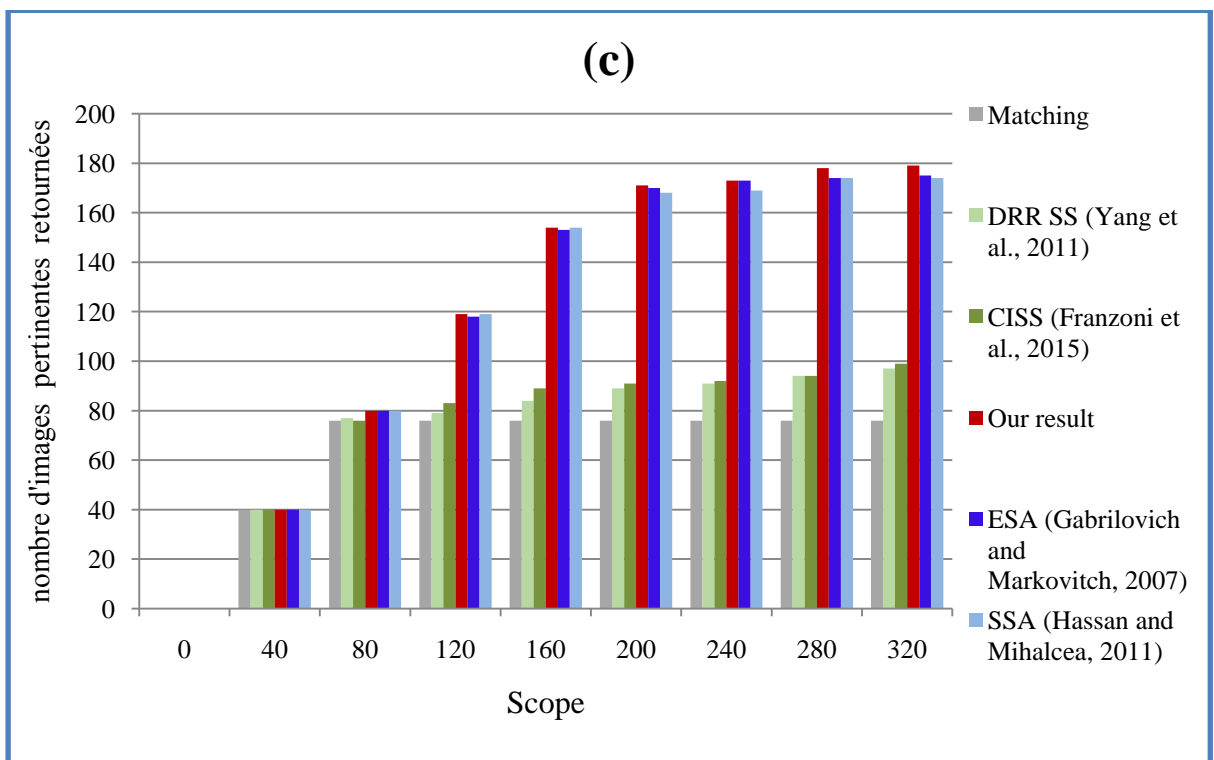
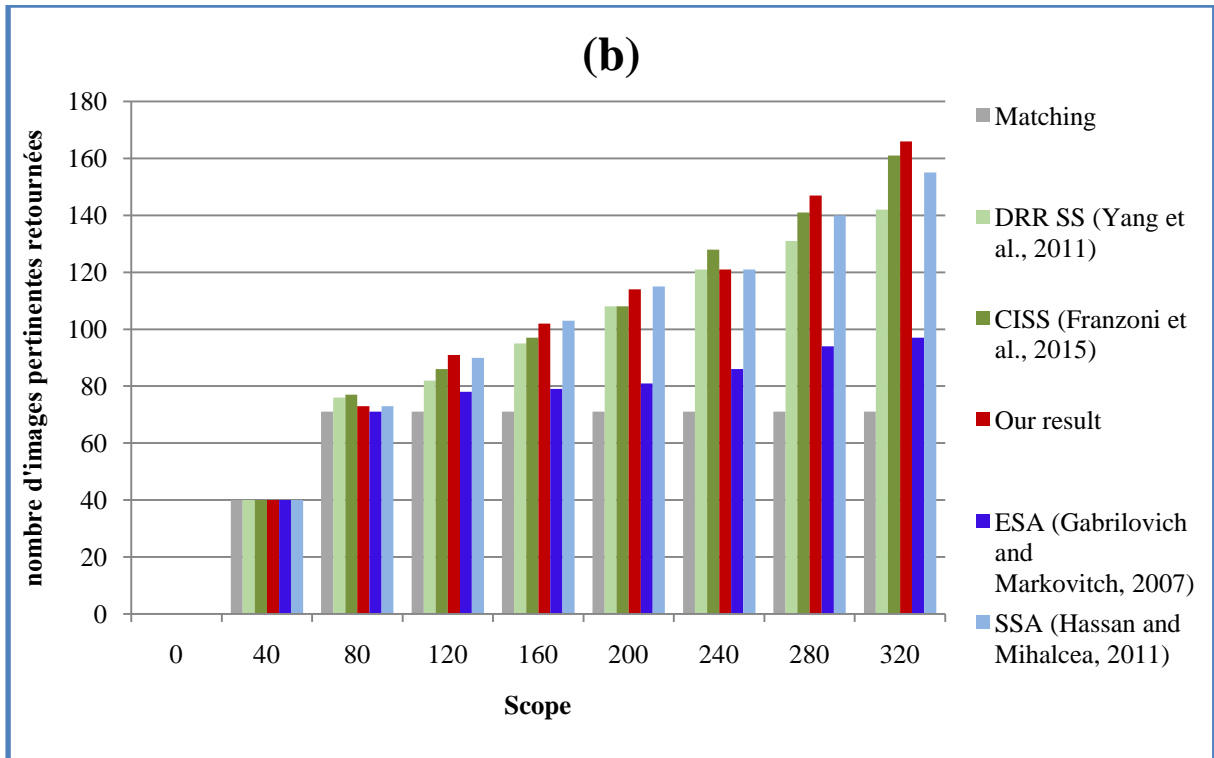
- La méthode de base de recherche d'images via la comparaison binaire.
- Deux méthodes basées corpus local « Diverse Relevance Ranking with Semantic Similarity (DRR SS) » présentée dans le travail (Yang et al., 2011), et « Context-based Image Semantic Similarity (CISS) » présentée dans (Franzoni et al., 2015). Ces deux méthodes exploitent la cooccurrence entre les concepts au sein des annotations d'une base d'images pour calculer la proximité sémantique entre les concepts.
- Deux méthodes basées Wikipedia « Explicit Semantic Analysis (ESA) (Gabrilovich and Markovitch, 2007) et « Salient Semantic Analysis (SSA) » (Hassan and Mihalcea, 2011). Ces

deux travaux focalisent sur la problématique de calcul de la proximité sémantique entre les concepts, indépendamment du domaine d'application. Pour ce faire, les auteurs ont utilisé le Wikipedia comme une source externe de connaissance. Ces deux méthodes sont très proches l'une à l'autre. Ainsi, nous avons implémenté leurs méthodes de calcul de proximité sémantique. Ensuite nous avons utilisé les résultats obtenus pour la recherche d'images, en utilisant notre moteur de recherche.

Le nombre de bonnes images retournées, qui ne sont pas annotées avec le concept requête, représente la valeur d'amélioration.

La figure .13 montre le résultat obtenu, en utilisant les requêtes : 'city', 'forest', 'sunset' et 'desert' respectivement. Cette figure trace le nombre d'images correctes récupérées, en termes de scope, par notre méthode et les autres méthodes. Autrement dit, c'est la précision.





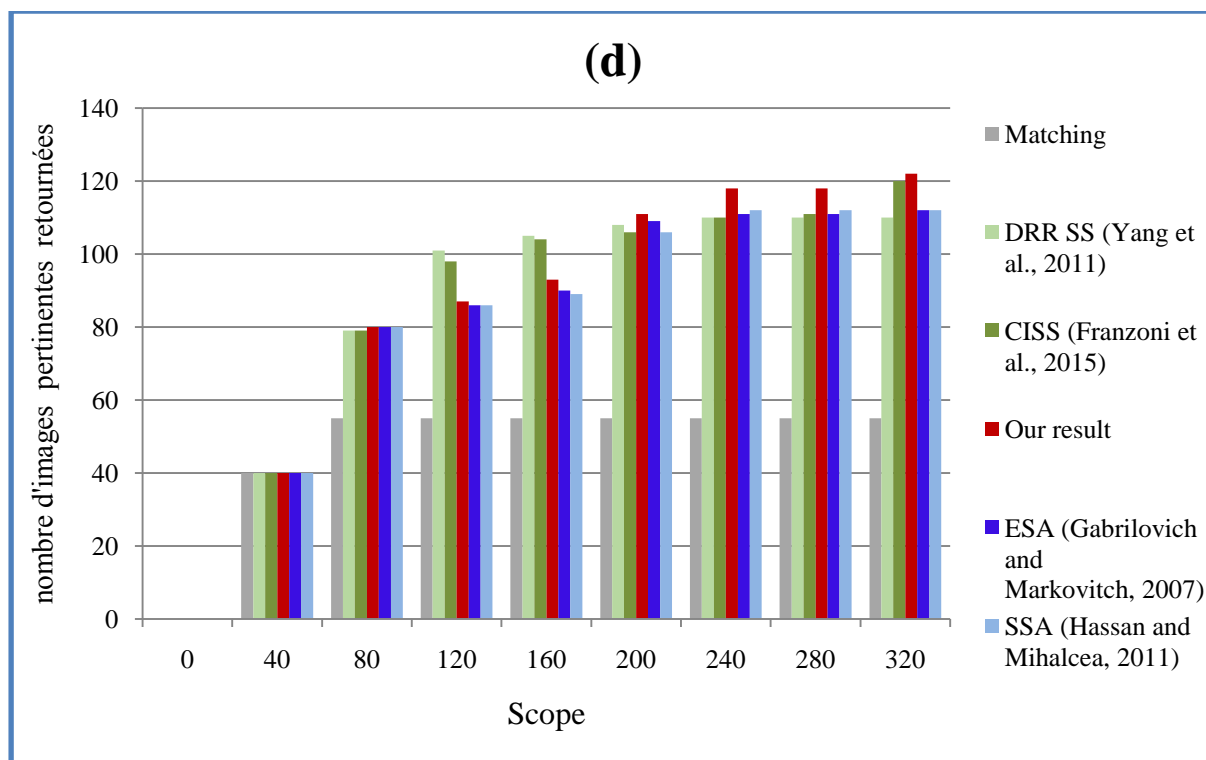


Figure 13. Une comparaison des résultats de recherche obtenus par les différentes méthodes, en utilisant les requêtes : ‘city’, ‘forest’, ‘sunset’ et ‘desert’ respectivement.

Selon les résultats obtenus, il est apparent que notre méthode est capable de détecter plus d’images pertinentes que la comparaison binaire et les deux méthodes utilisant des bases d’images locales. Par exemple, si nous prenons le concept ‘sunset’ comme requête, nous pouvons constater que le nombre d’images pertinentes retournées au scope 280 égale à 76 pour la comparaison binaire, 94 pour les deux méthodes DRR SS et CISS, alors qu’il est presque deux fois plus grand (c’est-à-dire 178) pour la méthode proposée. Également, la méthode proposée a rivalisé avec les méthodes basées-Wikipedia (méthodes ESA et SSA) et les a dépassées au scope 280. De plus, nous constatons qu’au scope 280, le nombre d’images pertinentes récupérées par notre méthode dépasse celui de la comparaison binaire, DRR SS, CISS, ESA et SSA, et ceci pour toutes les requêtes restantes.

Nous notons que la comparaison binaire a récupéré toutes les images de la base qui sont annotées avec le concept requête. Cependant, elle a ignoré un nombre important d’autres images pertinentes, car elles ne sont pas annotées explicitement avec le concept requête. En revanche, notre méthode prouve sa capacité à récupérer plus d’images pertinentes, même si le concept requête est absent dans leurs annotations.

De plus, notre méthode a dépassé les méthodes DRR SS et CISS pour deux raisons. Tout d'abord, elle a calculé la proximité sémantique entre les concepts à partir de Wikipedia, tandis que DRR SS et CISS ont exploité un corpus local (c.à.d. les annotations d'une base d'images). Cela confirme que l'utilisation d'un corpus local ne suffit pas pour calculer la proximité sémantique entre les concepts. Deuxièmement, notre méthode a utilisé le schéma de pondération TF_ICTF et la mesure de similarité cosinus alors que DRR SS a utilisé Google distance, qui est une distance plus appropriée pour calculer la similarité sémantique entre des concepts dans des structures hiérarchiques. La méthode CISS donne des résultats plus pertinents par rapport à la méthode DRR SS car elle a utilisé la confiance mutuelle MC (mutual confidence en anglais), une mesure statistique de proximité entre des concepts basée sur la probabilité conditionnelle.

En outre, notre système de recherche d'images a détecté plus d'images en silence lorsqu'il utilise nos résultats de calcul de proximité sémantique (c.à.d. notre matrice de proximité sémantique), que lorsqu'il utilise les résultats de la méthode ESA ou SSA. Cela confirme la capacité de notre méthode à mieux capturer la proximité sémantique entre les concepts.

V.4 Conclusion

Dans ce chapitre, nous avons testé l'efficacité de notre méthode pour réduire le silence dans un système de recherche d'images par le texte (TBIR), et ceci dans l'annotation comme dans la recherche, et nous avons rapporté les résultats expérimentaux. Ainsi, nous avons commencé par présenter la configuration expérimentale y compris la base d'images, le benchmark de proximité sémantique et les mesures de performance utilisées. Après, nous avons rapporté nos résultats de calcul de proximité sémantique ainsi que leur évaluation par rapport au jugement humain et par rapport à quelques travaux de la littérature. Par la suite, nous avons récapitulé nos résultats de recherche d'images et d'annotation, ainsi que leur comparaison par rapport à la recherche traditionnelle et quelques travaux récents de la littérature. Selon les résultats obtenus, il est clair que notre méthode est capable de détecter et de réduire le silence mieux que les autres méthodes. Autrement dit, notre méthode est plus performante et plus efficace. De plus, depuis cette présentation, il est clair que les contributions visées sont bien achevées. Ainsi, le calcul de proximité sémantique est plus fiable et plus performant par rapport à des méthodes de l'état de l'art, en raison du fait qu'il

était indépendant de toute base d'images locales, et que les résultats obtenus étaient plus proches du jugement humain. De plus, l'intégration de la proximité sémantique entre les concepts au sein du mécanisme de recherche a permis de localiser plus d'images pertinentes qui ont été complètement ratées par la comparaison binaire. En outre, le mécanisme de recherche proposé est plus souple et plus consistant. Plus souple du fait qu'il n'est pas limité à aucune relation prédéfinie mais plutôt reflète la sémantique derrière n'importe quelle relation, et plus consistant du fait que les formules de recherches développées sont indépendantes de tout type de relation.

Conclusion générale et perspectives

L'image est une source de connaissance très utilisée grâce à son expressivité. À cause de cela, les utilisateurs ont souvent besoin de rechercher des images qui répondent à un besoin particulier. Cependant, la popularité des appareils numériques ainsi que la démocratisation d'Internet ont conduit à une prolifération explosive du nombre d'images, que ce soit pour des collections personnelles, professionnelles ou dans le web. Par conséquent, la recherche manuelle des images est devenue laborieuse et très loin de satisfaire les besoins utilisateurs du point de vue précision et temps de réponse. A cet effet, des systèmes de recherche d'images ont été introduit afin d'automatiser cette tâche et satisfaire ces besoins.

Un système de recherche d'images par le texte (TBIR), dit Text-Based Image Retrieval en anglais, permet d'identifier des images pertinentes à une requête particulière, via une comparaison binaire entre la requête et les annotations associées aux images d'une base. Ainsi, une image annotée explicitement par la requête est considérée comme pertinente et donc retournée à l'utilisateur, et vice-versa. Toutefois, la comparaison binaire avec des annotations manuelles qui sont généralement incomplètes peut mener le système à ignorer plusieurs images pertinentes parce que tout simplement les concepts recherchés ne figurent pas explicitement dans leurs annotations. Ce problème, auquel nous nous attaquons dans cette thèse, est communément appelé « le silence ». Il dégrade considérablement les performances du TBIR.

Dans la littérature, plusieurs tentatives ont été faites pour minimiser le problème de silence. Certains travaux ont focalisé sur l'accomplissement des annotations manquantes. Par conséquent, quand la base d'images est très bien annotée, la comparaison binaire ne conduit pas au problème du silence. Alors que d'autres travaux se sont intéressés par la recherche. En fait, ils ont essayé de tirer profit de la sémantique associée aux images de tel sorte que le

système de recherche soit capable de détecter plus d'images pertinentes. Pour atteindre cet objectif, ils ont commencé par modéliser cette sémantique sous forme des relations sémantiques entre des concepts, puis ils l'ont exploité au moment de la recherche. Cependant, la majorité des méthodes proposées, que ce soit celles de l'annotation ou bien celles de la recherche, restent confrontées à certaines limitations importantes.

Pour les méthodes d'annotation, la phase d'apprentissage fait face à plusieurs lacunes qui peuvent influencer considérablement le processus d'apprentissage, et donc dégrader la qualité des résultats d'annotation. Ceci implique : la difficulté d'assurer une annotation complète et correcte des images d'apprentissage, la difficulté d'apprendre des concepts de haut niveau d'abstraction ayant des apparences visuelles différentes, la limitation à la relation caractéristiques visuelles-images et l'ignorance de la relation très importante concept-concept pour la plupart des méthodes proposées, le calcul de corrélation entre des concepts est limité à des corpus locaux et ne peut pas être généralisé pour des cas généraux et réels.

Pour les méthodes de recherche, la majorité des travaux proposés font face, à leur tour aussi, à plusieurs lacunes qui peuvent influencer négativement le mécanisme de recherche, et par conséquent dégrader la performance des résultats retournés. Ce qui se traduit par : la modélisation des relations sémantiques entre les concepts se limite à quelques types de relations seulement, les modèles sémantiques des relations ne reflètent pas la richesse sémantique existante dans la réalité, plusieurs images pertinentes restent écartées (silence) du fait que l'inférence se fait généralement sur des relations directes et explicites entre une requête et les autres concepts et ignore celles qui sont implicites et indirectes.

Compte tenu des lacunes citées ci-dessus, créer des bases d'images avec des annotations complètes, ou bien des mécanismes de recherches avec un degré de rappel satisfaisant est un défi persistant. Par conséquent, un TBIR reste confronté au problème de l'écartement de plusieurs images pertinentes, et donc le silence reste encore un problème ouvert. Pour cela, nous avons proposé, dans cette thèse, une nouvelle méthode permettant de réduire ce problème. La contribution majeure de cette méthode est de calculer la proximité sémantique entre les concepts et l'exploiter au moment de la recherche. La solution proposée opère en deux étapes : Premièrement, nous avons calculé la proximité sémantique entre les concepts. Pour ce faire, nous avons collecté un ensemble d'articles Wikipedia pour construire une source de connaissances externes. Ensuite, nous avons utilisé le schéma de pondération statistique TF_ICTF (Term Frequency_ Inverse Collection Term Frequency en anglais) pour calculer les poids des concepts dans ces articles. Les poids obtenus ont été utilisés comme

entrée pour calculer la similarité cosinus entre les concepts. Dans la deuxième étape, Nous avons intégré les résultats de proximité sémantique au sein du mécanisme de recherche.

L'utilisation de la proximité sémantique entre les concepts présente plusieurs avantages. Elle permet de refléter la sémantique entre les concepts sans se limiter à aucune relation sémantique prédéfinie. En outre, contrairement aux méthodes utilisant des corpus locaux pour capter la corrélation entre les concepts, notre méthode de calcul de proximité sémantique entre les concepts utilise des articles Wikipedia comme source de connaissances externes indépendantes de toute base d'images. Ceci, rend notre solution plus proche de la façon dont les êtres humains estiment la proximité sémantique entre les concepts du monde réel. De plus, la méthode statistique que nous avons adopté se caractérise par sa simplicité et sa performance. Un autre avantage majeur de notre méthode de calcul de proximité sémantique, c'est qu'elle est indépendante du domaine d'application. Comme nous l'avons appliqué pour résoudre le silence dans un TBIR, elle peut être appliquée pour d'autres domaines aussi.

Par ailleurs, la méthode proposée pour réduire le silence présente plusieurs avantages: elle est entièrement automatique et peut être appliquée soit dans la phase de recherche pour détecter plus d'images pertinentes, ou bien dans la phase d'annotation pour détecter et compléter des concepts pertinents manquants. De plus, le mécanisme de recherche adopté n'est plus une comparaison binaire, ni une inférence limitée sur des relations directes et explicites entre une requête utilisateur et les annotations des images, mais plutôt une inférence en fonction de la proximité sémantique entre ces deux éléments.

Bien que la solution proposée ait permis de minimiser le silence dans un système de recherche d'images par le texte, elle présente quelques limitations, à savoir :

- Dans la méthode que nous avons adopté pour calculer la proximité sémantique entre les concepts, nous n'avons exploité que les propriétés internes des articles Wikipedia. C'est-à-dire le contenu des articles (des statistiques sur le texte), tandis qu'il existe d'autres propriétés externes importantes qui peuvent être déduites ou être inférées à partir des articles. Comme instance, il y a le topique (ou le sujet) d'un article, sa catégorie dans la hiérarchie des catégories Wikipedia et ses liens avec d'autres articles. Ces propriétés importantes permettent d'analyser la sémantique et de localiser le contexte sémantique exact d'un concept, surtout qu'il existe des concepts avec différents contextes sémantiques (concepts anonymes).

- Notre système de recherche retourne parmi les images qu'il les considère comme pertinentes et qu'elles ont été en silence, des images bruits. Ceci s'explique par le fait que les

images bruits retournées sont annotées avec un concept anonyme à la requête (c.à.d. même concept que la requête mais dans un autre contexte sémantique) ou bien des concepts proches sémantiquement à un concept anonyme. Par exemple, si un utilisateur cherche des images de fruit ‘*avocat*’, alors le système peut retourner une image annotée avec le concept anonyme ‘*avocat*’ dans le contexte de la justice, comme il peut retourner des images annotées par ‘*tribunal*’.

Comme perspectives de notre travail, d'autres investigations peuvent être menées pour améliorer la méthode proposée à savoir :

A court terme :

- Essayer d'autres mesures autre que le cosinus de similarité de Salton, telle que ‘Second Order Co-occurrence Pointwise Mutual Information (SOCPMI)(Islam and Inkpen, 2006)’, et voir si ça donne des résultats de proximité sémantique encore plus mieux.
- Evaluer les résultats de proximité sémantique par rapport à d'autres benchmark de jugements humains appartenant à des différents endroits dans le monde, tel que le benchmark Mturk-771(Halawi et al., 2012), afin de vérifier si nos résultats de proximité sémantiques sont consistants.
- Mener plus d'expériences de recherche et d'annotation d'images sur des corpus plus grands, que ce soit pour le nombre d'images ou bien pour le nombre de concepts.

A moyen terme :

- Le traitement de requêtes composées de plusieurs concepts avec connecteurs logiques (par exemple : ou, non). Ainsi, le calcul de la pertinence d'une image par rapport à cette requête, peut être vu comme une mesure de similarité entre deux expressions logiques : la première est l'annotation d'une image, et la deuxième est la requête, en tenant compte la proximité sémantique entre les concepts des deux expressions et les connecteurs logique.
- Développer un mécanisme de raffinement de requête pour la prise en considération d'une requête visuelle composée de plusieurs images annotées.
- Pour minimiser le bruit et augmenter la précision, nous pensons à quelques solutions :
 - o Calculer la proximité sémantique entre les concepts en exploitant en plus des propriétés internes de Wikipedia (c.à.d. le texte d'un article), d'autres propriétés

externes comme la hiérarchie des catégories Wikipedia, sa structure d'hyperlien, les pages de désambiguïsation pour les polysémies.

- La prise en considération des préférences de l'utilisateur. Par exemple, l'historique de navigation de l'utilisateur. Ceci va permettre au système de recherche de sélectionner le contexte sémantique qui répond correctement aux besoins de l'utilisateur.
- Exploiter les caractéristiques visuelles des images
- Développer d'autres mécanismes plus sophistiqué pour inférer la pertinence des images, en exploitant toujours la proximité sémantique entre les concepts.

-

A long terme:

- Développer un système de recherche d'images général, qui consisterait en une hybridation de deux systèmes TBIR et CBIR. Ainsi, un utilisateur peut introduire une requête textuelle ou bien visuelle sous forme d'une image choisie parmi celles de la base. Donc, la requête visuelle est une image avec annotation. Pour mesurer la pertinence d'une image par rapport à cette requête, nous pensons à plusieurs alternatives. Par exemple, nous pouvons développer une mesure de distance combinée de deux distances : une distance visuelle entre les vecteurs des caractéristiques de bas niveau des deux images, et une distance sémantique entre leurs vecteurs d'annotation. Cette dernière distance sera calculée en fonction des valeurs de proximité sémantique entre les concepts des deux images, de telle sorte que, des grandes valeurs de proximité sémantique conduit à une petite distance entre les deux images, et vice versa. De plus, nous pensons également à un autre mécanisme de recherche basé graphe, où les nœuds représentent les différentes images d'une base, et les arcs représentent les distances entre les nœuds images. Ainsi, pour calculer la pertinence d'une image, nous devons inférer sur ce graphe.

- Développer une méthode d'annotation qui permet de faire l'apprentissage des concepts, et qui tient en compte la relation entre les caractéristiques visuelles des images d'apprentissage et les concepts correspondants, conjointement avec la proximité sémantique entre ces concepts.

- Appliquer la proximité sémantique entre concepts pour d'autres domaines, tel que : la classification, la détection, annotation des vidéos, Apprentissage multi-labels, les différentes tâches NLP, modélisation sémantique des phénomènes naturelles pour résoudre des problèmes du monde réel comme le diagnostic des maladies par exemple, les systèmes de

recommandation tel que la vente en ligne où le système recommande des articles pertinents à ce que veut acheter l'utilisateur.

A la fin de cette thèse, nous signalons que la modélisation de la sémantique entre des entités du monde réel, que ce soit via des relations sémantiques ou des proximités sémantiques, pour n'importe quelle application n'est pas encore parvenue (et ne parviendra jamais) à représenter parfaitement la richesse en relations sémantiques qui existent en réalité. Ainsi, l'automatisation de cette tâche reste un grand défi pour la communauté scientifique spécialisée.

Bibliographie

- Agarwal M and Maheshwari R. (2015) Co-occurrence of maximal Haar-like wavelet filters for CBIR. *International Journal of Signal and Imaging Systems Engineering* 8(5): 316-330.
- Ambika P and Samath JA. (2012) Ontology—Based semantic web CBIR by utilizing content and model annotations. *Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on.* IEEE, 453-457.
- Anderson JR. (1983) A spreading activation theory of memory. *Journal of verbal learning and verbal behavior* 22(3): 261-295.
- Aouicha MB and Taieb MAH. (2016) Computing semantic similarity between biomedical concepts using new information content approach. *Journal of biomedical informatics* 59: 258-275.
- Ayech MBH and Amiri H. (2016) A content-based image retrieval using PCA and SOM. *International Journal of Signal and Imaging Systems Engineering* 9(4-5): 276-282.
- Bakar SA, Hitam MS and Yussof WNJHW. (2013) Content-Based Image Retrieval using SIFT for binary and greyscale images. *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. TBD Melaka, Malaysia: IEEE, 83-88.
- Bao B-K, Ni B, Mu Y, et al. (2011) Efficient region-aware large graph construction towards scalable multi-label propagation. *Pattern Recognition* 44(3): 598-606.
- Belloulata K, Belallouche L, Belalia A, et al. (2014) Region based image retrieval using shape-adaptive dct. *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference on.* IEEE, 470-474.
- Bielza C, Li G and Larranaga P. (2011) Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning* 52(6): 705-727.
- Bröcker J. (2010) Regularized logistic models for probabilistic forecasting and diagnostics. *Monthly Weather Review* 138(2): 592-604.
- Budanitsky A and Hirst G. (2006) Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics* 32(1): 13-47.
- Chen L, Xu D, Tsang IW, et al. (2010) Tag-based web photo retrieval improved by batch mode re-tagging. *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. San Francisco, CA, 3440-3446.
- Cilibrasi RL and Vitanyi PM. (2007) The google similarity distance. *IEEE Transactions on knowledge and data engineering* 19(3).
- Claveau V and Nie J-Y. (2016) Recherche d'information et traitement automatique des langues: collaboration, synergie et convergence. *Traitement Automatique des Langues* 56(3).
- Contreras J, Benjamins VR, Blázquez M, et al. (2004) A semantic portal for the international affairs sector. *International Conference on Knowledge Engineering and Knowledge Management*. Springer, 203-215.
- Cox JJ, Miller ML, Minka TP, et al. (2000) The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. *IEEE transactions on image processing* 9(1): 20-37.

-
- Cui C, Ma J, Lian T, et al. (2015) Improving image annotation via ranking- oriented neighbor search and learning-based keyword propagation. *Journal of the Association for Information Science and Technology* 66(1): 82-98.
- Duygulu P, Barnard K, de Freitas JF, et al. (2002) Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. *European conference on computer vision (ECCV)*. Copenhagen, Denmark, 97-112.
- Enser PG, Sandom CJ and Lewis P. (2005) Surveying the reality of semantic image retrieval. *Lecture Notes in Computer Science* 3736: 177-188.
- Fabian M, Gjergji K and Gerhard W. (2007) Yago: A core of semantic knowledge unifying wordnet and wikipedia. *16th International World Wide Web Conference, WWW*. 697-706.
- Fan L and Li B. (2006) A hybrid model of image retrieval based on ontology technology and probabilistic ranking. *IEEE/WIC/ACM International Conference on Web Intelligence*. Hong Kong, 477-480.
- Fauzi F, Hong J-L and Belkhatir M. (2009) Webpage segmentation for extracting images and their surrounding contextual information. *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 649-652.
- Feng L and Bhanu B. (2011) Concept Learning with Co-occurrence Network for Image Retrieval. *Multimedia (ISM), 2011 IEEE International Symposium on*. IEEE, 428-433.
- Finkelstein L, Gabrilovich E, Matias Y, et al. (2001) Placing search in context: The concept revisited. *Proceedings of the 10th international conference on World Wide Web*. Hong Kong, 406-414.
- Franzoni V and Milani A. (2012) PMING Distance: A collaborative semantic proximity measure. *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 02*. IEEE Computer Society, 442-449.
- Franzoni V, Milani A, Pallottelli S, et al. (2015) Context-based image semantic similarity. *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. Zhangjiajie, China, 1280-1284.
- Gabrilovich E and Markovitch S. (2007) Computing semantic relatedness using wikipedia-based explicit semantic analysis. *IJCAI'07 Proceeding of the 20th International joint Conference on Artificial intelligence*. Hyderabad, India, 1606-1611.
- Gallas A, Barhoumi W, Kacem N, et al. (2015) Locality-sensitive hashing for region-based large-scale image indexing. *IET Image Processing* 9(9): 804-810.
- Ghamrawi N and McCallum A. (2005) Collective multi-label classification. *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 195-200.
- Gong T, Li S and Tan CL. (2010) A semantic similarity language model to improve automatic image annotation. *22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. Arras, France, 197-203.
- Gorisse D. (2010) Passage à l'échelle des méthodes de recherche sémantique dans les grandes bases d'images. Université de Cergy Pontoise.
- Guo Y and Gu S. (2011) Multi-label classification using conditional dependency networks. *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*. Barcelona, SPAIN, 1300-1305.
- Haddad H, Berrut C and Bruandet M-F. (1996) Un modèle Vectoriel de Recherche d'Informations Adapté aux Documents Vidéo. *CORESA*. 281-289.
- Halawi G, Dror G, Gabrilovich E, et al. (2012) Large-scale learning of word relatedness with constraints. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1406-1414.
- Hassan S and Mihalcea R. (2011) Semantic Relatedness Using Salient Semantic Analysis. *Proceeding of the 25th AAAI Conference on Artificial Intelligence*. San Francisco, California, 884-889.
- Haveliwala TH. (2003) Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on knowledge and data engineering* 15(4): 784-796.
- Islam A and Inkpen D. (2006) Second order co-occurrence PMI for determining the semantic similarity of words. *Proceedings of the International Conference on Language Resources and Evaluation*. 1033-1038.
- Jabeen S, Gao X and Andreae P. (2012) Harnessing wikipedia semantics for computing contextual relatedness. *PRICAI 2012: Trends in Artificial Intelligence*: 861-865.
-

- Jiang X and Tan A-H. (2006) Ontosearch: A full-text search engine for the semantic web. *AAAI*. 1325-1330.
- Joachims T, Finley T and Yu C-NJ. (2009) Cutting-plane training of structural SVMs. *Machine Learning* 77(1): 27-59.
- Kherfi M, Ziou D and Bernardi A. (2003) Atlas WISE: A Web-based image retrieval engine. *Proceedings of the International Conference on Image and Signal Processing*. 69-77.
- Kherfi ML, Ziou D and Bernardi A. (2004) Image retrieval from the world wide web: Issues, techniques, and systems. *ACM Computing Surveys (Csur)* 36(1): 35-67.
- Leacock C and Chodorow M. (1998) Combining local context and WordNet similarity for word sense identification. *WordNet: An electronic lexical database* 49(2): 265-283.
- Li J and Wang JZ. (2008) Real-time computerized annotation of pictures. *IEEE transactions on pattern analysis and machine intelligence* 30(6): 985-1002.
- Lin D. (1998) An information-theoretic definition of similarity. *Icml*. 296-304.
- Liu G-H and Yang J-Y. (2013) Content-based image retrieval using color difference histogram. *Pattern Recognition* 46(1): 188-198.
- Liu T-Y. (2009) Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval* 3(3): 225-331.
- Liu Y, Xu D, Tsang IW, et al. (2011) Textual query of personal photos facilitated by large-scale web data. *IEEE transactions on pattern analysis and machine intelligence* 33(5): 1022-1036.
- Lowe DG. (2004) Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2): 91-110.
- Makadia A, Pavlovic V and Kumar S. (2008) A new baseline for image annotation. *Computer Vision–ECCV 2008*: 316-329.
- Malone J, Stevens R, Jupp S, et al. (2016) Ten simple rules for selecting a bio-ontology. *PLoS Comput Biol* 12(2): e1004743.
- Manipoonchelvi P and Muneeswaran K. (2011) Significant region based image retrieval using curvelet transform. *Recent Advancements in Electrical, Electronics and Control Engineering (ICONRAEECE), 2011 International Conference on*. IEEE, 291-294.
- Manipoonchelvi P and Muneeswaran K. (2015) Significant region-based image retrieval. *Signal, Image and Video Processing* 9(8): 1795-1804.
- Manzoor U, Ejaz N, Akhtar N, et al. (2012) Ontology based image retrieval. *2012 International Conference for Internet Technology and Secured Transactions*. 288-293.
- Maree M and Belkhatir M. (2010) A Coupled statistical/semantic framework for merging heterogeneous domain-specific ontologies. *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*. IEEE, 159-166.
- Maree M and Belkhatir M. (2015) Addressing semantic heterogeneity through multiple knowledge base assisted merging of domain-specific ontologies. *Knowledge-Based Systems* 73: 199-211.
- Maree M, Belkhatir M, Fauzi F, et al. (2016) Multiple Ontology-Based Indexing of Multimedia Documents on the World Wide Web. *Intelligent Decision Technologies 2016*. Springer, 51-62.
- Medina LA, Fred AL, Rodrigues R, et al. (2012) Measuring Entity Semantic Relatedness using Wikipedia. *Proceeding of the International Conference on Knowledge Discovery and Information Retrieval (KDIR)*. Barcelona, SPAIN, 431-437.
- Metzler D and Manmatha R. (2004) An inference network approach to image retrieval. *CIVR*. Springer, 42-50.
- Mikolajczyk K and Schmid C. (2004) Scale & affine invariant interest point detectors. *International journal of computer vision* 60(1): 63-86.
- Miller GA. (1995) WordNet: a lexical database for English. *Communications of the ACM* 38(11): 39-41.
- Miller GA and Charles WG. (1991) Contextual correlates of semantic similarity. *Language and cognitive processes* 6(1): 1-28.
- Neal RM. (1993) Probabilistic inference using Markov chain Monte Carlo methods.
- Neelima N and Reddy ES. (2015) An improved image retrieval system using optimized FCM & multiple shape, texture features. *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*. Madurai, India, 1-7.

-
- Ni Y, Xu QK, Cao F, et al. (2016) Semantic documents relatedness using concept graph representation. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 635-644.
- Pakhomov S, McInnes B, Adam T, et al. (2010) Semantic similarity and relatedness between clinical terms: an experimental study. *AMIA annual symposium proceedings*. American Medical Informatics Association, 572.
- Pallottelli S, Franzoni V and Milani A. (2015) Multi-path traces in semantic graphs for latent knowledge elicitation. *Natural Computation (ICNC), 2015 11th International Conference on*. IEEE, 281-288.
- Parashar A. (2009) Region based image retrieval systems. *National Conf. on Signal and Image Processing Applications', IET Computer Vision*. IET, 55.
- Pesquita C, Faria D, Falcao AO, et al. (2009) Semantic similarity in biomedical ontologies. *PLoS Comput Biol* 5(7): e1000443.
- Popescu A, Moëllic P-A and Millet C. (2007) Semretriev: an ontology driven image retrieval system. *Proceedings of the 6th ACM international conference on Image and video retrieval(CIVR)*. Amsterdam, The Netherlands, 113-116.
- Rodrigues R, Filipe J and Fred AL. (2014) Semantic Relatedness with Variable Ontology Density. *Proceeding of the International Conference on Knowledge Discovery and Information Retrieval (KDIR)*. Rome, Italy, 554-559.
- Rubenstein H and Goodenough JB. (1965) Contextual correlates of synonymy. *Communications of the ACM* 8(10): 627-633.
- Strube M and Ponzetto SP. (2006) WikiRelate! Computing semantic relatedness using Wikipedia. *AAAI*. 1419-1424.
- Syam B and Srinivasa Rao Y. (2012) An effective similarity measure via genetic algorithm for Content-Based Image Retrieval with extensive features. *International Journal of Signal and Imaging Systems Engineering* 5(1): 18-28.
- Taieb MAH, Aouicha MB and Hamadou AB. (2013) Computing semantic relatedness using Wikipedia features. *Knowledge-Based Systems* 50: 260-278.
- Taieb MAH, Aouicha MB and Hamadou AB. (2014) A new semantic relatedness measurement using WordNet features. *Knowledge and information systems* 41(2): 467-497.
- Torralba A, Fergus R and Freeman WT. (2008) 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 30(11): 1958-1970.
- Tsochantaridis I, Joachims T, Hofmann T, et al. (2005) Large margin methods for structured and interdependent output variables. *Journal of machine learning research* 6(Sep): 1453-1484.
- Turney P. (2001) Mining the web for synonyms: PMI-IR versus LSA on TOEFL. *Machine Learning: ECML 2001*: 491-502.
- Vipparthi SK and Nagar S. (2015) Directional local ternary patterns for multimedia image indexing and retrieval. *International Journal of Signal and Imaging Systems Engineering* 8(3): 137-145.
- Wang H, Chia L-T and Gao S. (2010) Wikipedia-assisted concept thesaurus for better web media understanding. *Proceedings of the international conference on Multimedia information retrieval*. ACM, 349-358.
- Wang H, Jiang X, Chia L-T, et al. (2008) Ontology enhanced web image retrieval: aided by wikipedia & spreading activation theory. *Proceedings of the 1st ACM international conference on Multimedia information retrieval*. ACM, 195-201.
- Witten I and Milne D. (2008a) An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. *Proceeding of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy*, AAAI Press, Chicago, USA. 25-30.
- Witten IH and Milne DN. (2008b) An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. *Proceeding of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy*, AAAI Press, Chicago, USA.: 25-30.
- Wu L, Hoi SC, Jin R, et al. (2009) Distance metric learning from uncertain side information with application to automated photo tagging. *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 135-144.
-

- Wu L, Jin R and Jain AK. (2013) Tag completion for image retrieval. *IEEE transactions on pattern analysis and machine intelligence* 35(3): 716-727.
- Wu P, Hoi SC-H, Zhao P, et al. (2011) Mining social images with distance metric learning for automated image tagging. *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 197-206.
- Xu H, Pan P, Xu C, et al. (2016) Image auto-annotation via concept interdependency network. *Multimedia Tools and Applications* 75(11): 6237-6261.
- Yang K, Wang M, Hua X-S, et al. (2011) Tag-based social image search: Toward relevant and diverse results. *Social Media Modeling and Computing*. Springer, 25-45.
- Yang L, Jin R, Mummert L, et al. (2010) A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. *IEEE transactions on pattern analysis and machine intelligence* 32(1): 30-44.
- Yang X, Lv F, Cai L, et al. (2014) Adaptive learning region importance for region-based image retrieval. *IET Computer Vision* 9(3): 368-377.
- Yeh E, Ramage D, Manning CD, et al. (2009) WikiWalk: random walks on Wikipedia for semantic relatedness. *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*. Association for Computational Linguistics, 41-49.
- Yue J, Li Z, Liu L, et al. (2011) Content-based image retrieval using color and texture fused features. *Mathematical and Computer Modelling* 54(3): 1121-1127.
- Zha Z-J, Mei T, Wang J, et al. (2009) Graph-based semi-supervised learning with multiple labels. *Journal of Visual Communication and Image Representation* 20(2): 97-103.
- Zhang D, Islam MM, Lu G, et al. (2012a) Rotation invariant curvelet features for region based image retrieval. *International journal of computer vision* 98(2): 187-201.
- Zhang S, Huang J, Li H, et al. (2012b) Automatic image annotation and retrieval using group sparsity. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42(3): 838-849.
- Zhiqiang L, Werimin S and Zhenhua Y. (2009) Measuring semantic similarity between words using wikipedia. *International Conference on Web Information Systems and Mining*. Taiyuan, China: IEEE, 251-255.