



HAL
open science

Utilisation de structures sémantiques pour la recherche d'images sur Internet.

Adrian Popescu, Pierre-Alain Moellic, Ioannis Kanellos

► To cite this version:

Adrian Popescu, Pierre-Alain Moellic, Ioannis Kanellos. Utilisation de structures sémantiques pour la recherche d'images sur Internet.. 4ème Atelier ECOI: Extraction de COnnnaissance et Images, in conjunction with EGC 2008, Jan 2008, Sophia Antipolis, France. hal-02163614

HAL Id: hal-02163614

<https://hal.science/hal-02163614>

Submitted on 24 Jun 2019

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

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

Utilisation de structures sémantiques pour la recherche d'images sur Internet

Adrian Popescu*, **, Pierre-Alain Moëllic*
Ioannis Kanellos**

*CEA LIST, 18 route du Panorama, 92260 Fontenay aux Roses
{adrian.popescu, pierre-alain.moellic}@cea.fr
<http://www-list.cea.fr/>

**ENST de Bretagne, Technopôle Brest-Iroise, 29238 Brest
ioannis.kanellos@enst-bretagne.fr
<http://www.enst-bretagne.fr/>

Résumé. La recherche d'images (RIM¹) constitue une partie importante de la recherche d'information sur Internet. Malgré leur succès, les systèmes de RIM ont une série de limitations dont une partie découlent de leur insensibilité à la sémantique. Nous analysons ici l'introduction de structures sémantiques à grande échelle dans la recherche d'images sur Internet. Plus précisément, nous présentons quelques ressources utilisables, ainsi qu'un modèle d'agrégation de l'accès aux images par le texte et des techniques de traitement d'images. Enfin, nous illustrons nos propos avec quelques applications déjà implémentées ou en cours de développement. Une attention particulière est accordée au couplage des solutions proposées avec des usages existants ou potentiels. Les principaux avantages et difficultés liés à l'introduction d'une couche sémantique dans les systèmes de recherche d'images sont systématiquement mis en avant.

1 Introduction

La recherche de documents multimédia sur Internet est un domaine apparenté à la recherche de textes, mais elle se différencie sur un nombre de points importants. Ici, nous abordons le cas de la recherche d'images photographiques et articulons notre analyse autour de deux problèmes principaux. Premièrement, la recherche d'images à grande échelle, comme celle de données textuelles, est basée sur une exploitation du texte brut associé aux images. Cela entraîne une opacité totale des systèmes par rapport à la signification des requêtes et des réponses proposées par l'application. Deuxièmement, le paradigme de recherche multimédia actuel ne rend pas suffisamment compte des différences entre les types de documents traités impliquant une sous-exploitation de certaines techniques spécifiques aux types de données, qui pourraient améliorer le processus. Même si l'objet de la recherche est un ensemble d'images, les techniques de traitement d'images ne sont que très peu utilisées.

¹ RIM est un acronyme des auteurs pour Recherche d'IMages

Le premier des deux problèmes signalés peut se résoudre par l'introduction de structures sémantiques² dans l'architecture d'un système RIM. Les mêmes structures interviennent dans l'articulation de la recherche d'images par mots clé et la recherche basée sur le contenu visuel afin d'améliorer l'efficacité du dernier type d'accès aux données. Les bases de données textuelles structurées peuvent remplir plusieurs rôles dans les systèmes de recherche d'images. Dans ce papier nous illustrons leur utilité pour:

- une reformulation automatique des requêtes
- une présentation structurée des réponses
- une amélioration de l'interactivité des applications
- une meilleure précision des résultats
- une recherche par contenu visuel dans des régions conceptuellement cohérentes du corpus d'images

Les fonctions citées ci-dessus s'obtiennent au prix d'une complexification de l'architecture d'un système RIM. Il est cependant crucial d'intégrer ces structures sémantiques de façon à ne pas compromettre la facilité d'utilisation des applications de recherche. De plus, des ressources textuelles à très large échelle sont nécessaires afin de couvrir une partie significative de l'espace de requêtes dont Jansen et al. (2000) ont noté la grande complexité (requêtes variées en contenu et nombre de termes). Il est aussi nécessaire d'assurer une bonne qualité des connaissances incluses dans les bases de données structurées.

La suite de ce papier est structuré de la façon suivante: nous discutons tout d'abord de quelques travaux apparentés, puis la section 3 est dédiée à la présentation des structures sémantiques exploitées et à l'analyse de la relation entre les concepts inclus et leur représentation picturale. Les deux types d'accès aux données photographiques, par mots clef et par contenu visuel, ainsi qu'une possible agrégation, sont discutés dans la section 4. Avant de conclure, nous introduisons et évaluons quelques applications employant des ressources textuelles pour la RIM.

2 Travaux apparentés

Notre travail s'apparente à plusieurs domaines dont nous donnons ici un bref aperçu. En recherche d'images, Yang et al. (2001) incluent WordNet Miller (1990) dans un système afin de proposer des requêtes proches par rapport au concept courant. Wang et al. (2006) extraient la base taxinomique d'une ontologie des animaux de WordNet et la complètent avec des propriétés importées d'une autre ontologie. Leur structure comprend, à part les relations textuelles, des propriétés de bas niveau (comme la couleur ou la texture) caractérisant les classes dans la hiérarchie conceptuelle. Ruiz-Casado et al. (2007) quant à eux abordent la construction de ressources sémantiques à partir de Wikipedia.

O'Hare et al. (2005) et Ahern et al. (2007) s'intéressent à la recherche d'images possédant des coordonnées géographiques. Wilkins et al. (2005) et Quack et al. (2004) discutent le problème du passage à l'échelle dans les systèmes de recherche d'images par le contenu. Une discussion complète sur la RIM par le contenu augmentée par la sémantique se trouve dans Liu et al. (2007).

² Structure sémantique et ressource sémantique sont des termes désignant plusieurs types de bases de données textuelles structurées, comme: les thésauri, les hiérarchies conceptuelles ou les ontologies formelles.

3 Caractérisation des structures sémantiques

Les ressources linguistiques que l'on peut intégrer dans les systèmes de recherche d'images peuvent être caractérisées selon plusieurs critères. Nous retenons ici leur disponibilité, le degré de structuration et le domaine conceptuel des données incorporées.

3.1 Disponibilité des ressources

Une première approche pour l'introduction de ressources sémantiques et RIM vise à exploiter des données préexistantes et réutilisables (éventuellement après adaptation). Ces ressources assurent une couverture parfois insuffisante et il est nécessaire d'en créer de nouvelles. Étant donnée la diversité de l'espace des requêtes sur le Web, il est nécessaire de construire des structures à très large échelle afin de couvrir une partie significative des requêtes. La création manuelle de ressources étant trop coûteuse, une construction automatique doit être envisagée. Une première solution serait l'exploitation de documents textuels non structurés afin d'en extraire des connaissances. Cette approche s'avère inadaptée car les volumes de données qu'il est possible d'extraire, de qualité raisonnable, sont trop réduits comme l'indiquent Cimiano et al. (2004), Clough et al. (2005).

Alternativement, Ruiz-Casado et al. (2007) proposent l'utilisation de données semi structurées, comme les articles Wikipedia, pour construire des bases de connaissances. Cette approche est plus adaptée pour la RIM sur Internet et nous discutons dans la suite de cet article d'un cas d'application. D'autres bases de données structurées et spécifiques à des domaines existent et pourraient être exploitées, mais elles sont rarement libres de droits. IMDB³, une base de données cinématographique très détaillée, en est un parfait exemple.

3.2 Structuration des données dans les ressources

Les ressources linguistiques ont des degrés d'organisation variable qui influence grandement les possibilités d'exploitation des données. Un maximum de structuration logique est atteint avec les ontologies formelles Guarino (1998). Dans ces structures, il est possible de proposer une représentation compacte des connaissances et de raisonner automatiquement sur leur contenu. Malheureusement, deux problèmes subsistent : premièrement, les algorithmes de raisonnement automatique ne sont pas actuellement adaptés pour gérer des grands volumes de données ; deuxièmement, les contraintes imposées par le formalisme logique soutenant les ontologies sont parfois trop strictes pour modéliser la recherche d'information pour le grand public. Un bon exemple est l'impossibilité de définir un héritage multiple pour un concept. Les thésaurus ont une structure moins stricte que les ontologies formelles mais les possibilités de raisonnements sont plus réduites. La connaissance incluse doit être explicite afin de pouvoir en profiter, aussi, en contrepartie, on observe une plus grande flexibilité pour la définition des relations inter-conceptuelles. L'utilisation de thésaurus nous semble donc plus adaptée à la recherche d'images.

³ <http://imdb.com>

3.3 Données dans les structures sémantiques

L'espace des requêtes en RIM impose l'utilisation de ressources linguistiques comprenant des concepts appartenant à des domaines variés. Les ressources généralistes comme WordNet ou Cyc sont les seuls à remplir cette contrainte. Leur couverture est toutefois limitée et il est nécessaire de les compléter. Parmi les domaines pertinents pour la recherche d'images sur le Web, nous trouvons les noms de personnes célèbres et les entités géographiques Jansen (2000) qui sont insuffisamment renseignées. Cependant, il existe des bases de données utilisables permettant d'étendre WordNet à ces domaines.

3.4 Ressources exploitées

Nous utilisons plusieurs bases de connaissances pour améliorer le fonctionnement des applications de recherche d'images sur le Web que nous passons en revue dans les sous-sections suivantes.

3.4.1 WordNet

WordNet (Miller (1990)) est un réseau lexical construit manuellement par des experts comprenant plus de 100000 concepts nominaux structurés sous forme d'arbre. Les deux relations fondamentales dans WordNet sont la relation d'héritage conceptuel, déterminant la structure en profondeur et la synonymie, permettant de créer les atomes de la hiérarchie : les *synsets*. Parmi les propriétés importantes du réseau lexical, nous citons la séparation des sens pour les termes polysémiques et la possibilité d'inclure un héritage multiple pour un concept. Une adaptation de WordNet est nécessaire pour l'utiliser en RIM et nous décrivons ce processus en détail dans Popescu et Grefenstette (2008). Par exemple, il est nécessaire d'ordonner les concepts selon un critère de pertinence pour la recherche d'images. Nous avons choisi une mesure simple : la fréquence des termes sur Internet. Afin d'éviter les biais dus à la polysémie, les requêtes correspondant à chaque terme ont été reformulées utilisant le parent immédiat du concept.

3.4.2 Geonames

Geonames⁴ est une base de données géographiques comprenant plus de six millions d'entrées structurées sous la forme d'une hiérarchie à trois niveaux. Le plus général contient huit catégories (ex. : *parcs, constructions, montagnes, eaux* etc.). Le niveau intermédiaire en inclut 645 (sous concepts des catégories principales) comme : *monument, plage, ville, palais* ou *lac*. Les instances des 645 concepts intermédiaires forment le niveau le plus détaillé de Geonames (environ 6 millions de concepts). Pour chaque instance, il existe plusieurs informations disponibles : noms alternatifs, coordonnées géographiques, pays d'appartenance etc. L'introduction de l'information sur le placement de chaque entité crée une nouvelle dimension hiérarchique, caractérisée par la relation de méronymie entre les concepts, qu'il est possible d'exploiter. L'ordonnement des concepts dans Geonames est obtenu de manière similaire à celui de WordNet (fréquence du terme). La désambiguïsation est réalisée ici en employant dans la requête le terme et le nom de l'entité géographique englobante.

⁴ <http://geonames.org>

3.4.3 Base de données structurée extraite de Wikipedia

Les études des usages en recherche d'images sur Internet montrent qu'il y a un grand intérêt du public pour les photos de personnes célèbres. Avec moins de 20000 entrées, WordNet est clairement insuffisant dans ce domaine. Wikipedia constitue une source de données riche que nous avons utilisé pour extraire des informations concernant *les chanteurs* (plus de 6000 instances), *les acteurs* (plus de 13000 instances) et *les footballeurs* (plus de 26000 instances). L'intérêt principal des pages Wikipedia par rapport au texte libre est qu'elles incluent de l'information semi structurée facile à extraire. La description pour chaque type d'entité inclut :

- pour toutes les catégories : date de naissance (et, selon le cas, du décès), lieu de naissance, nationalité, célébrités associées
- *chanteurs* : groupes associés, genres abordés, nom réel, pseudonymes
- *acteurs* : films, rôles, prix obtenus
- *footballeurs* : club(s), poste(s)

De plus, chaque instance est rattachée à une des trois grandes catégories de célébrités (et, dans certains cas à des catégories plus précises comme *chanteur de jazz* ou *défenseur*), la base de données créée a une dimension hiérarchique. Il est ainsi possible de présenter des classes d'images structurées, de manière similaire à celle décrite et 3.4.1.

L'ordonnement des différentes instances de relations est réalisé en utilisant la fréquence jointe d'apparition de deux concepts sur le Web. Pour un acteur, on trouvera les acteurs associés, les rôles et les films les plus connus. Notons que la méthode d'extraction d'information mise en place pour les trois catégories citées est facilement adaptable à d'autres types de célébrités.

3.5 Concepts et représentation visuelle

Une partition possible des catégories donne les grandes classes suivantes : *concepts nominaux*, *concepts naturels* et *artéfacts* (Keil (1992)). La plupart des termes de la première classe sont difficiles à représenter directement dans les images. Le lecteur peut vérifier cette affirmation en essayant de s'imaginer quel serait le contenu des classes d'images pour des concepts comme : *vérité*, *association* ou *processus*. Parmi les concepts nominaux, les événements constituent une exception car il est possible d'en trouver des représentations cohérentes (il est facile d'avoir une image mentale d'*un match de football* ou d'*une grève*). Pour une autre partie des *concepts nominaux*, il y a des représentations culturellement encodées (ex. *justice*). La représentation des concepts nominaux dans les images passe par une interprétation. Il serait intéressant d'approcher leur description imagée mais ici on se concentre sur les termes directement représentables dans des images photographiques.

Les concepts naturels et les artéfacts sont, pour la plupart, représentables dans des photographies. S'ils ont une sous hiérarchie conceptuelle associée, il est possible de les représenter par l'intermédiaire de leurs sous-catégories. Si l'on considère la relation d'héritage conceptuel, les races comme *berger allemand*, *Norfolk terrier* ou *husky* forment une partie de la classe *chien*. La relation de méronymie est plus adaptée pour décrire d'autres concepts. Par exemple, *la France* peut être visualisée à travers des images *des villes* ou *des parcs naturels* contenus. A leur tour, *les villes* sont bien décrites par des concepts plus spécialisés (ex. *monuments*, images de *rues*). De même, les endroits les plus marquants représentent bien *les*

Utilisation de structures sémantiques pour la recherche d'images sur Internet

parcs naturels. Les ressources sémantiques décrites en section 3.4 incluent, selon le cas, des relations type – sous-type, de méronymie (ou autres) entre les concepts inclus.

3.6 Limites

L'introduction des ressources linguistiques en RIM n'est pas immédiate. Parmi les difficultés majeures rencontrées nous citons:

- la couverture conceptuelle des bases de connaissances : détermine la couverture de l'espace de requêtes. Premièrement, pour les requêtes portant sur un seul concept (ex. *Nice*), ce terme doit être renseigné dans la ressource. Deuxièmement, nous n'avons pas abordé le cas des requêtes contenant plusieurs termes (ex. *Nice EGC 2008*), majoritaires sur le Web (Jansen et al. (2000)). La définition d'une règle (ou de règles) de décomposition de ce type de requêtes utilisant l'héritage conceptuel est difficile.
- la qualité des ensembles d'images attachées aux concepts précis dans les hiérarchies conceptuelles. Nous avons montré dans Popescu et al. (2007b) que le bruit dans les réponses images est réduit quand on lance des requêtes avec *berger allemand* et *setter irlandais* pour définir la classe *chien* par rapport à une requête directe avec ce terme. Toutefois, il reste un certain nombre d'images non représentatives même parmi les images rattachées aux deux races de chiens citées.
- la qualité des connaissances incluses dans les ressources. Le principal désavantage de la constitution automatique de structures sémantiques est l'imprécision d'un nombre important des relations inter-conceptuelles découvertes (Ruiz-Casado et al. (2007)). Il est par exemple possible de trouver que *chien berger* hérite de *berger allemand* et non pas l'inverse. Un autre désavantage est l'incomplétude de la description des termes. Dans Wikipedia, les pages de certains acteurs sont plus complètes que d'autres Et, habituellement, on obtient le plus d'informations pour les acteurs les plus connus.
- la présentation des requêtes proches les plus adaptées. A ce stade de nos travaux nous n'avons pas encore abordé la notion de personnalisation des résultats. L'adaptation des résultats de la recherche compte tenu des centres d'intérêt de l'utilisateur est importante pour améliorer la qualité des résultats. Nous illustrons cela par deux exemples : un utilisateur passionné par les *chiens bergers* formule une requête avec *chien*. Dans ce cas on devrait favoriser la présentation des sous-types intéressants pour l'utilisateur même s'ils sont moins fréquents que d'autres. Un utilisateur italien demande de voir des images de *Zinedine Zidane* : il serait intéressant de lui proposer, en priorité parmi les requêtes proches, des noms de joueurs italiens associés à *Zidane*.

Malgré ces limites, l'utilisation des bases de connaissances présente un certain nombre d'avantages. De plus, une partie des limites ne sont pas insurmontables mais plutôt un effet de l'état actuel de notre travail. Nous pensons notamment à la personnalisation des résultats et au traitement des requêtes complexes. Dans la section 4 on regarde de plus près l'articulation entre les modalités d'accès aux images du Web et les structures sémantiques utilisées.

4 Types d'accès aux images sur Internet

Vu l'énorme volume de données sur le Web, il est nécessaire de proposer des modalités de sélection rapide de l'information intéressante pour l'utilisateur. Il existe deux principaux types d'accès aux données photographique : la recherche par mots clef et la recherche par le contenu visuel. Deux dimensions principales les différencient:

- la formulation de la requête: elle consiste respectivement en une chaîne de caractères et en une image en recherche par le contenu
- les réponses sont obtenues exploitant des descriptions différentes des données : les index textuels et les propriétés perceptuelles des images

Les moteurs de RIM existants implémentent la recherche par mots clef, tandis que la recherche par le contenu reste l'apanage des travaux de recherche et des prototypes issus de ces travaux. Dans cette section, nous présentons les deux types d'accès, ainsi qu'une possible agrégation. Nous présentons, dans les deux cas, l'utilité des ressources sémantiques.

4.1 Recherche par mots clef

Ce processus peut être décrit de la façon suivante: un utilisateur formule une requête composée d'un ou plusieurs termes et le système lui présente les images considérées les plus proches par rapport aux termes introduits. La liaison entre l'espace des requêtes et celui des réponses se réalise en exploitant la similarité entre les chaînes de caractères introduites par les utilisateurs et les termes associés aux images dans les index. Les systèmes actuels sont opaques par rapport à la signification de la requête (aucune analyse de son contenu n'est réalisée). Avec l'exploitation des ressources sémantiques, il devient possible de traiter la requête à un niveau plus symbolique et de rendre des résultats améliorés. Nous discutons ici quelques avantages de l'introduction des structures sémantiques en RIM, ainsi qu'une série de limites de cette approche.

4.1.1 Détection de l'ambiguïté

Une bonne partie des termes de la langue sont polysémiques et, par conséquent, les images associées à ces concepts sont différentes. Spärck-Jones (2007) souligne que les systèmes actuels gèrent mal l'ambiguïté des requêtes et propose des moyens pour un meilleur traitement de demandes d'information polysémiques. L'introduction d'une couche sensible à la polysémie, rend possible la présentation des sens d'un terme dans des classes différentes de réponses.

4.1.2 Reformulation automatique des requêtes

Rosch et al. (1976) distinguent trois niveaux de généralité pour les catégories linguistiques. WordNet recense jusqu'à 20 niveaux conceptuels entre les catégories les plus générales et les plus spécialisées. Geonames propose une structure à trois niveaux. L'utilisation des structures hiérarchiques permet la représentation des concepts sur les niveaux intermédiaires de la hiérarchie utilisant leurs sous-catégories. Par exemple quand un utilisateur demande des images pour *chien*, cette requête sera remplacée ou complétée automatiquement par ses sous-types (ex. *chien+berger allemand* ; *chien+setter irlandais*). Pour les requêtes localisées, l'utilisation de la reformulation permet de distinguer les différents noms géographiques ho-

monymes. Si l'utilisateur veut voir des images de *Montréal*, et plus précisément de l'église *Notre Dame* située dans cette ville, la reformulation automatique de la requête *Notre Dame* en *Notre Dame+Montréal* évitera la présentation d'images d'églises du même nom d'autres villes. Par expansion des requêtes, on élimine une partie des résultats non représentatifs et, en même temps, on propose une désambiguïsation des différents sens pour les termes polysémiques. De plus, les ensembles d'images sont conceptuellement et visuellement cohérents pour les concepts spécialisés dans les hiérarchies, permettant une agrégation efficace entre la recherche par mots-clef et celle par le contenu.

4.1.3 Structuration des résultats

Les moteurs de RIM existants proposent les résultats sous la forme d'une liste ordonnée par des critères d'adéquation des réponses à la requête et de popularité des images. La description des concepts par leurs sous-catégories facilite une présentation structurée des réponses, préservant en même temps les avantages offerts par les méthodes de classement actuelles.

4.1.4 Précision des réponses

Wang et al. (2006), Popescu et al. (2007b) montrent que la précision des résultats pour une requête dans Google Images se situe autour de 60% - 70%. Les stratégies d'expansion de requêtes proposées dans les deux papiers permettent une amélioration sensible de la précision dans les pages de réponses. Wang et al. (2006) combine les propriétés conceptuelles et les caractéristiques visuelles des objets afin d'obtenir l'amélioration, tandis que dans Popescu et al. (2007), seul l'héritage conceptuel est utilisé.

4.2 Recherche par contenu visuel

L'accès aux données à travers des images exemples est une technique introduite dans beaucoup de travaux de recherche (Liu et al. (2007)), mais qui tarde à s'imposer dans les systèmes de recherche photographique sur Internet comme une alternative ou un complément à l'interrogation par mots-clef. Cette situation est causée principalement par l'inadéquation des critères de similarité mis en oeuvre dans l'interrogation par contenu visuel et la similarité entre les images telle qu'elle est perçue par les humains. Le premier type de similarité se base sur des caractéristiques perceptuelles de l'image (couleur, texture), tandis que le deuxième est principalement conceptuel. Les deux types de similarité ne sont pas de la même nature et il serait difficile de les mettre sur le même plan. Nous proposons une approche simple, faisant de la recherche par contenu visuel un complément de l'interrogation par mots-clef. L'accès aux données par images exemples est proposé seulement dans des régions de la base de données ayant une cohérence conceptuelle (images qui ont en commun les mots clef de la requête initiale). De plus, avec la reformulation automatique, l'utilisateur est guidé vers des régions du corpus photographique définies par des concepts spécialisés (ayant une certaine cohérence visuelle). On contourne également un problème non négligeable de la recherche par le contenu visuel: la réduction de l'espace de recherche dans des bases de données à large échelle (Wilkins et al. (2005)). Dans le papier cité, cette réduction se basait sur une analyse des images, tandis qu'ici on tient compte de l'information textuelle déjà introduite par l'utilisateur pour limiter la région potentiellement intéressante.

L'agrégation entre les deux types d'interrogation présentée ici repose sur une supposition: une fois sélectionnée une image illustrant une certaine requête, l'utilisateur ne voudra pas voir des photographies avec un contenu conceptuellement différent. Si on pose comme requête une image de *Renault Twingo*, il est évident qu'on n'attend pas des images de *chevaux* ou de *couteaux*. Cela arrive souvent dans les systèmes qui exploitent principalement les caractéristiques visuelles⁵.

Nous avons présenté les modalités d'accès aux images, ainsi qu'une agrégation possible. Dans la section suivante, nous discutons quelques exemples de systèmes de recherche d'images reposant sur ce type d'approche.

5 Applications

Nous présentons ici quatre exemples d'applications, achevées ou en cours de développement, qui combinent l'utilisation de ressource sémantiques et la RIM par le contenu. Pour les systèmes fonctionnels, des évaluations et des références à des publications pertinentes sont proposés.

5.1 Recherche dans une base d'images structurée

Nous avons sélectionné une partie de WordNet (la sous hiérarchie des *animaux* contenant environ 1000 noeuds), et formé des requêtes avec les feuilles de cette hiérarchie et récupéré un maximum de 100 images d'Internet pour chaque concept feuille. Ensuite, la base a été filtrée pour éliminer les dessins et les images contenant des visages, et indexée en utilisant un descripteur de texture et de couleur. L'utilisateur peut introduire un mot-clef du domaine retenu et, s'il s'agit d'un concept ayant des héritiers, des images de ces derniers sont proposées. L'utilisateur peut cliquer sur les images pour réaliser une recherche par le contenu. On présente d'abord les images les plus proches dans la même classe feuille, mais il est possible de choisir des régions conceptuelles plus larges pour l'interrogation par le contenu. Une évaluation sur 30 concepts montre que la précision décroît quand on recherche des images similaires dans des régions de plus en plus larges. Cette application illustre bien la différence de performances entre une recherche par le contenu guidée par la sémantique des concepts et une interrogation exclusivement basée sur le contenu visuel. Une vidéo⁶ illustrant le comportement du système est également disponible. Pour plus de détails, nous renvoyons le lecteur à Popescu et al (2007).

5.2 Recherche conceptuelle d'images sur Internet

Inspirés par l'application de 5.1, nous avons utilisé l'intégralité de WordNet pour proposer un accès aux images d'Internet basé sur l'utilisation de la hiérarchie conceptuelle. Si on compare l'application, nommée Olive, à Google Image, on ajoute : une désambiguïsation des requêtes polysémiques; une reformulation automatique et une présentation de requêtes proches utilisant les connaissances dans WordNet. WordNet est exploité de la même manière que dans 5.1. Des différences notables sont toutefois à souligner : le nombre de requêtes trai-

⁵ Voir par exemple Alipr – <http://alipr.com>

⁶ <http://moromete.net/semretriev.wmv>

tées par les systèmes, 1000 dans 5.1, plus de 100000 ici et le fait que, pour Olive, la base de données image se remplit au fur et à mesure de l'utilisation du système. Il n'est donc plus possible de proposer un accès par le contenu pour tout le corpus, mais seulement aux niveaux des feuilles.

Des évaluations sur 40 concepts comparant notre système à Google Images montrent qu'on obtient une amélioration de la précision d'environ 15%, passant de 70% à 85%. En ce qui concerne la recherche par le contenu, nous avons comparé ce système à Cortina (Quack et al. (2004)), l'application de RIM par contenu visant le plus large volume de données (11 millions d'images) . 80 images représentant des concepts spécialisés ont été utilisées et la précision moyenne des résultats passe de 10% dans Cortina à 35% dans Olive. Un deuxième paramètre important qui est amélioré est le nombre de requêtes images pour lesquelles l'utilisateur considère qu'au moins une réponse est similaire à la question : moins de la moitié des requêtes pour Cortina, 95% des cas pour Olive. De plus, un test utilisateur réalisé sur un échantillon de 10 personnes confirme l'utilité de la proposition des requêtes proches et la préférence pour la présentation des réponses structurées conceptuellement contre la présentation de Google. Une vidéo⁷ illustrant le comportement du système est disponible. Pour plus de détails, nous renvoyons le lecteur à Popescu et Grefenstette (2008).

5.3 Recherche d'entités géographiques

Avec l'introduction de systèmes de géolocalisation (GPS) directement intégrés dans les appareils photo numériques, le volume d'images géo localisées devient important et il est intéressant d'exploiter l'information de localisation dans des applications de recherche d'images. Il y a un nombre croissant de travaux de recherche s'intéressant aux images géo référencées (Ahern et al. 2007) mais les résultats sont encore loin de témoigner d'une maturité des approches. Nous avons combiné Geonames, un service de visualisation de cartes de Yahoo!⁸, et la recherche par le contenu visuel pour construire ThemExplorer, un système de recherche d'images géographiques thématique.

Par rapport à l'application de référence (Ahern et al (2007)), il est notamment proposé d'explorer la carte choisissant un ou plusieurs sujets (on peut par exemple demander de voir que *les églises*, *les montagnes* ou *les églises et les monuments* dans une certaine région de la planète). De plus, une fois sélectionnée une entité précise, on propose une exploration de ces images utilisant le contenu visuel. L'application est encore en développement et il n'est pas possible de présenter des évaluations de ses performances. Toutefois, une vidéo de présentation⁹ est disponible.

5.4 Recherche de célébrités

La base de données structurée extraite de Wikipedia ciblant les noms de célébrités (section 3.4.3) est utilisable en recherche d'images sur Internet. Une présentation adaptée à chaque type de personne est rendue possible par l'exploitation des informations dans la ressource constituée au préalable. Ainsi, pour un acteur, il est possible de présenter des images relati-

⁷ <http://moromete.net/olive.wmv>

⁸ <http://tagmaps.research.yahoo.com/>

⁹ <http://moromete.net/themexplorer.avi>

ves à ses rôles les plus connus ou, pour un footballeur, des photos dans les clubs pour lesquels il a évolué. Une proposition de requêtes proches (noms de personnes, clubs, films, groupes musicaux) est aussi possible par l'utilisation de la ressource sémantique. Par rapport aux systèmes actuels, comme Ask¹⁰, on obtient ici une meilleure contextualisation et structuration des requêtes.

6 Conclusions

Notre travail s'intéresse à l'introduction de structures sémantiques et de techniques de traitement d'images dans la recherche d'images sur Internet. Cette approche vise à donner plus d'importance à l'utilisateur, en lui proposant une interaction enrichie, mais intuitive avec le système de RIM. L'exploitation de bases de connaissances aide l'utilisateur à s'orienter dans un espace conceptuel qu'il expliciterait difficilement autrement. L'agrégation de la recherche par mots-clef et de celle par le contenu proposée dans ce papier allie deux types de similarité, celle propre aux humains (conceptuelle) et celle propre aux machines (perceptuelle), dans le but de proposer une nouvelle dimension de navigation pour les requêtes ayant un grand nombre de réponses. En même temps, cette méthode simple d'agrégation contourne le problème du passage à l'échelle propre aux applications de recherche par le contenu.

Les résultats présentés dans ce papier nous encouragent de continuer dans la même ligne dans le futur. Un premier axe de travail est constitué par l'implémentation et l'évaluation des applications actuellement en développement. Deuxièmement, nous allons continuer à exploiter le contenu de Wikipedia pour d'autres domaines conceptuels, comme l'extraction de bases de données géographiques. Troisièmement, il est nécessaire d'analyser plus finement la relation entre la recherche par le contenu visuel et la recherche par mots-clef. Cela se traduit par une adaptation de la description de bas niveau de l'image au type de concept demandé. Par exemple, pour les objets manufacturés, l'utilisation de descripteurs locaux (comme les points d'intérêt) peut s'avérer utile.

Références

- Ahern, S., M. Naaman, R. Nair, J. Yang (2007) World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections. In Proc of. JCDL 2007.
- Cimiano, P., A. Pivk, L. Schmidt-Thieme, and S. Staab, S. (2004) Learning taxonomic relations from heterogeneous evidence. In Proc. of ECAI 2004, OLP Workshop.
- Clough, P., H. Joho, and M. Sanderson (2005), *Automatically Organising Images using Concept Hierarchies*, In Proc of ACM SIGIR -MIR Workshop.
- Guarino, N. (1998) *Formal Ontologies in Information Systems*. IOS Press.
- Jansen, J., A. Goodrum and A. Spink (2000) *Searching for multimedia: analysis of audio, video and image Web queries*. In World Wide Web Journal 3(4).
- Keil F. C. (1992) *Concepts, Kinds, and Conceptual Development*, Bradford Books, 1992.

¹⁰ <http://ask.com>

Utilisation de structures sémantiques pour la recherche d'images sur Internet

- Liu, Y., D. Zhang, G. Lu and W.-Y. Ma (2007) *A survey of content-based image retrieval with high-level semantics*, Pattern Recognition 40(1).
- Miller, G. A. (1990) *WordNet: An on-line lexical database*. Int. Journal of Lexicography 3, 4 (Winter 1990), 235-312.
- O'Hare, N., C. Gurrin, A. F. Smeaton, G.J.F Jones (2005) *Combination of content analysis and context features for digital photograph retrieval*. In Proc. of EWIMT 2005.
- Popescu, A., C. Millet, and P.-A. Moëllic (2007a) *Ontology Driven Content Based Image Retrieval*, In Proc. of ACM CIVR 2007.
- Popescu, A., G. Grefenstette, P.-A. Moëllic (2007b) *Improving Image Retrieval Using Semantic Resources*, Springer SCI series.
- Popescu, A., G. Grefenstette (2008) *Concept-based Internet Image Retrieval*, Article soumis pour LREC 2008.
- Quack, T., U. Monich, L. Thiele, and B. S. Manjunath (2004) *Cortina: A System for Large-scale, Content-based Web Image Retrieval*. In Proc. of ACM Multimedia 2004.
- Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem (1976) *Basic objects in natural categories*, Cognitive Psychology, 8.
- Ruiz-Casado, M. , E. Alfonseca and P. Castells (2007) *Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia*, Data and Knowledge Engineering, 61(3).
- Spärck-Jones, K., S. E. Robertson, M. Sanderson (2007) *Ambiguous Requests: Implications for Retrieval Tests, Systems and Theories*, ACM SIGIR Forum, December 2007.
- Wang, H., Liu, S. and Chia L.T. (2006) *Does ontology help in image retrieval?: a comparison between keyword, text ontology and multi-modality ontology approaches*, In Proc. of ACM Multimedia.
- Wilkins, P., P. Ferguson, A.F. Smeaton, C. Gurrin (2005) *Text Based Approaches for Content-Based Image Retrieval on Large Image Collections*. In Proc. of EWIMT.
- Yang, J., Wenyin, L., Zhang, H., Zhuang, Y. (2001) *Thesaurus-aided Approach for Image Browsing and Retrieval*, In Proc. of IEEE – ICME.