



Faculté des Nouvelles technologies de l'Informations et de la Communication
Département d'Informatique et de la Technologie de l'Information

Similarité sémantique entre concepts : ***Application à la recherche d'images***

Présenté par : Asma SIAGH, Chaima DEROUICHE.

Encadré par : Meriem KORICHI

Résumé : *Le défi des mesures de similarité sémantique consiste à trouver une méthode identique au raisonnement de l'être humain. L'utilisation d'ordinateurs pour quantifier et comparer les similitudes sémantiques est devenue un domaine important de la recherche dans divers domaines, y compris l'intelligence artificiel, la recherche d'information, traitement des langues naturels, ...etc.[1]*

Plusieurs efforts ont été fait pour définir la similarité sémantique. Les plus connus sont ceux qui utilisent une ontologie comme WordNet. C'est dans ce contexte qui s'inscrit notre travail, qui est devisé en deux parties: dans la première partie ,nous allons effectuer une comparaison pratique entre certains mesures de similarités sémantiques plus connus à base d'ontologie (Resnik, Lin et Jiang) et dans la deuxième partie, nous allons prouver les bénéfices d'application de ces mesures dans les moteurs de recherche d'images. Nous allons implémenter un moteur de recherche d'images en se basant sur les mesures de similarité étudiés.

Mots clés : *Mesures de similarité sémantique, Recherche d'images, WordNet, ImageNet, Resnik, Lin, Jiang.*

1. Introduction

La similarité sémantique est une métrique définie sur un ensemble de documents ou de termes. L'idée de la distance entre les documents et les termes est basée sur la ressemblance de leurs significations ou leurs contenus sémantiques, par opposition à la similitude qui peut être estimée en ce qui concerne leur représentation syntaxique (par exemple, leur format string) . Ce sont des outils mathématiques utilisés pour estimer la force de la relation sémantique entre les unités de la langue, des concepts ou des instances, à travers une description numérique obtenue en fonction de la comparaison des informations à l'appui de leur signification ou de décrire leur nature.

2. Similarité et proximité sémantique

Le terme similarité sémantique est souvent confondu avec le terme de proximité sémantique, mais la notion de similarité sémantique est plus spécifique que la proximité sémantique. La proximité comprend toute relation entre deux termes, comme l'antonymie , alors que la similarité sémantique ne comprend que la relation "est un". Par exemple, "une voiture" est similaire à "bus", mais est également liée à "route" et "conduite".

Une grande partie de la littérature utilise ces termes de façon interchangeable, avec des termes tels que la distance sémantique. En substance, la similarité sémantique, la distance sémantique et proximité sémantique tout cela signifie, "Combien de terme A à avoir avec le terme B?" La réponse à cette question est généralement un nombre compris entre -1 et 1, ou entre 0 et 1, où 1 signifie une similitude extrêmement élevé.

3. Approches de la similarité sémantique

Les approches de la similarité sémantique peuvent être classées en deux grandes catégories:

3.1 Similarité sémantique à base statistique

Les approches de similarité statistiques peuvent être apprendre à partir des données, ou des prédéfinies. De façon générale, ces approches construisent un modèle statistique de documents et l'utilisent à calculer la similitude. Plusieurs techniques ont été proposés dont nous pouvons cité : LSA (Latent semantic analysis) , PMI (Pointwise mutual

information), SOC-PMI (Second-order co-occurrence pointwise mutual information),....etc.

3.2 Similarité sémantique à base d'ontologie

3.2.1 Définition d'une ontologie

Les ontologies fournissent une spécification formelle d'une conceptualisation partagée. Elles sont exploitables par une machine et construisent à partir du consensus d'une communauté d'utilisateurs ou des experts de domaine. Les ontologies représentent une source de connaissance très fiable et structurée. Pour cette raison et grâce aux initiatives du Web sémantique, qui a apporté la création des milliers d'ontologies de domaine, les ontologies ont été largement exploitées dans les systèmes à base de connaissance, et plus précisément, pour le calcul de la similarité sémantique.[2]

3.2.2 Les composants d'une ontologie

Les composants d'une ontologie sont :

- **Les concepts** : une description d'une tâche, d'une fonction, d'une action, d'une stratégie ou d'un processus de raisonnement...etc.
- **Les relations** : Représentent un type d'interaction entre les notions d'un domaine.
- **Les axiomes** : Une ontologie est en outre composée d'axiomes qui forment des contraintes sémantiques pour le raisonnement et donnent un acompte d'une conceptualisation. Ils prennent la forme d'une théorie logique.

L'exemple paradigmatique d'une ontologie est **WordNet**.

3.2.3 Wordnet

Wordnet est une ontologie qui organise plus de 100,000 concepts anglais généraux, qui sont sémantiquement structurés. Elle regroupe les concepts en ensembles de synonymes appelés synsets. Wordnet fournit des définitions générales et courtes, et enregistre les différentes relations sémantiques entre ces ensembles de synonymes. Elle est particulièrement bien adaptée pour de mesures de similarité, car elle organise des noms et des verbes en hiérarchie de relation « est-un ».

Dans la littérature plusieurs mesures de similarité sémantique ont été proposées et la plupart d'entre eux ont été testés sur WordNet.

Les mesures de similarité sémantiques à base d'ontologie peuvent être classées en deux grandes catégories: approche basée sur les chemins et approche basée sur le contenu de l'information.

a. Mesures basées sur les chemins

L'idée principale de ce type de mesures est le fait que la similarité entre deux concepts est en fonction de la longueur du chemin reliant les concepts et leurs positions dans la taxonomie. Parmi ces mesures, nous pouvons citer: les mesures de ekar, de Cheng, de Cline, de Wu, et de Del Pozo. L'avantage principal des mesures basées sur les chemins est leur simplicité. Ils comptent seulement sur le modèle graphique d'une ontologie et dont l'évaluation exige un coût informatique bas (en comparaison avec les approches qui traitent des corpus de texte).

b. Mesures basées sur le contenu informationnel (IC)

Cette approche a pour objectif de surmonter les limites de l'approche basée sur les chemins, ceci en complétant la structure taxonomique d'une ontologie avec la distribution de l'information de concepts évalués dans des corpus d'apport. Il exploite la notion du contenu de l'information (IC), en associant des probabilités d'apparence à chaque concept dans la taxonomie, calculée à partir de leurs présences dans un corpus donné.

IC d'un terme A est calculé selon le log négatif de sa probabilité de présence. De cette manière, les mots peu fréquents sont considérés plus informatifs que les communs.

$$\mathbf{IC(A) = -\log(p(A))}$$

Parmi les mesures les plus connues de cette approche: La mesure de Resnik, la mesure de Lin et la mesure de Jiang.

- **La mesure de Resnik**

Resnik a été le premier qui a fusionner l'utilisation d'ontologie avec le corpus. Il définit la similarité sémantique entre deux concepts par la quantité d'information qu'ils partagent. Cette information partagée est égale au contenu informationnel du plus petit généralisant (LCS) – le concept le plus spécifique qui subsume les deux concepts dans l'ontologie.

$$\text{SimRes} = \text{IC}(\text{LCS}(\text{C1}, \text{c2}))$$

- **La mesure Jiang-Conrath**

La notion d'IC est également utilisée dans l'approche de la Jiang et Conrath. Cette approche soustrait la somme de l'IC de chaque concepts du double l' IC d'LCS de ces concepts .

$$\text{DisJC} = (\text{IC}(\text{C1}) + \text{IC}(\text{C2})) - 2(\text{IC}(\text{LCS}(\text{C1}, \text{C2})))$$

- **La mesure Lin**

La mesure de similarité décrite par Lin utilise les mêmes éléments que DisJC, mais d'une manière différente:

$$\text{SimLin} = \frac{2\text{IC}(\text{LCS}(\text{C1}, \text{C2}))}{\text{IC}(\text{C1}) + \text{IC}(\text{C2})}$$

4. La recherche d'images comme domaine d'application de la similarité sémantique

4.1 Définition

Un moteur de recherche d'images est un outil qui permet de localiser les images cherchées par un utilisateur en temps raisonnable.

Deux approches de recherches sont envisagées : recherche d'image par le texte et recherche d'image par le contenu visuel. Nous nous intéressons à la première catégorie où nous pouvons appliquer la similarité sémantique.

4.2 Recherche d'images par le text (TBIR) : (Textual Based Image Retrieval)

Ce mode permet à l'utilisateur de formuler sa requête en composant une requête textuelle. La recherche est effectuée en comparant les termes contenus dans la requête avec les mots clés ou les descripteurs textuels qui représentent les images de la base. L'avantage des systèmes textuels est qu'ils donnent la possibilité à l'utilisateur de poser des requêtes dans un langage de haut niveau lui permettant ainsi d'exprimer son besoin d'information facilement. Cependant la technique utilisée dans ces systèmes souffre de limitations importantes :

- Le texte dépend de la langue : par exemple si on écrit شجرة, toutes les images fournies seront celles qui sont nommées par le terme شجرة, sans prendre en considération d'autres images avec d'autres langues.

- Les résultats des interrogations ne satisfont pas le besoin d'utilisateur car les annotations textuelles dépendent de ce que l'annotateur peut saisir lors de la création de la base d'images. Ainsi que des utilisateurs différents peuvent décrire la même image d'une manière différente, par conséquent le moteur de recherche retourne des images qui ne correspondent pas à la requête (problème du bruit) et vice versa (problème de silence).

4.3 Recherche par le contenu sémantique SBIR (Semantic-Based Image Retrieval)

Dans la recherche textuelle classique (TBIR), les moteurs se limitent à retourner à l'utilisateur les images qui sont annotées explicitement avec les concepts présents dans la requête. Cependant, il a été noté qu'une image peut ne pas être annotée avec les concepts de la requête alors qu'elle est tout à fait pertinente à ces derniers. La recherche d'image à base de structure sémantique cherche à retrouver les images annotées avec les concepts présents explicitement dans la requête ainsi que celles annotées avec des concepts ayant une relation avec ces derniers. Cette relation est calculée à travers l'utilisation des mesures de similarité sémantique.

4.4 Architecture d'un moteur de recherche :

Le fonctionnement d'un moteur de recherche d'images se fait en deux phases : phase d'indexation (offline) et phase de recherche (online).

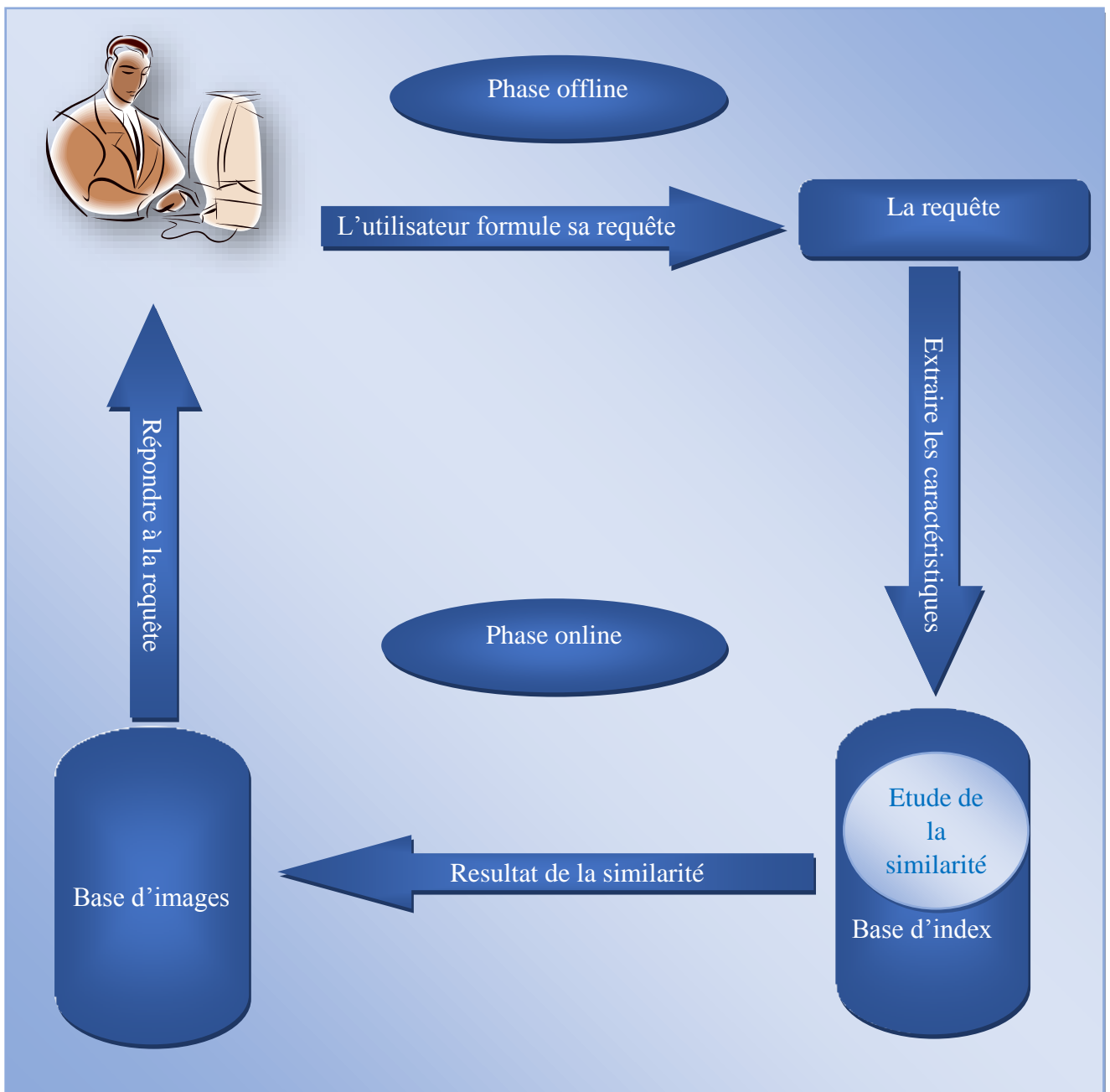


Figure 1: Architecture d'un moteur de recherche d'images

4.4.1 Phase offline(Indexation)

En phase d'indexation, chaque image est analysée et les concepts caractérisant son contenu informationnel sont extraits. Une fois qu'un ensemble de concepts a été constitué pour chaque image de la base, il est parfois nécessaire de les organiser de façon à pouvoir retrouver, ensuite, le plus rapidement possible des images recherchées. En effet, l'organisation des données n'est pas obligatoire, mais, dès lors que la base de recherche devient conséquente et que les méthodes de recherche utilisées nécessitent le stockage des données textuelles, elle devient vivement conseillée. Cette organisation consiste à la construction de structures de données, appelées index.

4.4.2 Phase de recherche

Une fois les images indexées, le problème est de pouvoir les retrouver simplement. La recherche d'images se traduit alors par la mise en correspondance des représentations sémantiques des images et d'une représentation sémantique de la requête. La mise en correspondance se fait grâce à des modèles de recherches.

4.4.2.1 Les modèles de recherches d'images

Il existe trois grandes catégories de modèles de recherche :

4.4.2.2 Modèle Booléen

Le premier et le plus simple des modèles, la requête s'exprime par des termes liés avec des opérateurs logiques (and, ou, not). Une image est sélectionnée si et seulement si elle satisfait l'expression booléenne.

4.4.2.3 Modèle probabiliste

Est une méthode probabiliste de représentation du contenu d'une image. Elle est utilisée pour exprimer une estimation de la probabilité de pertinence d'une image par rapport à une requête, et ainsi classer une liste des images dans l'ordre décroissant d'utilité probable pour l'utilisateur.

Le modèle probabiliste standard suppose que les images peuvent être classés en deux classes : celle des images pertinents, notée R, composée des images que l'utilisateur souhaite retrouver parmi l'ensemble des images disponibles, et celle des images non

pertinents, notée \bar{R} , composée du reste des images. Les ensembles R et \bar{R} sont donc disjoints et considérés comme deux variables aléatoires indépendantes.

4.4.2.4 Modèle vectoriel

Dans ce modèle, chaque image est représenté par un vecteur V , dont la dimension correspond au nombre des termes d'indexation. Chaque élément V_i du vecteur V consiste en poids associé au terme d'indice i et identifie le nombre d'occurrence du terme i . La composante du vecteur représente donc, le poids du mot i dans l'image. L'un du schéma de pondération les plus utilisés est le TF_IDF.

Telle que

$$\mathbf{TF} = 1/\text{taille du document}$$

$$\mathbf{IDF} = \mathbf{Log} \frac{|D|}{|\{d_i : i \in d_i\}|} \quad \text{Où } |D| \text{ est le nombre total de documents dans le}$$

corpus et le dénominateur désigne le nombre de documents où le terme i apparaît.

5. Présentation de notre travail

5.1 Comparaison entre les mesures de similarité sémantique: Resnik, Lin et Jiang

Ces mesures de similarité sont à base d'ontologie. Nous avons choisis l'ontologie Wordnet et puisque cette dernière est trop volumineuse, nous avons limité notre étude seulement à une partie. La figure suivante illustre cette partie.

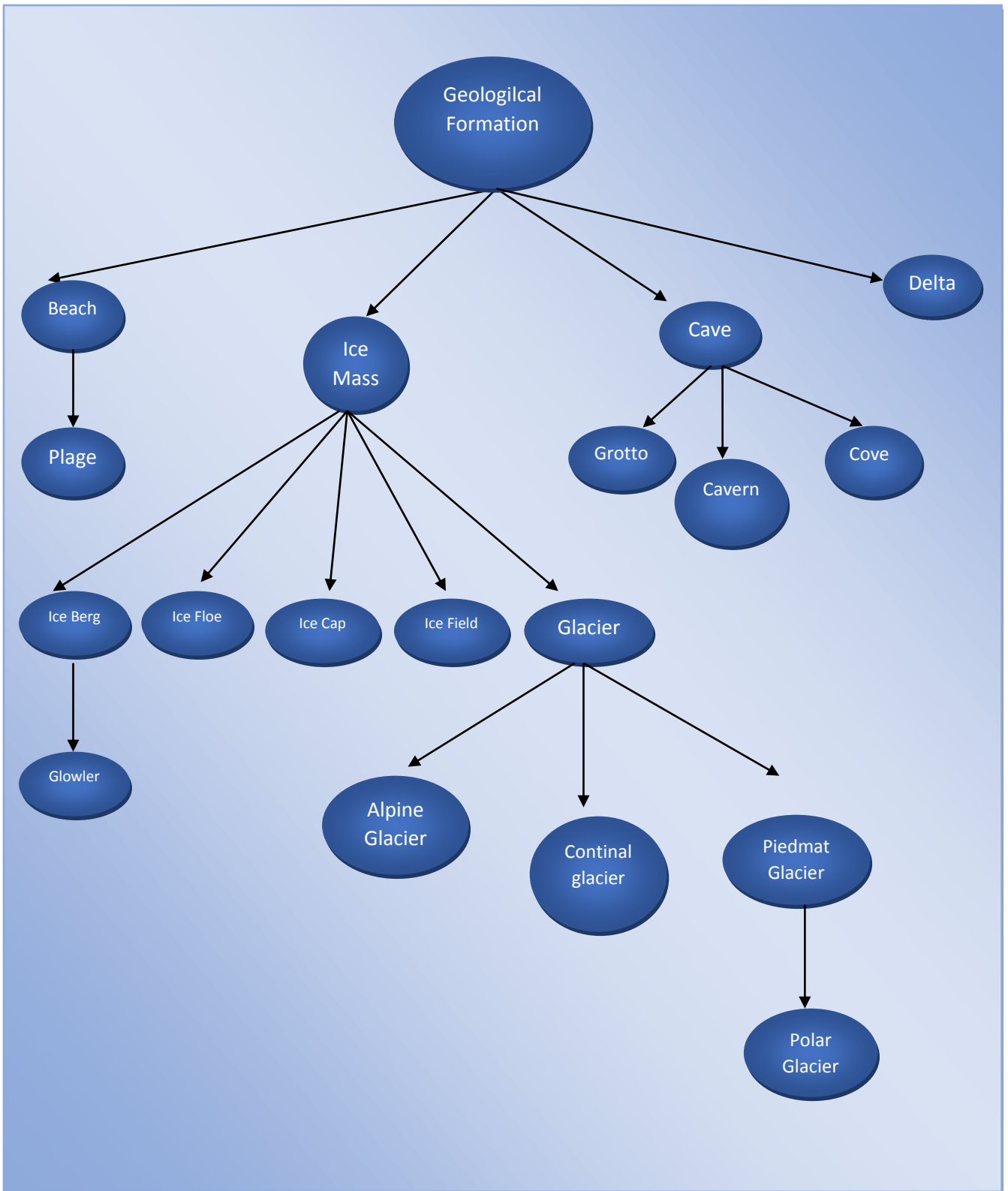


Figure 2: Notre Ontologie (une partie de WordNet)

Nous avons calculé la similarité sémantique entre chaque paire de concepts avec les trois mesures. Les résultats obtenus sont illustrés dans les tableaux suivants:

Matrice_Resnik	Geolo_Form	Beach	Plage	cave	cavern	Cove	grotto	delta	Ice_mass	Glacier	Alpine_glacier	Continental_glacier	Polar_glacier	Piedmat_glacier	Ice_berg	glower	Ice_cap	Ice_field	Ice_floe
Geolo_form	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
beach	0	2.48	2.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
plage	0	2.48	2.48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cave	0	0	0	1.39	1.39	1.39	1.39	0	0	0	0	0	0	0	0	0	0	0	0
cavern	0	0	0	1.39	2.48	1.39	1.39	0	0	0	0	0	0	0	0	0	0	0	0
cove	0	0	0	1.39	1.39	2.48	1.39	0	0	0	0	0	0	0	0	0	0	0	0
grotto	0	0	0	1.39	1.39	1.39	2.48	0	0	0	0	0	0	0	0	0	0	0	0
delta	0	0	0	0	0	0	0	2.48	0	0	0	0	0	0	0	0	0	0	0
Ice_mass glacier	0	0	0	0	0	0	0	0	0.54	0.54	0	0	0	0	0.54	0	0.54	0.54	0.54
Alpine_glacier	0	0	0	0	0	0	0	0	0.54	1.39	1.39	1.39	1.39	1.39	0.54	0.54	0.54	0.54	0.54
Continental_glacier	0	0	0	0	0	0	0	0	0	1.39	1.39	2.48	2.48	1.39	0.54	0.54	0.54	0.54	0.54
Polar_glacier	0	0	0	0	0	0	0	0	0	0.54	1.39	2.48	2.48	1.39	0.54	0.54	0.54	0.54	0.54
Piedmat_glacier	0	0	0	0	0	0	0	0	0	1.39	1.39	1.39	1.39	2.48	0.54	0.54	0.54	0.54	0.54
Ice_berg	0	0	0	0	0	0	0	0	0.54	0.54	0.54	0.54	0.54	2.48	0.54	0.54	0.54	0.54	0.54
glower	0	0	0	0	0	0	0	0	0	0.54	0.54	0.54	0.54	0.54	0.54	2.48	0.54	0.54	0.54
Ice_cap	0	0	0	0	0	0	0	0	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	2.48	0.54	0.54
Ice_field	0	0	0	0	0	0	0	0	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	2.48	0.54
Ice_floe	0	0	0	0	0	0	0	0	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	2.48

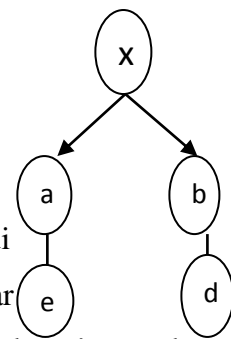
Tableau 1: Similarité sémantique entre concepts par l'utilisation de la mesure de Resnik.

Matrice_Lin	Geolo_Form	Beach	Plage	cave	cavern	cove	grotto	delta	Ice_mass	Glacier	Alpine_glacier	Continental_glacier	Polar_glacier	Piedmat_glacier	Ice_berg	glower	Ice_cap	Ice_field	Ice_floe
Geolo_form	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
beach	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
plage	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cave	0	0	0	1	0.72	0.72	0.72	0	0	0	0	0	0	0	0	0	0	0	0
cavern	0	0	0	0.72	1	0.56	0.56	0	0	0	0	0	0	0	0	0	0	0	0
cove	0	0	0	0.72	0.56	1	0.56	0	0	0	0	0	0	0	0	0	0	0	0
grotto	0	0	0	0.72	0.56	0.56	1	0	0	0	0	0	0	0	0	0	0	0	0
delta	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Ice_mass	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0.36	0	0.36	0.36	0.36
glacier	0	0	0	0	0	0	0	0	0.56	1	0.72	0.72	0.28	0.72	0.28	0.28	0.28	0.28	0.28
Alpine_glacier	0	0	0	0	0	0	0	0	0	0.28	1	0.56	0.56	0.56	0.22	0.22	0.22	0.22	0.22
Continental_glacier	0	0	0	0	0	0	0	0	0	0.28	0.56	1	1	0.56	0.22	0.22	0.22	0.22	0.22
Polar_glacier	0	0	0	0	0	0	0	0	0	0.28	0.56	1	1	0.56	0.22	0.22	0.22	0.22	0.22
Piedmat_glacier	0	0	0	0	0	0	0	0	0	0.28	0.56	0.56	0.56	1	0.22	1	0.22	0.22	0.22
Ice_berg	0	0	0	0	0	0	0	0	0.36	0.28	0.22	0.22	0.22	0.22	1	0.22	0.22	0.22	0.22
glower	0	0	0	0	0	0	0	0	0	0.28	0.22	0.22	0.22	0.22	0.22	1	0.22	0.22	0.22
Ice_cap	0	0	0	0	0	0	0	0	0.36	0.28	0.22	0.22	0.22	0.22	0.22	0.22	1	0.22	0.22
Ice_field	0	0	0	0	0	0	0	0	0.36	0.28	0.22	0.22	0.22	0.22	0.22	0.22	0.22	1	0.22
Ice_floe	0	0	0	0	0	0	0	0	0.36	0.28	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	1

Tableau 2: Similarité sémantique entre concepts par l'utilisation de la mesure de Lin.

Matrice_jiang	Geolo_Form	Beach	Plage	cave	cavern	cove	grotto	delta	Ice_mass	Glacier	Alpine_glacier	Continetal_glacier	Polar_glacier	Piedmat_glacier	Ice_berg	glower	Ice_cap	Ice_fie
Geolo_form	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
beach	0	0	0	4.97	4.97	4.97	4.97	3.02	3.87	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97
plage	0	0	0	3.87	4.97	4.97	4.97	4.97	3.02	3.87	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97
cave	0	3.87	3.87	0	1.1	1.1	1.1	3.87	1.93	2.77	3.87	3.87	3.87	3.87	3.87	3.87	3.87	3.87
cavern	0	4.97	4.97	1.1	0	2.2	2.2	4.97	3.02	3.87	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97
cove	0	4.97	4.97	1.1	2.2	0	2.2	4.97	3.02	3.87	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97
grotto	0	4.97	4.97	1.1	2.2	2.2	0	4.97	3.02	3.87	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97
delta	0	4.97	4.97	3.87	4.97	4.97	4.97	0	3.02	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97	4.97
Ice_mass	0	3.02	3.02	1.93	3.02	3.02	3.02	3.02	0	0.85	3.02	3.02	3.02	3.02	1.95	3.02	1.95	1.95
glacier	0	3.87	3.87	2.77	3.87	3.87	3.87	3.87	0.85	0	1.1	1.1	2.79	1.1	2.79	2.79	2.79	2.79
Alpine_glacier	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	3.02	1.1	0	2.2	2.2	2.2	3.89	3.89	3.89	3.89
Continetal_glacier	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	3.02	1.1	2.2	0	0	2.2	3.89	3.89	3.89	3.89
Polar_glacier	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	3.02	2.79	2.2	0	0	2.2	3.89	3.89	3.89	3.89
Piedmat_glacier	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	3.02	1.1	2.2	2.2	2.2	0	3.89	3.89	3.89	3.89
Ice_berg	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	1.95	2.79	3.89	3.89	3.89	3.89	0	3.89	3.89	3.89
glower	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	3.02	2.79	3.89	3.89	3.89	3.89	0	0	3.89	3.89
Ice_cap	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	1.95	2.79	3.89	3.89	3.89	3.89	3.89	3.89	0	3.89
Ice_field	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	1.95	2.79	3.89	3.89	3.89	3.89	3.89	3.89	3.89	0
Ice_floe	0	4.97	4.97	3.87	4.97	4.97	4.97	4.97	1.95	2.79	3.89	3.89	3.89	3.89	3.89	3.89	3.89	3.89

Tableau 3: Similarité sémantique entre concepts par l'utilisation de la mesure de jiang



5.2 Discussion

On remarque qu'avec l'utilisation de la mesure de Resnik, les concepts qui ont le même LCS, ont la même valeur de la similarité. Ceci est justifié par le fait que la mesure de Resnik ne dépend que du LCS. On a donné la même valeur de similarité entre a,b et d,e malgré que a,b sont plus proches de LCS que d,e, c'est l'inconvénient de cette mesure.

La méthode de Lin présente une petite amélioration de la méthode de Resnik par prendre en considération non seulement le contenu informationnel de LCS mais aussi le contenu informationnel de chaque concept, ce qui explique qu'on trouve la valeur de similarité entre le concept et lui-même=1 et elle diminue avec son père et ses frères. D'après les résultats qui ont fournis la méthode de Jiang, Il est intéressant de noter que ceci est une mesure de dissimilitude car plus les termes sont différents, plus la distance entre eux augmente et vice versa..

5.3 Notre moteur de recherche

Nous avons implémenté un moteur de recherche d'images par le contenu sémantique avec l'utilisation des mesures de similarité sémantique présentées. Notre objectif est de montrer les apports d'utilisation de la similarité sémantique au domaine de la recherche d'image.

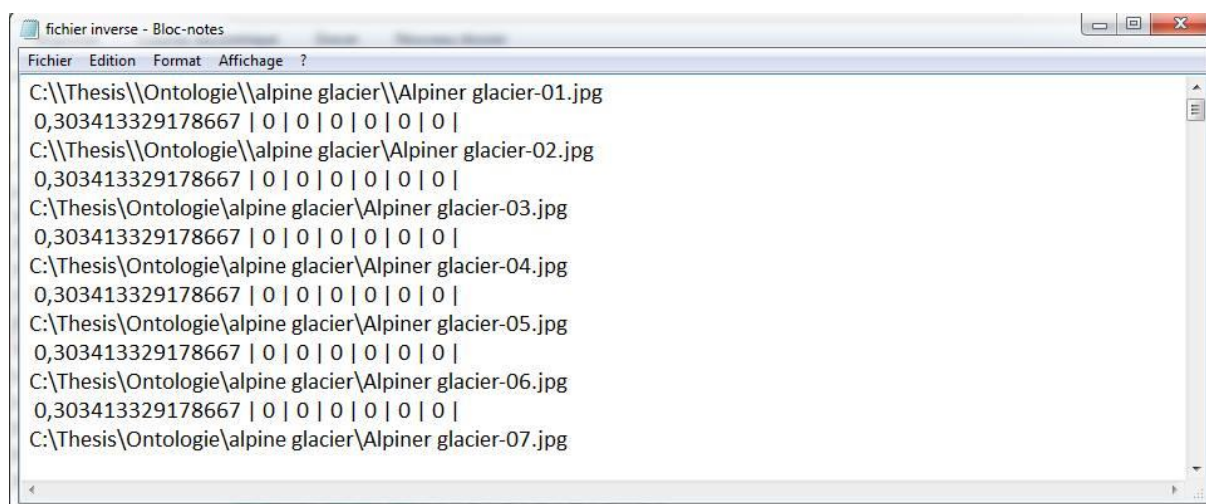
Pour nos expérimentation, nous avons choisi la collection d'images ImageNet. Ce choix est justifié par le fait que cette collection est annotée à base de l'ontologie WordNet. Comme modèle de recherche, nous avons utilisé le modèle vectoriel qui offre plus d'avantages par rapport aux autres modèles. La figure suivante illustre la page d'accueil de notre moteur.



Figure 3:Page d'accueil de notre moteur.

5.3.1 Phase d'indexation (extraction des caractéristiques)

Nous avons extrait les descripteurs de chaque image selon le modèle vectoriel. Pour chaque image, nous avons calculé son vecteur selon le mécanisme de pondération de termes (tf*idf) comme il est illustré dans la figure 4.



```
fichier inverse - Bloc-notes
Fichier Edition Format Affichage ?
C:\\Thesis\\Ontologie\\alpine glacier\\Alpiner glacier-01.jpg
0,303413329178667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
C:\\Thesis\\Ontologie\\alpine glacier\\Alpiner glacier-02.jpg
0,303413329178667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
C:\\Thesis\\Ontologie\\alpine glacier\\Alpiner glacier-03.jpg
0,303413329178667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
C:\\Thesis\\Ontologie\\alpine glacier\\Alpiner glacier-04.jpg
0,303413329178667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
C:\\Thesis\\Ontologie\\alpine glacier\\Alpiner glacier-05.jpg
0,303413329178667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
C:\\Thesis\\Ontologie\\alpine glacier\\Alpiner glacier-06.jpg
0,303413329178667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
C:\\Thesis\\Ontologie\\alpine glacier\\Alpiner glacier-07.jpg
```

Figure 4: Extraction des vecteurs de caractéristiques selon le modèle vectoriel

5.3.2 Phase de recherche

On demande à l'utilisateur de formuler une requête textuelle avec un ou plusieurs mots clés. Pour chaque mesure, on a identifié un seuil tels que les termes de chaque requête seront expansés par tous les concepts de l'ontologie qui ont une similarité sémantique plus grande que le seuil identifié. Le moteur ensuite calcul le descripteur de la nouvelle requête et le compare avec les descripteurs des images de la collection. Les résultats seront affichés selon leur similarité avec la requête. L'affichage est effectué aussi avec la méthode classique (TBIR) pour mieux comparer les résultats obtenus.

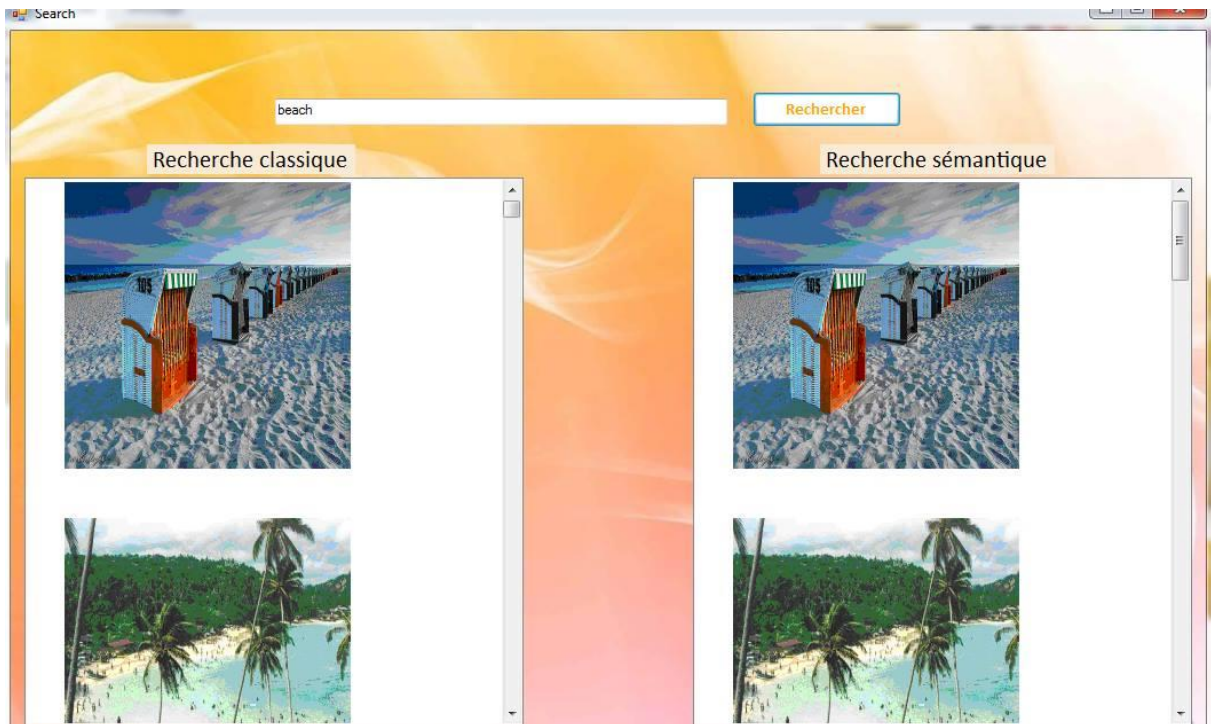


Figure 5: Affichage des résultats

6. Evaluation

Pour voir les apports d'utilisation de la similarité sémantique , nous allons comparer les résultats de recherche de notre moteur avec celles de la recherche classique (TBIR) , nous avons basé sur le critère du rappel.

Le Rappel (Re) mesure la proportion de bonnes images retournées à l'utilisateur parmi toutes les bonnes images dans la collection.

$$Re = \frac{\text{nombre des bonnes images}}{\text{nombre total des bonnes images}}$$

Nous avons lancé plusieurs requêtes avec la méthode classique de recherche et avec la méthode sémantique avec l'utilisation de la mesure de similarité puis nous avons calculé le rappel. Les résultats obtenus sont illustrés dans la figure 6.

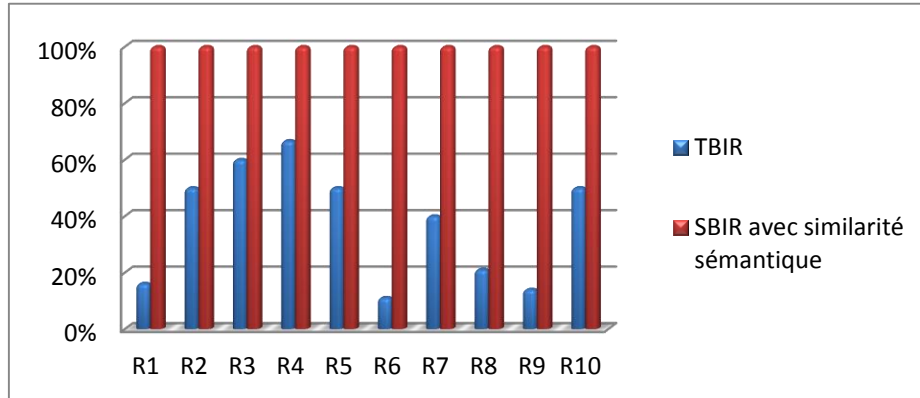


Figure 6: Comparaison entre TBIR et SBIR avec similarité sémantique

On remarque qu'avec l'utilisation de la méthode sémantique, le rappel est augmenté d'une façon très considérable. Ceci est justifié par le fait que la recherche est effectuée non seulement avec les termes initiaux de la requête mais aussi avec les termes qui ont un similarité sémantique élevé avec eux.

7. Conclusion

Dans ce travail, nous avons étudié c'est quoi une mesure de similarité sémantique et quelles sont ses approches. Nous avons concentré sur les approches à base d'ontologie et plus précisément sur celle à base du contenu informationnel (IC). Une comparaison expérimentale est effectuée entre les trois plus connus mesures: Resnik, Lin et Jiang.

Pour mieux comprendre l'utilité d'utilisation de la similarité sémantique, Nous avons pris la recherche d'image comme domaine d'application. Nous avons implémenté un moteur de recherche d'images en exploitant ces mesures de similarité et prouvé que le rappel est augmenté d'une façon considérable.

Références

- [1] David Sánchez et Montserrat Batet. " Ontology-based semantic similarity: A new feature-based approach" Expert Systems with Applications. 2012
- [2] Mohammed Hadj Taib." Ontology-based approach for measuring semantic similarity". Nov 2014 · Engineering Applications of Artificial Intelligence.
- [3] Thabet Slimani . " Description and Evaluation of Semantic similarity Measures Approaches".
- [4] Jorge, G., and Eduardo, M., 2008. "Web-Based Measure of Semantic Relatedness", 9th international conference on Web Information Systems Engineering, Springer-Verlag Berlin.
- [5] M. L. KHERFI and D.ZIOU. Image Retrieval from the World Wide Web: Issues, Techniques, and Systems.. 2004, ACM Computing Surveys.