

UNIVERSITÉ KASDI MERBAH OUARGLA

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

Département d'Informatique et des Technologies de l'Information



Mémoire

Master Académique

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Informatique Industrielle

Présenté par :

✉ Melle HAMROUNI Lamis

✉ Melle KADRI Amira

Thème

**Amélioration de la recherche d'images par
mesure de similarité sémantique entre mots**

Soutenu publiquement

Le: 07/06/2015

Devant le jury :

✉ M. BAKKARI Fouad	Président	UKM OUARGLA
✉ M. KHERFI Med Lamine	Examineur	UKM OUARGLA
✉ Mme. DEBBAGH Farah	Rapporteur	UKM OUARGLA

Année universitaire : 2014/2015

Remerciement

Avant tous, nous remercions ALLAH, pour nous avoir données la capacité pour réaliser ce travail.

*Un grand merci à notre encadreur Mme **Farah Debbagh** qui a fourni des efforts énormes, par ses conseils, ses orientations, et ses informations.*

*Un grand merci à **Mr Med Lamine Kherfi** pour leurs conseils, ses orientations, et ses informations, et leur soutien durant notre encadrement.*

*Un grand merci à **Mlle Meriem Korichi** pour leurs conseils, ses orientations, et ses informations, et leur encouragement durant notre encadrement.*

*Nous tenons à remercier les membres du jury et tous les enseignants de département **MATHEMATIQUE ET INFORMATIQUE.***

*Nous remercions particulièrement **Mr Med Lamine Kherfi** et **Mr Fouad Bakari.***

Enfin, nos plus chaleureux remerciements pour toute personne ayant participé de près ou de loin à la réalisation de ce modeste travail.

Résumé

Cette dernière décennie témoigne un accroissement exponentiel du volume des données multimédia tel que les images. A cet effet, l'introduction des méthodes de recherche d'images sur le web est devenue une nécessité considérable. Par conséquent, plusieurs types de recherche d'images sont apparus dans la littérature, la recherche par texte et la recherche par contenu.

Notre travail consiste à résoudre un des problèmes de la recherche par texte, nous nous sommes intéressés en particulier au problème du silence. Nous proposons comme solution trois méthodes, la première méthode, nous créons des relations entre l'image et les concepts, à partir du manque des relations entre l'image et certains concepts dans cette méthode on a construit une deuxième méthode qui permet de créer des relations entre les concepts eux mêmes, l'utilisation des deux méthodes précédentes présente notre troisième méthode : c'est la méthode de spreading activation. **Mots clés : recherche d'image par texte, matching (recherche binaire), silence, les relations sémantiques, relation entre l'images et les concepts, relation des concepts entre eux.**

Abstract

The last decade has seen an exponential increase growth in multimedia data volume such as images. Introducing methods of images search on the internet became considerably needed. Therefore several models in the literature researches have appeared as well as text researches and content ones.

Our study aims mainly at solving the problems of a text search, we have interested on the silence problem. So, we proposed three methods as solution the first method is to create a relation between the images and the concepts, due the lack of relations among images and some concepts we have constructed the second method which allow to create relations among the concepts themselves. And the use of the two previous methods provides the method of spreading activating. **Keywords: text image retrieval, matching (binary research), silence, semantic relationship, relationship between image and concepts, relationship between concepts.**

ملخص

عرفت العشرية الأخيرة نموا هائلا في حجم المعطيات و البيانات متعددة الوسائط كالصور , و كنتيجة لهذا النمو أصبح من الضروري إيجاد طرق بحث عن الصور , و نتيجة لذلك ظهرت طرق و أنواع بحث عدة في هذا المجال إما عن طريق كلمات مفتاحيه أو عن طريق محتوى الصور.

الهدف الرئيسي من عملنا في هذه المذكرة هي حل مشاكل البحث عن الصور عن طرق النص، و كان جل اهتمامنا هو معالجة مشكل الصمت. و كحل لهذه المشكلة العظيمة اقترحنا ثلاث تقنيات، تركز التقنية الأولى على إنشاء علاقات بين الصورة و المفاهيم. وانطلاقا من نقص في علاقات بعض المفاهيم بالصور في هذه التقنية أنشأنا التقنية الثانية و التي تسمح بخلق علاقات بين المفاهيم. وباستخدامنا التقنيتين السابقتين نحصل على التقنية الأخيرة ألا وهي تقنية "تفعيل الانتشار"

الكلمات المفتاحية : البحث عن صورة عن طريق نص البحث الثنائي, الصمت, العلاقات الدلالية, العلاقة بين الصور و المفاهيم, علاقات المفاهيم فيما بينهم.

Sommaire

Introduction générale

1.	Contexte et problématique :	1
2.	Objectifs :	2
3.	Organisation du mémoire :	2

Chapitre 01 :La recherche d'images

1	Introduction:	8
2	Définition d'une image :	8
3	Moteur de recherche d'images	8
4	Types de requêtes :	9
4.1	Requêtes textuelles	10
4.2	Requêtes visuelles :	10
4.2.1	Requête par image exemple :	10
4.2.2	Requête par sketch (esquisse) :	11
5	Types de moteurs de recherche :	11
5.1	Moteur de Recherche basé sur le contenu visuelle (CBIR):	11
5.1.1	Indexation et recherche d'images par le contenu visuel:	12
5.1.2	Les Avantages :	13
5.1.3	Les limites:	14
5.2	Moteur de Recherche d'images basé sur le texte :	14
5.2.1	Annotation des images:	15
5.2.2	Recherche d'images par mot-clé :	15
5.2.3	Recherche d'images par concept sémantique :	17
6	Conclusion :	18

Chapitre 02 :Problématique et Solution

1.	Introduction :	20
----	----------------------	----

2.	Les techniques utilisées pour résoudre le problème de silence:	20
2.1	Technique de relation entre l'image et les concepts(RIC):	21
2.2	Technique de relation entre concepts (RC):	22
6.1	Définition Spreading activation(SA):.....	24
2.2.1	Principe SA :	25
2.2.2	L'application de SA a notre problème :	26
3.	Conclusion :	26
 Chapitre 03 :L'évaluation		
1.	L'évaluation de techniques RIC, RC et SA.....	28
2.	Technique de relation entre l'image et les concepts (RIC):	28
3.	Technique de relation entre concept RC:	33
4.	Technique de SA :	39
5.	Discussion :	46
6.	Conclusion :	47
 Conclusion générale et perspective		
7	Conclusion générale et perspectives :	49
8	Bibliographie.....	Erreur ! Signet non défini.

Liste de figures

Figure 1 : Architecture générale de moteur de recherche d'images.....	9
Figure 2 : les requêtes textuelles	10
Figure 3 : les requêtes visuelles	11
Figure 4 : Architecture de moteur de recherche d'image par contenu visuel.....	13
Figure 5 : fossé sémantique.....	14
Figure 6 : moteur de recherche d'images par texte	15
Figure 7 : ambiguïté de recherche par le mot clé « avocat ».....	17
Figure 8 : Exemple d'une collection d'images.....	20
Figure 9 : l'image et son code html(contenu).	21
Figure 10 : la relation de l'image avec les concepts	22
Figure 11 : l'absence des relations entre les concepts.....	22
Figure 12 : création des relations entre les concepts.....	23
Figure 13 : le mécanisme de recherche d'une information dans la mémoire humaine.	25
Figure 14 : le principe de Spreading Activation	25
Figure 15 : l'application de spreading activation	Erreur ! Signet non défini.
Figure 16 : matrice <i>I</i>	28
Figure 17 : la technique de RIC	28
Figure 18:la relation entre l'image0 et les concepts.....	32
Figure 19 : la recherche par le concept sea	33
Figure 20:La Technique RC.....	34
Figure 21 : création de relation par la technique de RC.....	38
Figure 22:le résultat de recherche par concepts coast.....	38
Figure 23 Technique de Spreading Activation.....	39
Figure 24 : les relations créer par la technique de SA	44
Figure 25: Le résultat de recherche par concept mountain	45
Figure 26 : courbe illustre le rappel des les trois technique RIC,RC et SA	46
Figure 27 :courbe illustre la précision des trois technique RIC ,RC et SA	46

Liste des tableaux

Tableau 1 : les résultats de la technique RIC	29
Tableau 2 :les résultats de la recherche binaire , RIC et RC	35
Tableau 3:les résultats les techniques la recherche binaire, RIC, RC, SA.	40

Introduction générale

Introduction générale

1. Contexte et problématique :

« L'image vaut mieux que beaucoup de mots », Même si l'écriture est aujourd'hui le moyen de communication le plus commun entre les êtres humains, l'image reste le moyen d'expression le plus ancien et le plus expressif. Quelque soit le lieu ou le moment, le sens de l'image reste compréhensible par tout être humain.

Le nombre d'images dans le monde a augmenté, surtout après l'invention de la photographie en 1822 et celle de la cinématographie en 1895. Cet accroissement du nombre d'images suit une fonction exponentielle, avec un développement électronique et informatique croissant, la disponibilité des moyens de stockage à petit prix et des technologies de compression d'images.

Les estimations prévoient que la quantité d'images produites dans les quelques années à venir sera plus importante que la quantité d'images produites depuis le début de l'humanité.

Pour utiliser efficacement cette quantité d'image de manière facile et automatique, un système d'indexation et de recherche d'images est nécessaire. C'est pourquoi le sujet de la recherche d'images devient un sujet très actif dans la communauté internationale depuis plus d'une vingtaine d'années.

La recherche d'images consiste à établir une correspondance entre l'image disponible et celle recherchée par l'utilisateur.

On peut distinguer deux approches de recherche d'image: approche basé sur le contenu de l'image et approche basé sur le texte.

Dans la première approche, la recherche d'image se fait par extraction des propriétés visuelles de l'image comme la texture, la forme, la couleur...etc.

Cependant, dans la deuxième approche l'image est considérée comme une boîte noire. La recherche exploite l'information contextuelle autour de l'image telle que les annotations.

D'un point de vue sémantique des images, la recherche par le texte permet de capter cette sémantique mieux que la recherche par contenu visuel. Toutefois, elle reste insuffisante pour l'interprétation et la recherche. Elle souffre d'un handicap majeur qui est l'oubli ou le silence. Ce problème vient de fait que la recherche par texte n'arrive pas à retourner toutes les images

pertinentes, tout simplement parce qu'ils ne sont pas annotés explicitement avec les mots clés de la requête.

Dans notre travail, nous allons nous intéresser au problème de silence (oubli).

2. Objectifs :

L'objectif de ce travail est d'essayer d'améliorer la qualité de la recherche par le texte. Ce dernier présente plusieurs problèmes, tels que le problème de silence et le problème de bruit. Le problème de silence vient du fait, que le moteur de recherche par texte considère une image pertinente seulement si le mot requête se présente dans son annotation (index). Par conséquent, ce mécanisme pose un véritable problème: le moteur peut passer à côté de plusieurs images qui répondent à la requête et les ignore.

Nous tentons justement dans ce travail à pallier cette lacune, notre objectif est de développer un mécanisme de calcul de similarité sémantique entre les termes de requête et les termes d'indexation (termes d'annotation) des images de la base. Ensuite, les images seront retournées par ordre décroissant de similarité.

3. Organisation du mémoire :

Notre mémoire est organisé en trois chapitres. Incluant cette introduction générale et se termine par une conclusion générale : Dans Le **chapitre 1**, nous dressons un état de l'art du domaine de la recherche d'images. Nous commençons dans une première section, par la présentation des moteurs de recherches d'images, leurs architectures ainsi que ses manières d'interaction avec l'utilisateur. Ensuite dans une deuxième section, nous allons présenter les deux catégories principales des moteurs de recherches d'images et leur avantage et leur limite. Nous allons nous intéresser par un des problèmes des moteurs de recherche par le texte. Dans le **chapitre 2**, nous présentons en détail la définition du problème, et les techniques qu'on propose pour le résoudre.

Dans le **chapitre 3**, nous allons présenter l'évaluation des techniques, Nous allons évaluer ces techniques par une collection d'images et des concepts de la nature. Ensuite, nous allons présenter une analyse, une discussion et une comparaison de ces techniques. Dans la conclusion générale, nous présenterons les principaux points de ce travail et nos perspectives.

Chapitre 01 :

La recherche d'images

1 Introduction:

L'essor des dispositifs d'acquisitions d'images, des capacités de stockage, la baisse des coûts des matériels informatiques et la disponibilité des techniques de numérisation de haute qualité que nous observons ces dernières années, se traduit par une production permanente et considérable d'images numériques dans différents domaines, ce qui a conduit à un développement constant des bases de données d'images .

Pour l'utilisateur de ce type de bases de données, la recherche d'images est très problématique, elle nécessite une nouvelle technique de traitement des données. Dans ce contexte, aujourd'hui le domaine de recherche d'images est en plein essor. L'objectif principal de ce domaine de recherche est de développer des outils capables d'aider l'utilisateur à localiser les images recherchées en un temps raisonnable. Ces outils sont généralement appelés : moteurs de recherche d'images, ou encore systèmes de recherche d'images.

Les moteurs de recherche d'image sont des outils qui aident l'utilisateur à localiser rapidement des images, mais avant d'aborder la définition des moteurs de recherche d'image, tout d'abord c'est quoi une image ?

2 Définition d'une image :

Une image est une représentation visuelle de quelque chose, cela pourrait être une photographie, un graphique, un dessin...etc.

Dans le domaine de l'informatique les images sont appelées images numériques. La numérisation d'une image est la conversion de celle-ci de son état analogique en une image numérique représentée par une matrice bidimensionnelle de valeurs numériques $f(x, y)$ où :

- x, y : coordonnées cartésiennes d'un point de l'image.
- $f(x, y)$: niveau de gris en ce point.

3 Moteur de recherche d'images

Un moteur de recherche d'images est un outil ou un système permettant de localiser des images dans une base de données qui répondent à une requête ces moteurs de recherche sont caractérisés par une architecture.

L'architecture générale d'un moteur de recherche d'images comporte deux phases phase offline et phase online :

Phase offline :

Dans cette phase Les caractéristiques des images de la base sont extraites automatiquement et stocké aux descripteurs. Selon la méthode utilisée pour la recherche.

Phase online :

Dans cette étape, le système extrait le descripteur de la requêtes de l'utilisateur, ensuite il le compare avec le descripteur de la base d'images enfin, le système affiche les images résultats en fonction de valeurs de similarité ou de pertinence.

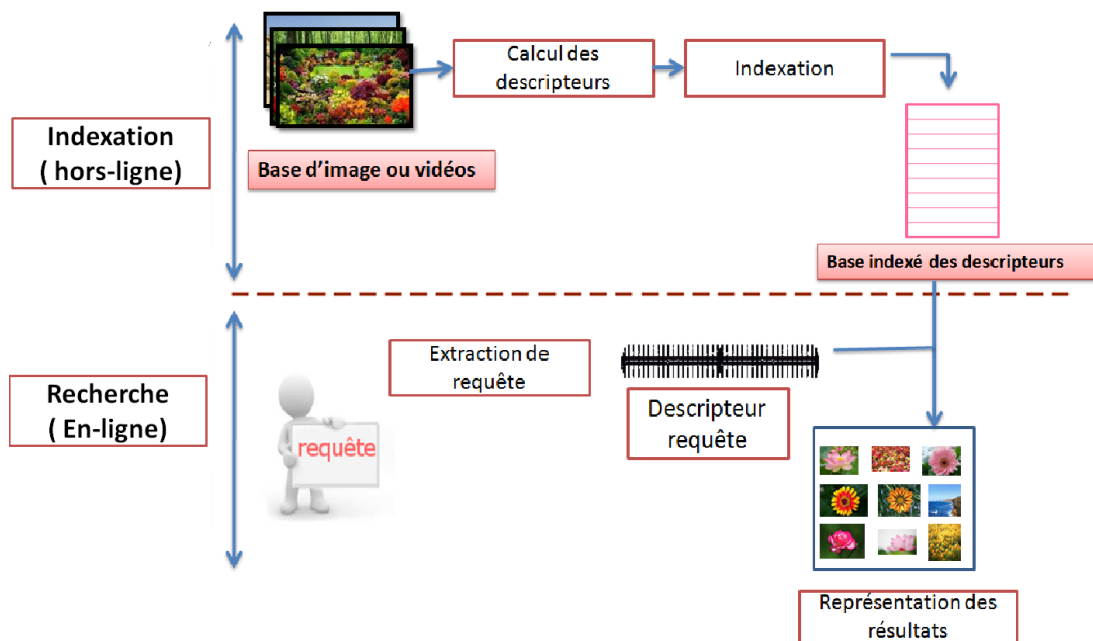


Figure 1 : Architecture générale de moteur de recherche d'images

4 Types de requêtes :

Le moteur donne la main à l'utilisateur afin qu'il formule sa requête. Il peut s'agir d'une zone de texte où l'usager entre les mots-clés décrivant ce qu'il cherche c'est ce qu'on a appel les requêtes textuelles, comme Il peut également s'agir d'un ensemble d'images dont l'usager peut choisir quelques unes comme exemples c'est ce qu'on a appel les requêtes visuelles. D'autres paradigmes de formulation de la requête sont également possible, mais Les types de

requêtes les plus utilisés sont les requêtes textuelles et les requêtes visuelles (image clé, image par sketch).

4.1 Requêtes textuelles

Les requêtes exprimées par une description textuelle sont généralement utilisées par des systèmes utilisant des mots-clés pour décrire l'image. L'utilisateur exprime ses besoins en fournissant des mots-clés, combinés à l'aide de connecteurs logiques tels que ET, OU et NON. Le texte peut être aussi des concepts d'une ontologie (LAOUBI, 2014).

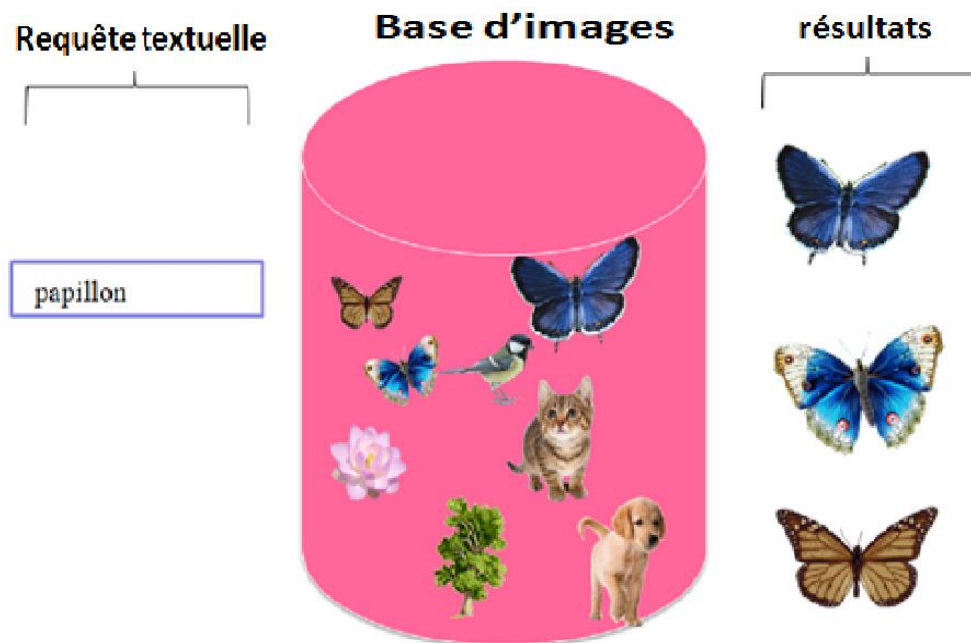


Figure 2 : les requêtes textuelles

4.2 Requêtes visuelles :

Les requêtes visuelles sont généralement exprimées par deux façons soit par image exemple ou par sketch (esquisse).

4.2.1 Requête par image exemple :

Les requêtes par exemple (ou bien images-clés), peuvent être exprimées selon trois scénarios :

- le système choisit quelques images au hasard dans la base d'images et les montre à l'utilisateur afin de choisir une requête image, ou bien
- l'utilisateur parcourt la base d'images et choisit une image-clé, ou encore
- l'usager donne son image requête. (TORJMEN, 2009)

4.2.2 Requête par sketch (esquisse) :

Dans ce type de requête le système donne à l'utilisateur la main de dessiner ou d'exprimé librement des images requêtes à l'aide des outils.



Figure 3 : les requêtes visuelles

5 Types de moteurs de recherche :

Selon le type de requête Les moteurs de recherche d'images peuvent être classifiés en deux grandes catégories : des moteurs basés sur le texte et des moteurs basés sur le contenu visuel.

5.1 Moteur de Recherche basé sur le contenu visuelle (CBIR):

Le système de recherche d'images orientée-contenu est basée sur un ensemble de caractéristiques de bas niveau (symbolique) d'une image telle que, les textures, la distribution

des couleurs, les formes, la brillance, et tout autre caractéristique de l'image que l'on peut imaginer. Les caractéristiques les plus utilisées sont les textures, couleurs et les formes (BEDOUHEN).

- **La couleur:**

La couleur est l'un des plus reconnaissables éléments du contenu visuel d'une image, c'est le plus utilisé dans la recherche image, il existe plusieurs distributeurs de couleur tel que : histogramme (BEN CHIKH & BEN ELZZINE, 2011).

- **La texture :**

Une définition formelle de la texture est quasiment impossible. Mais d'une manière générale la texture se traduit par un arrangement spatial des pixels que l'intensité; Ou les couleurs seules ne suffisent pas à décrire (BEN CHIKH & BEN ELZZINE, 2011).

- **La forme :**

La forme est utilisée pour caractériser les objets dans les images. On distingue deux catégories de descripteurs de formes : les descripteurs basés régions et les descripteurs basés frontières. Les premiers sont utilisés pour caractériser l'intégralité de la forme d'une région, Les seconds portent sur la caractérisation des contours de la forme (BEN CHIKH & BEN ELZZINE, 2011).

5.1.1 Indexation et recherche d'images par le contenu visuel:

La recherche par le contenu visuel consiste à caractériser le contenu visuel des images par des descripteurs visuels et d'effectuer des recherches par similarité visuelle à partir de ces descripteurs (BEDOUHEN).

Comme tout système de recherche d'information, les systèmes CBIR se basent sur une phase d'indexation et ensuite une phase de recherche.

5.1.1.1 La phase d'indexation (hors-ligne) :

Dans cette phase, le système extraites automatiquement les caractéristiques de l'image de la base et les stockées dans un vecteur numérique appelé descripteur visuel. Grâce aux

techniques de la base de données, on peut stocker ces caractéristiques et les récupérer rapidement et efficacement.

5.1.1.2 La phase de recherche (On-line) :

Dans cette phase, le système traite une ou plusieurs requêtes émises par l'utilisateur et lui donnent le résultat dans une liste d'images ordonnées, en fonction de la similarité entre leur descripteur visuel et celui de l'image requête en utilisant une mesure de distance (telle que distance de Minkowski, Euclidienne) (ABED & ZAOUI, 2011).

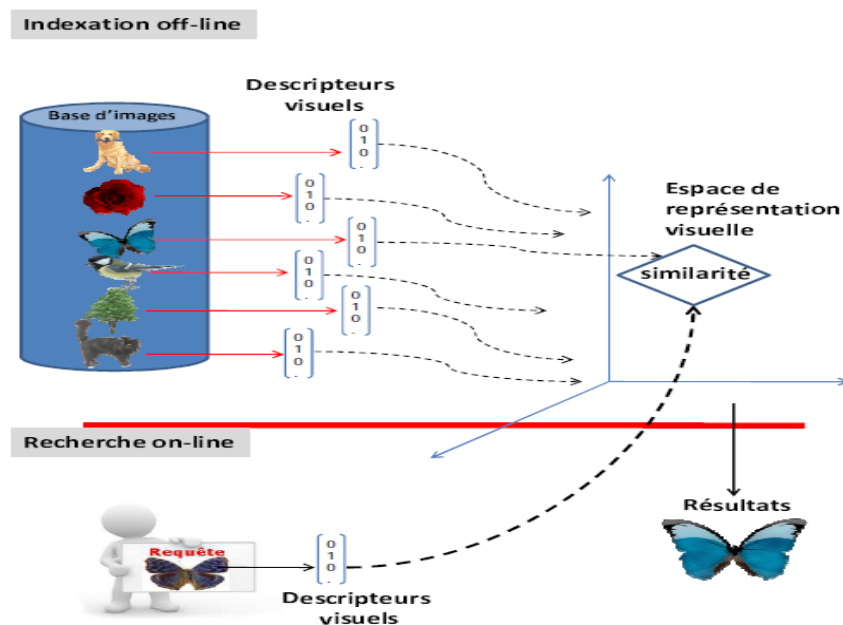


Figure 4 : Architecture de moteur de recherche d'image par contenu visuel.

La recherche par contenu visuel comporte un certain nombre d'avantages et pose certaines limites

5.1.2 Les Avantages :

L'utilisation de contenu visuel présente plusieurs avantages tel que:

-Le fait qu'elle peut être utilisée même si la BD ne comporte aucun texte.

- Elle s'applique bien aux images très complexes et celles qui contiennent une multitude d'objets qui ne peuvent être décrites avec du texte. (TORJMEN, 2009)

-le contenu des images est plus objectif que le texte (LAOUBI, 2014).

5.1.3 Les limites:

L'utilisation de contenu visuel pose plusieurs inconvénients nous pouvant citer:

- l'extraction des caractéristiques visuelles est coûteuse.
- le choix des meilleures caractéristiques visuelles à utiliser dans la phase de recherche.
- l'utilisateur doit faire plus d'effort si la requête doit être sous forme sketch.
- problème page de zéro : Si L'utilisateur ne peut pas trouver la bonne image qui décrit son besoin.
- la fossé sémantique :

La difficulté majeure des systèmes CBIR porte sur le plan sémantique ou leur efficacité est mise à mal par le fait qu'ils ne considèrent que des propriétés visuelles : par exemple la recherche d'images de "mer" et de "ciel" seront ramenées à des recherches de surfaces à texture bleue alors qu'il s'agit de concepts sémantiquement très différents. Ceci est connu sous le nom de fossé sémantique, entre les données numériques extraites des images et le contenu sémantique de ces mêmes images. (TORJMEN, 2009)

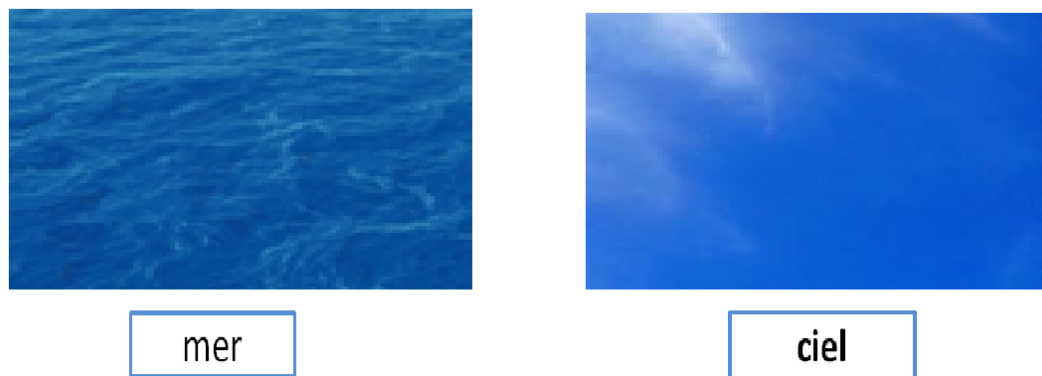


Figure 5 : fossé sémantique

5.2 Moteur de Recherche d'images basé sur le texte :

Le contexte d'une image concerne l'ensemble des informations autour d'une image permettant d'orienter sa signification ou sa description. La recherche basé sur le texte consiste à chercher dans l'environnement global de l'image sans aborder le contenu, l'image est considérée comme une boîte noire (LESLOUS & BOUCETTA, 2011).

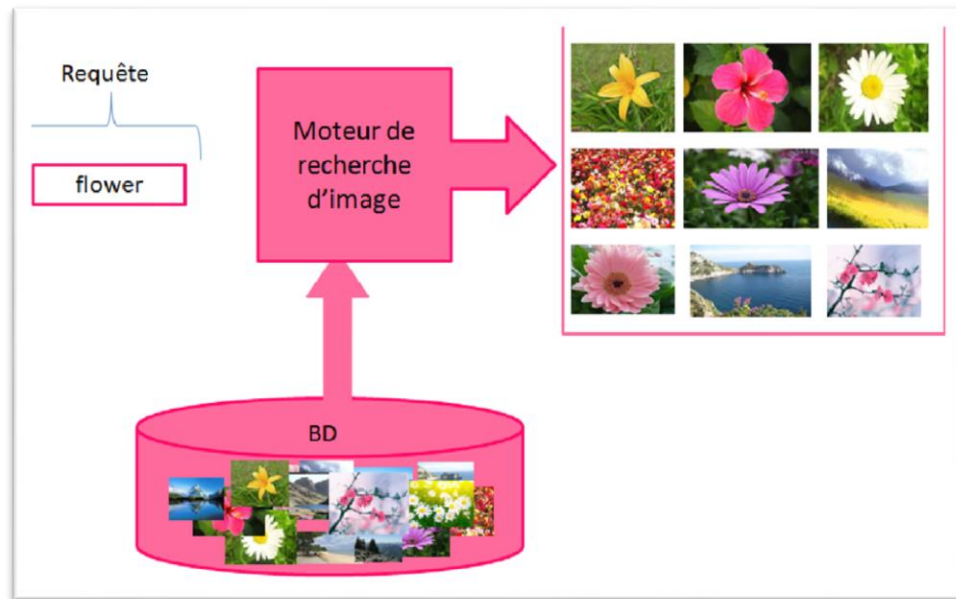


Figure 6 : moteur de recherche d'images par texte

Dans un système de recherche par le texte, la phase offline consiste à exploiter le texte de l'image pour construire son annotation. Par conséquent, la recherche exploite ces annotations pour localiser des images pertinentes à la requête utilisateur.

5.2.1 Annotation des images:

Une annotation est une note, une explication, ou tout sort d'information attachée à une source de connaissance (un document, phrase, un mot, image) sans toutefois être obligatoirement insérée dans cette dernière. Elle peut être réalisée en format papier qu'en format électronique (BEN CHIKH & BEN ELZZINE, 2011).

La recherche d'images par texte se fait soit par des mots clés ou par des concepts sémantiques

5.2.2 Recherche d'images par mot-clé :

5.2.2.1 Annotation par mot-clé :

Chaque image est annotée par une liste de mots-clés qui lui sont associés.

5.2.2.2 La Recherche:

Dans cette phase Le moteur présente à l'utilisateur les images qui sont indexés par ses mots-clés de requête. Une image est considérée pertinentes si le mot requête se présente dans son annotation. Autrement dit, il s'agit de faire une comparaison binaire 'matching' entre la requête et l'annotation.

5.2.2.3 Les Avantages :

L'utilisation de recherche par mot clé comporte plusieurs avantages nous pouvant citer:

-C'est une façon naturelle qui permet à l'utilisateur de s'exprimer comme il le fait dans la vie de tous les jours.

- elle peut de mieux capter les sémantiques des images. Imaginons par exemple un utilisateur qui est à la recherche d'images décrivant le mot «Joie», les techniques actuelles de recherche par le contenu ont énormément de difficulté à extraire un tel concept des images de façon automatique. Si l'on utilise le texte par contre, répondre à cette requête devient tout à fait possible en autant que certaines images soient annotées avec ce mot (KHERFI, 2008).

5.2.2.4 Les Limites :

Plusieurs limites se présentent pour ce type de recherche telles que:

- L'annotation des images représente une tâche longue et répétitive pour l'humain, surtout avec les bases d'images qui deviennent aujourd'hui de plus en plus grandes (BEN CHIKH & BEN ELZZINE, 2011).

- Les risques de bruits : fait référence aux documents non pertinents retrouvés par le système (BEN CHIKH & BEN ELZZINE, 2011).

- Les risques de silences: fait référence aux images pertinentes mais qui ne sont pas retrouvés par le système.

-les risques de fossé sémantique ou fausses combinaisons :

Polysémie, un mot clé peut désigner plusieurs objets différents sémantiquement et visuellement (ex : le terme « avocat » peut signifier la personne exerçant le métier d'avocat ou peut signifier le fruit « avocat »).



Figure 7 : ambiguïté de recherche par le mot clé « avocat ».

5.2.3 Recherche d'images par concept sémantique :

5.2.3.1 Annotation par des concepts:

Cette approche consiste à annoter les images par des concepts appartenant aux structures sémantiques par exemple: des ontologies ou par d'autre formalisme de représentation des connaissances (les réseaux de neurones, Arbres de décision, etc...).

5.2.3.2 La recherche d'image par concept sémantique:

L'utilisateur exprime ses besoins en fournissant un ou plusieurs concepts sémantiques, la recherche est effectuée dans ce cas, non seulement avec les concepts initiaux de la requête comme il est effectué dans la recherche textuelle classique, mais aussi avec les concepts qui ont une relation avec eux en exploitant la richesse sémantique qu'offrent les ontologies.

5.2.3.3 Les Avantages :

La recherche par concepts sémantiques présente un certain nombre d'avantages :

- ne pas se limiter a une comparaison binaire entre mots clés mais d'essayer de tirer profit de la sémantique associée aux images.
- C'est une façon naturelle qui permet à l'usager de s'exprimer comme il le fait dans la vie de tous les jours.
- Elle permet de réutiliser tout les techniques de recherche de texte, qui ont été développées au fil des années.
- elle permet de capter plus facilement les concepts sémantiques associés avec des images (LAOUBI, 2014).

5.2.3.4 Les limites :

La recherche par concepts sémantiques pose un certain nombre de limites tel que:

- cette technique devient inutilisable quand la collection ne contient aucun concept qui accompagne les images.
- cette technique est inutilisable même quand les images sont annotées avec du concept, cette annotation peut être très subjective. La même image peut être annotée avec des mots différents par des annotateurs différents (LAOUBI, 2014).

6 Conclusion :

Durant la décennie d'année le domaine de La recherche d'images est devient un sujet très important, La recherche d'image C'est une recherche utilisée pour localiser des images dans de grande base d'images en utilisant des systèmes. Les systèmes de recherches d'images sont classés en deux grandes familles : des systèmes de recherche basée sur le contenu visuel et des systèmes de recherche basée sur le texte. Chacune de ces familles présente des avantages et des limites.

Dans notre travail, nous nous intéressons aux limites de la deuxième famille en ce qui concerne spécialement le problème de silence ou d'oubli. Dans le chapitre suivant nous allons détailler ce problème, et décrire les différentes techniques qu'on a proposées pour le résoudre.

Chapitre 02 :

Problématique et Solution

1. Introduction :

Le principe d'un système de recherche d'image par le texte consiste à faire une comparaison binaire 'matching' entre la requête utilisateur et les annotations des images. Ainsi, une image est considérée pertinente si le mot requête se présente dans son annotation. Par conséquent, ce mécanisme pose un problème de silence ou d'oubli, en effet le moteur peut passer à côté de plusieurs images qui répondent à la requête et les ignore.

Prenons un exemple : Supposons que l'utilisateur formule sa requête avec le concept «flower », et supposons que notre collection est composée des images de la Fig. 8, où chaque image est annotée avec les concepts qui sont juste au dessous.



Figure 8 : Exemple d'une collection d'images.

Quand on fait une recherche binaire on obtient l'image 3 qui est annotée explicitement avec le concept « flower ». Les images 1 et 2 peuvent nous servir mais puisque le concept de requête ne se présente pas dans ses annotations donc ces deux images seront omises dans les résultats, c'est le problème de silence.

2. Les techniques utilisées pour résoudre le problème de silence:

Dans notre travail on va essayer de résoudre le problème de silence par trois techniques.

2.1 Technique de relation entre l'image et les concepts(RIC):

Aujourd'hui, on trouve presque toutes les pages web utilisant des images qui expriment leur contenu, ce dernier (code html) peut contenir des concepts décrivant ces images, comme il peut contenir des concepts qui ont une relation avec ces images.

A partir de cette idée on s'est basée.

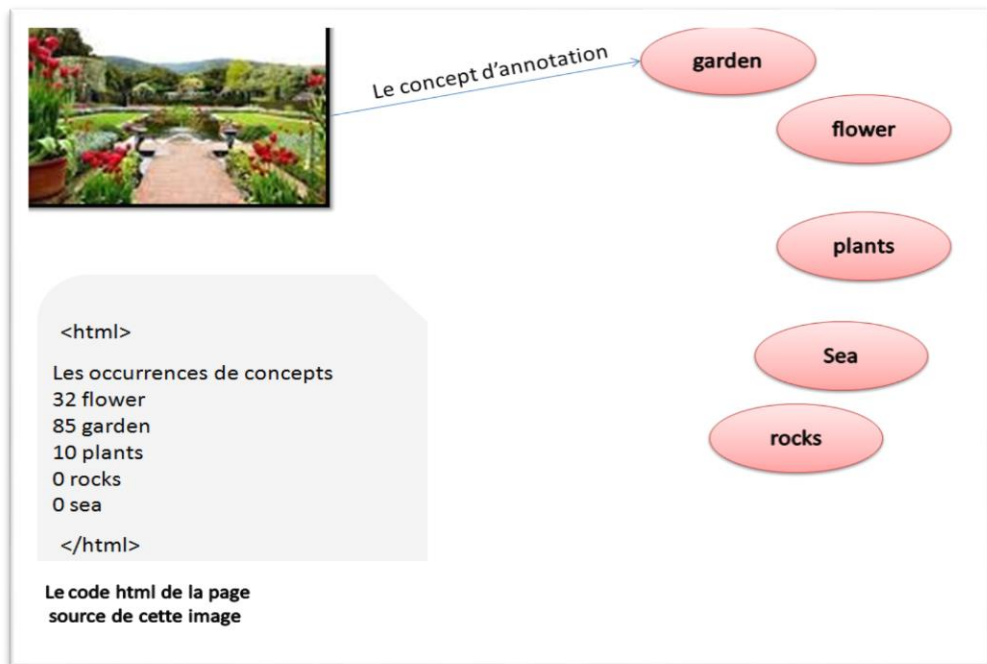


Figure 9 : l'image et son code html (contenu).

Notre hypothèse dans cette technique s'est, si un concept où plus apparaît dans le même contenu de la page (le code html de la page) d'une image, ceci est traduit par une relation entre ces concepts et cette image.

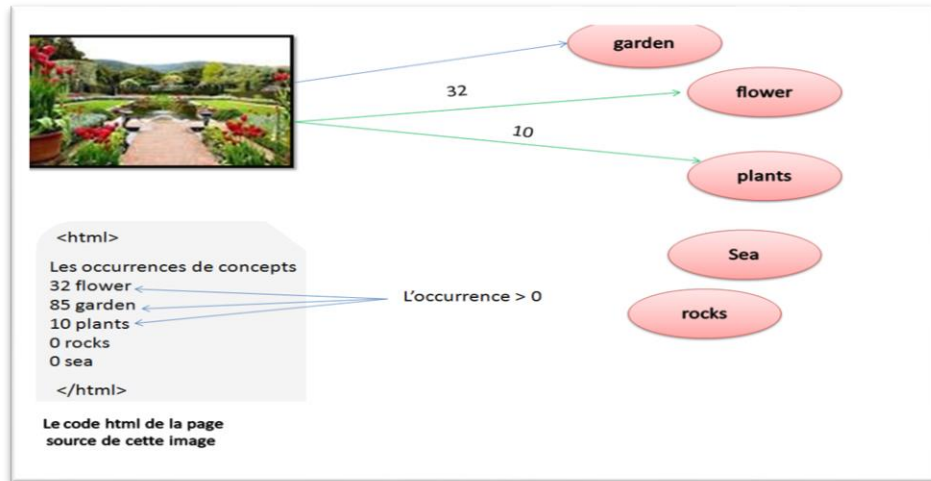


Figure 10 : la relation de l'image avec les concepts

Dans cette technique, à partir du code html de l'image, on a créé des relations entre les images et les concepts, mais on remarque que le code html seul dans certains cas ne permet pas d'avoir toutes les relations possibles, en effet, on peut avoir des images qui exprime des concepts, mais ces derniers n'apparaissent pas dans ses codes html ce qui fait que dans cette technique on ne peut pas avoir ces images dans les résultats de recherche.

Ce problème est probablement est du à l'absence de relation entre les concepts, c'est pour cela, on a pensé à créer des relations entre les concepts eux mêmes.

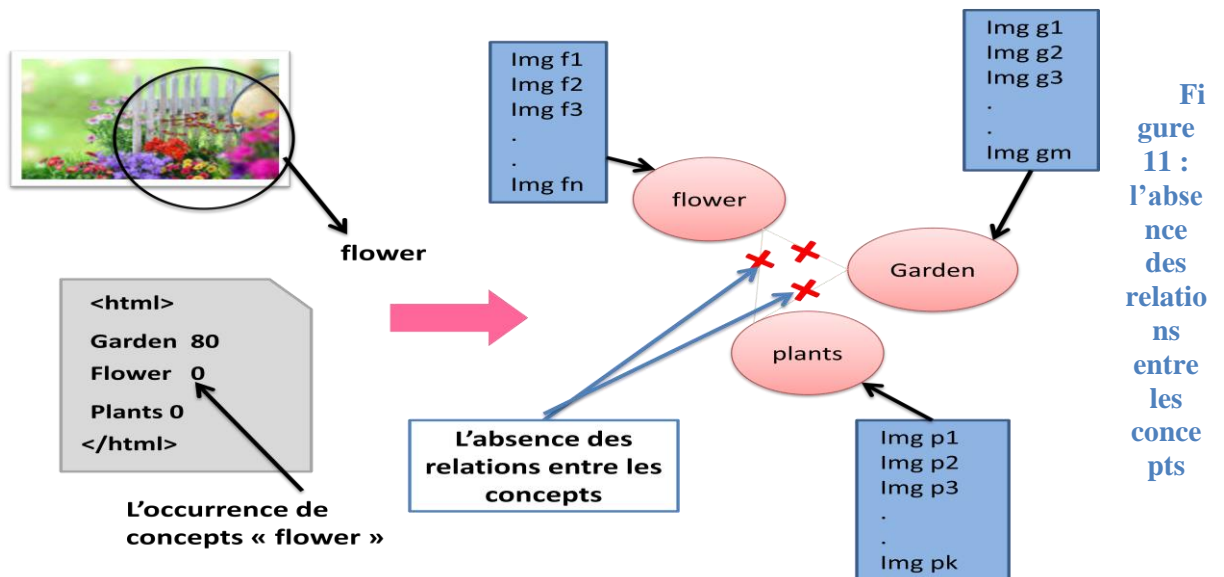


Figure 11 : l'absence des relations entre les concepts

2.2 Technique de relation entre concepts (RC):

Le départ de cette technique c'est aussi basé sur le code html, mais cette fois ci on s'est basée sur une autre hypothèse, si deux concepts apparaissent dans un code html d'une même image, c'est qu'il y'a une relation entre ces concepts.

De cette façon on a crée les relations entre les concepts,

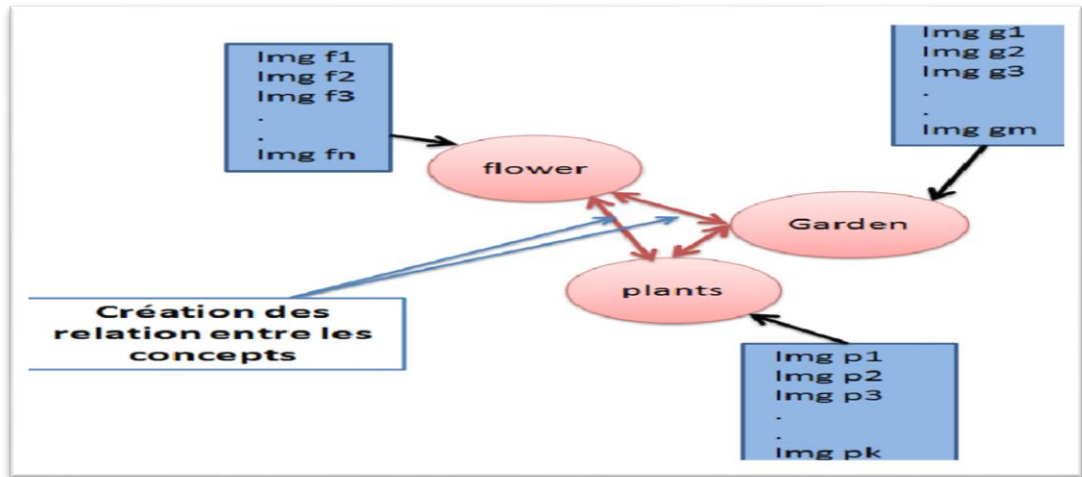


Figure 12 : création des relations entre les concepts

la technique RC a permet de crée des relations entre les concepts, ceci est traduit par une amélioration de résultats de recherche d'images et une résolution de problème présenter par la technique (RIC), en effet, l'image est apparu dans les résultats de recherche d'un concept même si ce dernier ne se présente pas dans son code de html il suffit juste d'avoir une relation entre ce concept et le concept qui annote cette image.

Mais, malgré l'amélioration qu'elle apporte la technique RC cette dernière reste insuffisante car elle pose des problèmes, d'un part les relations créer entre les concepts, certaines restent faibles ou nulles, d'un autre part l'image n'est pas mis en valeur, en effet les résultats des images sont affichées en fonction d'ordre de relations entre les concepts, on peut avoir une image du concept d'ordre supérieure moins significative que l'image du concept d'ordre inférieur.

Les problèmes présenté par les deux techniques précédente nous on poussé a pensé a La technique de spreading activation qui permet de créer des relations entre les concepts ou les

augmenté avec une considération de l'image, cette technique se base sur les deux techniques illustrées précédemment de sorte qu'à partir de la relation de l'image et les concepts réalisés par la technique de RIC et les relations entre les concepts réalisés par la technique de RC on aura une nouvelle relation de cette image avec les concepts, mais avant tout cela tout d'abord c'est quoi spreading activation ? elle s'inspire de quoi ? qu'elle est son principe, et son objectif ? tout ça et plus nous allons le détailler.

6.1 Définition Spreading activation(SA):

Spreading activation est un terme utilisé en psychologie, psycholinguistique et science cognitive¹.

Elle se réfère à un modèle décrivant le mécanisme de recherche dans les réseaux associatifs, réseaux de neurones² et les réseaux sémantiques³, elle repose sur le mécanisme de recherche d'une information stockée en mémoire humaine ou la mémoire est envisagée comme un réseau des idées ou de concepts, les concepts considérés comme des nœuds et les liens considérés comme des relations entre ces concepts.

La recherche dans la mémoire humaine est faite à partir de l'activation initiale par un ou plusieurs concepts (quand on pense à une idée automatiquement on pense aux autres idées les plus proches à l'idée initiale).

Prenons un exemple que si le concept pastèque est présenté dans un texte, les nœuds de concept rouge, vert, fruit et l'été sont probablement activés dépendent du contexte de ce concept dans le texte.

¹ https://en.wikipedia.org/wiki/Cognitive_psychology consulté le 4 mai 2015 09:12

² https://en.wikipedia.org/wiki/Artificial_neural_network consulté le 4 mai 2015 09:15

³ https://en.wikipedia.org/wiki/Semantic_network consulté le 4 mai 2015 09:20

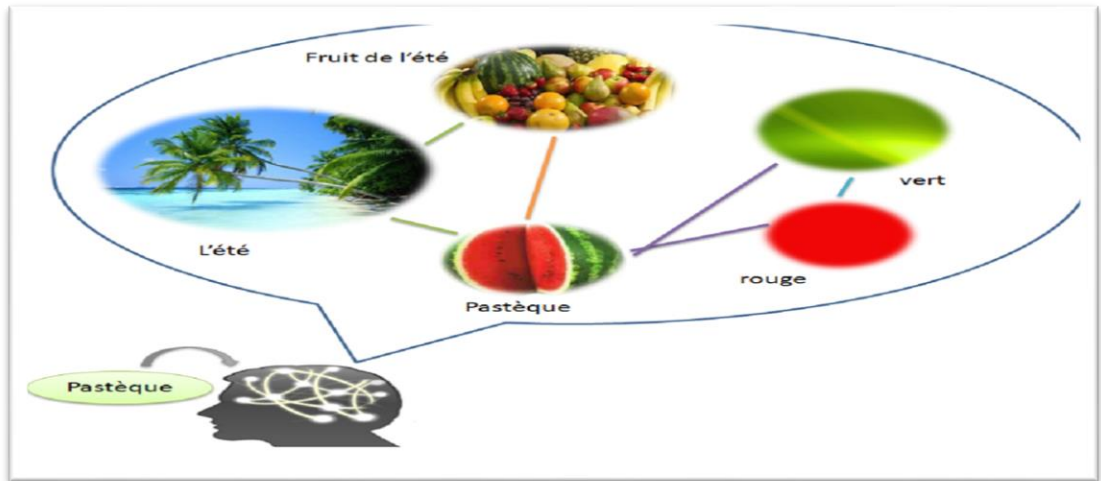


Figure 13 : le mécanisme de recherche d'une information dans la mémoire humaine.

2.2.1 Principe SA :

Le modèle (SA) dans sa forme pure est assez facile. Il se compose de nœuds reliés par des liaisons (réseau), les nœuds sont des objets ou des caractéristiques des objets du monde réel.

Les nœuds sont habituellement marqués avec le nom des objets et les Liens représentent les relations entre les nœuds. Dans le réseau un lien a généralement une direction, une étiquette, ou un poids attribué selon à une direction spécifique.

Le principe de SA est d'activer initialement un ou nœud ou plusieurs avec une valeur initiale, l'activation se propage vers les autres nœuds qui ont en relation avec les nœuds activés. La valeur initiale se diminue par un certain pourcentage (decay factor) dans l'opération de propagation dans le réseau.

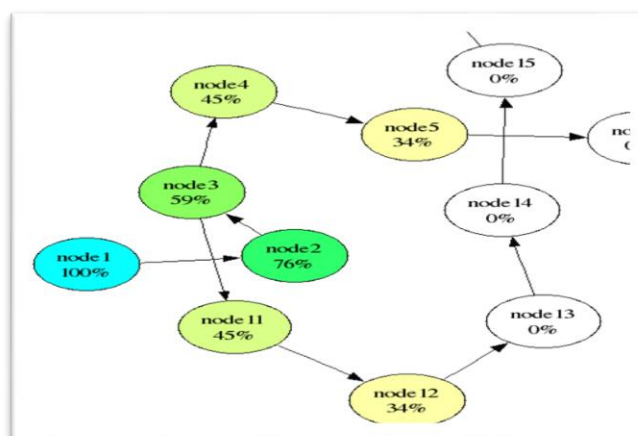


Figure 14 : le principe de Spreading Activation

2.2.2 L'application de SA a notre problème :

On va appliquer SA de sorte qu'elle permet la création ou l'augmentation relation entre l'image et des concepts se qui donne un réseau sémantique ou l'image peut avoir des relations avec plusieurs concepts.

Donc quand on fait une recherche à un concept les images qui sont en relation avec ce concept seront activées.

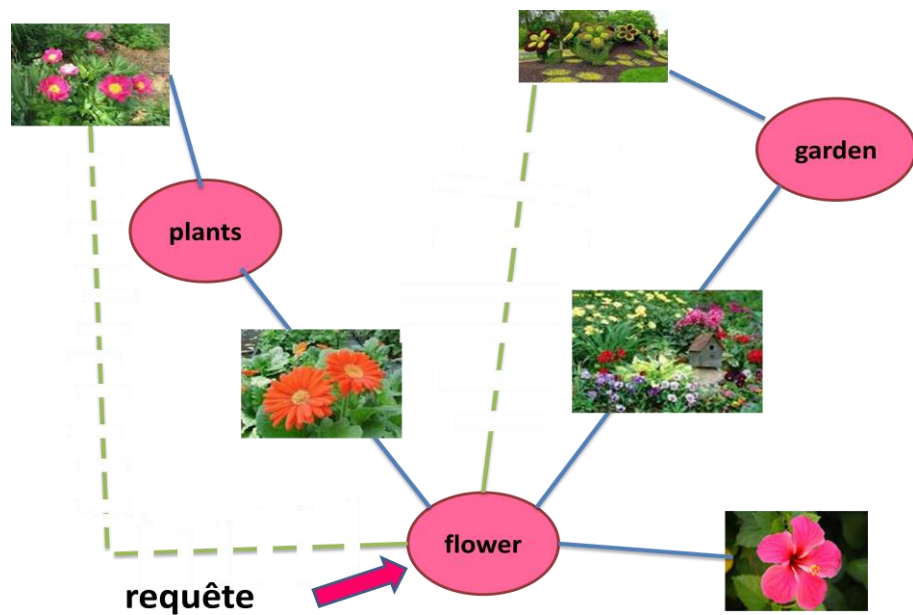


Figure 15 : l'application de spreading activation

3. Conclusion :

Dans ce chapitre nous avons présenté les différentes techniques proposées pour résoudre le problème de silence les trois techniques se base sur l'utilisation du code html de l'image dans la première technique (RIC) on a pu avoir des relations entre l'image et les concepts, avec la deuxième technique on a pu avoir des relations entre les concepts. Et avec SA on a aussi pu avoir des relations entre l'image et les concepts mais cette fois ci des relations plus fortes, cette technique utilise les deux techniques précédentes.

Dans le chapitre suivant nous présenterons l'évaluation de chaque technique.

Chapitre 03 : L'évaluation

1. L'évaluation de techniques RIC, RC et SA

Dans notre travail, on a essayé de résoudre le problème de silence avec une collection de 11 concepts de la nature, on a obtenu pour chaque concept 64 images télécharger depuis l'internet via l'outil de GSearchClient qui permet de télécharger des images et ramènent ces pages sources.

2. Technique de relation entre l'image et les concepts (RIC):

Pour chaque image de notre collection on à obtenus le code html, depuis ce dernier, on calcule la moyenne d'existence de chaque concept dans cette image.

De cette façon et après avoir ce calcule pour tous les codes html (images), on remplit la matrice *I* de 704 lignes (code html de la page source d'image) et de 11 colonnes (concepts) cette matrice signifie la relation entre l'image *i* et le concept *j*.

	Cp1	Cp 2	Cp 3	.	Cp k	.	Cp m	La somme
Img 1	207	0	2	705
Img 2	41	0	3	840
Img 3	70	10	0	620
.
Img i	220	6	1	500
.
Img n

207 Le nombre d'occurrence de Cp1 au le code HTML de l'image Img 1

705 est la somme des occurrence de les concepts au code HTML de l'image Img 1

Figure 16 : matrice *I*

Et après la construction de la matrice *I* on affiche les images en fonction d'ordre décroissant d'appartenance d'une image à un concept comme la figure ci dessous illustre.

	Cp1
Img 1	207/705 = 0.29
Img 2	41/840 = 0.04
Img 3	70/600 = 0.11
.	.
.	.
Img i	220 /500=0.4
.	.
Img n	.

Le tri des images selon les moyennes de Cp1 de façon décroissante

	Cp1
Img i	0.4
Img 1	0.29
Img 3	0.11
.	.
.	.
Img 2	0.04
.	.
Img n	.

Figure 17 : la technique de RIC

Le tableau ci-dessous présente les résultats de recherche obtenus par la technique RIC et la recherche binaire

Concept	Nombre d'images par matching	Seuil	Nombre d'images par le RIC	Nombre des bonnes images par RIC
«coast»	64	0.006	95	76
		0.008	94	77
		0.01	93	77
«flower»	64	0.01	304	215
		0.03	231	164
		0.05	177	149
«forest»	64	0.01	172	157
		0.02	148	138
		0.03	119	114
«garden»	64	0.007	184	160
		0.01	179	166
		0.03	127	122
«grass»	64	0.005	133	119
		0.007	122	109
		0.01	110	109
«mountain»	64	0.01	184	127
		0.02	153	132
		0.03	141	124
«plants»	64	0.01	212	200
		0.03	168	163
		0.06	150	147
«river»	64	0.01	186	136
		0.04	123	98
		0.07	97	85
«rocks»	64	0.01	174	116
		0.04	120	84
		0.07	89	75
«sea»	64	0.01	180	127
		0.04	129	105
		0.07	104	95
«tree»	64	0.01	246	191
		0.04	167	143
		0.07	125	105

Tableau 1 : les résultats de la technique RIC

Observation :

- On remarque que pour le concept «coast» avec le seuil 0.006 on obtient 13 images plus, avec 18 images bruit.

Avec le seuil 0.008 on obtient 13 images plus et 17 images de bruit, et avec le seuil 0.01 on obtient 14 images plus et 16 images de bruit.

- On remarque que pour le concept «flower» avec le seuil 0.01 on obtient 151 images plus, avec 89 images bruit.

Avec le seuil 0.03 on obtient 118 images plus et 49 images de bruit, et avec le seuil 0.05 on obtient 85 images plus et 28 images de bruit.

- On remarque que pour le concept «forest» avec le seuil 0.01 on obtient 93 images plus, avec 15 images bruit.

Avec le seuil 0.02 on obtient 74 images plus et 10 images de bruit, et avec le seuil 0.03 on obtient 50 images plus et 5 images de bruit.

- On remarque que pour le concept «garden» avec le seuil 0.007 on obtient 107 images plus, avec 13 images bruit.

Avec le seuil 0.01 on obtient 102 images plus et 13 images de bruit, et avec le seuil 0.03 on obtient 58 images plus et 5 images de bruit.

- On remarque que pour le concept «grass» avec le seuil 0.005 on obtient 55 images plus, avec 14 images bruit.

Avec le seuil 0.007 on obtient 45 images plus et 13 images de bruit, et avec le seuil 0.01 on obtient 35 images plus et 10 images de bruit.

- On remarque que pour le concept «mountain» avec le seuil 0.01 on obtient 63 images plus, avec 57 images bruit.

Avec le seuil 0.02 on obtient 68 images plus et 21 images de bruit, et avec le seuil 0.03 on obtient 60 images plus et 17 images de bruit.

- On remarque que pour le concept «plants» avec le seuil 0.01 on obtient 136 images plus, avec 12 images bruit.

Avec le seuil 0.03 on obtient 99 images plus et 5 images de bruit, et avec le seuil 0.05 on obtient 83 images plus et 3 images de bruit.

- On remarque que pour le concept «river» avec le seuil 0.01 on obtient 72 images plus, avec 50 images bruit.

Avec le seuil 0.04 on obtient 34 images plus et 25 images de bruit, et avec le seuil 0.07 on obtient 21 images plus et 12 images de bruit.

- On remarque que pour le concept «rocks» avec le seuil 0.01 on obtient 52 images plus, avec 58 images bruit.

Avec le seuil 0.04 on obtient 20 images plus et 36 images de bruit, et avec le seuil 0.07 on obtient 11 images plus et 14 images de bruit.

- On remarque que pour le concept «sea» avec le seuil 0.01 on obtient 63 images plus, avec 53 images bruit.

Avec le seuil 0.04 on obtient 41 images plus et 24 images de bruit, et avec le seuil 0.07 on obtient 31 images plus et 9 images de bruit.

- On remarque que pour le concept «tree» avec le seuil 0.01 on obtient 127 images plus, avec 55 images bruit.

Avec le seuil 0.04 on obtient 79 images plus et 24 images de bruit, et avec le seuil 0.07 on obtient 41 images plus et 20 images de bruit.

Analyse :

On constate que cette technique a amélioré la recherche par rapport à la recherche binaire avec tous les concepts on obtient plus d'images bonnes par rapport la technique de matching qui apporte 64 images par concept.

Les résultats obtenu par cette technique est en fonction de l'image et ses relations avec les concepts et pas limité d'annotation (recherche binaire), ce qui explique l'augmentation du nombre images bonnes par rapport la recherche binaire.

(Les relations sont obtenues par le code html de l'image comme déjà expliqué dans le chapitre plus haut).

Donc plus l'image à des relations avec plusieurs concepts plus à la chance d'apparence dans les résultats de recherche obtenus par ces concepts.

Comme par exemple l'image 0 de concept coast à des relations avec les concepts sea, rocks, river et plants donc dans les résultats de recherche obtenus par ces concepts on peut avoir cette image.

Le technique de « Relation entre les Images et les Concepts » (RIC)											
	coast	flower	forest	Garden	Grass	Mountain	Plants	River	Rocks	Sea	tree
Image 0	0.583	0	0.006	0	0	0	0.027	0.049	0.030	0.30	0

Le poids de relation entre l'image 0 et le concept « sea »

Figure 18:la relation entre l'image0 et les concepts.

Donc quand on fait la recherche par le concept sea on obtient cette image coast0

Requête par concepts « sea »

Image annotée par « coast »

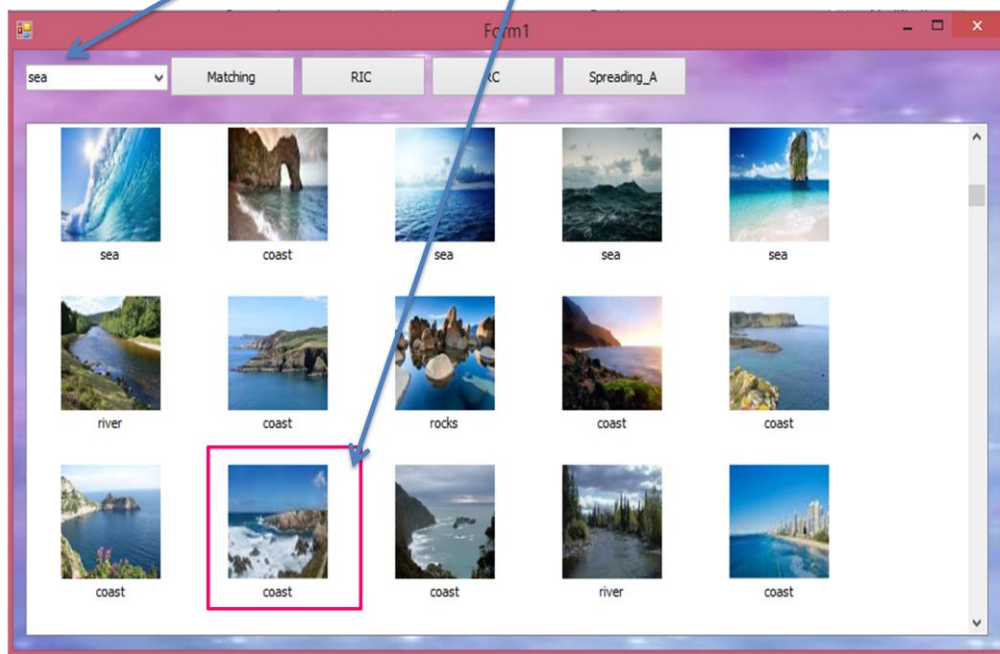


Figure 19 : la recherche par le concept sea

3. Technique de relation entre concept RC:

A partir de la matrice I , on construit la matrice $w_{[11*11]}$ qui représente les relations entre les concepts de manière que si la fréquence de concept $(cp_i) > 0$ et $(cp_j) > 0$ en la même ligne (images) dans la matrice I , autrement dit si le concept i et le concept j apparu dans la même images, on incrémente w_{ij} qui traduit la relation entre cp_i et cp_j , après avoir vérifier l'apparence de ces deux concepts avec toutes les images on met dans la case w_{ij} le nombre d'occurrence de cp_i avec cp_j divisé par le nombre d'occurrences de cp_i dans toutes les images. la figure ce dessous explique la technique RC

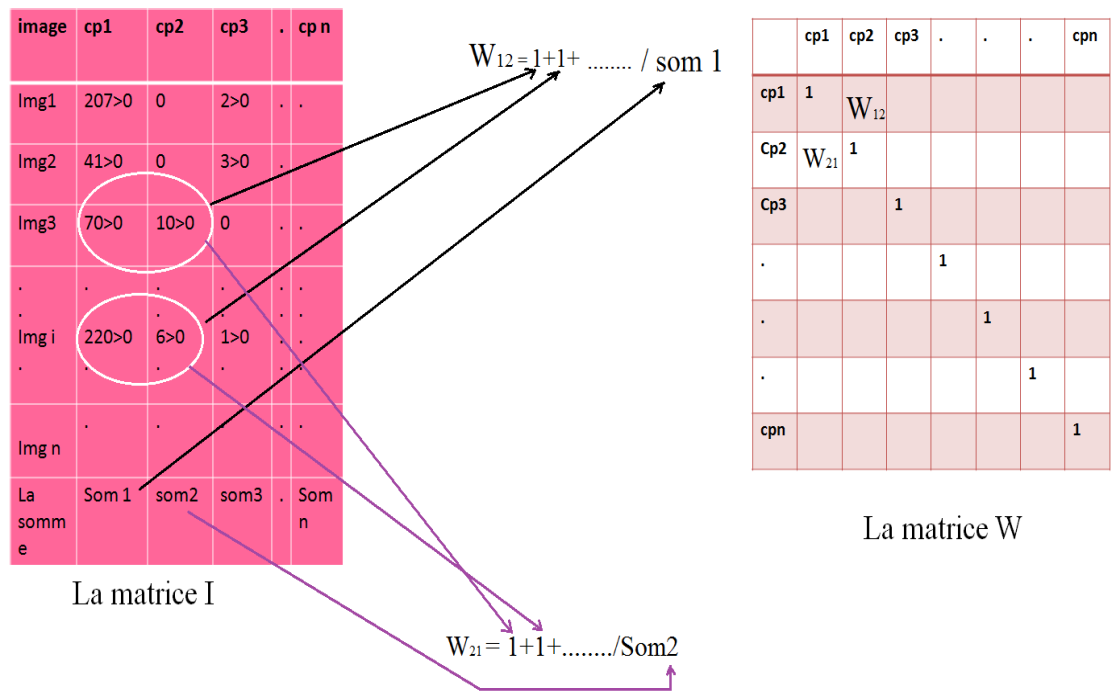


Figure 20:La Technique RC.

Et après la construction de la matrice $w_{[11*11]}$ les images est affiché-on fonction d'ordre décroissant des poids de relations entre un concept recherché et les autres concepts.

Ce tableau illustre les résultats obtenus, et la comparaison entre la recherche binaire RIC et RC

Concept	Images par matching	Seuil	images total par RIC	Les bonnes images par RIC	Images total par RC	les bonnes Images RC
Coast	64	0.006	95	77	646	208
		0.008	94	77	512	202
		0.01	93	76	448	198
Flower	64	0.01	304	215	256	156
		0.03	231	164	64	64
		0.05	177	149	64	64
Forest	64	0.01	172	157	384	343
		0.02	148	138	64	64
		0.03	119	114	64	64
Garden	64	0.007	184	160	192	190
		0.01	179	166	64	64
		0.03	127	122	64	64
Grass	64	0.005	133	119	384	200
		0.007	122	109	64	64
		0.01	110	109	64	64
Mountain	64	0.01	184	127	512	115
		0.02	153	132	64	64
		0.03	141	124	64	64
Plants	64	0.01	212	200	256	254
		0.03	168	163	64	64
		0.06	150	147	64	64
River	64	0.01	186	136	320	129
		0.04	123	98	64	64
		0.07	97	85	64	64
Rocks	64	0.01	174	116	448	144
		0.04	120	84	64	64
		0.07	89	75	64	64
Sea	64	0.01	180	127	64	64
		0.04	129	105	64	64
		0.07	104	95	64	64
Tree	64	0.01	246	191	576	260
		0.04	167	143	64	64
		0.07	125	105	64	64

Tableau 2 : les résultats de la recherche binaire, RIC et RC

Observation :

-On remarque de l'analyse de ce tableau que cette technique apporte plus d'images bonnes par rapport à la recherche binaire quand les seuils sont petits, mais pour les grands seuils, cette technique apporte le même nombre d'images obtenus par la recherche binaire.

-On remarque que pour le concept «coast» avec le seuil 0.006 on obtient 144 images plus par rapport la recherche binaire et 132 par rapport la technique de relation de l'image et les concepts, avec 438 images bruit.

Avec le seuil 0.008 on obtient 138 images plus par rapport la recherche binaire et 125 par rapport la technique de RIC plus et 310 images de bruit, et avec le seuil 0.01 on obtient 134 images plus par rapport la recherche binaire et 121 par rapport RIC et 250 images de bruit.

- On remarque que pour le concept «flower» avec le seuil 0.01 on obtient 92 images plus, par rapport la recherche binaire et 95 images moins par rapport la technique de relation de l'image et les concepts avec 100 images bruit.

- On remarque que pour le concept «forest» avec le seuil 0.01 on obtient 279 images plus, par rapport la recherche binaire et 186 images plus par rapport la technique de relation de l'image et les concepts avec 41 images bruit.

- On remarque que pour le concept «garden» avec le seuil 0.007 on obtient 127 images plus, par rapport la recherche binaire et 30 images plus par rapport la technique de relation de l'image et les concept, avec 1 image bruit.

- On remarque que pour le concept «grass» avec le seuil 0.01 on obtient 136 images plus, par rapport la recherche binaire et 81 images plus par rapport la technique de relation de l'image et les concepts, avec 184 images bruit.

- On remarque que pour le concept «mountain» avec le seuil 0.01 on obtient 51 images plus, par rapport la recherche binaire et 12 images moins par rapport la technique de relation de l'image et les concepts, avec 373 images bruit.

- On remarque que pour le concept «plants» avec le seuil 0.01 on obtient 191 images plus, par rapport la recherche binaire et 199 images moins par rapport la technique de relation de l'image et les concepts, avec 1 image bruit.
- On remarque que pour le concept «river» avec le seuil 0.01 on obtient 65 images plus, par rapport la recherche binaire et 7 images moins par rapport la technique de relation de l'image et les concepts, avec 191 images bruit.
- On remarque que pour le concept «rocks» avec le seuil 0.01 on obtient 80 images plus, par rapport la recherche binaire et 28 images plus par rapport la technique de RIC, avec 304 images bruit.
- On remarque que pour le concept «tree» avec le seuil 0.01 on obtient 196 images plus, par rapport la recherche binaire et 69 images moins par rapport RIC avec 316 images bruit.
- On remarque que pour le concept «sea» avec tout les seuils aucune amélioration ni pour la recherche binaire ni pour la technique RIC.

Analyse :

A partir de ce tableau, on remarque que cette technique a aussi amélioré la recherche par rapport à la recherche binaire, avec tous les concepts on obtient plus d'images bonnes.

Par rapport la recherche par la technique RIC l'amélioration c'est seulement pour les petits seuils.

Les résultats obtenus par cette technique et en fonction des relations entre les concepts et ne pas limiter par l'annotation (la recherche binaire).

Plus le concept a des relations fortes avec plusieurs concepts, plus le nombre d'images bonnes obtenues par ce concept est grand. Ce qui explique l'amélioration des résultats.

Prenons un exemple le concept coast a des relation avec les concept river, sea, rocks, grass, mountain , tree, forest ,flower.

Le résultats de RC											
	coast	flower	forest	Garden	Grass	Mountain	Plants	River	Rocks	Sea	tree
coast	1.0	0.008	0.011	0.007	0.005	0.011	0.007	0.014	0.013	0.014	0.012

Le poids de relation entre le concept « coast » et le concept « river »

Figure 21 : création de relation par la technique de RC

Donc quand on fait la recherche par le concept « river » on obtient l'image qui annoté par « coast »

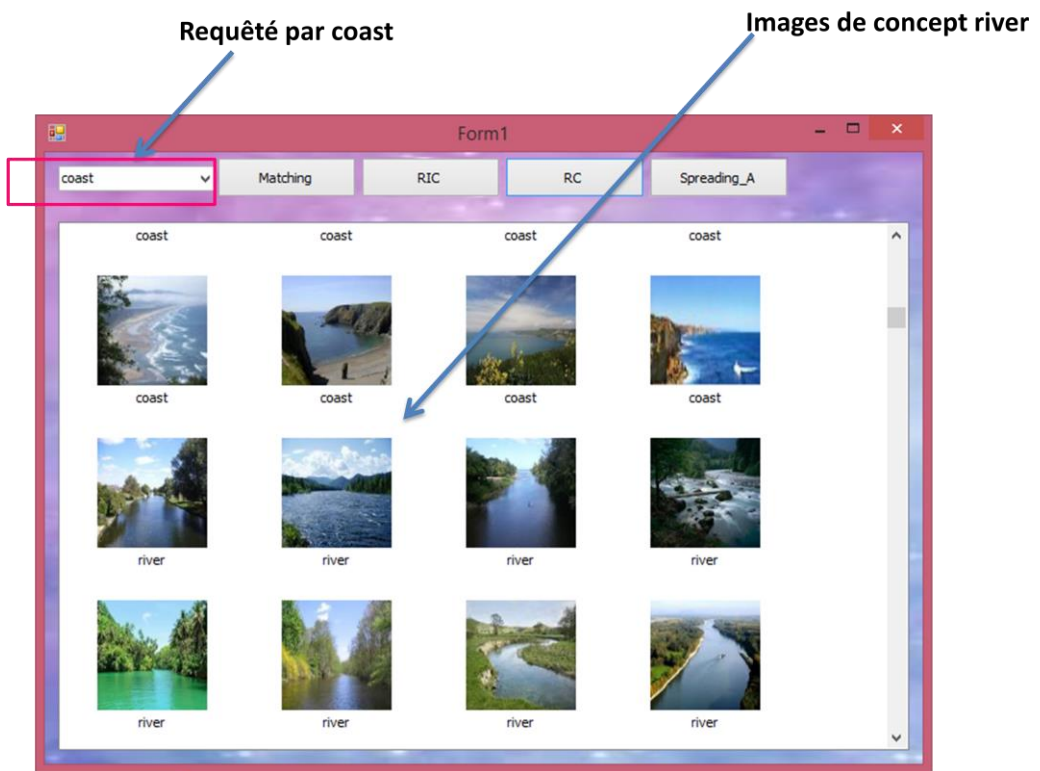


Figure 22:le résultat de recherche par concepts coast

4. Technique de SA :

La formule de SA : $\vartheta_i = [\varepsilon - w^t]^{-1} \cdot I_i$

$i = 0 \dots \dots \dots 704$

ε Est la matrice d'identité [11 * 11] ;

I_i C'est le vecteur de l'image i, qui représente la relation entre l'image i et les 11 concepts.

$w_{[11*11]}$ Matrice représente les relations entre les concepts, cette matrice est construite comme dans la technique RC.

ϑ_i C'est le nouveau vecteur de l'image i qui traduit l'augmentation des relations entre l'image i et les concepts à partir de relations entre les concepts (la matrice $w_{[11*11]}$) et la relation entre l'image i et les concepts(RIC).

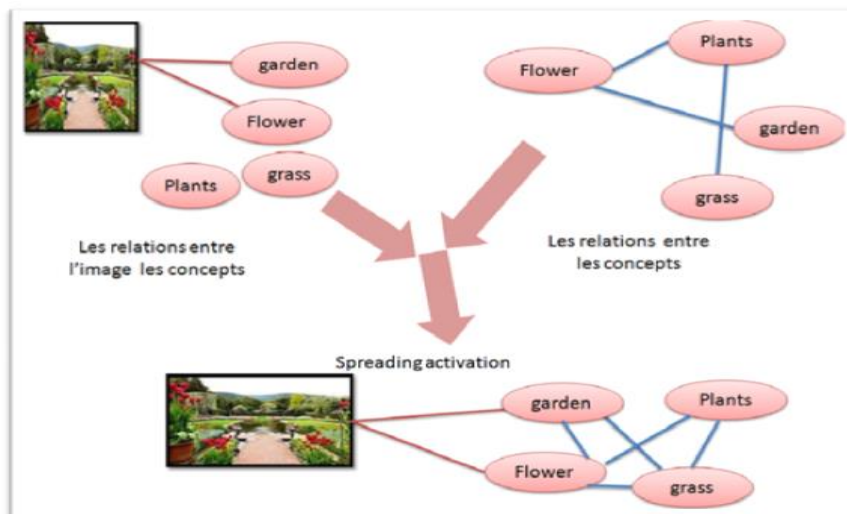


Figure 23 Technique de Spreading Activation

Et après la construction de la matrice ϑ de nouveaux vecteurs des images on affiche ces derniers de même manière (comme déjà illustré dans la technique RIC plus haut).

Ce tableau illustre les résultats obtenus, et la comparaison entre les techniques la recherche binaire, RIC, RC, SA.

Concept	images par matching	Seuil	Totale images par RIC	bonnes images par RIC	Total images par spreading	bonnes images par spreading	total par RC	bonnes Images par RC
Coast	64	0.006	95	76	139	104	646	208
		0.008	94	77	102	93	512	202
		0.01	93	77	97	85	448	198
Flower	64	0.01	304	215	512	214	256	156
		0.03	231	164	269	162	64	64
		0.05	177	149	200	114	64	64
Forest	64	0.01	172	157	435	307	384	343
		0.02	148	138	167	139	64	64
		0.03	119	114	153	130	64	64
Garden	64	0.007	184	160	445	255	192	190
		0.01	179	166	313	223	64	64
		0.03	127	122	150	121	64	64
Grass	64	0.005	133	119	413	322	384	200
		0.007	122	109	240	180	64	64
		0.01	110	109	178	142	64	64
Mountain	64	0.01	184	127	400	140	512	115
		0.02	153	132	178	124	64	64
		0.03	141	124	156	120	64	64
Plants	64	0.01	212	200	387	305	256	254
		0.03	168	163	188	180	64	64
		0.06	150	147	164	158	64	64
River	64	0.01	186	136	404	202	320	129
		0.04	123	98	139	106	64	64
		0.07	97	85	102	90	64	64
Rocks	64	0.01	174	116	375	143	448	144
		0.04	120	84	129	89	64	64
		0.07	89	75	101	78	64	64
Sea	64	0.01	180	127	249	176	64	64
		0.04	129	105	141	111	64	64
		0.07	104	95	108	93	64	64
Tree	64	0.01	246	191	567	240	576	367
		0.04	167	143	187	149	64	64
		0.07	125	105	144	123	64	64

Tableau 3:les résultats les techniques la recherche binaire, RIC, RC, SA.

Observation :

- On remarque que pour le concept «coast» avec le seuil 0.006 on obtient 40 images plus par rapport la technique de matching et 28 images plus par rapport le RIC et 74 images moins par rapport le RC et, avec 35 images bruit.

Avec le seuil 0.008 on obtient 27 images plus par rapport la technique de matching et 47 images plus par rapport le RIC et 80 images moins par rapport le RC, avec 9 images bruit.

Avec le seuil 0.01 on obtient 27 images plus par rapport la technique de matching et 14 image plus par rapport le RIC 88 images moins par rapport le RC, avec 6 images bruit.

- On remarque que pour le concept «flower» avec le seuil 0.01 on obtient 150 images plus par rapport la technique de matching et 21 images moins par rapport le RIC 58 images plus par rapport le RC, avec 296 images bruit.

Avec le seuil 0.03 on obtient 98 images plus par rapport la technique de matching et 2 images moins par rapport le RIC, avec 107 images bruit.

Avec le seuil 0.05 on obtient 50 images plus par rapport la technique de matching et 35 images moins par rapport le RIC, avec 86 images bruit.

On remarque que pour le concept «forest» avec le seuil 0.01 on obtient 243 images plus par rapport la technique de matching et 150 images plus par rapport le RIC 36 images moins par rapport le RC, avec 128 images bruit.

Avec le seuil 0.02 on obtient 75 images plus par rapport la technique de matching et 1 image plus par rapport le RIC , avec 28 images bruit.

Avec le seuil 0.03 on obtient 66 images plus par rapport la technique de matching et 16 images plus par rapport le RIC, avec 23 images bruit.

- On remarque que pour le concept «garden» avec le seuil 0.007 on obtient 191 images plus par rapport la technique de matching et 95 images plus par rapport le RIC 65 images plus par rapport le RC ,avec 167 images bruit.

Avec le seuil 0.01 on obtient 168 images plus par rapport la technique de matching et 66 images plus par rapport le RIC , avec 90 images bruit.

Avec le seuil 0.03 on obtient 57 images plus par rapport la technique de matching et 1 image moins par rapport le RIC , avec 29 images bruit.

- On remarque que pour le concept «grass» avec le seuil 0.005 on obtient 258

images plus par rapport la technique de matching et 203 images plus par rapport le RIC 123 images plus par rapport le RC, avec 78 images bruit.

Avec le seuil 0.007 on obtient 116 images plus par rapport la technique de matching et 81 images plus par rapport le RIC, avec 60 images bruit.

Avec le seuil 0.01 on obtient 78 images plus par rapport la technique de matching et 33 images plus par rapport le RIC, avec 35 images bruit.

- On remarque que pour le concept «mountain» avec le seuil 0.01 on obtient

176 images plus par rapport la technique de matching et 113 images plus par rapport le RIC 125 images plus par rapport le RC ,avec 158 images bruit.

Avec le seuil 0.02 on obtient 60 images plus par rapport la technique de matching et 8 images moins par rapport le RIC , avec 52 images bruit.

Avec le seuil 0.03 on obtient 56 images plus par rapport la technique de matching et 4 images moins par rapport le RIC , avec 31 images bruit.

- On remarque que pour le concept «plants» avec le seuil 0.01 on obtient 241

images plus par rapport la technique de matching et 105 images plus par rapport le RIC 54 images plus par rapport le RC ,avec 14 images bruit.

Avec le seuil 0.03 on obtient 116 images plus par rapport la technique de matching et 17 images plus par rapport le RIC , avec 7 images bruit.

Avec le seuil 0.05 on obtient 94 images plus par rapport la technique de matching et 11 images plus par rapport le RIC , avec 5 images bruit.

- On remarque que pour le concept «river» avec le seuil 0.01 on obtient 138

images plus par rapport la technique de matching et 73 images plus par rapport le RIC 66 images plus par rapport le RC ,avec 158 images bruit.

Avec le seuil 0.04 on obtient 42 images plus par rapport la technique de matching et 8 images plus par rapport le RIC , avec 33 images bruit.

Avec le seuil 0.07 on obtient 26 images plus par rapport la technique de matching et 5 images plus par rapport le RIC , avec 12 images bruit.

- On remarque que pour le concept «rocks» avec le seuil 0.01 on obtient 79 images plus par rapport la technique de matching et 27 images plus par rapport le RIC une images moins par rapport le RC ,avec 158 images bruit.

Avec le seuil 0.04 on obtient 25 images plus par rapport la technique de matching et 5 images plus par rapport le RIC , avec 33 images bruit.

Avec le seuil 0.07 on obtient 14 images plus par rapport la technique de matching et 3 images plus par rapport le RIC , avec 12 images bruit.

- On remarque que pour le concept «sea» avec le seuil 0.01 on obtient 112 images plus par rapport la technique de matching et 49 images plus par rapport le RIC et 112 images plus par rapport le RC, avec 73 images bruit.

Avec le seuil 0.04 on obtient 47 images plus par rapport la technique de matching et 6 images plus par rapport le RIC , avec 29 images bruit.

Avec le seuil 0.07 on obtient 29 images plus par rapport la technique de matching et 2 images moins par rapport le RIC , avec 11 images bruit.

- On remarque que pour le concept «tree» avec le seuil 0.01 on obtient 176 images plus par rapport la technique de matching et 49 images plus par rapport le RIC 127 images moins par rapport le RC ,avec 229 images bruit.

Avec le seuil 0.04 on obtient 85 images plus par rapport la technique de matching et 6 images plus par rapport le RIC , avec 38 images bruit.

Avec le seuil 0.07 on obtient 59 images plus par rapport la technique de matching et 18 images plus par rapport le RIC, avec 21 images bruit.

Analyse :

A partir de ce tableau, on constate que cette technique a amélioré la recherche par rapport à la recherche binaire la recherche par la technique RIC, avec tous les concepts on obtient

plus d'images bonnes, mais pour la technique RC apporte plus d'images bonnes dans certains seuils.

La technique de SA a permis à l'image d'avoir de nouvelles relations avec les concepts de sorte qu'elle crée les relations absentes ou l'augmente si c'était faible. Donc quand on fait une recherche à un concept les images qui sont en relation avec ce concept seront activées et ce qui augmente le nombre de bonnes images.

Comme par exemple l'image 0 de concept coast à des relations plus fortes qu'avant avec les concepts sea, rocks, river et plants et création des relations par avec grass et mountain donc dans les résultats de recherche obtenus par ces concepts on peut avoir cette image.

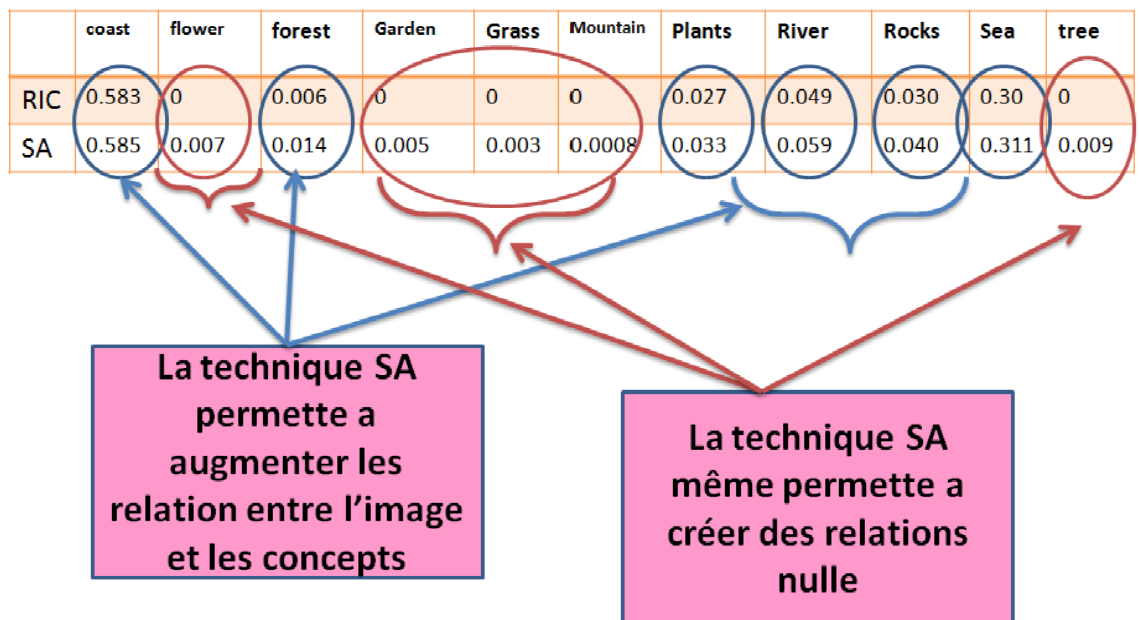


Figure 24 : les relations créées par la technique de SA

Donc quand on fait une recherche par le concept « mountain » l'image coast0 qui annotée par « coast » apparaît dans les résultats.

Requête par concept « mountain »

image annoté par coast apparu dans le résultat de recherche par concept « mountain »

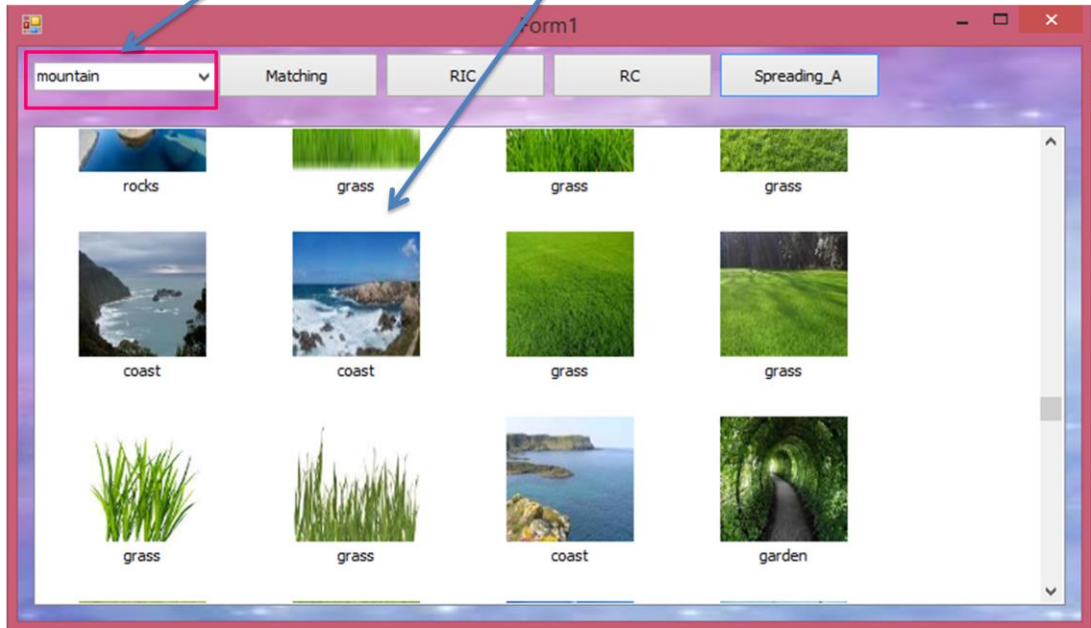


Figure 25: Le résultat de recherche par concept mountain.

5. Discussion :

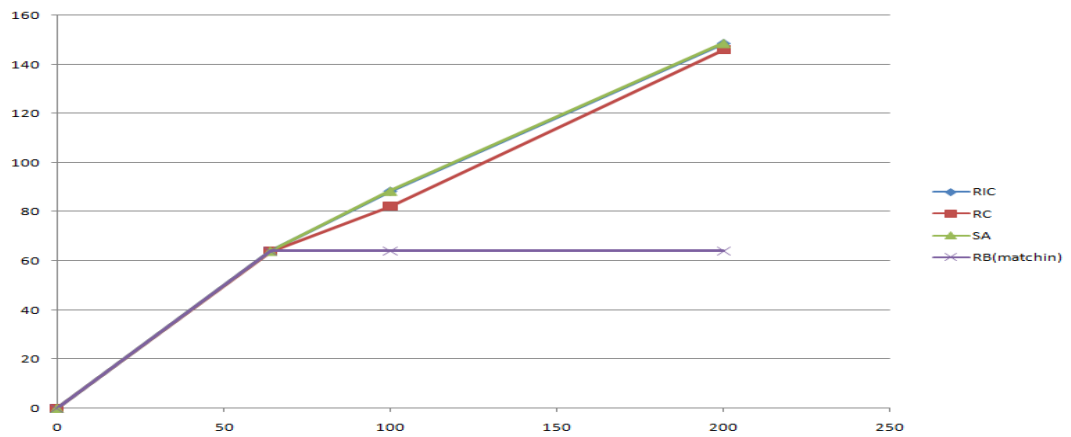


Figure 26 : courbe illustre le rappel des techniques RIC, RC, RB et SA

Rappel : bonnes images.

Scope : nombre des images retournées.

A partir de l'analyse des trois techniques on constate que les résultats de recherche se sont améliorés par rapport à la technique recherche binaire « matching ».

Les deux techniques RIC et SA ont apporté en générale un nombre d'images bonnes (rappel) très près.

La technique RC a apporté moins d'images bonnes par rapport aux deux techniques RIC et SA.

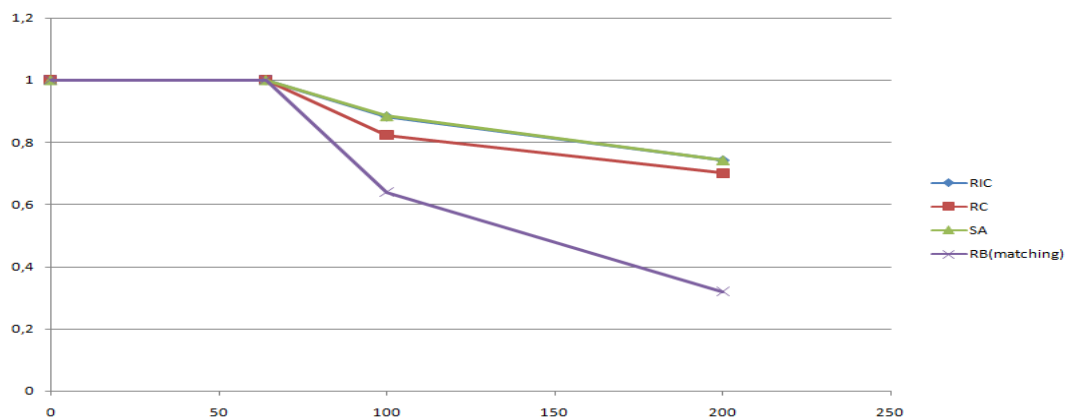


Figure 27 : courbe illustre la précision des techniques RIC, RB, RC et SA

Scope : nombre d'image retourné.

Précision=nombre d'image bonne/scope.

On remarque en général que les trois technique on donné une précision mieux que la précision par la technique RB

La technique RC a donné une précision moins que les deux techniques RIC et SA.

Les deux techniques RIC et SA on donné en générale une précision très prés.

6. Conclusion :

Dans ce chapitre Nous avons présenté en détaille comment on a exploité les relations par ces trois techniques ainsi que leurs évaluations. Nous avons présenté aussi la comparaison des techniques et ces apports par rapport à la recherche binaire.

Conclusion générale et perspective

7 Conclusion générale et perspectives :

Depuis quelques années, les bases d'images numériques connaissent un essor considérable. Leur facilité d'acquisition et de stockage les rendent très attractives pour des applications diverses. Il en résulte une production permanente et considérable d'images numériques dans différents domaines, cependant pour utiliser ces bases d'images de manière rapide et efficace un moteur de recherche ou un système de recherche d'images est nécessaire. Les moteurs de recherche d'image sont classés en deux grandes classes : des moteurs de recherche basés sur le texte, des moteurs de recherche basés sur le contenu visuel.

Dans le cadre de notre travail, nous nous sommes intéressés par la première classe. Notre but était de tenter de résoudre le problème du silence (oubli) qui est un des problèmes de la recherche par le texte, le problème de silence vient de fait que le moteur n'arrive pas à retourner tous les images pertinents à l'utilisateur, tout simplement parce qu'ils ne sont pas annotés explicitement avec les mots clés de la requête. Nous avons tenté de résoudre ce problème par trois techniques :

la première technique RIC(relation image et concept), nous avons créé des relations entre l'image et les concepts à partir de la fréquence d'occurrences des concepts au contexte de l'image , en général cette technique a amélioré la recherche, et à un certain degré elle a permis de résoudre le problème ,cependant, cette technique reste insuffisante, car le contexte tout seul ne permet pas de donner toutes les relations possibles entre l'image et les concepts ceci est traduit par l'absence de certaines relations entre les concepts, alors nous avons essayé de résoudre ce problème par une deuxième technique RC(relation entre concept) qui permet de créer des relations entre les concepts eux même, cette technique a amélioré la recherche, et à un certain degré, permet de résoudre le problème mais elle reste insuffisante, car cette technique ne met pas l'image en valeur, et les relations entre certains concepts reste faible. C'est pour cela, on a pense à la technique de SA (spreading activation) qui permis à augmenter des relations faibles ou les créer avec une considération de la valeur de l'image.

La technique SA a permet de l'image de créer des relations plus avec les concepts ceci est traduit par une augmentation de résultats bonnes mais certaine de ces relations on contribué a la présence d'un nombre de bruit dans les résultats.

En général, on a pu résoudre le problème de silence par ces trois techniques, les résultats de recherche se sont améliorés par rapport à la recherche binaire 'matching' avec la présence d'un certain nombre de bruit par les trois techniques.

Les perspectives de notre travail sont :

-améliorer plus la technique de SA par une considération d'un autre facteur et non pas seulement par la relation de l'image et les concepts, et les relations entre les concepts.

- résoudre le problème de silence par la recherche de plusieurs concepts et ne pas se limiter par un seul concept.

8 Bibliographie

ABED, H., & ZAOUI, L. (2011). Système D'Indexation et de Recherche d'Images par le contenu. *Système D'Indexation et de Recherche d'Images par le contenu* , 3.

BEDOUHEN, S. (s.d.). Recherche d'images par le contenu. *recherche d'imagepar contenu* . memoire de magestère,Automatiqu,genie electrique, Tizi Ouezo: Mouloud Feaoun.

BEN CHIKH, N., & BEN ELZZINE, R. (2011, 06). LA RECHERCHE D'IMAGES PAR LA SEMANTIQUE. *LA RECHERCHE D'IMAGES PAR LA SEMANTIQUE* . Memoire de Master informatique acadimique ,departement mathématique at informatique, Ouargla: Unvercité kasdi merbah Ouargla.

LAOUBI, A. (2014, 06 19). Recherche d'images sémantique basée sur la sélection automatique des concepts. *Mémoire MASTER ACADEMIQUE* . Ouargla, Département d'informatique et des technologies de l'information, Algérie : UNIVERSITE KASDI MERBAH OUARGLA.

LESLOUS, O., & BOUCETTA, M. (2011, 07 27). recherche d'images par le contenu visuel : recherche d'images par le contenu visuel:. *Memoire de mastère* . Ouargla, departemenet mathématique et informatique, Ouargla: Univercité kasdi merbah .

TORJMEN, M. (2009, 09 04). Approches de Recherche Multimedia dans des Documents Semi-Structurées :Utilisation du contexte textuel et structurel pour la s'élection d'objets multimedia. *Approches de Recherche Multimedia dans des Documents Semi-Structurées :Utilisation du contexte textuel et structurel pour la s'élection d'objets multimedia* . Toulouse, l'Universit'e Paul Sabatier, Toulouse: l'Universit'e Paul Sabatier.