

**RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE**  
**MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE**  
**SCIENTIFIQUE**

**UNIVERSITÉ MOULOUD MAMMERRI DE TIZI-OUZOU**

**FACULTÉ DE GÉNIE ÉLECTRIQUE ET D'INFORMATIQUE**  
**DÉPARTEMENT D'INFORMATIQUE**



# **Mémoire de fin d'études**

**En vue de l'obtention du Diplôme de  
Master en Informatique**

## **Thème**

# **Recherche d'images basée sur le texte et le contenu**

**Proposé et dirigé par :**  
Mme Amirouche

**Présenté par :**  
- Mr. MEHADDI Farid

**PROMOTION : 2011-2012**

# Remerciements

Je tiens à remercier vivement et à témoigner ma sincère gratitude à ma promotrice madame AMIROUCHE pour tout ce qu'elle m'a apporté comme aides, connaissances et conseils.

Je tiens à remercier tous nos enseignants qui ont contribué à notre formation.

Je tiens à remercier également les membres du jury qui me font l'honneur de juger mon travail.

Je tiens à remercier toutes les personnes qui m'ont aidés de près ou de loin.

# Sommaire

Introduction générale.....	1
Chapitre I : recherche d'information.....	3
I. Introduction.....	4
II. Notions de base.....	5
1. Document.....	5
2. Requête.....	5
3. Pertinence.....	5
III. Processus de recherche d'information.....	6
1. Le processus d'indexation.....	7
A. Indexation manuelle.....	7
B. Indexation semi-automatique.....	7
C. Indexation automatique.....	7
i. L'extraction des mots-clés du document.....	7
ii. L'élimination des mots vides.....	7
iii. La normalisation des mots-clés du document.....	8
iv. La pondération des mots-clés.....	8
2. L'appariement document-requête.....	9
3. Reformulation de la requête.....	10
IV. Les modèles théoriques de la RI.....	10
1. Le modèle booléen.....	10
2. Le modèle vectoriel.....	11
3. Le modèle probabiliste.....	12
V. Modèles étendus.....	13
1. Le modèle booléen étendu.....	13
2. Le modèle LSI (Latent semantic indexing).....	14
VI. Évaluation des SRI.....	15
1. Rappel et précision.....	15
VII. Conclusion.....	16
Chapitre II : Indexation et recherche d'images par le contenu.....	17
I. Introduction.....	18

II.	Architecture du système d'indexation et recherche d'images.....	19
III.	Types de requêtes.....	20
IV.	Descripteurs visuels.....	21
	1. Les descripteurs de couleurs.....	21
	A. L'histogramme.....	21
	B. Les espaces de couleur.....	22
	C. Les moments statistiques.....	23
	D. Description par couleurs dominantes.....	24
	2. Les descripteurs de texture.....	24
	A. Matrices de cooccurrences.....	24
	B. Les filtres de gabor.....	26
	3. La forme.....	26
	4. Segmentation et points d'intérêt.....	26
	A. Segmentation.....	27
	B. Points d'intérêt.....	28
V.	Mesures de similarité.....	29
	1. Intersection d'histogramme.....	29
	2. Distance euclidienne.....	29
	3. Distances entre distributions.....	29
VI.	Conclusion.....	30
<b>Chapitre III : indexation et recherche d'images en combinant le texte et le contenu.....</b>		<b>31</b>
I.	Introduction.....	32
II.	Système de recherche d'image par le contenu.....	32
	1. Les descripteurs utilisés.....	32
	A. Les moments statistiques.....	32
	B. Les couleurs dominantes.....	32
III.	Système de recherche d'image par le texte.....	33
	1. Processus d'indexation.....	33
	2. Les fichiers inversés (Inverted files).....	34

<b>IV.</b>	<b>Système de recherche par le texte et l'image.....</b>	<b>35</b>
<b>V.</b>	<b>La méthode combinant textes et images.....</b>	<b>36</b>
<b>VI.</b>	<b>Conclusion.....</b>	<b>37</b>
	<b>Chapitre IV : implémentation et évaluation expérimentale.....</b>	<b>38</b>
<b>I.</b>	<b>Introduction.....</b>	<b>39</b>
<b>II.</b>	<b>Architecture du processus d'indexation et de recherche par le contenu et le texte .....</b>	<b>40</b>
	<b>1. Architecture du processus d'indexation.....</b>	<b>40</b>
	<b>2. Présentation des différents modules.....</b>	<b>41</b>
	<b>3. Processus de recherche.....</b>	<b>41</b>
<b>III.</b>	<b>Expérimentations de la recherche par image exemple.....</b>	<b>42</b>
	<b>1. Les bases d'images utilisées.....</b>	<b>42</b>
	<b>A. La base de Wang.....</b>	<b>42</b>
	<b>B. La base de Coil.....</b>	<b>42</b>
	<b>2. Résultats obtenus.....</b>	<b>43</b>
<b>VII.</b>	<b>Illustrations de la recherche par couleurs et/ou texte.....</b>	<b>45</b>
<b>VIII.</b>	<b>Conclusion.....</b>	<b>46</b>
	<b>Conclusion générale.....</b>	<b>47</b>
	<b>Annexe.....</b>	<b>48</b>
	<b>Bibliographie.....</b>	<b>54</b>

# **Introduction générale**

## **Contexte du travail**

Depuis une décennie, les bases de documents multimédia se multiplient et leur taille augmente de manière spectaculaire. Parmi les medias les plus stockées à l'heure actuelle, on trouve en première position le texte et les images.

En effet l'explosion d'Internet et aussi le développement à grande échelle de la photographie numérique, il n'est pas rare d'avoir des bases d'images numériques contenant plusieurs milliers et même plusieurs dizaines de milliers d'images. Du coup plusieurs base d'images annotées en vu le jour dans un contexte extrêmement récent. Ce qui a conduit au développement de techniques de recherche d'images. On peut distinguer deux principales techniques de recherche d'images.

La première technique est la technique de la recherche d'images basée sur le texte. Elle remonte à la fin des années 70 et elle est due à la gestion de base de données communauté. Dans cette approche, les images sont d'abord annotées avec le texte, et puis les techniques de recherche sur le texte peuvent être appliquées. Mais cette approche à pas mal d'inconvénients, parce que d'une part les images ne sont pas toujours annotées, et d'autre part leur annotation manuelle peut s'avérer très coûteux en temps, mais aussi peut être subjective.

La deuxième approche concerne le contenu de l'image telle que la couleur et la texture. Cette approche, connue sous le nom en l'anglais de « Content-based Image Retrieval » ou CBIR, a été proposée au début des années 90 et vient de la communauté de vision par ordinateur. Le principal inconvénient des systèmes de CBIR est le manque des aspects sémantiques. Qui provient de l'absence de coïncidence entre des informations qui peuvent être extraites à partir des données visuelles et des interprétations que l'utilisateur assigne à la même donnée dans une même situation. En effet, en ajoutant des caractéristiques au niveau bas, on utilise des concepts sémantiques pour classer et identifier les images, mais ces concepts sont encore difficiles pour extraire automatiquement à partir du contenu de l'image.

Afin de surmonter les inconvénients des deux approches, certains chercheurs ont étudié les possibilités d'associer le texte au contenu d'image dans la recherche. On a proposé plusieurs systèmes qui peuvent les lier.

## **Objectif**

Notre objectif dans ce mémoire est d'étudier l'état de l'art des méthodes et des systèmes déjà existants à l'heure actuelle. Et de construire un système qui puisse associer le contenu visuel et le texte de l'image.

## **Contenu du rapport**

Ce mémoire se concentre sur la recherche d'information par croisement du média texte et image, il comporte quatre chapitres décrits comme suit :

Le premier chapitre présente un état de l'art sur la recherche d'information, dans lequel, nous présentons les notions et concepts de base de la recherche d'information, et nous décrivons le processus de la RI. Par la suite, les principaux modèles qui sont à la base de la majorité des systèmes actuels sont décrits.

Dans le deuxième chapitre nous présentons un état de l'art sur la recherche d'images par le contenu, CBIR.

Le troisième chapitre présente notre système de recherche d'image par le contenu, ainsi que par le texte, et l'association du texte et du contenu.

Enfin, dans le quatrième chapitre nous procédant à l'implémentation et l'évaluation de notre système de recherche d'image par le contenu et le texte.

# **Chapitre I**

## **Recherche d'information**

## **I. Introduction :**

La recherche d'information (RI) est un domaine de l'informatique qui s'intéresse à l'acquisition, l'organisation, le stockage, la recherche et la sélection d'information [Tebri ,04].

La RI est apparu au début des années 1950, peu après l'invention des ordinateurs. Le nom de « recherche d'information » (information retrieval) fut donné par Calvin N. Mooers en 1948 pour la première fois quand il travaillait sur son mémoire de maîtrise. La première conférence dédiée à ce thème International Conference on Scientific Information s'est tenue en 1958 à Washington. On y comptait les pionniers du domaine, notamment, Cyril Cleverdon, Brian Campbell Vickery, Peter Luhn.

L'apparition de nouveaux documents multimédias (non-textuelle), notamment d'images, a créé de nouveaux besoins concernant la recherche. Dans notre nous nous focalisons sur la recherche d'images qui peut être réalisée de deux manières différentes :

Une recherche basée sur le contexte : ceci consiste à ne pas utiliser les informations de l'image elle-même, mais plutôt l'ensemble des informations entourant l'image telles que les légendes d'image.

La recherche basée sur le contenu (CBIR : Content Based Image Retrieval) qui utilise des descripteurs physiques de bas niveau tels que la couleur, la texture...etc.

Dans ce premier chapitre nous décrivons les concepts de base de la RI, le processus de RI, et les principaux modèle de RI, et nous terminons par les mesures d'évaluation d'un SRI.

## **II. Notions de base**

Un SRI a pour rôle de sélectionner, à partir d'un besoin en information, les documents qui peuvent l'intéresser l'utilisateur. Cette définition fait apparaître trois notions clés qu'il convient de préciser : documents, requête, pertinence.

### **1. Document :**

Le document constitue l'information élémentaire d'une collection de documents. Un document peut être un texte, une page WEB, une image, une bande vidéo. Dans notre contexte, nous appelons document toute unité qui peut constituer une réponse à un besoin en information exprimé par un utilisateur.

### **2. Requête :**

La requête constitue le besoin d'information de l'utilisateur, qu'il écrive sous Plusieurs formes.

### **3. Pertinence :**

La notion de pertinence est très complexe, car elle est fortement subjective, c'est-à-dire qu'elle dépend de l'utilisateur. En distingue deux types de pertinence. La pertinence system et la pertinence utilisateur. La pertinence système est l'évaluation par un système de l'adéquation entre des documents et une requête. Par contre la pertinence utilisateur se traduit par des jugements de pertinence sur les documents fournis en réponse à une requête [Simonnot, 02].

### III. Processus de recherche d'information :

L'utilisateur traduit ses besoins de façon structurée sous forme de requête qu'il transmet au système de recherche d'information (SRI). Celui-ci doit retourner à l'utilisateur le maximum de documents pertinents par rapport à son besoin. Pour cela, le SRI met en relation les documents du corpus d'une part et la requête utilisateur d'autre part [Dahak, 05].

Les différentes étapes du processus de RI, sont représentées schématiquement par le processus en U qui est illustré par la **figure 1.1** et qui sont :

- Les opérations d'analyse, d'indexation et d'appariement qui permettent globalement de traiter la requête et les documents dans le but de sélectionner des documents à présenter à l'utilisateur.
- La fonction de reformulation de la requête, qui intervient en réponse aux résultats obtenus.

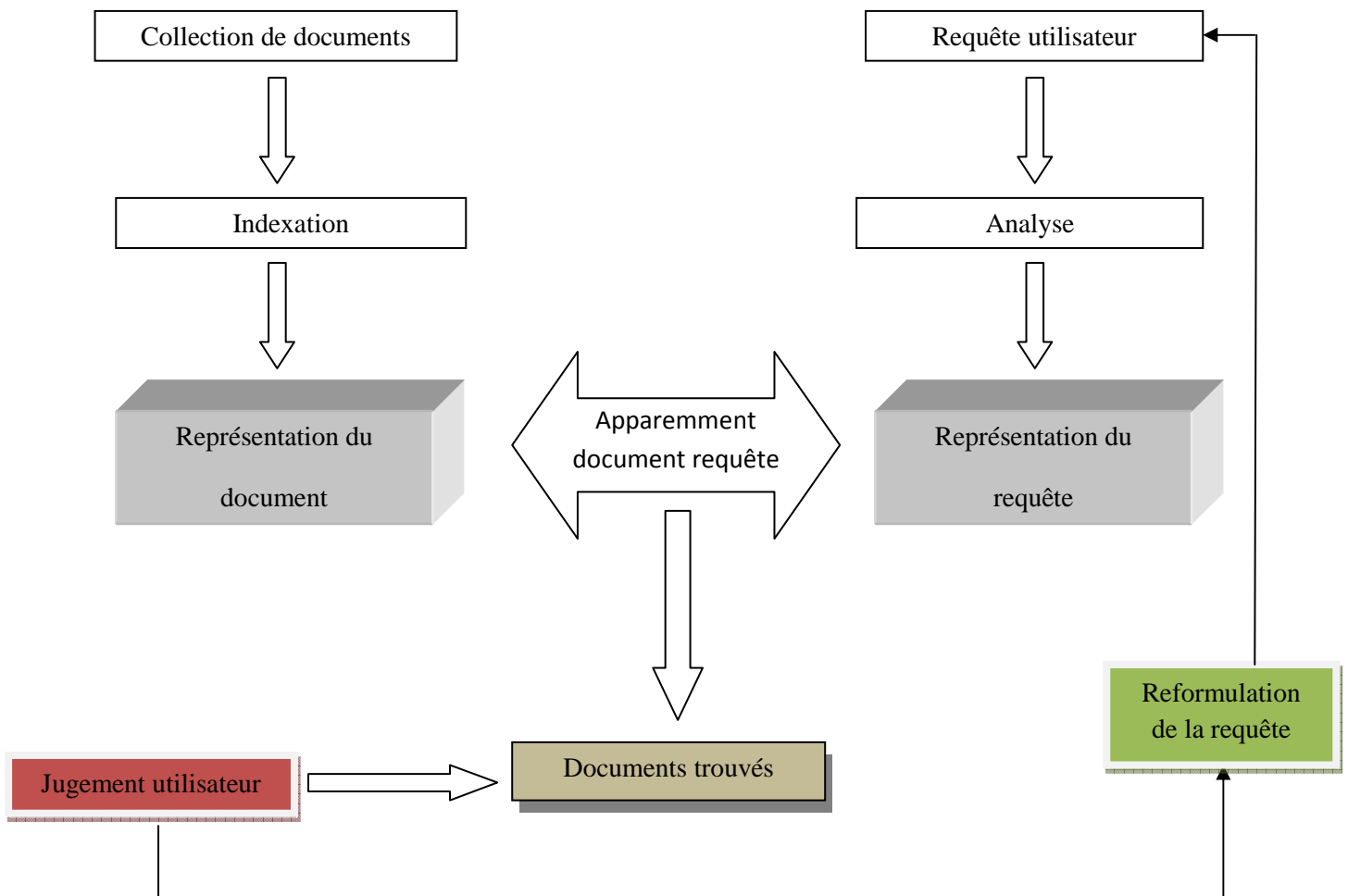


Figure 1.1 : processus de recherche d'information (processus en U)

## **1. Le processus d'indexation**

L'indexation est une étape primordiale dans un SRI, et elle consiste à analyser chaque document de la collection afin de créer un ensemble de mots-clés aussi appelés descripteurs. Un descripteur peut-être soit un mot simple ou un groupe de mots. Ces derniers aussi descripteurs seront par la suite exploités dans le processus d'interrogation. Elle peut être effectuée d'une façon manuelle, semi-automatique ou automatique.

### **A. Indexation manuelle :**

L'indexation manuelle est faite par un expert du domaine ou par un documentaliste, qui se charge de décrire le contenu sémantique d'un document, selon ses connaissances propres. Cette approche peut être subjective car elle dépend de l'indexeur et de ses connaissances du domaine. De plus, le temps nécessaire à sa réalisation est très important dans le cas de corpus de textes volumineux [Boubekeur, 08].

### **B. Indexation semi-automatique :**

Dans le cas d'une indexation semi-automatique [Maniez, 91], les indexeurs utilisent un thésaurus<sup>1</sup> ou une base terminologique, qui est une liste organisée de descripteurs (mots clés) obéissants à des règles terminologiques propres et reliés entre eux par des relations sémantiques.

### **C. Indexation automatique :**

L'indexation automatique est un processus complètement informatisé dont le but est d'extraire les mots-clés, l'indexation automatique est basée sur les étapes suivantes : analyse lexicale, la normalisation des mots-clés, élimination des mots vides, pondération des mots-clés. Ces étapes sont décrites qui suit [Maron et Kuhns, 60] :

#### **i. Analyse lexicale**

L'extraction des mots-clés aussi Appelée Tokenization en anglais, permet d'extraire un ensemble de termes simples par une analyse lexicale permettant de reconnaître les espaces de séparation des mots, des chiffres, les ponctuations, etc.

#### **ii. L'élimination des mots vides**

Cette étape consiste à supprimer Les mots vides pour améliorer la fiabilité et les performances du SRI. Les mots vides sont des termes non significatifs (pronoms personnels, prépositions,...) ou mots athématiques (les mots qui peuvent se retrouver dans n'importe quel document parce qu'ils exposent le sujet mais ne le traitent pas, comme par exemple contenir, appartenir, etc.). L'élimination des mots vides se base sur l'utilisation :

- D'une liste de mots vides (aussi appelée anti-dictionnaire),
- En définissant un seuil minimum et un seuil maximum de fréquence des termes, les termes dépassant un certain le seuil maximum ou inférieur au seuil minimum dans la collection.

<sup>1</sup> Un thésaurus est un ensemble hiérarchique de termes clés représentant des concepts d'un domaine particulier. Ils sont organisés en thèmes et possèdent des liens sémantiques entre eux : synonymie, équivalence, terme spécifique (lien vers un concept de sens plus précis), terme général (lien vers un concept de sens plus large).

### **iii. La normalisation des termes**

Cette étape consiste à effectuer un traitement au niveau de chacun des mots afin de regrouper un ensemble de mots significativement identiques ainsi dans l'index ne sont conservées que les formes normalisées, ce qui offre un gain de place appréciable. Appelée aussi « traitement morphologique des mots-clés » [Denoyer, 04]. Parmi ses traitements :

- Le stemming consiste à regrouper sous un même identifiant des mots dont la racine est commune (par exemple, « voisine » et «voisin »).
- la lemmatisation consiste à regrouper des mots dont la signification est la même alors même que leurs racines sont différentes (par exemple, « maison » et « baraque»). C'est une tâche plus complexe que le stemming et qui repose habituellement sur l'utilisation de grandes bases de connaissances.

### **iv. La pondération des mots-clés**

La pondération des termes consiste à affecter à chaque terme un poids qui représente son degré d'importance dans le document où il apparait. Ce poids permet de distinguer les termes entre eux.

Les techniques de pondération, sont basées sur les facteurs tf et idf.

- **TF (Term Frequency)** : Cette mesure a été introduite pour tenir compte de la fréquence d'un terme dans un document. L'idée sous-jacente est que plus un terme est fréquent dans un document plus il est important dans sa description. La mesure tf représente une pondération locale d'un terme dans un document.
- **IDF (Inverse Document Frequency)** : ce facteur mesure la fréquence d'un terme dans toute la collection, c'est une mesure de "pondération globale". En effet, un terme fréquent dans la collection est moins discriminant pour un document qu'un terme moins fréquent. La force de discrimination d'un terme pour un document est inversement proportionnelle à sa fréquence. Elle sera mesurée par sa fréquence documentaire inverse IDF. Il est exprimé comme :  $\log(N/n_i)$ , avec N est la taille (nombre de documents) de la collection et  $n_i$  le nombre de documents contenant le terme  $t_i$ . La combinaison des deux mesures (tf et idf) donne une bonne approximation de l'importance du terme dans le document, particulièrement dans les corpus de documents de tailles homogènes. Les fonctions de pondération sont souvent référencées sous le nom de TFIDF.

## 2. L'appariement document-requête

Après le processus d'indexation, le processus d'appariement document requête permet d'assigner un score de pertinence à un document vis-à-vis d'une requête. Cette valeur est calculée à partir d'une fonction ou d'une probabilité de similarité notée RSV (D, Q) (Retrieval Status Value), où Q est une requête et D un document. Cette valeur permet d'ordonner les documents fournis en résultat.

## 3. Reformulation de la requête

L'utilisateur formule ses besoins sous forme d'une requête pour le SRI, il peut arriver que le SRI ne renvoie pas les résultats qui l'intéressent. Pour remédier à ce problème, une étape de reformulation de la requête est souvent nécessaire pour retrouver plus de documents pertinents.

La reformulation de la requête peut être automatique (cas expansion de la requête) ou manuelle (cas de la réinjection de pertinence). Dans le premier cas, l'utilisateur n'intervient pas. La reformulation de la requête est faite à partir d'un thésaurus qui définit les relations entre les différents termes de l'index et permet de sélectionner de nouveaux termes à ajouter à la requête initiale.

Par contre dans la reformulation manuelle de la requête. On présente à l'utilisateur une liste de documents jugés pertinents par le SRI comme réponse à la requête initiale. Après les avoir examinés, l'utilisateur indique ceux qu'il considère comme pertinents. L'idée principale de la réinjection de pertinence est de d'utiliser les termes importants appartenant aux documents jugés pertinents par l'utilisateur pour construire une nouvelle requête, et de renforcer l'importance de ces termes dans la nouvelle requête.

## IV. Les modèles théoriques de la RI

Le modèle joue un rôle central dans la Recherche d'Information dans la mesure que c'est lui qui détermine le comportement clé d'un SRI. Il doit fournir une formalisation du processus de recherche d'information et accomplir plusieurs rôles dont le plus important est de fournir un cadre théorique pour la modélisation de la mesure de pertinence [Sauvagnat, 05].

Il doit créer une représentation interne pour un document ou pour une requête basée sur ces termes et définir une méthode de comparaison entre une représentation de document et une représentation de requête afin de déterminer leur degré de correspondance (ou similarité). Dans ce qui suit nous allons définir les principaux modèles de RI.

### 1. Le modèle booléen

Le modèle booléen est une méthode ensembliste de représentation du contenu d'un document. Ce modèle de recherche est basé sur l'algèbre de boole, un document est représenté par un ensemble de termes, la requête est exprimée par une expression booléenne de mots assemblés par les opérateurs logique : ET, OU, NON.

La similarité entre un document et une requête est définie par :

$$\begin{aligned} \text{rsv}(q, d) &= 1 \text{ si } d \text{ appartient à l'ensemble décrit par la requête} \\ \text{Rsv}(q, d) &= 0 \text{ sinon} \end{aligned}$$

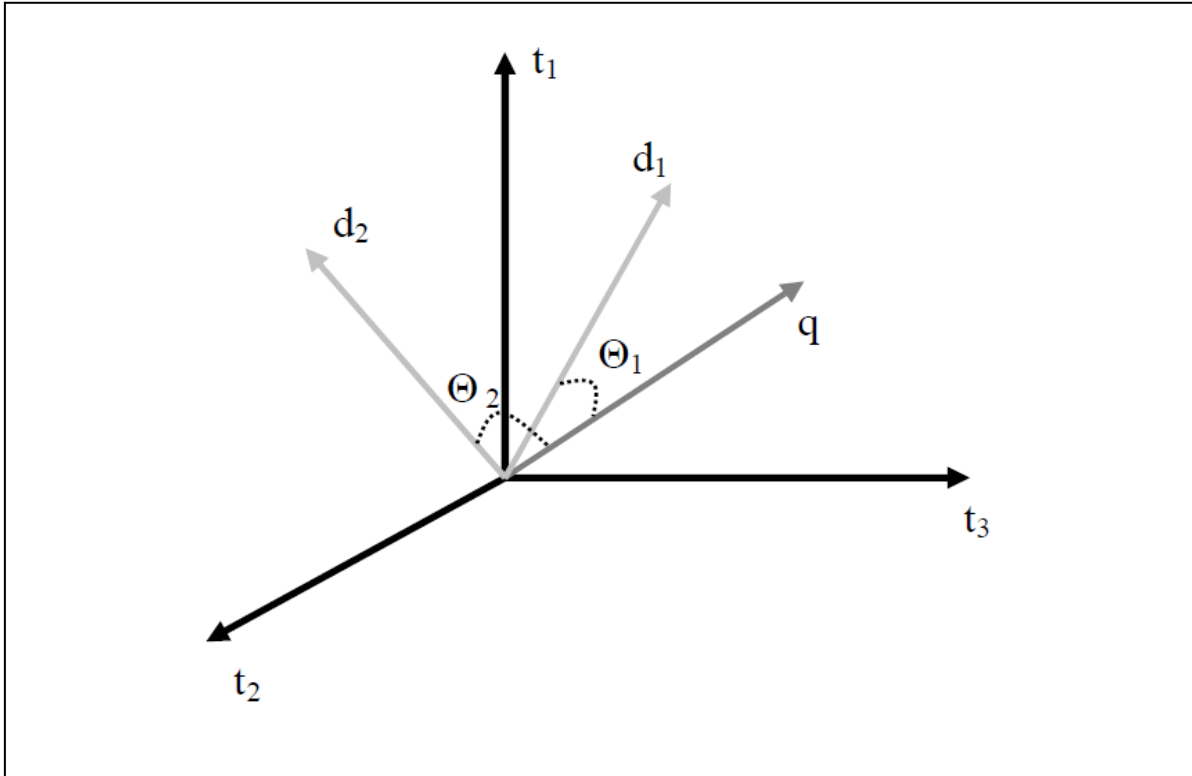
Le principal inconvénient avec le modèle booléen c'est qu'il n'effectue pas la pondération des termes (descripteurs des documents), ce qui n'implique aucun ordre pour les documents sélectionnés.

### 2. Le modèle vectoriel

Le modèle vectoriel a été créé pour compenser la limitation de la pondération binaire des modèles booléens. Ce modèle propose de représenter les documents et les requêtes par des vecteurs d'indexation dans un espace engendré par les termes d'indexation. Le modèle vectoriel représente les requêtes et les documents sous forme de vecteurs dans un même espace vectoriel.

Chaque document est représenté par un vecteur  $d_j = \{w_{1,j}, w_{2,j}, w_{3,j}, \dots, w_{n,j}\}$  représente le poids des termes dans le document,  $n$  étant le nombre total de termes de l'index, et la requête aussi est vue comme un vecteur  $q = \{w_{1,q}, w_{2,q}, w_{3,q}, \dots, w_{n,q}\}$ .

La **figure 1.2** montre un exemple d'espace vectoriel composé des trois termes  $t_1, t_2, t_3$ . Les index de deux documents  $d_1$  et  $d_2$  et une requête sont représentés dans cet espace. Étant donné ces deux vecteurs, la pertinence est traduite en une similarité vectorielle : un document est d'autant plus pertinent à une requête que le vecteur associé est similaire à celui de la requête, c'est-à-dire autant que la distance et l'angle entre les vecteurs documents et requêtes sont petits.



**Figure 1.2 : Représentation vectorielle de deux documents (d1 et d2) et d'une requête (q) dans un espace composé de trois termes (t1, t2, t3)**

Les principales mesures de similarités d'évaluation sont :

**Le produit scalaire :**  $RSV(\vec{d}, \vec{q}) = \sum_{i=1}^t w_{i,j} * w_{i,q}$

**Mesure de cosinus :**  $RSV(\vec{d}, \vec{q}) = \frac{\vec{d}_j * \vec{q}}{|\vec{d}_j| * |\vec{q}|} = \frac{\sum_{i=1}^t w_{i,j} * w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} * \sqrt{\sum_{i=1}^t w_{i,q}^2}}$

### 3. Le modèle probabiliste :

Le modèle probabiliste utilise un modèle mathématique fondé sur la théorie de la probabilité [Ben auouicha, 09], Ce modèle tente d'estimer la probabilité qu'un document  $d$  soit pertinent pour une requête  $q$ , notée :  $P(\text{pert} / d, q)$ . Cette approche est justifiée dans le Probability Ranking Principle (PRP) formulé par Robertson [Robertson, 77]. On distingue deux classes de documents pour une requête : les pertinents (Pert) et les non pertinents (Npert), ainsi deux mesures de probabilité seront calculées :

- $P(\text{Pert}_q / d)$  : probabilité que  $d$  soit dans la classe des Pert.
- $P(\text{Npert}_q / d)$  : probabilité que  $d$  soit dans la classe des Npert.

Un document sera sélectionné si  $P(\text{pert} / d) > P(\text{Npert} / d)$ , ce qui est équivalent à ordonner les documents par rapport à un degré de vraisemblance :

$$RSV(q, d) = P(\text{Pert} / d) / P(\text{Npert} / d). (1)$$

En appliquant la règle de Bayes sur la formule (1), cela donnera :

$$RSV(q, d) = P(d / \text{Pert}) / P(d / \text{Npert}). (2)$$

Pour estimer les probabilités  $P(d / \text{Pert})$  et  $P(d / \text{Npert})$  un document sera décomposé en un ensemble d'événements. Chaque événement dénotera la présence ou l'absence d'un terme dans un document, c'est le modèle de recherche indépendant "Binary Independence Retrieval" (BIR) qui suppose l'indépendance des termes des documents. La formule (2) Devient :

$$RSV(q, d) = \sum_{i=1}^t \log \frac{P(t_i / \text{Pert})}{P(t_i / \text{Npert})}$$

Avec :

- $P(t_i / \text{Pert}) = r_i / R$
- $P(t_i / \text{Npert}) = (n_i - r_i) / (N - R)$

Où

$r_i$  : est le nombre de documents pertinents dans lesquels le terme  $t_i$  apparaît.

$R$  : est le nombre de documents pertinents pour la requête.

$n_i - r_i$  : est le nombre de documents non pertinents dans lesquels le terme  $t_i$  apparaît.

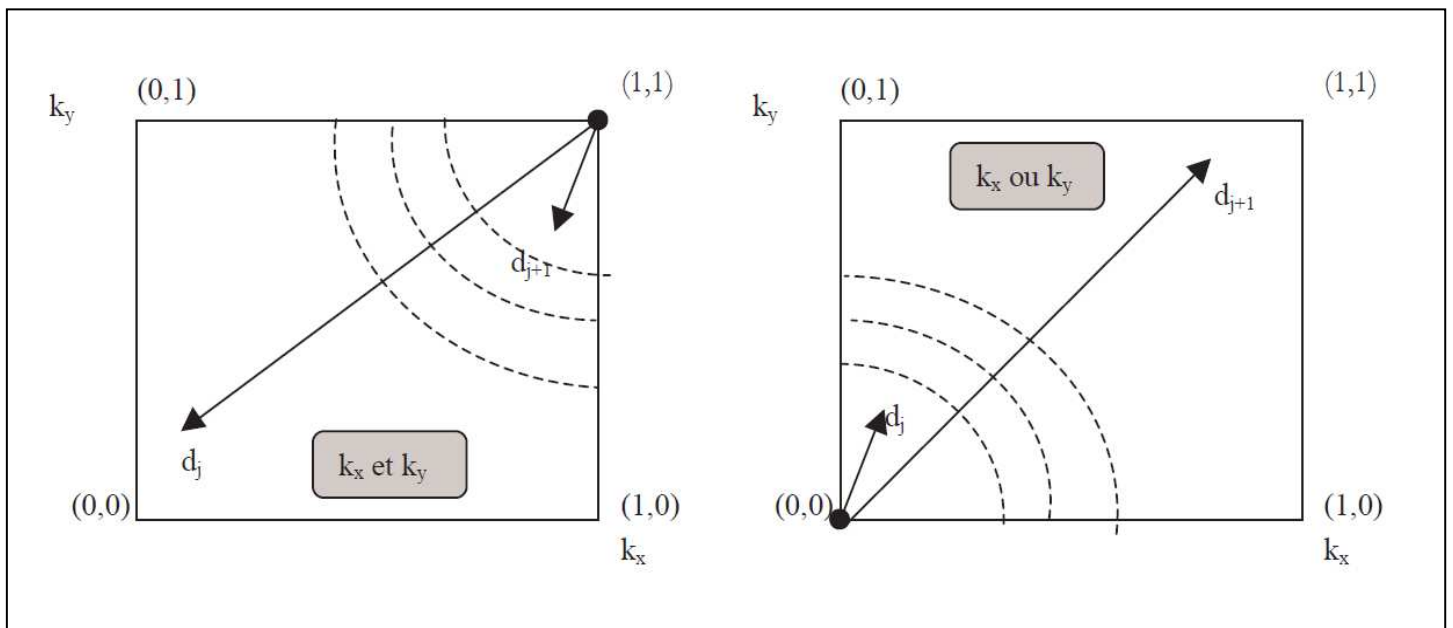
$N$  : est le nombre total de documents dans la collection.

## V. Modèles étendus

### 1. Le modèle booléen étendu

Introduit en 1983 par Salton, Fox, and Wu, le modèle booléen étendu peut être vu comme une combinaison du modèle vectoriel et du modèle booléen pour la requête car plus simple à formuler pour ce modèle, comme dans le modèle vectoriel les documents sont représentés sous forme de vecteurs.

Prenons un exemple simple dans la **figure 1.3** avec seulement deux termes dans la requête,  $k_x$  et  $k_y$ . On peut représenter les requêtes et les documents dans un espace à deux dimensions :



**Figure 1.3 :** Logique booléenne étendue en considérant un espace composé de deux termes  $k_x$  et  $k_y$ .

Pour la requête ( $k_x$  ou  $k_y$ ), le point (0,0) correspond à la situation où aucun des termes  $k_x$  et  $k_y$  n'est présent dans le document, cette situation est donc la moins désirée. Pour la requête ( $k_x$  et  $k_y$ ), le point (1,1) représente la situation où les termes  $k_x$  et  $k_y$  sont présents dans le document, il s'agit du cas le plus désiré. On peut donc conclure que les requêtes avec l'opérateur **ou** doivent ranger les documents dans l'ordre décroissant de leur distance au point (0,0) et que les requêtes avec l'opérateur **et** doivent ranger les documents dans l'ordre croissant de leur distance au point (1,1).

Sur cette base, les scores de similarité document - requête sont ainsi calculés :

$$RSV(\vec{d}, q_{ou}) = \sqrt{\frac{x^2 + y^2}{2}}$$

$$RSV(\vec{d}, q_{et}) = 1 - \sqrt{\frac{(1-x)^2 + (1-y)^2}{2}}$$

Le modèle p-norm généralise cette notion de distance en incluant non seulement les distances euclidiennes mais aussi les p-distances, avec  $1 \leq p \leq \infty$ . La valeur de p est indiquée au moment de la requête [Sauvagnat, 05]. Si m est le nombre de termes dans la requête, les fonctions de similarité deviennent alors:

$$RSV(\vec{d}, q_{ou}) = \left( \frac{x_1^p + x_2^p + \dots + x_m^p}{m} \right)^{\frac{1}{p}}$$

$$RSV(\vec{d}, q_{et}) = 1 - \left( \frac{(1-x_1)^p + (1-x_2)^p + \dots + (1-x_m)^p}{m} \right)^{\frac{1}{p}}$$

## 2. Le modèle LSI (Latent semantic indexing)

Les autres modèles présentés dans ce chapitre utilisent les mots-clés pour représenter le contenu d'un document (ou d'une requête). On se pose souvent la question sur la qualité de cette représentation. En effet, les mots-clés ne sont pas une représentation idéale. Le but de la LSI est de transformer une représentation par des mots-clés en une autre représentation qui est "meilleure". Le mot "meilleur" est compris dans le sens suivant: les documents et les requêtes sémantiquement similaires seront plus proches avec la représentation transformée qu'avec les mots-clés. La transformation par LSI est comme suit:

- Au début, chaque document et requête est représenté comme un vecteur de mots-clés.
- LSI utilise la SVD (singular value decomposition) pour créer un nouvel espace vectoriel:

$$\mathbf{X} = \mathbf{T}_0 \mathbf{S}_0 \mathbf{D}_0$$

Où X est la matrice de document-terme originale (de taille  $t \times d$ )

$\mathbf{T}_0$  est une matrice  $t \times m$

$\mathbf{S}_0$  est une matrice  $m \times m$  diagonale (seulement les éléments en diagonal sont non-nuls)

$\mathbf{D}_0$  est une matrice  $m \times d$ .

La valeur de m est choisi comme une valeur  $\leq \min(t,d)$ . En plus, on trie les valeurs dans  $\mathbf{S}_0$  dans l'ordre décroissant. Il existe juste une seule décomposition de cette façon.

On pense que la représentation par des mots-clés contient beaucoup de bruits. Typiquement, ces bruits se retrouvent dans les dimensions de  $\mathbf{S}_0$  qui ont des valeurs faibles. Ainsi, la technique de LSI veut supprimer ces dimensions de valeurs faibles (ou de les ramener à la

valeur 0), ce qui ramène les dimensions de  $S_0$  à  $k$ , et cette matrice réduite est notée par  $S$ . En conséquence, les matrices  $T_0$  et  $D_0$  nettoyées deviennent  $T$  et  $D$ .

La matrice reconstituée correspondrait à une matrice de document-terme nettoyé. Mais la LSI ne fait pas cette reconstitution. Plutôt, les matrices décomposées vont rester. En particulier  $S$  correspond à un nouvel espace vectoriel. Des exemples montrent que certains documents sémantiquement similaires se seront rapprochés dans ce nouvel espace.

Quand une requête est soumise, elle est aussi traduite dans ce nouvel espace. Dans [Deerwester & al, 90], une requête est transformée d'abord en un pseudo-document comme suit:

$$D = X_q' T S^{-1}$$

Où  $X_q$  est le vecteur de mots-clés de la requête (dans l'espace des mots-clés). Ensuite, ce pseudo-document est ajouté dans la matrice  $D$  comme un nouveau "document". Le calcul de similarité entre chaque paire de documents peut se faire par:

$$D S^2 D'$$

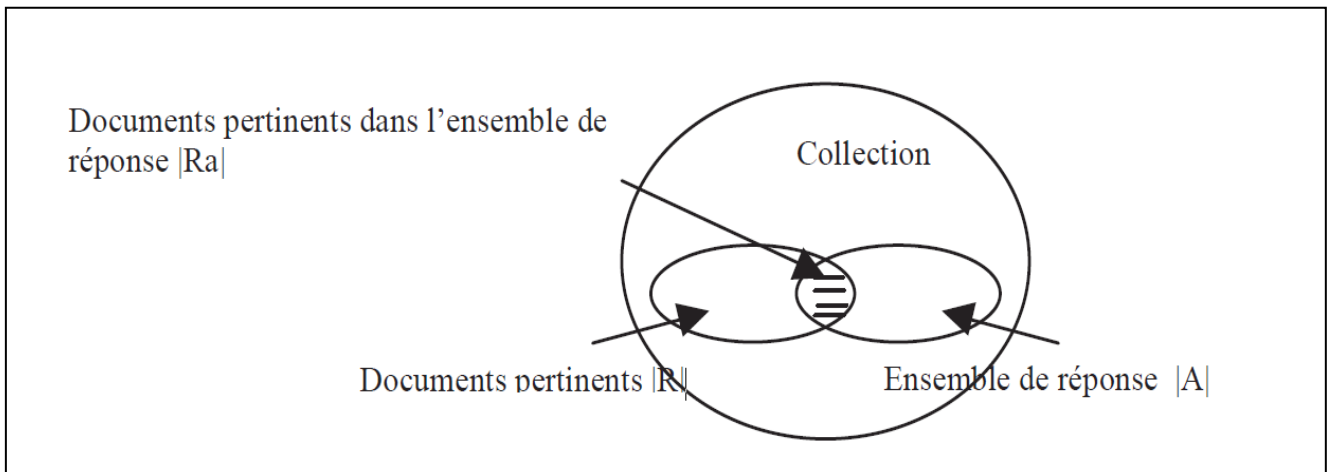
Ainsi, après ce calcul, on peut connaître la similarité de ce pseudo-document (ou la requête) avec tous les autres documents. Ce modèle a montré des performances très intéressantes. Pour un corpus de petite ou moyenne taille, la performance est très supérieure au modèle vectoriel classique, et est un des meilleurs modèles. Quand la taille de corpus augmente, la différence avec les autres modèles classiques semble diminuer.

## VI. Évaluation des SRI

### 1. Rappel et précision

Lorsque nous utilisons un SRI on s'attend un ensemble de réponses (sous forme de documents). À partir de l'ensemble de réponses obtenues mis en regard de l'attente de l'utilisateur, on peut mesurer les performances de notre SRI mis en œuvre pour retrouver des documents. Les critères de mesure des performances sont le **rappel** et la **précision**.

Soient  $|A|$  le nombre de documents renvoyés par un système pour une requête donnée,  $|R|$  le nombre de documents pertinents dans la collection pour cette requête et  $|Ra|$  le nombre de documents pertinents renvoyés par le système [Sauvagnat, 05].



**Figure 1.4 : Précision et Rappel [Sauvagnat, 05]**

L'objectif principal des systèmes de recherche d'information est d'une façon générale de retrouver tous les documents pertinents, et rejeter tous les documents non pertinents. Cet objectif est évalué par les mesures de rappel et de précision (**Figure 1.5**). Le rappel donne donc, une indication sur le nombre de documents pertinents trouvés par rapport au nombre total de documents pertinents pour la requête et la précision donne une indication sur la proportion de documents pertinents renvoyés.

$$rappel = |Ra| / |R| \quad precision = |Ra| / |A|$$

**Figure 1.5 : formule du rappel et de la précision**

## VII. Conclusion

Dans ce chapitre nous avons présenté quelques concepts fondamentaux de la RI, le processus et les modèles les plus utilisés en RI, enfin l'évaluation d'un SRI.

Les documents électroniques existe sur plusieurs forme (texte, son, image, vidéo). Mais cependant la recherche d'informations a été orientée pendant longtemps vers les informations textuelles. La raison fondamentale a été que la majorité des documents étaient composés de texte. C'est pour ça que la recherche d'information textuelle a évolué bien avant que les autres formes de recherche d'information.

Les medias qui nous intéressent dans notre travail sont le texte et l'image, dans le chapitre suivant nous présentons un état de l'art sur la recherche d'images fixes.

# **Chapitre II**

## **Indexation et recherche d'image par le contenu**

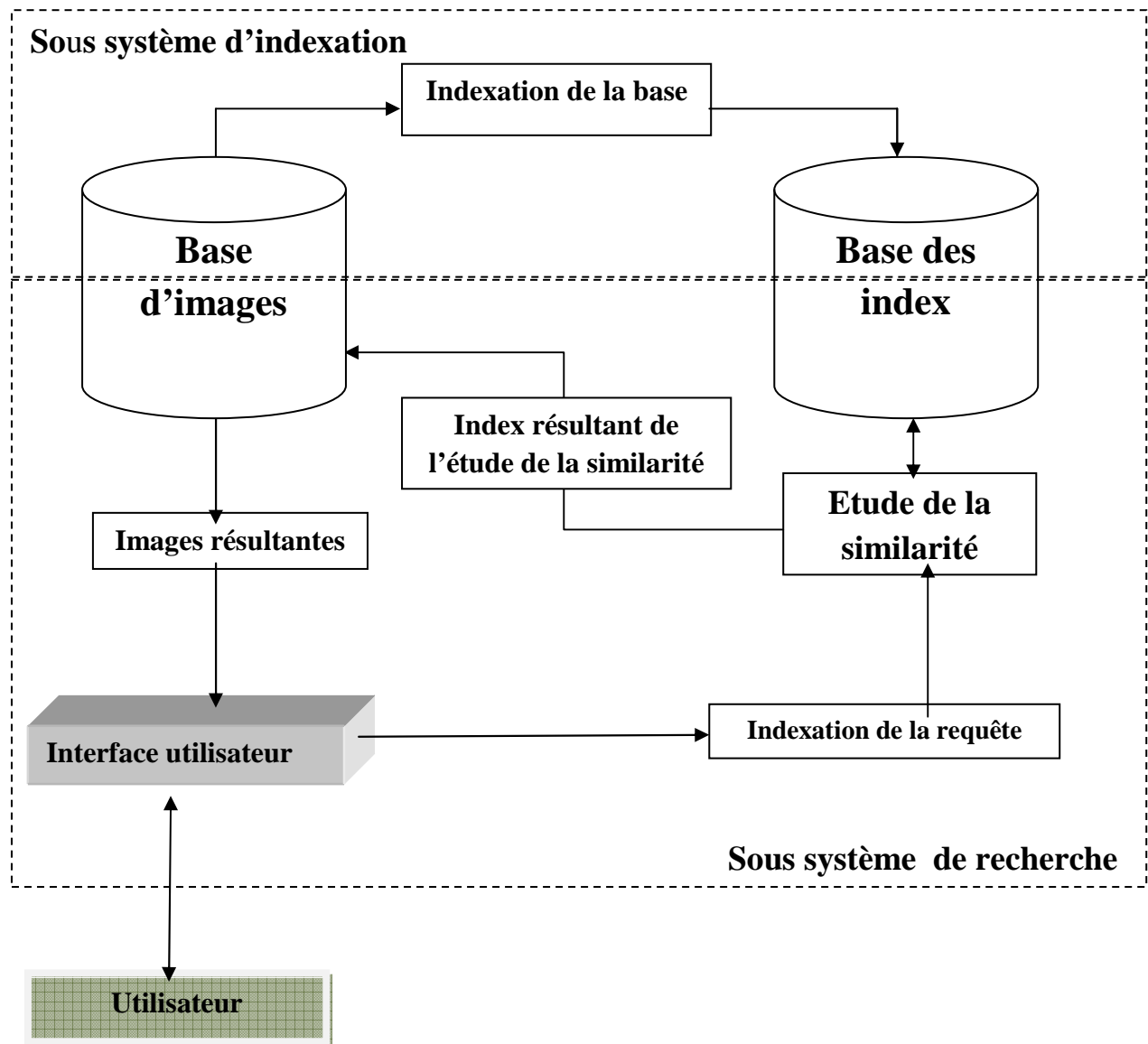
## **I. Introduction**

La recherche d'images par le contenu (CBIR : Content Based Image Retrieval) est un champ de recherche très étudiée à cause de la profusion d'images numériques ainsi que la richesse sémantique de l'image. Les méthodes traditionnelles (basées sur le texte) de la recherche des images se basent sur l'annotation textuelle et manuelle de l'image. Il y a deux inconvénients. Premièrement, le coût élevé de l'annotation manuelle des milliers d'images. Deuxièmement, la richesse sémantique d'une image est difficile à décrire et des gens différents peuvent décrire la même image de différents points de vue (facteur de subjectivité), Pour franchir ces inconvénients, la recherche des images par le contenu (CBIR) est une nouvelle approche qui permet la recherche des images directement et automatiquement basée sur les caractéristiques visuelles du contenu de l'image comme la couleur, la texture, la forme.

Dans ce chapitre nous allons décrire le processus de recherche d'indexation d'images par le contenu, ainsi que les différentes caractéristiques visuelles, et nous terminerons par quelques mesures de similarités.

## II. Architecture du système d'indexation et recherche d'images

La **figure 2.1** présente l'architecture du système de recherche d'images par le contenu.



**Figure 2.1 : L'architecture du système d'indexation et recherche d'images**

Ce système s'exécute en deux étapes : l'étape d'indexation et l'étape de recherche.

Dans l'étape d'indexation, des caractéristiques sont automatiquement extraites à partir de l'image et 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.

Dans l'étape de recherche, le système prend une ou des requêtes à l'utilisateur et lui donne le résultat correspond à 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.

### III. Types de requêtes

D'après [Cox et al, 00], un utilisateur peut avoir différents objectifs lorsqu'il effectue une recherche d'images. Il peut rechercher :

- une image en particulier (et aucune autre). Par exemple, une iconographe d'une agence de presse va rechercher le portrait officiel du président de la République, ou bien un particulier va rechercher la photographie du sphinx qu'il a prise lors de son dernier voyage en Egypte.
- une catégorie d'images. Par exemple, des beaux couchers de soleil.
- une image avec des caractéristiques non-définies. Par exemple, l'utilisateur recherche un fond d'écran avec des couleurs pastel. Dans ce troisième cas, l'utilisateur n'a pas une idée précise de ce qu'il veut. Il parcourt les images sans objectif précis, et peut même changer d'idée au cours de la recherche.

Pour trouver l'information qu'il recherche, l'utilisateur dispose de plusieurs types de moyens :

**Les requêtes par exemple visuel (query by example) :** L'utilisateur fournit un exemple, le système en extrait des descripteurs visuels, puis recherche les images ayant des descripteurs similaires. L'exemple peut être :

- **Une image :** la requête de l'utilisateur est composée d'une image entière.
- **Une région :** l'utilisateur recherche des régions d'images qui ressemblent à la région de l'image sélectionnée par l'utilisateur.
- **Une composition logique de catégories de régions :** l'utilisateur choisit les régions parmi des régions construites à partir d'un résumé (thésaurus photométrique) des régions de la base [Fauqueur, 03].
- **Une composition spatiale de régions :** l'utilisateur recherche une image ayant la région en haut de la même couleur que l'objet A, et la région à droite de la même texture que l'objet B [Ma & Manjunath, 99].
- **Un croquis dessiné par l'utilisateur :** le croquis est un cas particulier de requête par l'exemple, car dans ce cas là, le principal descripteur utilisé est la forme [Schomaker et al, 99].

**Les requêtes directes (direct query) :** Les requêtes directes sont composées de descripteurs visuels. On distingue :

- **Les requêtes par caractéristiques visuelles :** (par exemple, 25% de rouge, 30% de bleu et 45% de vert).

- **les requêtes par combinaison pondérée de caractéristiques visuelles** : (par exemple, la couleur a un poids de 75% et la texture de 25%)

## **IV. Descripteurs visuels**

Puisque l'on souhaite construire des systèmes de recherche d'images qui soient utilisables par l'être humain, et que c'est l'être humain qui donne un sens à ce qu'il voit, il peut être intéressant de s'inspirer du système de perception humaine pour choisir les espaces visuels afin de s'approcher le plus possible de sa compréhension de l'image, et ainsi réduire le fossé sémantique.

### **1. Les descripteurs de couleurs :**

La couleur est très souvent le premier descripteur qui est employé pour la recherche d'images. Plusieurs travaux ont déjà prouvé qu'il s'agit d'un descripteur efficace [Smeulders, 00].

Il existe de nombreuses possibilités d'attributs pour caractériser la couleur : l'histogramme, les moments couleur... Avec la même méthode, si on change l'espace de couleur, il peut donner des informations différentes de l'image.

#### **A. L'histogramme**

La signature la plus courante pour traduire l'information de couleur est l'intersection d'histogramme [Swain&Ballard, 91]. Les histogrammes sont faciles et rapides à calculer, et robustes à la rotation et à la translation. Cependant il y a quatre problèmes en utilisant d'histogrammes pour l'indexation et la recherche d'images [Gong, 98] :

Premièrement, ils sont de grandes tailles, donc par conséquent il est difficile de créer une indexation rapide et efficace.

Deuxièmement, ils ne possèdent pas d'informations spatiales sur les positions des couleurs. Dans certains cas, il y a des images différentes mais ces images ont les mêmes histogrammes.

Troisièmement, ils sont sensibles à de petits changements de luminosité. C'est-à-dire, c'est difficile pour comparer des images similaires dans des conditions différentes.

Quatrièmement, on ne peut pas faire la comparaison partielle des images (objet particulier dans une image), puisque on doit calculer globalement l'histogramme sur toute l'image. Et il y a certaines de travaux pour améliorer ces problèmes.

Il y a deux approches suivies par les chercheurs : l'une ajoute des informations spatiales à l'histogramme, l'autre utilise d'autres espaces de couleurs qui se rapprochent plus, de la perception humaine.

La première approche ajoute des informations spatiales aux histogrammes. Dans [Stricker, 04], les auteurs ont divisé une image en cinq blocs fixés superposés et ils ont extrait les trois premiers moments d'inertie de chaque bloc pour créer un vecteur de descripteurs. [Pass, 96] ont ajouté de la cohérence spatiale dans des histogrammes. Un pixel est cohérent s'il appartient à une région validée par la segmentation et incohérent autrement (soient les pixels situés hors des régions segmentées). Les valeurs d'un histogramme sont divisées en deux

classes en fonction des pixels correspondants à ces valeurs: classe cohérente et classe incohérente. La comparaison entre deux histogrammes devient la comparaison entre les valeurs d'histogrammes dans les classes correspondantes. [Huang Huang, 97] ont proposé le corrélogramme et l'auto-corrélogramme.

## B. Les espaces de couleur

Les espaces de couleurs se basent sur la perception de couleur de l'humain. De nombreux espaces colorimétriques parmi eux :

**L'espace RVB** est très simple à utiliser, car c'est celui employé par de nombreux appareils de capture d'images qui effectuent leurs échanges d'informations uniquement en utilisant les triplets (R, V, B). On parle d'espace colorimétrique orienté matériel. Cependant, ces trois composantes sont fortement corrélées (par exemple, si l'on diminue la composante verte, la teinte paraît plus rouge), l'espace RVB est sensible aux changements d'illumination, et ne correspond pas au processus de perception humaine.

**L'espace XYZ** s'appuie sur des couleurs primaires imaginaires afin d'élargir l'espace colorimétriques, de ne plus avoir à manipuler des coordonnées négatives. Elles sont notées X pour l'axe du rouge et Y pour l'axe du vert puis Z pour l'axe du bleu et sont obtenues par une transformation linéaire, ce qui signifie que le nouvel espace est proportionnel à l'espace RVB d'origine. L'espace XYZ n'est rien d'autre qu'un modèle particulier de RVB.

### ■ Conversion d'une image couleur en niveau de gris :

La conversion d'une image couleur en une image en niveau de gris se fait par suite à l'application de la fonction suivante :

$$\text{Avec :} \quad I(i, j) = \frac{[R(i,j) + V(i,j) + B(i,j)]}{3}$$

$I(i, j)$  : Niveau de gris du pixel situé à la ligne  $i$  et à la colonne  $j$ .

$R(i, j)$  : Intensité de la couleur rouge du pixel  $(i, j)$ .

$V(i, j)$  : Intensité de la couleur vert du pixel  $(i, j)$ .

$B(i, j)$  : Intensité de la couleur bleu du pixel  $(i, j)$ .

La C.I.E (commission international de l'éclairage) propose, de caractériser l'information de luminance (la valeur de gris) d'un pixel par deux formules :

Dans sa recommandation 709, qui concerne les couleurs « vraies » ou naturelles :

$$\text{Gris} = 0.2125 \cdot \text{Rouge} + 0.7154 \cdot \text{Vert} + 0.0721 \cdot \text{Bleu}$$

Dans sa recommandation 601 pour les couleurs non linéaires, c'est-à-dire avec correction du gamma (image vue à partir d'un écran vidéo) :

$$\text{Gris} = 0.299 \cdot \text{Rouge} + 0.587 \cdot \text{Vert} + 0.114 \cdot \text{Bleu}$$

Ces formules rendent compte de la manière dont l'œil humain perçoit les trois composantes, rouge, vert et bleu, de la lumière. Pour chacune d'elles, la somme des trois coefficients vaut 1. On remarquera la forte inégalité entre ceux-ci : une lumière verte apparaît plus claire qu'une lumière rouge, et encore plus qu'une lumière bleue.

**L'espace HSV** (Hue-Saturation-Value) (aussi connu sous le nom de système de cône Hexagonal) sépare les informations relatives à la teinte (Hue), la saturation (Saturation) et l'intensité (Value). Cet espace est plus intuitif à utiliser car il correspond à la façon dont nous percevons les couleurs. La teinte décrit la couleur (rouge, vert...), la saturation décrit l'intensité de la couleur, et la valeur décrit la luminosité de la couleur. La composante H de l'espace HSV offre une certaine invariance. Une étude récente [Lim & Lu, 03] compare six espaces colorimétriques et montre que l'espace HSV est le plus efficace pour la recherche d'images par le contenu, cependant cet espace n'est pas perceptuellement uniforme.

**Autres espaces** sont **LUV** [Wyszeski & Stiles, 82], **HVC** et **Lab**. Ils sont représentés par la distance euclidienne entre deux représentations de couleurs correspondantes aux différences de perception visuelle entre deux couleurs.

### C. Les moments statistiques

La méthode d'histogramme utilise la distribution complète de la couleur. On doit stocker de nombreuses données. Au lieu de calculer la distribution complète, dans les systèmes de recherche d'images, on calcule seulement des dominantes caractéristiques de couleur tel que l'espérance, la variance et d'autres moments. Dans [Stricker, 95] les auteurs ont prouvé que les méthodes utilisées des moments statistiques marchent plus vite et donnent des résultats meilleurs que les méthodes d'histogrammes. L'espérance, la variance, les moments d'ordre 3 peuvent également être calculés sur chaque composante couleur par la formule suivante :

$$E_i = \frac{1}{N} \sum_{j=1}^N p_{ij}$$

$$\delta_i = \left( \frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^2 \right)^{1/2}$$

$$s_i = \left( \frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^3 \right)^{1/3}$$

Où  $E_i$  est l'espérance,  $\delta_i$  est la variance,  $s_i$  est le troisième moment qui correspond à composante couleur  $i$  ( $i$  est R, V, B de l'espace RVB ou T, S, V de l'espace TSV).

## D. Description par couleurs dominantes

Ce descripteur fournit une description compacte des couleurs dominantes de l'image ainsi que leurs proportions [DON, 06]. Il est défini par :

$$F = \{(c_i, p_i, v_i, s)\}, (i = 1, 2, \dots, N)$$

Où N désigne le nombre de couleurs dominantes,  $c_i$  la valeur de la  $i^{\text{ème}}$  couleur dominante,  $p_i$  un poids exprimant sa fréquence relative d'apparition dans l'image et  $v_i$  sa variance. La cohérence spatiale  $s$  représente l'homogénéité spatiale des couleurs dominantes dans l'image. Pour le calcul des couleurs dominantes, il faut choisir un espace de couleurs uniforme au niveau de la perception humaine.

## 2. Les descripteurs de texture

La texture est la répétition d'éléments de base construits à partir de pixels qui respectent un certain ordre. L'aléatoire joue un rôle particulier dans les textures. On peut distinguer deux types extrêmes de textures, entre lesquels se positionnent toutes les textures :

- les textures régulières, Elles peuvent être décrites par des approches fréquentielles (spectres de Fourier, ondelettes de Gabor) ou des approches structurelles dans lesquelles on associe un motif et des règles de placement sur un pavage régulier.
- les textures aléatoires, Elles peuvent être décrites par des lois statistiques (matrices de cooccurrences, matrice de longueurs de plages) sur les distributions.

### A. Matrices de cooccurrences :

Haralick [Haralick, 73] a proposé une méthode en se basant sur les matrices de cooccurrences de niveaux de gris. Elle est probablement une des méthodes les plus célèbres pour analyser la texture. La matrice  $P_{\text{Har } d, \alpha}(g, g')$  compte les nombres de paires de pixel  $(m, n)$  et  $(m', n')$  dans une image qui a une valeur d'intensité  $g$  et  $g'$  avec une distance  $d$  dans une direction  $\alpha$ . Dans [Lan, 99] un exemple a été montré comme suit : Une région de 15 pixels quantifiés sur 8 niveaux de gris.

1	2	1	3	4
2	3	1	2	4
3	3	2	1	1

Tab 2-1 : *Région de 15 pixels*

Matrice de cooccurrence avec  $d = 1$  et  $\alpha = 0$ (un pixel vers le bas)

<b>b</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>a</b>								
<b>0</b>	0	0	0	0	0	0	0	0
<b>1</b>	0	1	2	0	0	0	0	0
<b>2</b>	0	1	0	2	0	0	0	0
<b>3</b>	0	0	1	1	0	0	0	0
<b>4</b>	0	1	0	0	1	0	0	0
<b>5</b>	0	0	0	0	0	0	0	0
<b>6</b>	0	0	0	0	0	0	0	0
<b>7</b>	0	0	0	0	0	0	0	0

Tab 2-2 : *Matrice de cooccurrence avec  $d = 1$  et  $\alpha = 0$ (un pixel vers le bas)*

Afin d'estimer la similarité entre les matrices de cooccurrences, Haralick a proposé 14 caractéristiques statistiques extraites à partir de cette matrice. Actuellement, seulement les quatre caractéristiques les plus appropriées sont largement utilisées : l'énergie, l'entropie, le contraste et le moment inverse de différence.

$$\text{L'énergie (T}_1\text{)} : \sum_i \sum_j P_d^2(i, j)$$

$$\text{L'entropie (T}_2\text{)} : -\sum_i \sum_j P_d(i, j) \log P_d(i, j)$$

$$\text{Le contraste (T}_3\text{)} : \sum_i \sum_j (i - j)^2 P_d(i, j)$$

$$\text{Le moment inverse de différence (T}_4\text{)} : \sum_i \sum_j \frac{P_d(i, j)}{|i - j|^2}, i \neq j$$

## **B. Les filtres de gabor :**

L'un des filtres le plus utilisé est le filtre de Gabor. Ce filtre n'est qu'une fréquence pure modulée par une gaussienne, c'est-à-dire, un filtre passe bande avec une enveloppe gaussienne. Ce filtre est très répandu du fait de sa propriété de résolution optimale conjointe en fréquence et en temps. En plus, des études physiologiques sur les mammifères ont montré qu'on peut assimiler le fonctionnement de certains neurones du cortex visuel à ce type de filtre.

### **3. La forme**

Pour extraire les descripteurs d'une forme, la première chose à faire est de définir sa fonction caractéristique. En général, elle est représentée sous la forme d'un masque dans lequel chaque pixel est représenté par le numéro de la région dans laquelle il appartient. C'est à partir de cette fonction, que sont calculés la plupart des descripteurs de formes, soit à partir de la région entière, soit à partir des contours seulement.

On peut citer quelques descripteurs classiques :

- les moments d'inertie. Ils ont la propriété d'être invariants par rotation. Ils décrivent bien l'allongement de formes régulières comme des ellipses ou des distributions gaussiennes. Ils sont plus ambigus sur des formes complexes.
- la recherche des boîtes englobantes ou minimales de la région.
- les moments invariants (nommés moments de Hilbert), invariants par translation, rotation et changement d'échelle.
- les polygones de Guzman (1968). Cette approche consiste à envelopper l'objet à reconstruire dans des boîtes de formes de plus en plus précisément adaptées ;
- descripteurs de Fourier.

La forme est intéressante pour retrouver certains concepts qui ne peuvent l'être autrement. Prenons le mot ballon, par exemple. Il n'y a pas de couleurs qui puissent caractériser un ballon puisque cet objet peut être de toutes les couleurs, par contre, il a une forme très caractéristique. La forme peut avoir certains avantages également sur la texture. Par exemple, une panthère peut avoir un pelage tacheté ou bien uni. La texture extraite de deux images représentant une panthère pourra donc être très différente, par contre, la forme reste identique. La forme est donc une information discriminante qui peut être utile pour réduire le fossé sémantique.

### **4. Segmentation et points d'intérêt**

L'extraction de descripteurs visuels sur l'image entière (descripteurs globaux) permet de réduire le nombre de calculs nécessaires, la taille de la base de données ainsi que le coût des recherches des images les plus similaires. Cependant, l'approche globale ne permet pas une recherche efficace d'objets (au sens large) dans l'image. A l'inverse, les descripteurs extraits d'une partie de l'image (descripteurs locaux) sont efficaces, mais coûteux. Les descripteurs locaux peuvent être :

- des régions de l'image obtenues soit par segmentation de l'image entière, soit par recherche de régions d'intérêt,
- des points d'intérêt.

## A. Segmentation

La segmentation d'image consiste à séparer en régions homogènes les divers composants visibles dans une image. L'humain sait naturellement séparer des objets dans une image. Pour cela, il se base notamment sur des connaissances de haut niveau qui lui permettent de détecter ce qui l'intéresse dans l'image. En traitement du signal, on caractérise une région comme étant un ensemble de points (pixels) ayant des propriétés communes d'intensité, de texture, de couleur..., qui la différencient des régions voisines. Il y a de nombreuses méthodes de segmentation, on distingue cependant deux grandes familles d'algorithmes :

- la segmentation par approche «contour» ou «frontière» (edge-based segmentation). Un contour est une frontière entre deux milieux différents (2 couleurs, 2 niveaux de gris). On utilise souvent l'image dérivée pour mettre en évidence les variations de niveau. Quand il y a une forte variation, cela est souvent dû à un changement de teinte, d'éclaircissement ou de texture, et de façon générale à un changement de région. Un algorithme classique d'extraction de contours est l'algorithme de Canny [Canny, 86].
- la segmentation par approche «région» (region-based segmentation). Le principe est de trouver les régions en regroupant les pixels ayant des caractéristiques similaires et en séparant ceux qui sont différents (techniques division-fusion, par accroissements de régions, par statistiques bayésienne...).

Une des difficultés de la segmentation est de savoir en combien de régions doit être découpée une image, car certaines images peuvent être segmentée en peu de régions (par exemple, un portrait sur fond uniforme), alors que d'autres nécessitent plus de régions. A quel niveau une segmentation doit-elle s'arrêter?

[Da Rugna & Konik, 04] compare des méthodes classiques de segmentation pour l'indexation d'images (clustering, mean shift, approche morphologique ou multirésolution). Leurs résultats montrent qu'il n'existe pas de segmentation générique ou meilleure que les autres, mais qu'en fonction de la tâche à accomplir certaines méthodes sont plus efficaces que d'autres.



**Figure 2.2: Exemple de segmentation par approche région.**

[Barnard et al, 03] propose une étude comparative des techniques de segmentation pour la reconnaissance d'objets dans les images. Pour cette tâche, il est nécessaire d'obtenir des segmentations qui ne découpent pas un objet en plusieurs régions. Par exemple, dans la figure 2.2, il n'est pas possible de mélanger les parties blanches et noires du pingouin pour obtenir un objet pingouin. Une partie du ciel ou d'un champ peut être séparée en plusieurs régions. Cependant, les régions obtenues à partir de segmentation sur des images décrites par des traits visuels de bas niveaux ne correspondent pas souvent à des objets sémantiques.

Les résultats d'auto-annotation montrent que l'algorithme normalized cuts donne de meilleurs résultats que mean shift et que la segmentation de BLOBWORLD [Datta et al, 05]. C'est pourquoi [Barnard et al, 03] propose d'associer les régions obtenues par segmentation bas niveau avec les mots prédits afin de faire une segmentation prenant en compte des traits bas niveaux et de haut niveaux. Pour cela, il regarde les probabilités a posteriori des mots dans les régions adjacentes : plus la valeur du produit scalaire des deux probabilités de régions adjacentes est grande plus il y a de chance de fusion des deux régions. La comparaison des fusions avec celles réalisées par des êtres humains montrent que cette fusion permet dans une certaine mesure de créer des segmentations plus pertinentes.

Il existe également des approches par segmentation où les régions d'images sont déterminées a priori. Elles supposent par exemple que l'objet intéressant est présent au centre de l'image ou au contraire impose une grille où toutes les régions de l'image ont la même aire. Pour une recherche sémantique d'images, ce type de grille ne nous paraît pas être intéressant, car plusieurs objets peuvent se retrouver dans la même région. De plus, les descripteurs de formes et de textures ne peuvent pas être utilisés.

### **B. Points d'intérêt**

Les points d'intérêt d'une image sont les points qui seront trouvés similaires dans les images similaires. Une manière de les déterminer est de prendre en compte les zones où le signal change. Par exemple, les points d'intérêt peuvent être les coins, les jonctions en T ou les points de fortes variations de texture.

Trois types d'approche pour l'extraction de points d'intérêt :

- approches contours : les contours d'une image sont d'abords détectés, puis les points d'intérêt sont extraits le long des contours en considérant les points de courbures maximales ainsi que les intersections de contours.
- approches intensité : la fonction d'intensité est utilisée pour extraire directement des images les points de discontinuité.
- approche à base de modèle : les points d'intérêt sont identifiés dans l'image par mise en correspondance de la fonction d'intensité avec un modèle théorique.

Les approches intensité sont les plus utilisées, car elles sont indépendantes des contours et du type de points d'intérêt. On peut citer au moins deux avantages des points d'intérêt par rapport aux régions : ils ne nécessitent pas de chaînages pour détecter les contours des régions, et on peut les extraire efficacement de la plupart des images.

## V. Mesures de similarité

Nous allons décrire maintenant des mesures de similarité dans le cadre de la recherche d'images similaires. Ces mesures sont également utilisables pour la recherche de régions similaires, mais aussi dans le cas de la classification supervisée ou non-supervisée d'images.

Il existe un grand nombre de mesures de similarité. Certaines sont des distances, c'est-à-dire des mesures qui ont les propriétés de non-négativité, réflexivité, symétrie et qui respectent l'inégalité triangulaire. Certaines mesures sont spécifiques aux histogrammes ou aux distributions. Dans [Puzicha et al, 99], une étude comparative de 9 mesures est proposée. Nous donnons ci-après quelques unes des mesures classiques.

### 1. Intersection d'histogrammes

Si les images sont indexées par des histogrammes, les distances géométriques s'appliquent. Cependant, il est possible de définir des mesures de similarité propres à cette représentation. Ainsi, l'intersection d'histogramme est une des plus anciennes techniques développées pour la comparaison d'histogrammes. Elle a été proposée par Swain et Ballard [Swain & al. 91] dans le cadre de l'indexation couleur :

$$\text{Inter sec}(I_1, I_2) = \frac{\sum_{i=1}^N \min(I_1(i), I_2(i))}{\sum_{i=1}^N I_2(i)}$$

### 2. Distance euclidienne

Une distance classique est la **L<sub>p</sub> norme** (ou distance de Minkowski) :

$$L_p - \text{norme}(\vec{x}, \vec{y}) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

Où p est le paramètre de la norme. La L1 norme est appelée la distance de Manhattan ou distance City-Block. La L2 norme est la distance euclidienne.

### 3. Distances entre distributions

L'image peut être considérée comme une variable aléatoire dont les vecteurs d'attributs des pixels sont les réalisations. Le problème de mesure de similarité se ramène alors à déterminer

si les réalisations correspondant aux deux images sont issues de la même, distribution de probabilités. Nous parlons alors d'approche statistique de la mesure de similarité. Issue de la théorie de l'information, la divergence de Kullback-Leibler [Puzicha et al, 97] permet de mesurer la dissimilarité (entropie mutuelle) de deux distributions de probabilités :

$$d_{Kull}(I_1, I_2) = K(I_1/I_2) = \sum_{i=1}^N I_1(i) \log \frac{I_1(i)}{I_2(i)}$$

## VI. Conclusion

Dans ce chapitre, nous avons introduit quelques techniques fondamentales de l'indexation et la recherche d'images par le contenu. Les caractéristiques visuelles générales de l'image sont les plus utilisées pour cet effet, comme la couleur, la forme, et la texture. D'un autre côté, Les méthodes de recherche d'images basées seulement sur les descripteurs visuels ne sont pas efficaces car elles présentent encore un certain nombre de faiblesses, notamment le fossé sémantique qui existe entre les propriétés visuelles extraites et les concepts utilisateurs.

Dans le prochain chapitre nous allons présenter un autre type de processus de recherche d'image qui permettra d'associer le contenu visuel de l'image au texte qui peut entourer l'image.

# **Chapitre III**

## **Indexation et recherche d'images en combinant le texte et le contenu**

## I. Introduction

Ce chapitre se concentre sur la recherche d'images basées sur le texte (Text-Based Image Retrieval) ou TBIR et la recherche d'images basées sur le contenu visuel (Content Based Image Retrieval) ou CBIR. Et la manière pour combiner les deux systèmes de recherche d'images basées sur le texte et le contenu.

Premièrement, nous allons décrire le système de recherche d'images basé sur le contenu, et puis, nous allons décrire le système de recherche d'images basées sur le texte. Troisièmement, nous allons utiliser parallèlement les deux approches.

## II. Système de recherche d'image par le contenu

Les systèmes de recherche d'images par le contenu permettent de rechercher les images d'une base d'images en fonction de leurs caractéristiques visuelles, ce que nous décrirons dans ce qui suit.

### 1. Les descripteurs utilisés :

Comme nous l'avons vu dans le chapitre précédent, la comparaison directe entre les images n'est pas envisageable, d'où la nécessité d'en extraire des descripteurs visuels.

Dans notre travail, nous sommes centrés sur les descripteurs de couleur qui sont les couleurs dominantes et les moments statistiques dans l'espace de couleur RVB.

#### A. Les moments statistiques

C'est une approche qui a été utilisée avec succès sur plusieurs systèmes de recherche d'information, elle consiste comme nous l'avons cité dans le chapitre précédent à calculer trois valeurs qui sont : l'espérance  $\mu_i$  qui représente la couleur moyenne de l'image.  $\sigma_i$  La variance qui représente le contraste d'une image, plus la variance est grande plus l'image est contrastée. Le moment de troisième ordre  $s_i$  caractérise la quantité de lumière dans une image. Où  $i$  représente une valeur du canal (RVB).

La distance entre une image requête  $I$  et une image de la base  $H$  en utilisant les moments statistiques de couleur est la suivante :

$$Dist_{Mom} = \left( \sum \left| (\mu_i(I) - \mu_i(H)) + (\sigma_i(I) - \sigma_i(H)) + (s_i(I) - s_i(H)) \right|^2 \right)^{1/2}$$

#### B. Les couleurs dominantes

Pour chaque image de la base d'images, nous extrayons les couleurs dominantes et nous les assignons comme les signatures d'une image. Dans notre travail, nous avons choisi de trouver les couleurs dominantes de 12 couleurs candidates : rouge, orange, jaune, vert, cyan, bleu, violet, magenta, rose, noir, gris, blanc.

La sélection de ses couleurs dominantes se base sur la perception et la représentativité des couleurs par l'homme. En effet ces couleurs ont été sélectionnées de par la facilité qu'ont les utilisateurs à les différencier, et en même temps ce sont les plus représentatives d'une image naturelle. De plus le nombre de couleurs ne doit pas être trop important, pour gagner en espace mémoire.

Pour chaque image couleur nous calculons la fréquence d'une couleur en correspondant les couleurs de l'image aux couleurs sélectionnées avec 10% d'erreur.

Dans notre représentation de la couleur, chaque couleur  $\{C_i\}$  à un pourcentage de couleur  $\{P_i\}$  qui lui correspond ou  $i$  représente le nombre de couleurs de la requête. La similarité entre une ou plusieurs couleurs requête  $I$  et une image de la base  $H$  en utilisant la formule suivante :

$$SIM_{COL} = \sum | P_i (I) - P_i (H) |$$

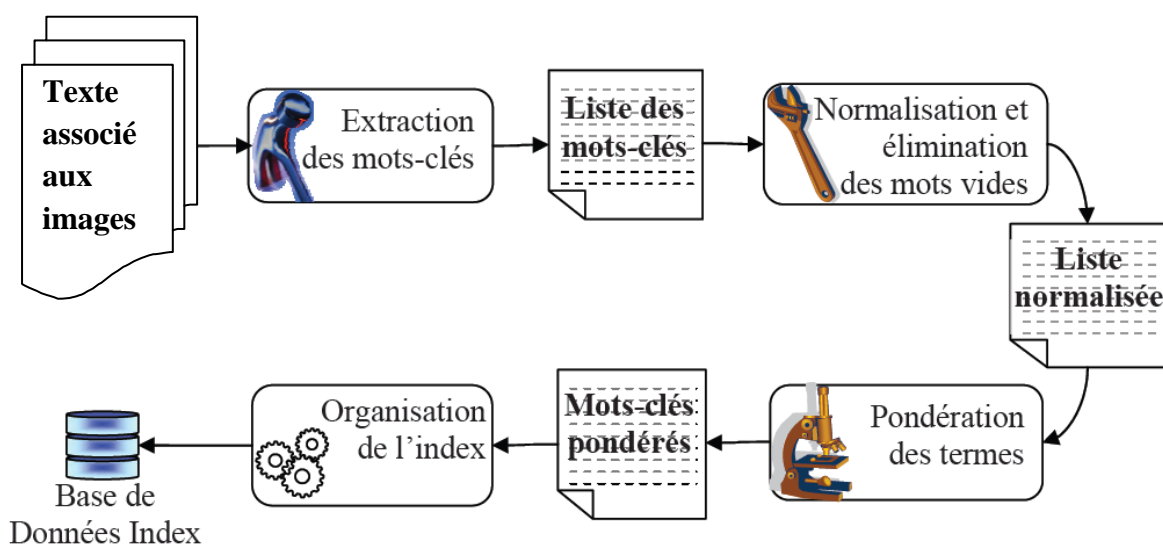
### III. Système de recherche d'image par le texte :

Pour la recherche d'images par le texte, nous utilisons une approche d'indexation et de recherche classique, du fait que la recherche d'image ici est basée uniquement sur le texte.

#### 1. Processus d'indexation :

Dans le premier chapitre nous avons décrit les différentes étapes du processus d'indexation par le texte qui sont : Extraction des mots-clés, Normalisation et mots-clés élimination des mots vides, Pondération des termes, que nous avons schématisés dans la **Figure 3.1**.

Quant à la structure du fichier index nous avons choisis le fichier inversé (inverted file) que nous décrivons plus bas.



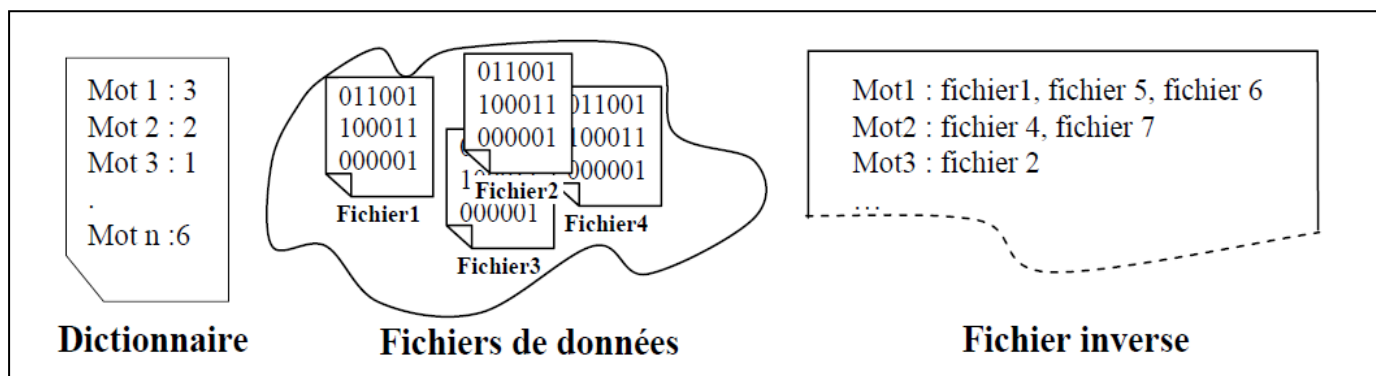
**Figure 3.1 : Étapes du Processus d'indexation**

## 2. Les fichiers inversés (Inverted files) :

La structure de fichier inversé est à la base de tous les systèmes de recherche d'information. Un système à base de fichiers inversés contient trois composants principaux :

- **Un dictionnaire** : Le fichier dictionnaire contient tous les mots ou groupes nominaux spécifiques pouvant servir de mots-clés pour l'indexation et la recherche dans l'ensemble des fichiers à traiter. A chaque entrée du dictionnaire est associé le nombre de fois où l'entrée apparaît dans l'ensemble documentaire.
- **Un fichier de hachage** : Ce fichier contient pour chaque entrée du dictionnaire une liste décrivant dans quel fichier apparaît cette entrée. Cette méthode permet de restreindre l'étude sur les fichiers qui nous intéressent et pas les autres. A signaler que dans certains cas, la position dans le fichier est aussi stockée.
- **Les fichiers de données** : Qui représentent les documents du corpus à indexer.

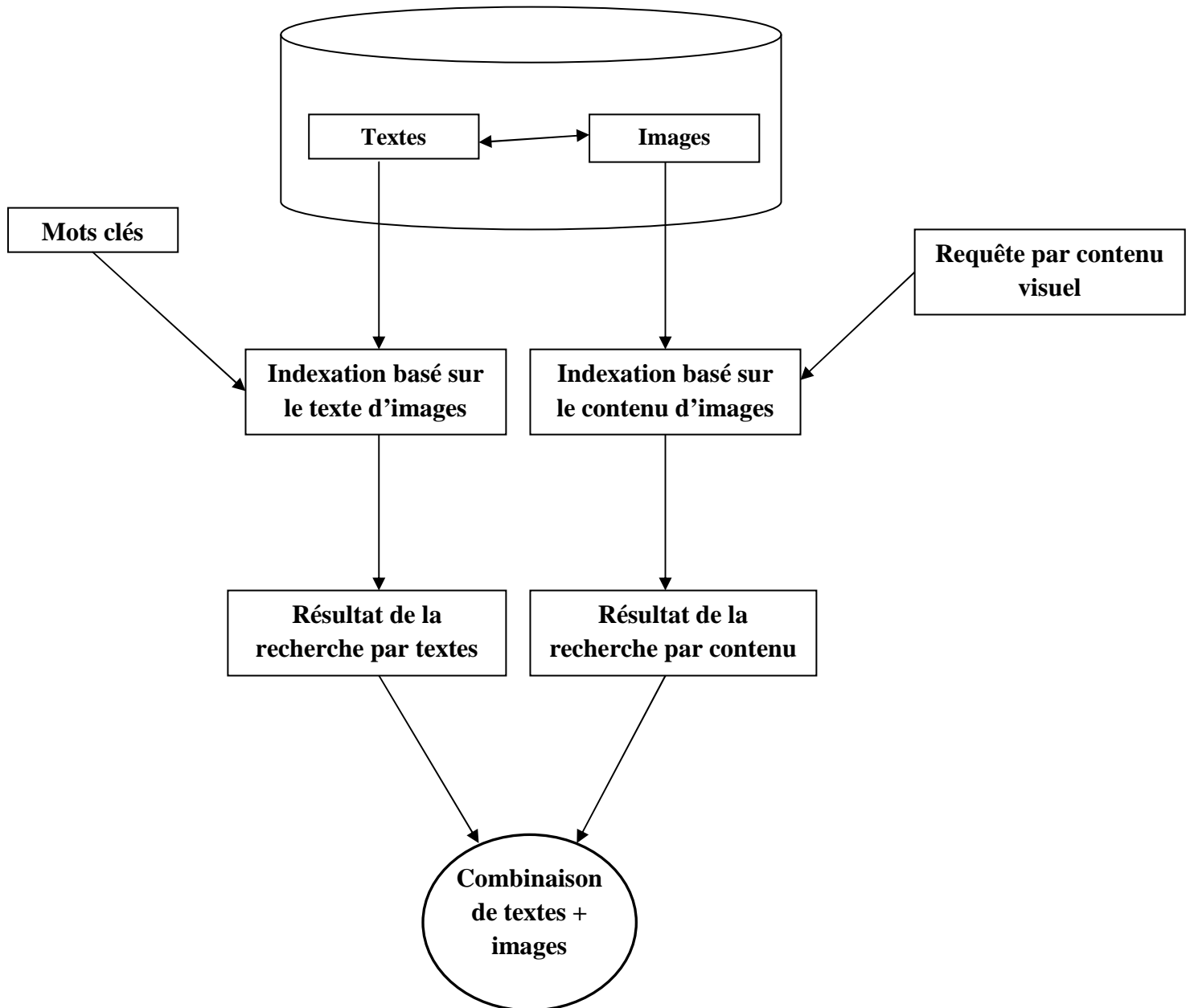
La **Figure 3.2** est une représentation d'un système à base de fichiers inversés.



**Figure 3.2 : Système à base de fichiers inversés [Dahak, 05]**

## IV. Système de recherche par le texte et l'image

Dans le système que nous avons construis, une requête peut être lancée avec trois méthodes comme décrit dans la **figure 3.3** :



**Figure 3.3 : Système recherche par le texte et l'image**

- L'utilisateur peut formuler une requête sous forme d'image exemple. Dans ce cas le moteur de recherche va calculer les similarités entre l'image exemplaire et les images dans la base d'images en utilisant la fonction de similarité et Il donne une liste ordonnée des résultats.
- L'utilisateur peut formuler une requête sous forme de texte, le moteur de recherche va calculer les similarités entre les mots clés apparaissant dans le texte et le texte qui entoure l'image. et Il donne une liste ordonnée des résultats.
- L'utilisateur peut formuler une requête sous forme de texte et par une ou plusieurs couleurs que nous avons décrits précédemment, le moteur de recherche va combiner les deux méthodes, et va mélanger les deux listes de résultats. et la réarranger en utilisant une méthode qui est détaillé plus bas.  
Enfin, le moteur de recherche donne les résultats à l'utilisateur par une liste d'images ordonnées.

## V. La méthode combinant textes et images

Pour une meilleure performance de la recherche, nous présentons une méthode qui combine le texte et le contenu d'images. Ce système utilise le système de reclassement afin de combiner les résultats de recherche des deux moteurs. Étant donné une requête basée sur du texte et le contenu, les deux moteurs de recherche vont retourner un ensemble de résultats images. Ensuite en combinant les deux ensembles de résultats avec leurs poids, nous allons retourner un résultat commun [HO Vi Dai & al, 09].

$$\text{Score}(Q,D) = W1 * \text{Score}(Q_{\text{texte}}, D_{\text{texte}}) + W2 * \text{Score}(Q_{\text{visuel}}, D_{\text{visuel}})$$

**W1** et **W2** sont des paramètres de réglages.

## **VI. Conclusion**

Dans ce chapitre nous avons présenté notre système de recherche d'image par le contenu, ainsi que notre système de recherche et d'indexation par les media texte et image. Et nous avons présenté une méthode qui permet de combinait les deux.

L'approche que nous avons présenté permet de combinait les scores de la recherche d'images basées sur le texte et le contenu pour amélioré le résultat final.

Dans le prochain chapitre nous avons présenté l'implémentions des différents module qui participes à l'indexation ainsi que nous faisant une petite expérimentation de notre système de recherche et d'indexation par le contenu et le texte.

# **Chapitre VI**

## **Implémentation et évaluation expérimentale**

# **I. Introduction**

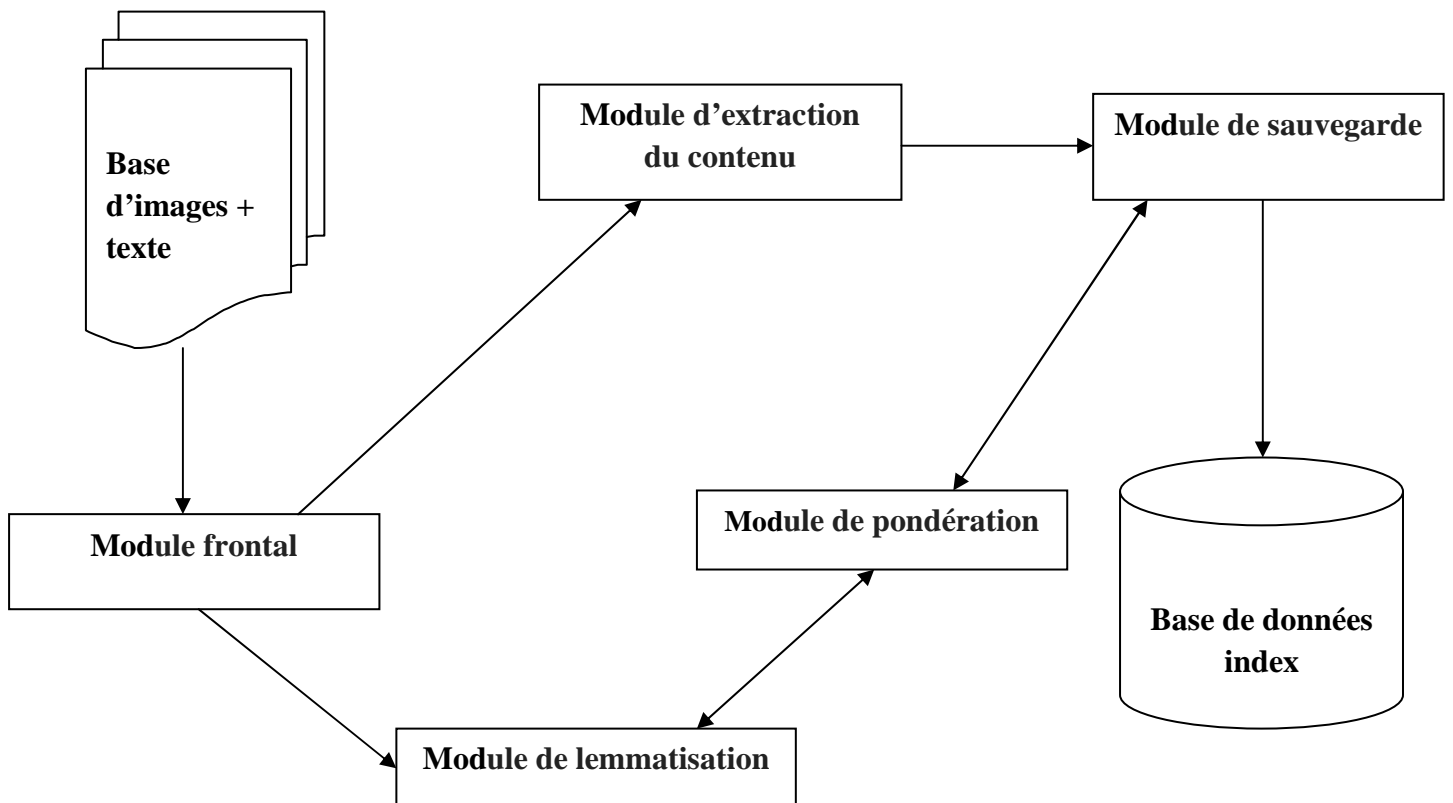
Dans ce chapitre il sera question des étapes suivi pour l'implémentation de nos différents modules d'indexation ainsi que du processus de recherche dans ses derniers.

Comme déjà décrit dans le chapitre III le processus d'indexation et de recherche d'images par le contenu et le texte ce fait par deux étapes la première est l'étape d'indexation du contenu et du texte. Ce qui donne naissance à la deuxième étape qui est le processus de recherche qui peut ce faire soit : Par image exemple (query by exemple), par le choix d'une ou plusieurs couleurs dominantes par l'intermédiaire d'une palette de couleurs, par le texte, par le texte et une ou plusieurs couleurs dominantes. Dans ce qui suit nous allons décrire avec plus de détail les deux étapes.

## II. Architecture du processus d'indexation et de recherche par le contenu et le texte :

### 1. Architecture du processus d'indexation :

Nous présentons dans un premier temps l'architecture de notre système d'indexation comme il est démontré par **Figure 4.1**. Ensuite l'implémentation des différents modules le constituant.



**Figure 4.1 : Architecture du processus d'indexation**

## 2. Présentation des différents modules :

**Module frontal :** ce module permet de récupérer la liste de fichiers d'un répertoire ainsi que les titres de chaque fichier.

**Module de lemmatisation :** le module de lemmatisation élimine les mots vides et procèdent au stemming des mots clés.

**Module de pondération :** le module de pondération procède à la pondération tf-idf de la liste des mots clés normalisés.

**Module d'extraction du contenu :** permet d'extraire le contenu visuel que ça soit les moments statistiques, ou le calcul de la fréquence des couleurs dans l'image.

**Module de sauvegarde :** ce module interagit avec la base de données index et sauvegarde la liste des termes pondérés, et les descripteurs visuel, ainsi que les fichiers qui leurs sont associés.

## 3. Processus de recherche

Le processus de recherche peut ce faire de trois façons différentes :

### A. Requête exemple :

L'utilisateur formule sa requête sous forme d'image a partir de celle-ci sont extraie le contenu visuel qui sont les moments statistiques, par la suite un module de calcul de similarité va calculer la distance entre l'image requête et la base de donnée index.

### B. Requête texte :

L'utilisateur formule sa requête sous forme de texte celle-ci passe par les étapes d'indexation qui sont la lemmatisation et la pondération par la suite un module d'apparement calcul la le score d apparement document requêté.

### C. Requête par une ou plusieurs couleurs:

L'utilisateur formule sa requête sous forme de texte et choisi une ou plusieurs couleurs parmi une palette de 12 couleur qu'on a site dans le chapitre III.

### III. Expérimentations de la recherche par image exemple :

#### I. Les bases d'images utilisées

Pour évaluer et valider notre système, nous avons utilisé deux bases d'images. Ces bases d'images sont disponibles sur Internet librement, et elles possèdent déjà des classes définies où chaque image n'appartient qu'à une seule classe.

##### A. La base de Wang

La base d'images de Wang est un sous-ensemble de la base d'images Corel. Cette base d'images contient 1000 images naturelles en couleurs. Ces images ont été divisées en 10 classes, chaque classe contient 100 images. L'avantage de cette base est de pouvoir évaluer les résultats. Cette base d'images a été utilisée pour faire des expériences de classification. Un exemple de chaque classe peut être vu sur la figure 4.2 [Deselaers, 03]. Cette base d'images a été créée par le groupe du professeur Wang de l'université Pennsylvania State et est disponible à l'adresse : <http://wang.ist.psu.edu/>. Chaque image dans cette base d'images a une taille de  $384 \times 256$  pixels ou  $256 \times 384$  pixels.

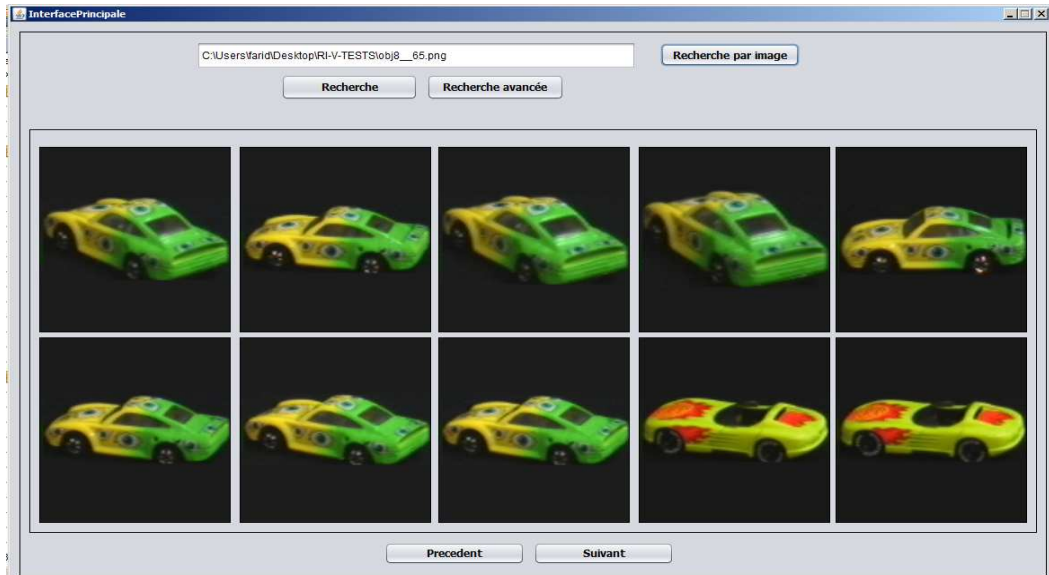


Figure 4.2 : base d'images de Wang

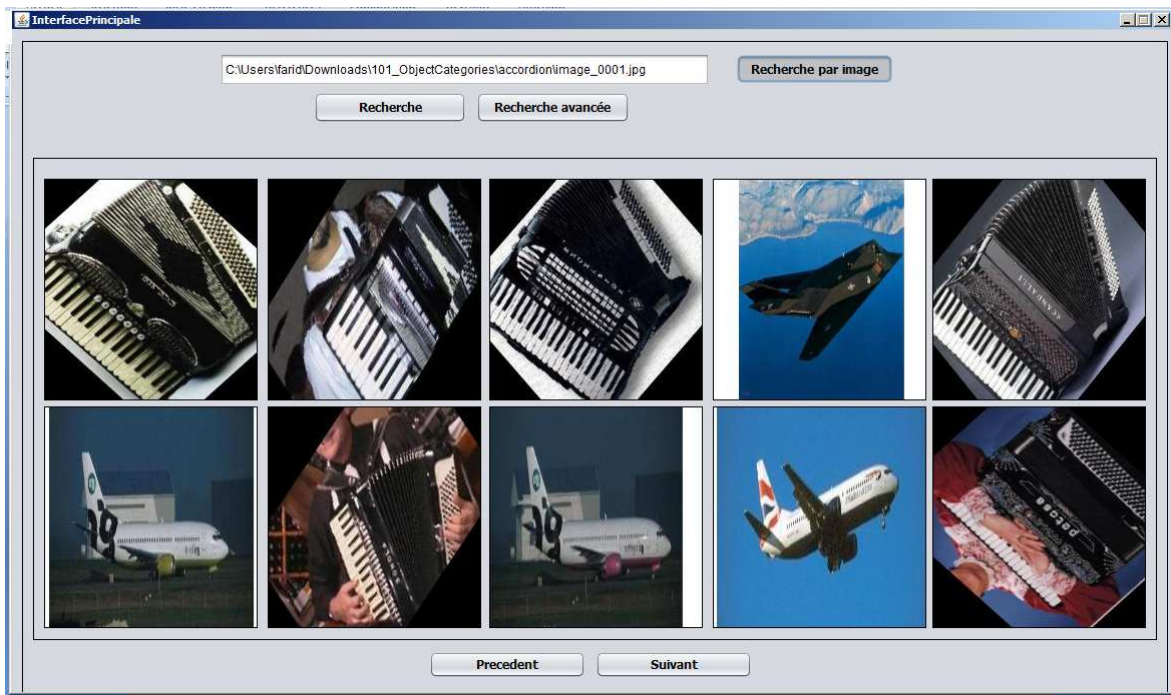
##### B. Coil (Columbia Object Image Library)

Cette base d'images est très connue pour la reconnaissance des objets. Il y a deux bases d'images COIL : COIL-20 qui contient des images en niveaux de gris prises à partir de 20 objets différents et COIL-100 qui contient des images en couleurs prises à partir de 100 objets différents. Les deux bases d'images consistent en des images prises à partir des objets 3D avec des positions différentes. La base COIL-100 a 7200 images en couleurs (100 objets x 72 images/objet). Chaque image a une taille de  $128 \times 128$  pixels. Tous les objets utilisés dans COIL-100 sont montrés dans la figure 4.1 [Deselaers, 03]. La base COIL-20 à 1440 images en niveaux de gris (20 objets x 72 images/objet). Chaque image a une taille  $128 \times 128$  pixels. Tous les objets utilisés dans COIL-20 sont dépeints sur la figure 4.3 [Deselaers, 2003]. Ces bases





**Figure4.5 : Quelques résultats avec la base de Coil**



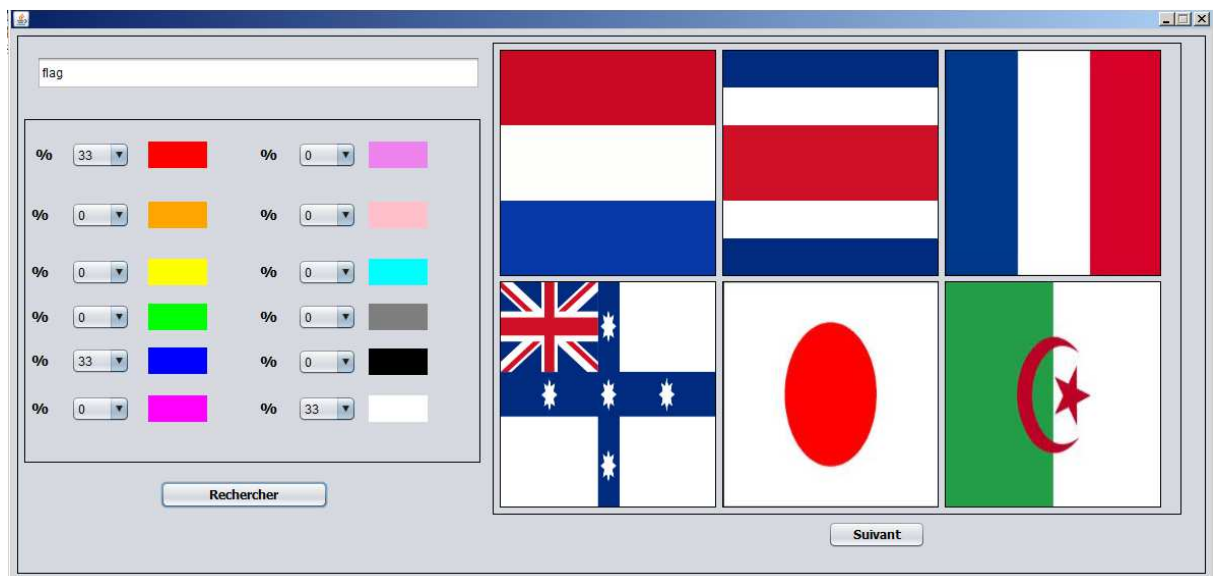
**Figure4.5 : Quelques résultats avec la base de Wang**

	Base d'images de Coil	Base d'images de Wang
<b>Nombre d'images pertinentes dans images</b>	<b>18</b>	<b>14</b>
<b>Précision</b> ( $ Ra / A $ ) $ A =20$	<b>90%</b>	<b>70%</b>
<b>Rappel (<math> Ra / R </math>)</b> $ R =72$ pour la base de Coil $ R =100$ pour la base de Wang	<b>25%</b>	<b>14%</b>

**Tableau 4.1: Les valeurs de rappel et précision des bases des Wang et Coil, avec un nombre d'images retrouvées est de 20.**

#### **IV. Illustrations de la recherche par couleurs et/ou texte:**

Dans le but d'illustrer notre recherche par couleur nous avons choisi une base d'images composées de 100 drapeaux, la figure 4.6 présente quelques résultats obtenu :



**Figure 4.6 : Quelques résultats avec la recherche basée sur la couleur et le texte**

## **IV. Conclusion**

Nous avons présenté dans ce chapitre l'implémentation et l'évaluation de notre recherche par le contenu et/ou bien par texte. Nous avons illustré notre recherche en utilisant des bases d'image disponible sur internet ainsi qu'une base d'image qui a été constituer par nos soins, dans le but est de montré les possibilités d'une association du texte d'image au contenu visuel que nous avons représenté par la couleur dans notre travail.

# Conclusion générale

Dans cette étude, nous nous sommes proposé de mesurer dans quelle proportion l'utilisation conjointe du texte et des images améliorent les résultats des systèmes pour la tâche de recherche de documents présentant ces deux modalités. Nous avons de plus voulu répondre à la question de savoir dans quelle mesure la méthode d'indexation par texte et contenu apporte un gain significatif lors de la fusion de ces deux formes d'information.

La première étape consiste à donner une représentation du contenu textuel, dans ce but nous avons utilisé un processus d'indexation que nous avons mis en œuvre, puis nous avons procédé à l'indexation du contenu visuel, à cet effet nous avons décrit plusieurs descripteurs de couleur, texture, forme. Notre choix s'est porté sur deux descripteurs de couleur qui sont, les moments statistiques et les couleurs dominantes, nous avons utilisé les moments statistiques pour la recherche par image, contre aux couleurs, leurs utilisations furent dans la construction d'une palette de couleurs qui permet de choisir une ou plusieurs couleurs dans le but d'améliorer la recherche textuelle. En effet notre système de recherche permet de combiner les résultats de la recherche textuelle et par couleurs, pour permettre une recherche plus précise.

À l'issue des travaux menés dans le cadre de ce mémoire, nous dégagons quelques perspectives qui peuvent être envisagées. La première et que nous envisageons de enrichir notre système en d'autres caractéristiques visuelles et d'autres espaces de couleurs.

De plus un système de bouclage de pertinence semble une solution intéressante pour réduire le fossé sémantique. Mais aussi prendre en compte l'aspect sémantique dans la recherche textuelle pour améliorer les résultats de la recherche textuelle. Enfin nous pouvons également étudier et tester différentes mesures de similarités afin de les comparer et déduire celles possédant les meilleures performances.

# **Annexe**

## **Systèmes de recherche d'images combinant texte et images**

## **I. Quelques systèmes de recherche d'images**

Nous allons maintenant décrire quelques systèmes de recherche d'images. Nous avons choisi ces systèmes d'abord en fonction de leur popularité, puis en fonction de leur capacité à combiner texte et images. Bien entendu il existe de très nombreux autres systèmes, mais nous ne pouvons pas tous les citer. Les systèmes sont ordonnés en fonction de la date du premier article conséquent parlant de ce système. Cette date ne traduit pas l'année du début du projet (en général, elle se situe deux ou trois années avant), mais elle donne une idée de l'ordre dans lequel les systèmes ont été développés. Cet ordre est important, car les anciens systèmes informatiques n'ont pas les mêmes capacités en calcul et en mémoire que les systèmes actuels.

Avant de rentrer dans le détail de quelques uns de ces systèmes, notons que certains sont des systèmes dit «académiques», que d'autres sont ou sont devenus des systèmes commerciaux (QBIC, IMAGESEEKER, VIRAGE [Bach et al, 96], que d'autres encore peuvent être vus comme des systèmes proposant des outils d'extraction et de manipulation d'informations visuelles (GIFT, PICSOM, PHOTOBOOK...). Notons aussi que nous nous intéresserons surtout à savoir s'ils utilisent des informations textuelles, et si oui, comment ils les utilisent. De plus, nous indiquons dans le tableau 2.1 si le système propose une recherche et une indexation textuelle (colonne W), si le système combine les informations textuelles et visuelles (colonne &), et s'il utilise le bouclage de pertinence (colonne BP). Pour remplir les colonnes, nous avons utilisé les informations dans les articles cités en référence, il se peut que dans des travaux ultérieurs certaines fonctions aient été ajoutées.

L'un des premiers systèmes de recherche d'image par le contenu visuel est le système QBIC (Query By Image Content) [Flickner et al, 95]. QBIC utilise des traits de couleur, de texture et de forme. Il permet des requêtes par image-exemple, directes ou par croquis. Il permet également de faire de la navigation avec bouclage de pertinence (BP). Le texte associé aux images, comme le nom de l'auteur ou du média, est seulement utilisé pour faire un filtrage des résultats (par exemple, seules les images associées à tel auteur sont affichées), et non pas pour améliorer sémantiquement les résultats.

### **1. Le système CHABOT**

Le système CHABOT [Ogle & Stonebraker, 95] propose de stocker les histogrammes de couleur et le texte associé dans une base de données relationnelle. Il permet de réaliser des recherches telles que : «rechercher des images un peu rouges réalisées au bord du lac Tahoe». Ces recherches sont exprimées sous la forme de requêtes dans la base relationnelle. Il ne fusionne donc pas réellement les informations textuelles et visuelles.

### **2. Le système PICTION**

Le système PICTION [Srihari & Burhans, 94] semble être le premier système à vraiment proposer de combiner texte et information visuelle. Le texte est extrait de la légende associée aux images d'articles de presse. Une hiérarchie visuelle combinant à la fois des informations de structures textuelles extraites de WordNet [Fellbaum, 98] et à la fois des informations de couleur, de texture et de forme est également proposée.

### **3. Le système WEBSEER**

Le système WEBSEER [Frankel et al, 96] est un système construit pour rechercher des images sur le web. Les images sont classées en deux classes «photographies» et «dessins» à l'aide d'arbres de décision sur les descripteurs de couleur. Le système utilise le texte des pages web afin d'annoter automatiquement les images. Des méta-données contenues dans l'en-tête des images telles que le type de fichier, la date, la taille de l'image... sont également prises en compte. Ce système permet d'effectuer des recherches du type : «rechercher des photographies de portrait de Rebecca De Mornay ayant telles caractéristiques de couleur». Les informations visuelles et les informations textuelles ne sont donc pas vraiment combinées.

### **4. Le système VISUALSEEK**

Le système VISUALSEEK [Smith & Chang, 96] a la particularité de rechercher des régions d'intérêt, et de prendre en compte leurs positions ainsi que leur taille lors des requêtes. Il n'utilise pas d'information textuelle.

### **5. PHOTOBOOK**

PHOTOBOOK [Pentland et al, 96] recherche les images en comparant leur descripteurs spécialement projetés dans un espace adapté au type de recherche (reconnaissance de visages, de formes, de textures...) effectuée. Il permet de faire des recherches du type : «montre-moi des images qui ont la même apparence que cette image». Dans la version de l'article [Pentland et al, 96], le système PHOTOBOOK propose d'utiliser le texte associé aux images pour permettre des requêtes du type : «montre moi des images qui sont annotées de façons similaires à cette image, mais prise à Boston». Les informations visuelles et les informations textuelles ne sont donc pas vraiment combinées. Cependant, PHOTOBOOK propose un module appelé FOUREYES [Minka & Picard, 96] qui est capable d'apprendre des concepts au fur et à mesure de l'interaction avec l'utilisateur. Les images sont segmentées en 16X16 blocs. Avant l'intervention de l'utilisateur, le système précalcule des similarités entre blocs d'images et les regroupe. Puis l'utilisateur annote certains des blocs avec des mots tels que building. Cette technique est très coûteuse car elle nécessite l'annotation manuelle de régions d'images.

### **6. Le système MARS**

Le système MARS (Multimedia Analysis and Retrieval System) [Rui, 97] est un des premiers systèmes de recherche d'images par le contenu à utiliser le bouclage de pertinence. Les images sont segmentées en blocs par une grille de 5X5. De chaque bloc, sont ensuite extraits des descripteurs de couleur et de texture (ondelettes). Le système propose de retrouver et d'ordonner les résultats à l'aide d'un modèle booléen. Il n'utilise pas d'information textuelle.

### **7. Le système IMAGEROVER**

Le système IMAGEROVER [La Cascia et al, 98] a pour objectif de combiner les informations textuelles d'une page web avec le contenu visuel des images. Dans ce système, chaque image est indexée par un vecteur global concaténant les vecteurs visuels (réduits par ACP) et textuels (réduits dans l'espace latent).

## **8. Le système NETRA**

Le système NETRA [Ma & Manjunath, 99] utilise des descripteurs de couleur, de texture, de forme, une segmentation en régions et les relations spatiales entre ces régions. La localisation spatiale d'une région est mesurée par deux rectangles (intérieur et extérieur). Il n'utilise pas d'information textuelle.

## **9. Le système BLOBWORLD**

Le système BLOBWORLD [Carson et al, 99] propose de segmenter les images en regroupant les pixels qui ont des couleurs et des textures similaires à l'aide d'un algorithme de clustering basé sur l'estimation de distributions gaussiennes par EM (voir annexe B.2.4 page □). Ce type de segmentation doit permettre de retrouver plus efficacement les objets dans les images. Des expériences sont réalisées sur le corpus Corel et montrent que retrouver par blobs similaires est plus efficace que retrouver les images ayant des histogrammes de couleurs et de textures similaires. Au niveau de l'interface, BLOBWORLD permet à l'utilisateur de sélectionner la région qui l'intéresse, et de retrouver les images qui possèdent des blobs similaires. L'information textuelle ajoutée aux images de Corel est utilisée pour vérifier les catégories sémantiques lors des tests et non pas pour améliorer l'efficacité du système.

## **10. Le système PICHUNTER**

Le système PICHUNTER [Cox et al, 00] est un système qui utilise les probabilités bayésiennes pour prédire quel est l'objectif des utilisateurs en fonction de leurs actions. Pour la phase de bouclage de pertinence, l'algorithme essaye de maximiser l'information (au sens de l'entropie) obtenu par l'utilisateur à chaque itération. Les résultats obtenus sont comparés à ceux obtenus par des expériences de psychovision. Le système utilise des annotations «cachées» pour améliorer les résultats. Le principe est que le texte associé aux images est utilisé pour améliorer les résultats renvoyés, mais il n'est pas proposé à l'utilisateur de faire des recherches textuelles.

## **11. Le SIMPLICITY**

Dans SIMPLICITY (Semantics-sensitive Integrated Matching for Picture Libraries) [Wang et al, 01], la requête est un ensemble de régions et la distance entre deux ensembles est la somme des distances entre les régions appariées, pondérées par un score de satisfaction attribué à l'appariement. La distance ne tient pas compte de l'information spatiale, mais la méthode permet la mise en correspondance d'une région d'une image à plusieurs régions de l'autre image. Ce système n'utilise pas d'information textuelle. Par contre, il possède un module appelé A-LIP (voir ci-après) qui permet de faire de l'auto-annotation d'images.

## **12. Le système IKONA**

Le système IKONA [Boujemaa et al, 01] fait suite au système SURFIMAGE [Nastar et al, 98] du groupe IMEDIA d'unité INRIA-Rocquencourt. Ce système propose de rechercher des images par similarité visuelle (couleur, texture, forme). Il propose également des recherches par sélection de régions, et d'utiliser le bouclage de pertinence. Dans [Boujemaa et al, 01], la

combinaison avec le texte est mis dans les perspectives d'évolution du système. Cela a notamment été le cas dans la thèse de Marin Ferecatu [Ferecatu, 05]. Il part du principe que si une image a été annotée par des mots-clés, le système peut utiliser ces mots-clés pour une recherche plus rapide. Cependant, comme les mots-clés peuvent être ambigus, ils proposent de propager les mots-clés à l'aide d'informations entre les termes extraites de WordNet [Fellbaum, 1998]. Plusieurs combinaisons des informations textuelles et visuelles sont comparées en association à des phases de bouclages de pertinence.

### **13.Le système PICSOM**

Le système PICSOM [Laaksonen et al, 02] utilise des cartes auto-organisatrices (Self-Organising Map) [Kohonen et al, 00] pour classer les images. Il fournit en plus un grand nombre d'outils pour la classification et la recherche d'images par le contenu. Ce système n'a donc pas pour vocation de combiner information textuelle et visuelle. Notons cependant que dans [Viitaniemi & Laaksonen, 05], ce système est utilisé pour l'auto-annotation d'images. Pour cela, pour chaque mot, le système construit deux ensembles d'images, les images annotées par ce mot et les autres, et les utilisent pour apprendre à classifier. Les images de test sont ensuite annotées en fonction de leur classement dans la classe pertinent ou non pertinent pour un mot donné.

### **14.Le système A-LIP**

Le système A-LIP (Automatic Linguistic Indexing of Pictures) [Li & Wang, 03] est un module du système SIMPLICITY. Il propose d'annoter les images à l'aide de modèles de Markov cachés. A partir des descripteurs des images d'apprentissage correspondant à un certain concept, le système construit des modèles des images sur plusieurs niveaux de résolution. Les modèles ainsi construits pour chaque concept forment un dictionnaire de concepts. Lorsqu'une nouvelle image doit être annotée, la vraisemblance entre les descripteurs (multirésolution) de l'image et le modèle est calculée. Les mots associés aux concepts les plus significatifs sont associés à l'image pour permettre une recherche textuelle ultérieure.

### **15.Le système GIFT**

Le système GIFT (GNU Image-Finding Tool) anciennement appelé VIPER est un système de recherche d'images par le contenu libre disponible sur « <http://gnu.org/software/gift> ». Il propose un ensemble d'outils d'extraction de descripteurs visuels des images. Il n'a pas été conçu pour utiliser l'information textuelle. Cependant, comme ce système est un système où chacun peut proposer des contributions, il n'est pas exclu de rajouter des capacités de gestion des informations textuelles.

### **16.Le système IMAGESEEKER**

Le système IMAGESEEKER est un système commercial qui permet de faire des recherches textuelles et visuelles. Cependant disponible sur « <http://www.ltutech.com> », il semble que les informations textuelles et visuelles ne soient pas vraiment combinées pour améliorer la pertinence des indexations.

## II. Remarques

Si nous observons l'évolution des systèmes de recherche d'images par le contenu, nous remarquons que ces systèmes sont de plus en plus complexes. Les premiers systèmes proposaient de rechercher des images visuellement similaires en réalisant une mesure entre les histogrammes de couleurs des images entières, puis des régions d'images. Maintenant, les améliorations sont obtenues en utilisant des systèmes d'apprentissage de plus en plus complexes (systèmes bayésiens, cartes auto-organisatrices, modèles de Markov cachés...) qui sont nécessaires pour fusionner les différentes informations visuelles et/ou textuelles. Il est assez surprenant de constater que depuis 2003, il n'y a que peu de nouveaux systèmes complets de recherche d'images par le contenu. La plupart des travaux proposent d'améliorer une partie d'un système, mais pas un nouveau type de système.

Il y a finalement peu de systèmes qui combinent vraiment les informations textuelles et visuelles (PICTION, FOUREYES, IMAGEROVER, PICHUNTER, A-LIP). De plus, la plupart sont des modèles d'auto-annotation d'images, et non pas des systèmes de recherche d'images complets. Le système PICHUNTER propose une combinaison originale des informations textuelles et visuelles, car il utilise le texte associé aux images pour améliorer les similarités entre les images, mais l'utilisateur ne peut pas effectuer de recherche textuelle et ne sait pas que le système utilise ce type d'information. Il y a que peu de systèmes complets qui combinent vraiment les informations textuelles et visuelles, mais il existe cependant quelques travaux prometteurs parmi lesquels [Haddad & Mulhem, 01] et [ Belkhatir et al, 2005].

# Bibliographie

[Tebri ,04] Tebri H. Formalisation et spécification d'un système de filtrage incrémental d'information. Thèse de doctorat de l'université Paul Sabatier, Toulouse, 2004.

[Sauvagnat,05] Sauvagnat K. *Modèle flexible pour la Recherche d'Information dans des corpus de documents semi-structurés*, Thèse en vue de l'obtention du Doctorat de l'Université Paul Sabatier, 2005.

[Boubekeur, 08] Boubekeur.F *Contribution à la définition de modèles de recherche d'information flexibles basés sur les CP-Nets*. Thèse en vue de l'obtention du Doctorat de l'Université Paul Sabatier, 2008.

[Haralick, 73] Haralick R.M., Shanmugam K., Dinstein I. (1973), "Textural Features for Image Classification", *IEEE Trans. On Systems, Man, and Cybernetics*, Vol. SMC-3, No. 6, November 1973, pp. 610-621.

[Dahak, 05] Fouad DAHAK. Indexation des documents Semi-Structurés, Proposition d'une approche basée sur le fichier inversé et le "Trie". Thèse en vue de l'obtention du Magister de l'Institut National de formation en Informatique (I.N.I), 2005.

[Maron et al, 60] Maron, M., and Kuhns, J. On relevance, probabilistic indexing and information retrieval. *Journal of the Association for Computing Machinery* 7 (1960), pages 216–244.

[Denoyer, 04] Denoyer L., Apprentissage et Inférence Statistique dans les Bases de Documents Structurés : Application aux Corpus de Documents Textuels. Thèse de Doctorat en Informatique de l'Université de Paris 6. Décembre 2004.

[Smeulders, 00] Smeulders A.W.M., Worring M., Santini S., Gupta A., Jain R.(2000), "Content-Based Image Retrieval at the End of the Early Years". *IEEE*

[Swain & al. 1991] Swain (M.J.) et Ballard (D.H.). Color indexing. *International Journal of Computer Vision*, vol. 7, n\_ 1, 1991, pp. 11-32.

[Stricker, 95] Stricker M., Orengo M.(1995), "Similarity of Color Images", *In Proceedings of SPIE, Vol. 2420(Storage and Retrieval of Image and Video Databases III)*, SPIE Press, Feb 1995.

[Cox et al, 00] Cox, J., Miller, M. L., Minka, T. P., Papathomas, T. V., & Yianilos, P. N. 2000. The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychological Experiments. *IEEE Transactions on Image Processing*, 9(1), 20-37.

[Fauqueur, 03] Fauqueur, Julien. 2003. *Contributions pour la Recherche d'Images par Composantes Visuelles*. Thèse de doctorat, Université de Versailles Saint-Quentin en Yvelines, INRIA-Rocquencourt, Projet IMEDIA.

- [Ma & Manjunath, 99] Ma, W.-Y., & Manjunath, B. S. 1999. NeTra: A Toolbox for Navigating Large Image Databases. *Multimedia Systems*, 7(3), 184-198.
- [Schomaker *et al*, 99] Schomaker, L., de Leau, E., & Vuurpijl, L. 1999. Using Pen-Based Outlines for Object-Based Annotation and Image-Based Queries. *Pages 585-592 of: Proceedings of International Conference on Visual Information and Information Systems*. London, UK: Springer-Verlag.
- [Canny, 86] Canny, J. 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679-698.
- [Da Rugna & Konik, 04] Da Rugna, J., & Konik, H. 2004. Étude comparative de méthodes de segmentation dans une approche orientée indexation. *In: Actes du Congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA)*.
- [Barnard *et al*, 03] Barnard, K., Duygulu, P., Guru, R., Gabbur, P., & Forsyth, D. 2003. The effects of segmentation and feature choice in a translation model of object recognition. *Computer Vision and Pattern Recognition*, 675-682.
- [Datta *et al*, 05] Datta, R., Li, J., & Wang, J. Z. 2005. Content-Based Image Retrieval - Approaches and Trends of the New Age. *In: Proceedings of International Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia*.
- [Puzicha *et al*, 99] Puzicha, J., Rubner, Y., Tomasi, C., & Buhmann, J. M. 1999. Empirical Evaluation of Dissimilarity Measures for Color and Texture. *Pages 1165-1172 of: Proceedings of IEEE International Conference on Computer Vision (ICCV)*.
- [Swain & al, 91] Swain, M. J., & Ballard, D. H. 1991. Color indexing. *International Journal Computer Vision*, 7(1).
- [Puzicha *et al*, 99] Puzicha, J., Rubner, Y., Tomasi, C., & Buhmann, J. M. 1999. Empirical Evaluation of Dissimilarity Measures for Color and Texture. *Pages 1165-1172 of: Proceedings of IEEE International Conference on Computer Vision (ICCV)*.
- [Bach *et al*, 96] Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Jam, R., & Shu, C.F. 1996. The virage image search engine: An open framework for image management. *Pages 76-87 of: Proceedings of SPIE Storage and Retrieval for Still Image and Video Databases IV*.
- [Flickner *et al*, 95] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. 1995. Query by Image and Video Content: The QBIC System. *Computer*, 28(9), 23-32.
- [Ogle & Stonebraker, 95] Ogle, V. E., & Stonebraker, M. 1995. Chabot: Retrieval from a Relational Database of Images. *Computer*, 28(9), 40-48.
- [Srihari & Burhans, 94] Srihari, R. K., & Burhans, D. T. 1994. Visual semantics: extracting visual information from text accompanying pictures. *Pages 793-798 of: National Conference on Artificial Intelligence (AAAI)*.

- [Fellbaum, 98] Fellbaum, C. (ed). 1998. *WordNet - An Electronic Lexical Database*. Bradford books.
- [Frankel *et al*, 96] Frankel, C., Swain, M. J., & Athitsos, V. 1996. *WebSeer: An Image Search Engine for the World Wide Web*. Tech. rept. Chicago, IL, USA.
- [Smith & Chang, 96] Smith, J. R., & Chang, S.-F. 1996. VisualSEEK: A Fully Automated Content-Based Image Query System. *Pages 87-98 of: ACM Multimedia*.
- [Pentland *et al*, 1996] Pentland, A., Picard, R. W., & Sclaroff, S. 1996. Photobook: content-based manipulation of image databases. *International Journal of Computer Vision*, **18**(3), 233-254.
- [Minka & Picard, 96] Minka, T., & Picard, R. 1996. Interactive learning using a 'society of models'. *Pages 447-452 of: Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [Rui, 97] Rui, Y. 1997. A relevance feedback architecture in content-based multimedia information retrieval systems. *In: Proceedings IEEE Workshop Content-Based Access of Image and Video Libraries*.
- [La Cascia *et al*, 98] La Cascia, M., Sethi, S., & Sclaroff, S. 1998 (6). Combining Textual and Visual Cues for Content-based Image Retrieval on the World Wide Web. *In: IEEE Workshop on Content-based access of Image and Video Libraries*.
- [Ma & Manjunath, 99] Ma, W.-Y., & Manjunath, B. S. 1999. NeTra: A Toolbox for Navigating Large Image Databases. *Multimedia Systems*, **7**(3), 184-198.
- [Carson *et al*, 99] Carson, C., Thomas, M., Belongie, S., Hellerstein, J. M., & Malik, J. 1999. Blobworld: A system for region-based image indexing and retrieval. *In: International Conference on Visual Information Systems*.
- [Cox *et al*, 00] Cox, J., Miller, M. L., Minka, T. P., Papatomas, T. V., & Yianilos, P. N. 2000. The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychological Experiments. *IEEE Transactions on Image Processing*, **9**(1), 20-37.
- [Wang *et al*, 01] Wang, J. Z., Li, J., & Wiederhold, G. 2001. SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(9), 947-963.
- [Boujemaa *et al*, 01] Boujemaa, N., Fauqueur, J., Ferecatu, M., Fleuret, F., Gouet, V., Saux, B. Le, & Sahbi, H. 2001. IKONA: Interactive Generic and Specific Image Retrieval. *In: International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR)*.
- [Nastar *et al*, 98] Nastar, C., Mitschke, M., Meilhac, C., & Boujemaa, N. 1998. SurfImage: a flexible content-based image retrieval system. *In: The 6<sup>th</sup> ACM International Multimedia Conference (MM'98)*.

[Ferecatu, 05] Ferecatu, M. 2005. *Recherche d'images par retour de pertinence actif utilisant des descripteurs visuels et textuels*. Thèse de doctorat, Université de Versailles Saint-Quentin-en-Yvelines, Versailles.

[Laaksonen *et al*, 02] Laaksonen, J., Koskela, M., & Oja, E. 2002. PicSOM - Self-organizing image retrieval with MPEG-7 content descriptions. *IEEE Transactions on Neural Networks, Special Issue on Intelligent Multimedia Processing*, **13**(4), 841-853.

[Kohonen *et al*, 00] Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V., & Saarela, A. 2000. Self organization of a massive document. *IEEE Transactions on Neural Networks*, **11**(3), 574-585.

[Viitaniemi & Laaksonen, 05] Viitaniemi, V., & Laaksonen, J. 2005 (November). Keyword-detection approach to automatic image annotation. *In: Proceedings of 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT)*.

[Li & Wang, 03] Li, J., & Wang, J. Z. 2003. Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**(9), 1075-1088.

[Haddad & Mulhem, 01] Haddad, H., & Mulhem, P. 2001. Utilisation de la Fouille de Données Images pour l'Indexation Automatique des images. *In: INFORSID'2001*.

[Belkhatir *et al*, 05] Belkhatir, M., Mulhem, P., & Chiaramella, Y. 2005. A Full-Text Framework for the Image Retrieval Signal/Semantic Integration. *Pages 113-123 of: DEXA*.