



HAL
open science

Un modèle vectoriel relationnel de recherche d'information adapté aux images

Jean Martinet

► **To cite this version:**

Jean Martinet. Un modèle vectoriel relationnel de recherche d'information adapté aux images. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université Joseph Fourier (Grenoble I), 2004. Français. NNT: . tel-01693438

HAL Id: tel-01693438

<https://theses.hal.science/tel-01693438>

Submitted on 26 Jan 2018

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

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

UNIVERSITÉ JOSEPH FOURIER – GRENOBLE I
U.R.F. INFORMATIQUE ET MATHÉMATIQUES APPLIQUÉES

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER – GRENOBLE I

Discipline : Informatique

Présentée et soutenue publiquement le 22 décembre 2004 par

Jean Martinet

TITRE

**Un modèle vectoriel relationnel
de recherche d'information
adapté aux images**

Directeurs de thèse : Yves Chiaramella et Philippe Mulhem

Composition du jury :

Présidente : Mme Catherine GARBAY
Rapporteurs : M. Jean-Marie PINON
: M. Mohand BOUGHANEM
Examineurs : M. Yves CHIARAMELLA
: M. Philippe MULHEM

Thèse préparée au sein de l'équipe MRIM du Laboratoire CLIPS-IMAG
(Communication Langagière et Interaction Personne-Système)
Université Joseph Fourier – Grenoble I

Remerciements

C'est un grand plaisir pour moi de remercier toutes les personnes qui ont permis à ce travail d'être ce qu'il est.

Je remercie tout d'abord Mme Catherine Garbay, Directrice de Recherches au CNRS, qui m'a fait l'honneur de présider le jury de cette thèse.

Je remercie M. Jean-Marie PINON, Professeur à l'INSA de Lyon ainsi que M. Mohand BOUGHANEM, Professeur à l'Université Paul Sabatier de Toulouse pour avoir accepté de rapporter ce travail, ainsi que pour l'intérêt qu'ils ont manifesté à l'égard de ce travail de thèse.

Je remercie M. Yves Chiamella, Professeur à l'Université Joseph Fourier, et M. Philippe Mulhem, Chargé de Recherches au CNRS, pour avoir co-encadré mon travail, et pour leur aide précieuse, leur patience, et leur support inestimable durant ce travail, depuis le début de mon DEA. Je remercie M. Iadh Ounis, Maître de Conférences à l'Université de Glasgow, pour avoir encadré une partie de ce travail lors de mon séjour à l'Université de Glasgow.

Je remercie tous les membres de l'équipe MRIM pour leur accueil dans l'équipe, et pour leurs questions pertinentes lors des réunions, qui ont fait avancer ce travail.

Je remercie tous les membres du laboratoire CLIPS pour leur accueil.

Je remercie C.J. "Keith" van Rijsbergen, Professeur à l'Université de Glasgow ainsi que tous les membres du IR Group pour leur accueil au sein du groupe. Mentions spéciales pour Jana Urban, Robert Villa et Iain Campbell.

Je remercie tous les thésards l'équipe MRIM et du labo CLIPS, notamment Quick, Domi, Dima, Stéph A., Lizbeth, Moh, Mbarek, Leila, Caro T., Helga, Razan et Quoc, Daniel, Niko, Daniel, Federico et les autres, pour les bons moments passés au labo. Je remercie Caro G. pour les pauses thés. Mention spéciale pour mes prédécesseurs Mathias et Hatem pour avoir tracé la route.

Je remercie tout spécialement Domi, Béranger, Quick, et Marlène pour les relectures attentives de ce document.

Je remercie tou(te)s les participant(e)s à l'évaluation pour le temps qu'il ont généreusement consacré à cliquer sur des images.

Je remercie tous ceux qui donnent son âme à la cafétéria.

Je remercie surtout ma famille : Laura (et Yoann !) et Eugène, Laetitia, Emeric, et très spécialement ma mère, pour leur soutien constant à travers ces longues années.

Je remercie beaucoup Marlène pour son soutien indéfectible et inconditionnel lors de ma phase terminale, et pour avoir supporté mon mauvais caractère lorsque j'étais égoïste et irritable comme je ne l'ai jamais été.

Je remercie enfin toutes les personnes que j'ai oublié de remercier ici.

Résumé

La démocratisation des systèmes d'acquisition d'images numériques, combinée à l'essor du Web, nécessite la définition de modèles et de systèmes de Recherche d'Information (RI) capables d'accéder rapidement aux images recherchées par des utilisateurs dans de grands volumes de données. L'objectif de ce travail est de définir un modèle de RI adapté aux images, en intégrant une sémantique riche pour la représentation de ces données visuelles et des requêtes des utilisateurs, tout en permettant l'application à de grands corpus. Notre proposition fusionne le modèle vectoriel de RI – modèle éprouvé en RI textuelle – avec le formalisme des Graphes Conceptuels (GC), en se fondant sur l'utilisation de graphes étoiles (GC élémentaires constitués d'une relation connectée à des concepts représentant les objets de l'image) comme descripteurs d'images. Une pondération originale de ces graphes étoiles (fondée sur la taille et la position des objets, ainsi que sur l'hétérogénéité des images) est définie et validée expérimentalement avec soin. Nous montrons que l'intégration des relations dans le modèle vectoriel par le biais des graphes étoiles augmente la précision du système, et la rend comparable à celle des systèmes basés sur la projection de graphes, tout en diminuant le temps de traitement des requêtes des utilisateurs.

Mots-clés : Recherche d'Information, Indexation d'Images, Modèle Vectoriel, Graphe Conceptuel, Pondération, Évaluation.

Abstract

The increase of digital image acquisition devices, combined to the growth of the Web, requires the definition of Information Retrieval (IR) models and systems providing fast access to images searched by users among large amounts of data. The aim of our work is to define an IR model suited to images, integrating a rich semantics for representing these visual data as well as users' queries, while allowing the application to large corpora. Our proposition merges the vector space model of IR – a widely tested model in textual IR – with Conceptual Graphs (CG) formalism, based on the use of star graphs (a star graph is an elementary CG made up of a single relation connected to some concepts representing image objects) as image descriptors. A novel weighting scheme (based on objects size and position, as on image heterogeneity) is defined and carefully validated in the experiments. We show that integrating relations into the vector space model through star graphs increases the system precision, and that the results are comparable with graph projection systems, while decreasing the processing time for users' queries.

Keywords : Information Retrieval, Image Indexing, Vector Space Model, Conceptual Graph, Weighting Scheme, Evaluation.

Table des matières

I	Introduction, problématique et état de l'art	1
1	Introduction	3
1.1	Contexte : la Recherche d'Information	3
1.2	Problématique	4
1.2.1	Spécificités de l'image	5
1.2.2	Vers des descriptions relationnelles	8
1.3	Objectifs	9
1.4	Notre approche	10
1.5	Plan de la thèse	12
2	De la recherche de texte...	15
2.1	Modalités d'accès à l'information	15
2.2	Modélisation d'un SRI	19
2.2.1	Indexation des documents	20
2.2.2	Interrogation du système	22
2.2.3	Modélisation	24
2.3	Évaluation et comparaison de systèmes	25
2.4	Le modèle vectoriel	27
2.5	Pondération des termes dans le modèle vectoriel	30
3	... à la recherche d'images	35
3.1	Signal et sens : le fossé sémantique	35
3.2	Les approches signal	36
3.2.1	Modèle de documents	36
3.2.2	Modèle de requêtes	40
3.2.3	Modèle de correspondance	43
3.2.4	Discussion	44
3.3	Les approches symboliques basées sur les mots clés	45
3.4	Les graphes conceptuels pour la recherche d'images	48
3.4.1	Intérêts des graphes conceptuels pour la RI	48
3.4.2	Limites	55
3.5	Mise en perspective de notre travail	58

II Un modèle vectoriel relationnel de recherche d'images 59

4 Description du modèle proposé 61

4.1	Rappel des objectifs	61
4.2	Modèle de représentation des documents et des requêtes	63
4.2.1	Modèle d'image physique	63
4.2.2	Interprétation des objets images	66
4.2.3	Les graphes étoiles comme vocabulaire d'indexation	67
4.2.4	Compromis entre pouvoir expressif et complexité	70
4.3	Processus d'indexation	73
4.4	Fonction de correspondance	75
4.5	Conclusion	76

5 Pondération des graphes étoiles 79

5.1	Hypothèses relatives à l'importance des objets images	82
5.1.1	La taille	83
5.1.2	La position	84
5.1.3	L'éclatement	85
5.1.4	L'hétérogénéité	87
5.2	Modélisation de l'importance des objets images	89
5.2.1	Le paramètre taille	90
5.2.2	Le paramètre position	91
5.2.3	Le paramètre éclatement	92
5.2.4	Le paramètre hétérogénéité	94
5.2.5	Pondération des objets images	96
5.3	Pondération de graphes étoiles	96
5.3.1	Importance locale d'un graphe étoile dans un document	96
5.3.2	Importance globale d'un graphe étoile dans la collection	98
5.3.3	Poids d'un graphe étoile	99
5.4	Conclusion	99

III Mise en œuvre : Expérimentation et validation 101

6 Validation de la pondération 103

6.1	Description de l'expérimentation	104
6.1.1	La collection d'images	104
6.1.2	Les participants	107
6.1.3	Déroulement	108
6.2	Interprétation des données collectées	110
6.2.1	Modélisation des données	110
6.2.2	La taille	114
6.2.3	La position	114

6.2.4	L'éclatement	116
6.2.5	L'hétérogénéité	116
6.2.6	Combinaison de critères	117
6.3	Questionnaire	119
6.4	Nouvelle formule de pondération des objets images	121
6.5	Discussion	122
6.6	Conclusion	122
7	Evaluation du système	123
7.1	Collection de tests	123
7.2	Evaluations comparatives	126
7.2.1	Comparaison avec les mots clés	126
7.2.2	Comparaison avec la projection de graphes	127
7.2.3	Impact de la stratégie de pondération	130
7.3	Conclusion	131
IV	Conclusion générale	133
8	Conclusion	135
8.1	Synthèse et contributions	135
8.2	Perspectives	137
V	Annexes et bibliographie	139
A	Autres modèles de recherche d'information	141
A.1	Modèle booléen	141
A.2	Modèle probabiliste	143
A.3	Modèle logique	145
A.4	Discussion sur ces modèles	146
B	Le formalisme des graphes conceptuels	147
B.1	Support	148
B.1.1	Les concepts et les relations	148
B.1.2	Les signatures et les graphes étoiles	149
B.1.3	Les référents	150
B.2	Graphes conceptuels	151
B.3	Opérations canoniques sur les graphes conceptuels	152
B.4	Projection de graphes	153
	Bibliographie	166

Table des figures

1.1	Situation de notre approche.	11
1.2	Exemples de graphes étoiles mettant en relation des concepts.	12
2.1	Architecture d'un SRI.	19
2.2	Représentation vectorielle de deux documents (D_1 et D_2) et d'une requête (Q) dans un espace composé de trois termes (t_1 , t_2 et t_3).	29
2.3	Loi de Zipf et conjecture de Luhn.	32
3.1	Deux images susceptibles d'être décrites par les mêmes mots clés.	48
3.2	Deux graphes conceptuels décrivant respectivement les images 1 et 2 de la figure 3.1.	49
3.3	Un graphe conceptuel de requête.	50
3.4	Les concepts et les relations sont organisés en treillis de types de concepts (gauche) et treillis de relations (droite).	51
3.5	Projection de graphes conceptuels.	52
3.6	Vue globale d'une image dans le modèle EMIR ²	54
4.1	Exemple de segmentation en régions et en objets images.	68
4.2	Graphes étoiles de relations unaires, binaire et ternaire, de gauche à droite.	69
4.3	Représentation vectorielle de l'index d'un document.	70
4.4	Relation entre l'expressivité du formalisme et la complexité des traitements pour les mots clés, les graphes étoiles et les graphes conceptuels.	72
4.5	Vecteur $\vec{D}1'$ après expansion du document.	75
5.1	A quel point le graphe étoile est-il un bon descripteur de l'image ?	80
5.2	Exemple de deux images entre lesquelles la taille de l'objet image varie.	84
5.3	Exemple de deux images entre lesquelles la position de l'objet image varie.	85
5.4	Exemple de deux images entre lesquelles l'éclatement de l'objet image varie.	86
5.5	Exemple de deux images d'hétérogénéités différentes.	89
5.6	Distributions de probabilités (2D) : (a) uniforme, (b) par paliers, (c) linéaire, (d) non linéaire.	92
6.1	16 situations logiques d'apparition d'un objet image (symbolisé ici par les disques noirs), selon les 2 modalités de chacun des 4 critères.	105
6.2	Taux de corrélation des données en fonction du nombre de participants.	108

6.3	Interface d'évaluation.	109
6.4	Treillis de Gallois associé aux critères mis en jeu dans les combinaisons de situations.	111
6.5	Exemple de distribution binomiale $B(30,0.8)$	112
6.6	Distribution des choix selon le type d'objet pour le critère <i>taille</i>	115
6.7	Distribution des choix selon le type d'objet pour le critère <i>position</i>	115
6.8	Distribution des choix selon le type d'objet pour le critère <i>éclatement</i>	116
6.9	Distribution des choix selon le type d'objet pour le critère <i>hétérogénéité</i>	117
7.1	Aperçu des images de la collection de tests utilisée.	124
7.2	Un exemple de requête de la collection de tests.	125
7.3	Décomposition d'un graphe conceptuel en graphes étoiles.	125
7.4	Courbes de rappel/précision pour notre système et le système basé sur les mots clés.	127
7.5	Courbes de rappel/précision pour notre système et le système basé sur la projection de graphes.	128
7.6	Courbes de rappel/précision pour notre système et le système basé sur la projection de graphes. A gauche : 29 requêtes sans relation, à droite : 9 requêtes avec relations.	129
B.1	Les concepts et les relations sont organisés en treillis de types de concepts (à gauche) et treillis de relations (à droite).	149
B.2	Situation de la base canonique et des graphes canoniques et conceptuels.	150

Liste des tableaux

5.1	Relation entre la taille apparente d'un objet et sa profondeur.	83
6.1	Modalités de variation des 4 critères.	104
6.2	Moyennes (et écarts-types) des critères pour les 30 participants.	113
6.3	Probabilités estimées (et écarts-types) pour tous les types d'objets, et probabilités associées au test T de Student (H_0).	113
6.4	Probabilités estimées pour les variables mettant en jeu 2 critères.	118
6.5	Probabilités estimées pour la variable $Y_{\{TPEH\}}$ mettant en jeu les 4 critères.	119
6.6	Rangs moyens des réponses au questionnaire.	120
6.7	Rangs moyens (et écarts-types) des classements des 4 critères par ordre d'importance.	121
7.1	Précisions moyennes et gains relatifs pour 3 stratégies de pondération différentes.	130

Première partie

Introduction, problématique et état de l'art

Chapitre 1

Introduction

S'il est important de savoir modéliser, transmettre, et stocker de l'information, il est également important de permettre aux utilisateurs d'un Système d'Information de localiser rapidement une information recherchée. Initialement réduite à une version électronique d'un document *papier* – un simple document textuel atomique [Sal71], la notion de document électronique a beaucoup évolué. Cette notion est maintenant étendue à tous les types de média : texte, image, son ou vidéo, ainsi qu'aux documents structurés composés de plusieurs de ces médias.

1.1 Contexte : la Recherche d'Information

La Recherche d'Information (RI) est un domaine de recherche en informatique qui s'attache à définir des modèles et des systèmes dont le but est de faciliter l'accès à un ensemble de **documents** sous forme électronique (corpus), afin de permettre à un utilisateur de retrouver ceux qui sont **pertinents** pour lui, c'est-à-dire ceux dont le **contenu** correspond le mieux à son **besoin d'information** à un moment donné. Le domaine de la RI est donc centré autour des utilisateurs, notamment par le biais de la notion de *pertinence*, c'est-à-dire l'adéquation pour l'utilisateur entre le contenu d'un document et l'information recherchée.

Avec l'émergence des réseaux électroniques et le développement du Web, la quantité d'information disponible pour les utilisateurs est sans cesse croissante, et la RI doit faire

face à de nouveaux défis d'accès à l'information, à savoir retrouver une information dans un espace diversifié et de taille considérable.

1.2 Problématique

La simplicité du langage des images en fait un média d'expression et de communication privilégié entre les humains. Ce média inclut notamment :

- les bases d'agences photographiques,
- les images médicales (e.g. radiographies, IRM),
- les images d'archives (e.g. musées, patrimoine culturel),
- les images géographiques, topographiques et cartographiques,
- les photos aériennes et satellitaires (e.g. météorologiques),
- les logos publicitaires,
- les icônes informatiques,
- les timbres poste,
- les pictographies et graphiques figuratifs,
- les icônes religieux,
- les dessins et croquis,
- les tableaux et peintures artistiques,
- les photographies personnelles.

Cette liste, sans être exhaustive, donne une idée de l'étendue des domaines d'application du média image. Nous nous intéressons plus particulièrement dans notre travail au cas des photographies personnelles, dont l'intérêt a connu un essor considérable ces dernières années. Avec la démocratisation des systèmes personnels d'acquisition d'images numériques comme les appareils photo numériques, le nombre d'images existantes – et disponibles sur les réseaux électroniques par le biais du Web, est de plus en plus grand. Selon une étude menée à Berkeley [LV03], il existerait en 2003 environ 900 milliards de photographies personnelles. De plus, toujours d'après cette étude, la production annuelle de photographies

personnelles serait d'environ 75 milliards par an, et ce chiffre va en augmentant. Les progrès technologiques et la mise à disposition de ces grands volumes d'information nécessitent des moyens d'accès adaptés.

1.2.1 Spécificités de l'image

Notre travail porte sur la définition d'un modèle de RI adapté au média image. Les modèles de RI les plus répandus sont destinés à la recherche d'information textuelle ; l'application de la RI aux images demande une attention particulière dans la définition de la représentation interne au système du contenu des images et du besoin d'information des utilisateurs, ainsi que dans la mise en correspondance de ces représentations, pour produire le résultat escompté, qui est la liste des images satisfaisant le besoin des utilisateurs.

Lorsque l'on s'intéresse au média image, la définition du modèle de représentation de document pose davantage de problèmes que pour le texte, car la sémantique n'est pas intrinsèquement exprimée dans les pixels des images. Le domaine de la Recherche d'Image (RIIm) doit donc faire face à une caractéristique importante des images qui est le manque d'un langage de représentation explicite permettant d'en exprimer la sémantique. Cet écart entre les pixels et leur signification est appelé le **fossé sémantique**. Smeulders décrit dans [SWS⁺00] ce fossé sémantique de la manière suivante :

"Le fossé sémantique est le manque de concordance entre l'information que l'on peut extraire des données visuelles et l'interprétation des mêmes données qu'en fait un utilisateur dans une situation donnée¹."

Ce fossé sépare donc le niveau d'abstraction auquel un utilisateur exprime son besoin d'information à travers une requête d'une part, et le niveau d'abstraction auquel de l'information peut être automatiquement extraite d'une image d'autre part.

Les approches sur lesquelles se basent les Systèmes de Recherche d'Images (SRIm) se

¹"The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation".

distinguent selon leur niveau de représentation du contenu des images :

- **le niveau signal** : l'image est considérée comme une matrice brute de pixels, et l'opération de recherche est effectuée selon une similarité visuelle sur des traits de bas niveau, comme les couleurs ou les textures présentes dans les images,
- **le niveau symbole** : l'indexation et la recherche sont fondées sur une interprétation sémantique du contenu de l'image.

Ces deux approches sont fondamentalement différentes du fait qu'elles intègrent ou non une sémantique liée au contenu de l'image, et les modèles et systèmes qui en découlent le sont également. Alors que les approches signal caractérisent automatiquement le contenu brut des images à l'aide d'histogrammes de couleurs ou de textures, les approches symboliques interprètent ce contenu brut, et le décrivent dans un langage ou formalisme de plus haut niveau.

Le problème de la description sémantique des documents, bien qu'il se pose également pour le texte, se révèle avec davantage d'acuité dans le cas des images, en raison de la **subjectivité de leur interprétation**. Cette interprétation est souvent manuelle ou assistée, et donc à la fois subjective et incertaine [ZD69] : une même image peut être perçue très différemment selon le contexte (situation, lieu géographique, etc.) et selon les personnes, en fonction de leur culture, leur éducation, ou leurs affinités. Par exemple, la nature subjective de l'interprétation est liée à l'aspect culturel de la perception. Citons l'expérience de Hudson en 1960 sur la perception de la profondeur² chez des personnes issues de groupes ethniques d'Afrique du Sud, décrite dans l'ouvrage de Messaris [Mes94]. Cette expérience a démontré que l'inférence de la profondeur en fonction d'indices visuels comme l'occlusion, la perspective, la taille relative et la hauteur des objets dans le champs visuel dépend du niveau de scolarisation et de l'origine sociale des sujets. Par ailleurs, l'interprétation peut se situer au niveau dénotatif (référence à une description objective des objets qui apparaissent dans une image, comme *arbre*, *bateau* ou *maison*) ou au niveau connotatif

²La profondeur est la distance virtuelle entre un objet physique représenté dans une image et l'œil de celui qui le voit.

(interprétation subjective que l'on peut en faire, comme une émotion suggérée). Il découle du fossé sémantique et de cette subjectivité d'interprétation un problème non trivial pour les approches symboliques de détermination automatique ou semi-automatique du contenu sémantique. La difficulté réside notamment dans les points précisés dans les paragraphes suivants.

Le choix des descripteurs

Il s'agit de déterminer les éléments de l'image représentatifs de son contenu, selon un certain niveau de granularité de description et une certaine exhaustivité d'indexation. Cela consiste à définir le vocabulaire d'indexation (voir la partie 2.2), qui peut être simplement constitué de mots clés décrivant les objets de l'image. Dans ce cas, il est nécessaire de les extraire à travers un processus de segmentation (voir la partie 3.3), qui consiste à délimiter les régions d'intérêt de l'image et à leur associer une étiquette.

La pondération des descripteurs

Comme dans le cas de la RI textuelle, la pondération des mots clés vise à leur donner plus ou moins d'importance selon leur intérêt. Alors que les approches classiques de RI textuelle se basent sur le nombre d'occurrences d'un mot clé pour établir son importance dans un document, les différentes occurrences des objets dans une image ont des caractéristiques picturales individuelles, qu'il est nécessaire de considérer lors de la définition d'un poids.

Exemple 1 *Dans le cas où le mot "bateau" apparaît trois fois dans un texte, l'importance de bateau dans ce texte est directement reliée au nombre 3, et on n'indique pas qu'une occurrence est plus importante que l'autre. Par contre, dans le cas où trois bateaux apparaissent dans une image, en plus du fait qu'ils sont au nombre de 3, il est également nécessaire de tenir compte de leurs "importances" individuelles, et de les inclure dans le calcul du poids du descripteur bateau pour cette image.*

Il est nécessaire de considérer cette différence majeure entre le texte et l'image dans une stratégie de pondération dédiée aux images.

1.2.2 Vers des descriptions relationnelles

La nature bidimensionnelle des images induit des aspects perceptifs et des relations sémantiques et spatiales, qu'il est nécessaire de prendre en compte afin de représenter plus précisément le contenu des images. L'intégration de descriptions *relationnelles* a fait l'objet de nombreux travaux sur des formalismes riches, aptes à modéliser différentes facettes du contenu des images. Le formalisme des **graphes conceptuels** [Sow84] est l'un de ceux qui ont été utilisés pour représenter le contenu des images dans le cadre de la recherche d'information [Che92, Khe95, BFM⁺97, CMF96, MMFB97, Gen00, MLG00], et spécialement pour la recherche d'images [Mec95, MBC95, CM97, Oun98, OP98, TADC03, MLLK03, DM04]. Les graphes conceptuels sont des graphes orientés bipartis constitués de nœuds concepts et de nœuds relations interconnectés. Les concepts et les relations sont organisés dans des treillis de types contenant les liens de généralité et de spécificité entre les types. L'utilisation de ce formalisme permet une grande expressivité du langage d'indexation, et par conséquent une grande précision de recherche. Les images et les requêtes sont représentées par des graphes conceptuels ; la décision de pertinence d'une image pour une requête est basée sur l'opération de *projection de graphe*, qui consiste à rechercher dans le graphe d'une image un sous-graphe de même structure que le graphe de la requête, dont les nœuds sont éventuellement plus spécifiques (voir l'annexe B). Ces travaux ont validé le fait que l'intégration des relations de différentes natures dans la représentation des images et des requêtes permet d'obtenir des résultats de meilleure qualité. Cependant, ces approches rencontrent un certain nombre de problèmes inhérents à l'utilisation de l'opération de projection de graphes, comme ceux que nous énonçons :

- L'opération de projection est un isomorphisme de graphe, qui est un problème NP-complet [CM92, MC96], dont les implantations mettent en œuvre des algorithmes non

polynomiaux [Gui96]. En conséquence, quand les graphes sont complexes et quand la base de documents est grande, le problème du temps de réponse du système se pose, ce qui est un facteur limitant de sa performance.

- Le résultat de l'opération de projection est binaire : il existe ou non une projection du graphe requête dans le graphe document. Ainsi, un document satisfaisant *partiellement* la requête est considéré comme non pertinent. Avec une telle correspondance, les documents partiellement pertinents ne sont pas sélectionnés, et la réponse risque d'être vide en cas de requête complexe [MLLK03].
- Un autre problème de la correspondance binaire pour la RI est qu'elle ne permet pas d'ordonner les réponses du système par ordre de pertinence décroissante. Quand le nombre de réponses est grand, il est souhaitable de les ordonner de telle sorte que les réponses les plus pertinentes soient les premières dans la liste. Ce problème, ainsi que le précédent, est le même que pour le modèle booléen de RI, qui ne permet qu'une évaluation binaire de la pertinence (voir l'annexe A). La distinction n'est pas faite entre un cas où l'un des deux graphes, document ou requête, est entièrement contenu dans l'autre (exhaustivité ou spécificité du document par rapport à la requête [SM83], cas favorable de la recherche), et le cas où les graphes de document et de requête sont identiques (identité du document et de la requête, cas idéal de la recherche).

1.3 Objectifs

De nombreuses études sur les attentes des utilisateurs vis-à-vis des systèmes de gestion de collections de photographies personnelles [Rod99, SCI02, RW03, MLLK03] mettent en avant le besoin des utilisateurs de pouvoir gérer leurs collections en utilisant la sémantique ; les besoins principaux concernent la recherche et l'organisation des photographies. L'objectif de notre travail est de définir une approche de RI répondant aux attentes des utilisateurs :

- Le modèle doit intégrer une sémantique riche dans les représentations des documents et des requêtes afin que les utilisateurs puissent interroger le système en formulant des

- requêtes dont l'expressivité est suffisante.
- Le modèle doit permettre d'évaluer une correspondance partielle entre la requête et les documents afin de rendre l'utilisation du système souple et ainsi de résoudre le problème de la correspondance binaire.
 - Par ailleurs, il est nécessaire que l'indexation permette de représenter les importances des descripteurs dans les index en les pondérant, de manière à pouvoir fournir un classement des documents selon leur degré de pertinence estimée. Cette pondération, pour qu'elle soit consistante avec la perception des utilisateurs de l'importance des descripteurs, doit être définie et validée selon leurs jugements.
 - Le système doit pouvoir gérer une grande quantité de données, tout en garantissant un temps de réponse acceptable, selon les recommandations d'utilisabilité et de qualité des systèmes interactifs [CMK66, Nie94].

1.4 Notre approche

Nous situons nos travaux au niveau symbolique de RI_m : nous nous intéressons à des représentations d'images inspirées du formalisme des graphes conceptuels. Afin de résoudre les problèmes de complexité et de rigidité énoncés plus haut liés à l'utilisation de l'opération de projection pour la RI, nous proposons de fusionner ces représentations avec un modèle classique de RI. Parmi les modèles existants de RI, le modèle vectoriel est l'un de ceux qui a permis d'obtenir de bons résultats à partir de connaissances atomiques telles que les mots clés utilisés classiquement pour le texte. Le modèle vectoriel consiste à représenter un document par un vecteur dans un espace dont les dimensions correspondent à un ensemble de termes (mots clés pour le texte) servant de descripteurs du contenu des documents. Chaque coordonnée du vecteur document dans cet espace dénote l'importance du terme correspondant dans ce document ; les requêtes sont représentées selon le même principe. La sélection des documents pertinents pour une requête consiste à comparer le vecteur requête à l'ensemble des vecteurs documents de la base. Le résultat est la liste triée des documents qui

obtiennent une valeur de comparaison élevée (supérieure à un seuil). Ce modèle a permis d'atteindre de bons résultats en terme de précision moyenne³ lors de différentes campagnes d'évaluation comme TREC⁴. De plus, la simplicité de représentation des documents et des requêtes fait que ce modèle est facile à mettre en œuvre. Si un modèle vectoriel étendu per-

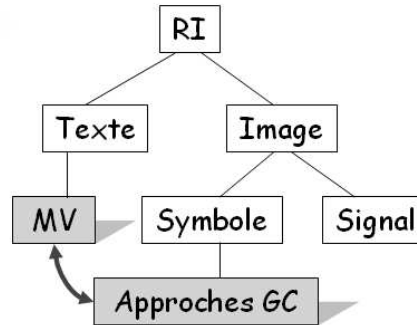


FIG. 1.1 – Situation de notre approche.

met de prendre en compte les caractéristiques de représentation de connaissances comme les graphes conceptuels, alors nous pouvons espérer obtenir des systèmes à la fois aptes à représenter des images de manière précise, et à les rechercher de manière efficace – en temps de réponse et en qualité.

La figure 1.1 situe précisément notre approche par rapport aux éléments décrits plus haut : nous proposons une fusion du modèle vectoriel (MV) de recherche d'information textuelle avec une approche de recherche symbolique d'image fondée sur le formalisme des graphes conceptuels (GC). Nous proposons pour cela de représenter les documents et les requêtes par des vecteurs dans un espace constitué de **graphes étoiles**. Ce sont des graphes conceptuels élémentaires contenant une seule relation, et les concepts qui lui sont attachés (voir la figure 1.2). Ils représentent les briques de base constitutives des graphes conceptuels. Le chapitre 4 décrit plus précisément cette notion de graphes étoiles, et la manière de la

³Voir les détails sur l'évaluation des systèmes dans la partie 2.3.

⁴TREC = *Text REtrieval Conference*. URL : <http://trec.nist.gov/>.

mettre en œuvre dans notre contexte.

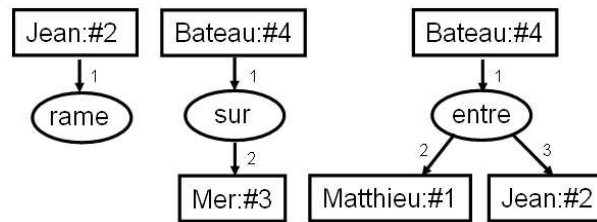


FIG. 1.2 – Exemples de graphes étoiles mettant en relation des concepts.

Le cœur de notre proposition réside dans l'utilisation des graphes étoiles comme vocabulaire d'indexation du modèle vectoriel. Chaque dimension de l'espace est un graphe étoile, ce qui permet une indexation relationnelle plus riche que les simples mots clés. Afin de représenter dans les index les importances des dimensions, nous définissons une méthode de pondération originale dédiée aux graphes étoiles, qui met en jeu :

- D'une part : l'importance locale d'un graphe étoile dans l'image, qui dépend elle-même de l'importance des concepts attachés à la relation de ce graphe étoile. Pour cela, nous nous sommes intéressés à la définition de l'importance des objets apparaissant dans une image en fonction de critères géométriques et topologiques des objets et de l'image.
- D'autre part, l'importance globale d'un graphe étoile dans la collection, basée sur une étude de sa distribution dans la collection.

1.5 Plan de la thèse

Nous traitons le sujet de la manière suivante. Le **chapitre 2** présente la problématique générale de la recherche d'information textuelle, et introduit le vocabulaire de base de l'étude des systèmes de recherche d'information. Ce chapitre présente également le modèle vectoriel classique, qui constitue la base de nos travaux. Le **chapitre 3** présente un état de l'art en recherche d'images selon le point de vue du niveau de représentation du contenu, en

distinguant les approches signal et les approches symboliques. Cet état de l'art comprend une description de l'utilisation des graphes conceptuels pour la recherche d'images. Les **chapitre 4 et 5** constituent notre proposition. Le chapitre 4 donne une description du modèle vectoriel relationnel proposé, qui consiste en une fusion du modèle vectoriel de RI et du formalisme des graphes conceptuels, par le biais des graphes étoiles qui constituent le vocabulaire d'indexation. Ce modèle est décrit selon son modèle de représentation des documents et des requêtes, son modèle de connaissances, et son modèle de correspondance. De la même manière que dans le modèle vectoriel classique, les termes d'indexation sont pondérés afin de refléter leur importance dans les index. La définition d'une stratégie de pondération pour les graphes étoiles fait l'objet du **chapitre 5**. Cette pondération, inspirée des approches textuelles, est fondée sur deux sources : l'importance du terme dans l'image et l'importance du terme dans la collection ; l'importance du terme dans l'image étant une fonction de l'importance des objets mis en relation par le graphe étoile. Ce chapitre propose une modélisation des critères perceptifs de l'importance des objets apparaissant dans les images. Le **chapitre 6** traite de l'évaluation de cette modélisation auprès d'utilisateurs, qui a permis de valider statistiquement l'utilité des critères perceptifs définis. Il décrit le protocole expérimental, et présente l'analyse des données collectées. Le modèle vectoriel proposé a été implanté, et le système a fait l'objet d'une évaluation sur une collection de tests, présentée dans le **chapitre 7**. Ce chapitre analyse trois aspects du système : il compare notre système à une approche non relationnelle basée sur des mots clés, ainsi qu'à une approche fondée sur les graphes conceptuels, et examine l'impact de la fonction de pondération sur les résultats du système. Enfin, le **chapitre 8** conclut ce travail en mettant en avant notre contribution, et en présentant les différentes perspectives qui en découlent.

Chapitre 2

De la recherche de texte...

2.1 Modalités d'accès à l'information

Les différentes manières d'accéder aux documents d'une base documentaire peuvent être analysées selon le mode d'interaction entre l'utilisateur et le système. Waterworth et Chignell [WC91] ont identifié les trois dimensions suivantes de modalité d'accès à l'information :

- **l'objectif** : quel est le but de l'utilisateur ? Ce but peut être la découverte de la base documentaire ou bien la recherche d'une information précise. L'objectif identifie l'état cognitif de l'utilisateur selon son besoin d'information. Par exemple, un utilisateur cherchant dans sa collection de photographies personnelles une image de vacances dont il se souvient est à la recherche d'une information précise. Le but d'un SRI est alors d'aider l'utilisateur à rechercher une information spécifique.
- **la responsabilité** : qui effectue la tâche ? La responsabilité fait référence à l'acteur de la tâche, c'est-à-dire le système ou l'utilisateur. Un utilisateur recherchant une information à l'aide d'un moteur de recherche sur le Web formule une requête, et c'est le système qui effectue l'opération de recherche et de présentation du résultat à l'utilisateur. Dans ce cas, c'est le SRI qui est responsable de la tâche.
- **la méthode d'interaction** : comment l'utilisateur interagit-il avec le système ? L'interaction peut se faire en saisissant une requête (e.g. un moteur de recherche sur le

Web) ou bien en sélectionnant un lien de type hyperlien¹, comme sur le Web. C'est le premier cas qui concerne les SRI : un SRI analyse la requête et fournit un résultat qui est la liste des documents pertinents.

Un utilisateur peut accéder aux documents stockés dans le système selon différentes modalités. Les trois dimensions citées plus haut permettent d'analyser les modalités d'interaction. Nous distinguons l'*exploration*, la *navigation* et l'*interrogation* du système à l'aide de requêtes [Sma94, SWS⁺00], que nous décrivons ci-dessous.

Exploration

L'utilisateur peut découvrir la base documentaire en l'**explorant**. L'exploration consiste à parcourir les documents de la base de manière aléatoire, séquentielle, ou hiérarchique, dans le but de la **découvrir** ou d'en appréhender le contenu. C'est la manière la plus intuitive d'accéder à l'information. L'objectif dans l'exploration est donc la découverte de la base documentaire.

Imaginons une bibliothèque dont les livres ne sont ni classés, ni organisés d'une quelconque manière : un lecteur désireux de découvrir le contenu de la bibliothèque peut circuler dans les rayons de livres, en décidant d'en consulter certains au gré de ses déplacements. Nous pouvons également faire une analogie avec l'explorateur de documents de *Microsoft* [®] *Windows* [®], qui permet de découvrir de manière hiérarchique arborescente les documents stockés dans un ordinateur personnel. Dans ces deux cas, il incombe à l'utilisateur de cheminer à travers la base, la responsabilité lui revient donc. L'interaction se fait par sélections successives de vues en relation les unes avec les autres. La méthode d'interaction est donc de type sélection de liens (hiérarchiques).

¹Smith et Weiss [JS88] ont défini ni l'hyperlien comme une approche de gestion de l'information dans laquelle les données sont stockées dans des nœuds en réseau, connectés par des liens.

Navigation

L'utilisateur peut également **naviguer** dans la base documentaire [Sma94, Chi97], c'est-à-dire se déplacer de document en document en suivant des liens. Le but de l'exploration est de découvrir la base documentaire, ou bien de rechercher une information précise.

Ces liens peuvent être contenus dans les documents, donc prévus à l'avance par l'auteur du document, ou bien déterminés par le système – par exemple des liens de similarité. Dans ce dernier cas, le système fournit un outil de navigation qui présente à l'utilisateur le document courant ainsi qu'un ensemble de documents dans son voisinage. L'utilisateur navigue dans la base en ouvrant sur l'espace des documents des fenêtres successives, jusqu'à ce qu'il trouve l'information recherchée [Chi97]. Au début de cette interaction, les documents présentés à l'utilisateur sont des points d'entrée dans la base documentaire – par exemple un échantillon de documents représentatifs du contenu de la base, une classification de termes (comme l'annuaire du moteur de recherche *Google*² par exemple) ou bien le résultat d'une requête préalable. La responsabilité de la tâche est partagée entre le système et l'utilisateur : l'utilisateur sélectionne les liens qui l'intéressent, le système calcule et affiche le résultat, et une nouvelle itération commence, jusqu'à la satisfaction de l'utilisateur.

Interrogation

La troisième modalité d'accès à l'information est l'interrogation du système à l'aide de requêtes. Écartons tout d'abord le domaine des Systèmes de Gestion de Bases de Données (SGBD), dans lesquels l'utilisateur accède généralement aux documents via leurs attributs **externes**, comme leur titre, date de création ou nom d'auteur. Les requêtes soumises à un SGBD portent généralement sur ces attributs. De manière complémentaire, les SRI permettent d'accéder aux documents via leurs attributs **internes**, c'est-à-dire par leur contenu. Cependant, la frontière entre SGBD et SRI n'est pas clairement établie, un SGBD pouvant servir de base pour implanter un SRI, en définissant des attributs de type internes au sein

²Google Directory est disponible à l'URL suivante : <http://directory.google.com/>

du SGBD [MP04] ou en utilisant le prédicat *LIKE* en SQL. Les requêtes soumises à un SRI portent sur le contenu des documents, et peuvent prendre la forme d'une expression formulée dans le langage d'interrogation du système, ou bien d'un ou de plusieurs documents de la base lorsqu'il s'agit de **requête par l'exemple** (ou QBE, "Query By Example").

Plus finement, les requêtes formulées dans le langage d'interrogation peuvent être catégorisées [CMM⁺00] selon qu'elles sont thématiques ou précises :

- Les **requêtes thématiques** portent sur une catégorie ou un thème donné, comme : *"photos d'alpinisme"*, *"textes concernant la politique"* ou *"vidéo sur le football"*. Ces requêtes permettent à l'utilisateur de s'informer du domaine général sur lequel porte la requête.
- Les **requêtes précises** regroupe les requêtes descriptives qui décrivent partiellement ou exhaustivement le contenu des documents. Par exemple, dans le domaine de la recherche d'images, il arrive que l'utilisateur recherche une image qu'il connaît en la décrivant. La requête peut alors contenir une description des éléments de l'image, éventuellement leur disposition dans l'image comme : *"photo de Victor Hugo assis, accoudé à une table"*.

Cependant cette terminologie n'est pas définitivement établie. Cox [CMM⁺00] distingue trois classes générales de recherche selon le but de l'utilisateur, qui sont la **recherche ciblée** – qui correspond à une interrogation précise, la **recherche par catégorie** – qui correspond à une interrogation thématique, et la **recherche ouverte** – qui correspond à une navigation dans laquelle besoin d'information de l'utilisateur est susceptible de changer au cours de l'interaction avec le système en fonction des résultats successifs présentés à l'utilisateur.

Discussion

Ces modalités, bien que distinctes, peuvent être combinées au sein d'un même système. Par exemple, l'interrogation et la navigation sont présentées dans [Chi97] comme des modalités d'accès complémentaires pour la recherche d'information, pour combiner les avantages

de chaque modalité.

Des trois modalités d'accès à l'information que sont l'exploration, la navigation et l'interrogation, c'est l'interrogation qui concerne plus particulièrement les SRI [BYRN99]. Un SRI doit être capable de fournir à l'utilisateur ce type de modalité d'accès aux documents de la base. Pour cela, se pose le problème de la représentation interne du contenu des documents et des requêtes ; comme dans d'autres domaines de l'informatique, afin de décrire, d'analyser, d'évaluer et de comparer des SRI, il est nécessaire de les modéliser.

2.2 Modélisation d'un SRI

La figure 2.1 décrit l'architecture générale d'un SRI ; elle présente les éléments principaux mis en jeu dans un SRI.

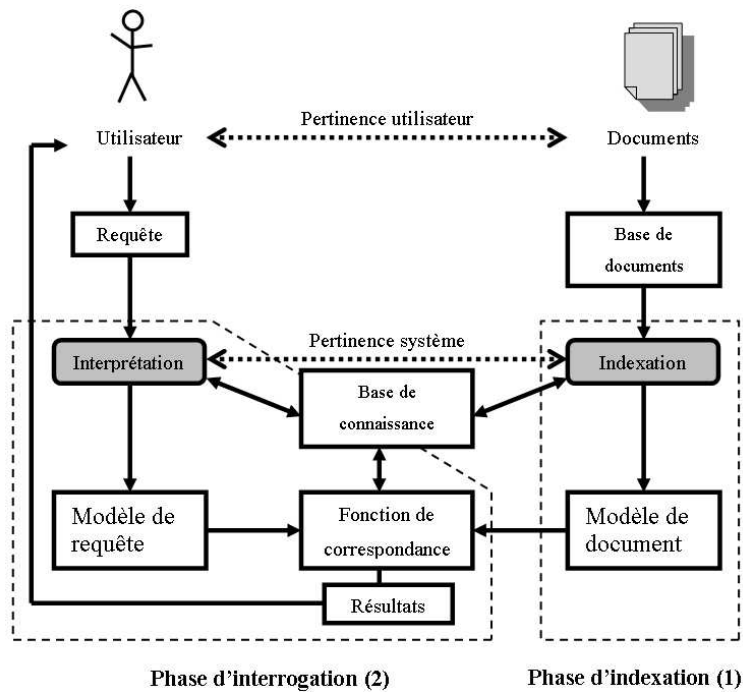


FIG. 2.1 – Architecture d'un SRI.

On distingue classiquement la **pertinence utilisateur**, qui est le jugement apporté par

l'utilisateur sur le document, en fonction de son besoin d'information, et la **pertinence système**, qui correspond à la valeur de correspondance entre le document et la requête, calculée par le SRI. La satisfaction de l'utilisateur est liée à la correspondance entre ces deux pertinences. La figure 2.1 met également en évidence les deux phases de mise en oeuvre d'un SRI, qui sont l'indexation et l'interrogation :

1. L'indexation consiste à extraire et à représenter sous la forme d'un index électronique directement manipulable par le système le **contenu sémantique des documents**. On appelle contenu sémantique du document une représentation du sens associé au contenu brut du document (selon le modèle de requête), qui correspond lui-même au niveau *signal* (c'est-à-dire la simple séquence des chaînes de caractères d'un texte ou la matrice de pixels d'une image).
2. L'interrogation est l'interaction d'un utilisateur final avec le SRI, une fois le contenu sémantique des documents représenté de manière interne sous forme d'index. L'utilisateur exprime son besoin d'information sous la forme d'une requête, qui est interprétée selon le modèle de requête, et le système évalue la pertinence des documents par rapport à cette requête par l'intermédiaire de la fonction de correspondance (et éventuellement d'une base de connaissances). La réponse à ce besoin est la liste des documents qui obtiennent une valeur de correspondance élevée. Cette liste est généralement triée par ordre de valeur de correspondance décroissante – c'est-à-dire du plus pertinent au moins pertinent – et présentée à l'utilisateur.

Les parties 2.2.1 et 2.2.2 portent sur l'indexation et l'interrogation, qui sont les aspects fonctionnels d'un SRI. Ensuite, la partie 2.2.3 décrit les éléments de base sur lesquels ils sont modélisés.

2.2.1 Indexation des documents

La norme AFNOR NF Z 47-102 1996 définit l'indexation de la manière suivante.

Définition 1 (Indexation) *L'indexation est l'opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse.*

Le but de l'indexation d'un document est d'en extraire le sens et de représenter le contenu *nécessaire et suffisant* pour qu'il soit retrouvé par un utilisateur. Autrement dit, un document doit être retrouvé par le système *si et seulement si* son contenu correspond au besoin de l'utilisateur. Ainsi, le choix des descripteurs de documents détermine la qualité de la représentation. Les approches classiques de RI textuelle sont basées sur les mots clés comme descripteurs du contenu des documents ; le contenu d'un document textuel est ainsi exprimé sous la forme d'un ensemble de mots clés jugés représentatifs de ce contenu. Ces mots clés, qui constituent le **vocabulaire d'indexation**, sont un sous-ensemble des mots apparaissant dans les documents. En effet, seuls les mots porteurs de sens sont habituellement retenus dans le vocabulaire d'indexation : les mots outils de la langue comme “*de*”, “*le*”, “*par*”, qui sont recensés dans une base appelée anti-dictionnaire, ne sont pas conservés dans le vocabulaire d'indexation. Par ailleurs, il est courant de “raciniser” (ou lemmatiser) les mots, c'est-à-dire de détecter les variantes orthographiques, comme les pluriels, ou les conjugaisons des verbes, et de ramener les variantes orthographiques d'un même mot à une racine commune [SM83]. Ces traitements sont effectués à l'aide d'algorithmes comme ceux de Lovins [Lov68] ou de Porter [Por80] pour l'anglais, ou celui de Namer [Nam00] pour le français. Par exemple, les mots “*construction*”, “*construirez*” et “*constructible*” ont pour racine commune “*constru*”.

Une technique utilisée pour affiner l'indexation est l'expansion de documents. Elle est notamment utilisée par Singhal et Pereira [SP99] sur des documents textuels issus de transcriptions automatiques de la parole. Les données étant bruitées en raison d'erreurs de reconnaissance (remplacement, délétion ou insertion), la qualité de l'index est affectée, et par conséquent l'efficacité de la recherche est diminuée. Les auteurs étendent les documents

avec les termes apparaissant dans les plus proches voisins d'un document – déterminés à l'aide de la fonction de correspondance, en utilisant la formule de Rocchio [Roc71], et limitent ainsi la perte d'efficacité du système.

Dans un SRI, les descripteurs d'un document peuvent tous être considérés comme ayant une importance égale, c'est-à-dire aussi représentatifs les uns que les autres de la sémantique de ce document. Il s'agit alors d'une pondération booléenne [BYRN99] : seule la présence ou l'absence des termes dans les documents importe. Cependant, en réalité, les descripteurs n'ont généralement pas tous la même importance dans les documents ; afin de pouvoir exprimer ces importances relatives et ainsi affiner l'indexation, les descripteurs sont **pondérés** selon qu'ils sont jugés plus ou moins représentatifs du contenu d'un document. Nous détaillons plusieurs méthodes de pondération au travers du modèle vectoriel de recherche d'information dans la partie 2.5.

2.2.2 Interrogation du système

L'interrogation du système comprend la formulation du besoin d'information d'un utilisateur à travers une requête, la traduction de la requête dans une représentation interne définie par le modèle de requête, et la comparaison de la requête aux index des documents du corpus par la fonction de correspondance, et la présentation des résultats (voir la figure 2.1).

La formulation du besoin d'information correspond à l'expression de la représentation mentale qu'a l'utilisateur, de son besoin d'information. Cette expression varie d'une part avec la connaissance qu'a l'utilisateur du domaine de l'information recherchée, et d'autre part avec son aptitude à définir son besoin. En conséquence, il y a souvent une importante perte d'information entre le besoin et son expression. Cette expression est interprétée selon le modèle de requête. La fonction de correspondance établit une valeur de correspondance de chaque document du corpus pour la requête : l'évaluation de la pertinence système des documents du corpus est effectuée à cette étape cruciale. Les résultats sont l'ensemble des documents ou des références de documents qui obtiennent une valeur de correspondance

élevée (souvent fixée à partir d'une valeur seuil). Les documents retrouvés par le système sont triés par ordre de pertinence décroissante, et présentés à l'utilisateur. La fonction de correspondance a donc pour rôle de favoriser la sélection des documents les plus pertinents pour l'utilisateur.

Éventuellement, le système peut proposer à l'utilisateur de lui indiquer parmi les résultats les documents qu'il estime pertinents et ceux qu'il estime non pertinents, afin de retourner de nouveaux résultats en fonction des indications de l'utilisateur. Ce processus interactif, appelé **bouclage de pertinence**³ (BP) [Roc71, SB90], a pour but d'affiner la recherche et d'améliorer la qualité des résultats. Il est aussi possible d'avoir recours à des processus d'expansion automatique des requêtes, dont nous donnons deux exemples : l'un est basé sur l'utilisation de thésaurus construits automatiquement [QF93] et l'autre utilise un thésaurus construit manuellement [MTT99]

La première approche est celle de Qiu et Frei présentée dans [QF93]. Elle consiste à considérer que le *contexte* d'un mot est composé du mot lui-même, ainsi que de tous les mots qui co-occurrent souvent avec ce mot dans les documents du corpus. Leurs travaux sont basés sur un thésaurus généré automatiquement à partir du corpus. Ce thésaurus exprime des similarités entre les termes d'indexation. Étant donné ce thésaurus de similarité, l'expansion de requêtes consiste en deux étapes :

1. Calculer une similarité entre la requête et chaque terme corrélé à au moins l'un des termes de la requête. Ce calcul se base sur un calcul de cosinus entre le vecteur requête et le vecteur du terme dans le thésaurus.
2. Étendre la requête avec les termes ayant la plus grande similarité avec la requête. Les termes ajoutés sont pondérés en fonction de leur similarité avec la requête originale.

Cette expansion de requêtes permet une amélioration de l'ordre de 20% à 30% de la qualité des résultats sur des collections de test classique de RI textuelle.

³**Relevance feedback** en anglais

La seconde approche est celle de Mandala, Takunaga et Tanaka, qui décrivent dans [MTT99] l'utilisation d'un thésaurus construit manuellement – Wordnet, pour l'expansion automatique de requêtes. Dans ce cas, les similarités entre termes sont calculées à partir des chemins qui mènent d'un terme à un autre, et non à partir de connaissances statistiques du corpus. Une fois ces similarités calculées, le principe est similaire à celui de Qiu et Frei. L'utilisation de ce principe a été effectuée sur un corpus de plusieurs centaines de milliers de documents, et les résultats d'un système vectoriel s'en trouvent améliorés de 10%.

2.2.3 Modélisation

De manière formelle, un modèle de SRI est constitué d'un modèle de documents, un modèle de requêtes, un modèle de correspondance, et une représentation des connaissances, que représente la figure 2.1.

- **Le modèle de documents** (aussi appelé langage d'indexation) exprime le contenu sémantique des documents dans un formalisme de représentation des connaissances (qui peut être très simple comme celui des mots clés, ou beaucoup plus complexe à partir de formalismes de représentation de connaissances). Le choix de ce formalisme est fondamental, car il détermine la qualité de la représentation interne des documents, et donc la qualité de la recherche et des résultats. Cette modélisation doit offrir un compromis entre deux besoins non compatibles, qui sont la compacité de la représentation pour qu'elle puisse être traitée efficacement par un système informatique, et l'expressivité du formalisme pour qu'elle soit précise et exhaustive, afin d'exprimer aussi fidèlement que possible le contenu des documents.
- **Le modèle des requêtes** exprime le contenu sémantique du besoin d'information de l'utilisateur. Ce formalisme détermine la précision de la définition du besoin. Dans la plupart des approches existantes pour le texte, le modèle de requêtes utilise le même formalisme de représentation des connaissances que le modèle de documents pour des raisons de simplicité et de cohérence.

- **Le modèle de correspondance** entre une requête et un document formalise le degré de similarité entre les modèles de la requête et du document. Comme dit plus haut, la fonction de correspondance évalue cette similarité afin de déterminer la pertinence des documents pour cette requête.
- **Une représentation des connaissances** du domaine considéré, afin de satisfaire aux contraintes des modèles de documents et de requêtes, et de prendre en compte des connaissances externes. Cette représentation des connaissances peut inclure un thésaurus composé des termes apparaissant dans l’ensemble des documents, reliés entre eux par des liens de généralité/spécificité ou de synonymie. Par exemple, une requête portant sur “voiture” pourra retourner des documents contenant le terme “automobile”. L’utilisation d’un thésaurus permet d’augmenter le nombre de réponses du système, en incluant dans le résultat des documents contenant des mots reliés aux mots de la requête.

Ces quatre éléments décrivent un **modèle de recherche d’information**. Un SRI est une instance d’un modèle de recherche d’information, et implante donc ces quatre éléments.

2.3 Évaluation et comparaison de systèmes

Des SRI basés sur des modèles différents donnent des résultats différents. Il est difficile de procéder à une démarche analytique formelle pour évaluer différents SRI, en raison du nombre et de la complexité des éléments à prendre en compte. Il est dès lors courant de procéder à une démarche expérimentale d’évaluation – approche dite *boîte noire* – en utilisant une collection de tests, qui contient un ensemble de documents, un ensemble de requêtes, et la liste des documents pertinents pour chacune de ces requêtes établie par des experts ayant une bonne connaissance de la collection et du domaine⁴. Un système est évalué en comparant les réponses du système pour toutes les requêtes de la collection de tests

⁴Soulignons que les collections de tests fournissent des jugements de pertinence binaires : un document est considéré comme pertinent ou non pertinent pour une requête, sans mesure du degré de pertinence.

avec les documents pertinents. L'objectif est d'estimer la correspondance entre la pertinence utilisateur et la pertinence système. En raison des imprécisions inhérentes à la formulation du besoin de l'utilisateur, à la représentation interne des documents et à la mise en correspondance, ces deux pertinences ne sont pas confondues. La comparaison des réponses d'un système pour une requête avec les réponses idéales permet d'évaluer les deux métriques suivantes [Sal71] :

- Le **rappel** mesure la capacité du système à retrouver *tous* les documents pertinents pour une requête q , c'est-à-dire la proportion de documents pertinents retrouvés parmi tous les documents pertinents dans la base :

$$rappel_q = \frac{\|R_q \cap P_q\|}{\|P_q\|}$$

où R_q est l'ensemble des documents retrouvés par le système pour la requête q , et P_q est l'ensemble des documents pertinents de la collection pour cette requête. Cette notion est associée au **silence**, qui correspond à des documents pertinents non retrouvés par le système : $silence_q = 1 - rappel_q$.

- La **précision** mesure la capacité du système à ne retrouver *que* les documents pertinents pour une requête q , c'est-à-dire la proportion de documents pertinents retrouvés parmi tous les documents retrouvés par le système :

$$precision_q = \frac{\|R_q \cap P_q\|}{\|R_q\|}$$

Cette notion est associée au **bruit**, qui correspond à des documents non pertinents retrouvés par le système : $bruit_q = 1 - precision_q$.

Ces valeurs sont calculées pour une requête donnée, mais il est possible de définir la valeur de P_q pour plusieurs requêtes. Ceci n'est possible que si les résultats sont triés (voir la partie 2.2). Il est possible de faire une moyenne de la précision du système pour un échantillon de valeurs de rappel – typiquement 11 valeurs – pour obtenir une mesure scalaire de la qualité du système. Cependant il est plus courant d'établir pour chaque requête une courbe de la précision en fonction du rappel. La moyenne de ces courbes pour un ensemble

de requêtes constitue un profil visuel de la qualité du système. L'avantage est que ce profil visuel informe du comportement du système pour toutes les valeurs de rappel, notamment les petites valeurs de rappel, ce qui correspond aux premiers documents retrouvés.

De plus, il est intéressant de comparer les systèmes en observant les valeurs de rappel et précision pour les quelques premiers résultats comme par exemple la précision après que le système a retrouvé 10 documents, qui indique le nombre de documents pertinents renvoyés par le système dans les 10 premiers résultats.

Un bon système donne de bons taux de précision et de rappel en même temps. Un système dont la précision vaut 1 pour toutes les valeurs de rappel est un système idéal qui retrouve *tous* les documents pertinents, et *uniquement* les documents pertinents – la pertinence système et la pertinence utilisateur sont alors confondues.

D'autres mesures concernant la facilité d'utilisation du système par un utilisateur [CMK66] sont souvent négligées, comme la présentation de l'interface d'interrogation, l'effort cognitif que doit fournir l'utilisateur pour utiliser le système, ou bien la rapidité du système pour fournir une réponse. Le temps de réponse est en effet une caractéristique fondamentale d'un système interactif. Pour que le système soit utilisable, le temps de réponse doit être inférieur à 10 secondes [Nie94]. Au delà, l'attention de l'utilisateur n'est plus focalisée sur le système. Une étude de Bouch, Kuchinsky et Bhatti sur la qualité des services Web [BKB00] stipule qu'au delà de 8 ± 2 secondes de temps de réponse, les utilisateurs considèrent que le système est lent. Cette nécessité de bonnes performances quantitatives est une raison qui motive nos choix dans l'utilisation du modèle vectoriel de RI.

2.4 Le modèle vectoriel

Les approches de RI textuelle développées depuis 30 ans [Sal71] se basent sur les mots clés utilisés comme descripteurs du contenu des documents. L'énorme avantage de cette approche est qu'elle peut être automatisée, donc rapide et applicable à de très grands corpus, comme nous le constatons avec les chercheurs sur le Web qui sont capables de retrouver

des documents parmi des centaines de millions en quelques secondes [BP98]. Les mots clés des documents représentent du sens de manière intrinsèque. Par exemple, un document dans lequel apparaît le mot *alpinisme* est considéré comme un document *à propos* d'alpinisme. Par conséquent, ce document est pertinent pour une requête qui porte sur l'*alpinisme*. Nous mettons ici en évidence le fait que la base de travail d'un SRI textuel est la même que celle de l'utilisateur : les mots de la langue naturelle. Les mots clés sont extraits automatiquement, et pondérés automatiquement pour leur donner plus ou moins d'importance selon qu'ils sont plus ou moins représentatifs du contenu des documents. Les différents modèles de RI textuelle dits *classiques* parmi lesquels figurent les modèles booléen, vectoriel et probabiliste, sont basés sur cette même notion de vocabulaire d'indexation à base de mots clés. La différence se situe dans les modèles de documents et de requêtes, et dans l'évaluation de la correspondance entre un document et une requête. Pour les lecteurs intéressés, l'annexe A décrit brièvement les modèles booléen, probabiliste et logique. Nous présentons ci-dessous le modèle vectoriel, qui est la base de nos travaux, en le décrivant selon les modèles de documents, de requêtes, de correspondance et de connaissances.

Le modèle vectoriel [Sal71] représente les documents et les requêtes par des vecteurs d'un espace à n dimensions, les dimensions étant constituées par les termes du vocabulaire d'indexation. L'index d'un document d_j est le vecteur $\vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{n,j})$, où $w_{i,j} \in [0, 1]$ dénote le poids du terme t_i dans le document d_j . Le poids d'un terme dénote son intérêt dans le document. Cet aspect du modèle fait l'objet de la partie suivante 2.5. Une requête est également représentée par un vecteur $\vec{q} = (w_{1,q}, w_{2,q}, \dots, w_{n,q})$, où $w_{i,q} \in [0, 1]$ est le poids du terme t_i dans la requête q . La figure 2.2 montre un exemple d'espace vectoriel composé des trois termes t_1, t_2 et t_3 . Les index de deux documents D_1 et D_2 et une requête Q sont représentés dans cet espace.

La fonction de correspondance mesure la similarité entre le vecteur requête et les vecteurs documents. Une mesure classique est le cosinus de l'angle formé par les deux vecteurs :

$$\text{correspondance}(d_j, q) = \cos(\vec{d}_j, \vec{q}) \quad (2.1)$$

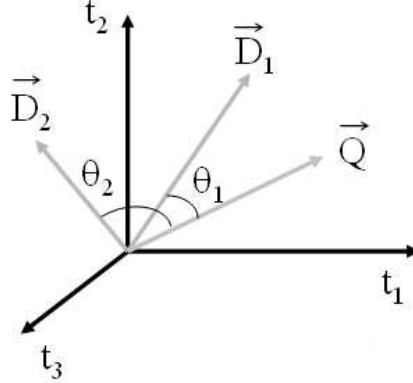


FIG. 2.2 – Représentation vectorielle de deux documents (D_1 et D_2) et d’une requête (Q) dans un espace composé de trois termes (t_1 , t_2 et t_3).

où $\cos(\vec{d}_j, \vec{q})$ est le cosinus de l’angle formé par les vecteurs \vec{d}_j et \vec{q} :

$$\cos(\vec{d}_j, \vec{q}) = \frac{\vec{d}_j \cdot \vec{q}}{\|\vec{d}_j\| \cdot \|\vec{q}\|} = \frac{\sum_{i=1}^n w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^n w_{i,j}^2} \times \sqrt{\sum_{i=1}^n w_{i,q}^2}}$$

Plus deux vecteurs sont similaires, plus l’angle formé est petit, et plus le cosinus de cet angle est grand. Dans l’exemple de la figure 2.2, le document D_1 est plus similaire à la requête que le document D_2 . A l’inverse du modèle booléen, la fonction de correspondance évalue une correspondance partielle entre un document et une requête, ce qui permet de retrouver des documents qui ne satisfont la requête qu’approximativement. Les résultats peuvent donc être ordonnés par ordre de pertinence décroissante. La formulation de requêtes sous la forme d’expressions logiques comme “*documents concernant le sport, mais pas le football*” n’est toutefois pas possible dans le modèle vectoriel.

Eventuellement, le modèle peut intégrer un thésaurus dans la base de connaissances (voir la partie 2.2.3) afin de tenir compte des relations de synonymie entre les mots clés lors de l’évaluation de la correspondance, et ainsi augmenter le nombre de réponses.

Le modèle vectoriel est fondé sur une hypothèse implicite d’indépendance des mots clés, donc d’orthogonalité des dimensions de l’espace. En effet, l’équation 2.1 ne prend pas en

compte les dépendances éventuelles entre des mots clés, ce qui a fait l'objet de critiques du modèle [RW86]. Cependant, dans la pratique, il semble que la prise en compte des dépendances ne permette pas d'améliorer notablement la qualité du modèle [BYRN99] : en raison du caractère local de ces dépendances, le fait de les considérer de manière globale pour la base documentaire est susceptible, au contraire de l'effet attendu, de dégrader le modèle. Par ailleurs, une autre critique du modèle vectoriel concerne l'absence de base théorique forte dans la représentation des documents et des requêtes, et de la fonction de correspondance, contrairement par exemple au modèle probabiliste (voir l'annexe A). Cependant, lors de différentes campagnes d'évaluation comme TREC, le modèle vectoriel a donné de bons résultats à la fois :

- **qualitatifs** : le modèle vectoriel est au moins aussi bon en qualité que les autres modèles classiques [BYRN99],
- **quantitatifs** : les performances en temps de réponse et la qualité des résultats restent bonnes même quand le nombre de dimensions est grand (plusieurs milliers) et le nombre de documents est grand (plusieurs centaines de milliers).

C'est pour ces raisons que nous avons choisi d'utiliser le modèle vectoriel dans nos travaux.

2.5 Pondération des termes dans le modèle vectoriel

Le poids d'un terme dénote son intérêt dans l'index, qui dépend du fait que le terme est important dans le document, et du fait que le terme permet de distinguer un document des autres dans la base. Les méthodes de pondération les plus largement utilisées pour le texte sont basées sur des variantes de la formule $tf.idf$ (ou *term frequency* \times *inverted document frequency*) [Sal71, SM83, SB88, BYRN99]. La valeur tf mesure l'importance d'un terme dans un document. Habituellement, la valeur tf est définie en fonction du nombre d'occurrences du terme dans le document, et peut être normalisée pour ne pas donner plus d'importance à

un document long qu'à un document court⁵. Il existe différentes manières de calculer le *tf* d'un terme, et de normaliser sa valeur par rapport à la taille du document.

Exemple 2 *Le tf d'un terme t_i dans un document d_j peut être défini comme :*

$$tf_{i,j} = \frac{freq_{i,j}}{\max_k freq_{k,j}} \quad (2.2)$$

où $freq_{i,j}$ est le nombre d'occurrences du terme t_i dans le document d_j , et $\max_k freq_{k,j}$ est le nombre d'occurrences du terme le plus fréquent dans le document d_j .

La valeur *idf* est une mesure de la discriminance du terme [SWY75, SB88] (aussi appelée *pouvoir de résolution* [vR79], ou *informativité*), c'est-à-dire de la capacité du terme à différencier les documents du corpus. Plus précisément, la notion de discriminant renvoie à la qualité d'un terme qui distingue bien un document des autres documents. Un terme qui apparaît seulement dans un petit nombre de documents permet bien de discriminer les documents. Au contraire, un terme qui apparaît dans un grand nombre, voire dans tous les documents d'une collection n'est pas discriminant en raison de sa distribution uniforme.

Exemple 3 *La formule la plus répandue de calcul de la discriminance d'un terme t_i est la suivante [Sal71] :*

$$idf_i = \log \left(\frac{N}{n_i} \right) \quad (2.3)$$

où N est le nombre total de documents dans le corpus et n_i le nombre de documents dans lesquels le terme t_i apparaît. Lorsque ce terme apparaît dans beaucoup de documents, la valeur idf_i est petite. Elle appartient à l'intervalle $[0, \log(N)]$.

Cette méthode de pondération est donc fondée sur deux sources, l'une représentant l'importance du terme dans le document, l'autre représentant la discriminance du terme dans la collection. Les méthodes de pondération les plus courantes et les plus efficaces pour les documents textuels [Sal71, SM83, SB88, BYRN99] utilisent une combinaison des *tf* et *idf*

⁵Cependant une telle normalisation est inutile quand la fonction de correspondance est le cosinus, car cette fonction ne considère que l'angle entre les vecteurs, indépendamment de leur longueur.

comme dans les formules $w_{i,j} = tf_{i,j} \times idf_i$ pour les documents, et $w_{i,q} = idf_i$ pour les requêtes. Il existe d'autres approches pour évaluer la discriminance d'un terme t_i , parmi lesquelles nous citons les deux suivantes :

- Une approche consiste à calculer la similarité moyenne deux à deux des documents avec et sans t_i , la différence de ces deux valeurs indique si le terme t_i aide ou non à séparer les documents les uns des autres. Lorsque c'est le cas, le terme t_i obtient une grande valeur de discriminance. Salton, Wong et Yang présentent dans [SWY75] une étude de la corrélation entre la qualité de l'indexation et la densité de l'espace. L'idée des auteurs est de minimiser dans l'espace d'indexation la similarité entre les documents, augmentant ainsi la précision du résultat. En effet, quand la différence moyenne entre les documents est grande, un document retrouvé par le système est retourné à l'utilisateur sans nécessairement retourner également les documents considérés abusivement comme étant dans le voisinage de ce document.
- L'autre approche est basée d'une part sur la loi de Zipf [Zip32, Zip49], et la conjecture de Luhn [Luh57].

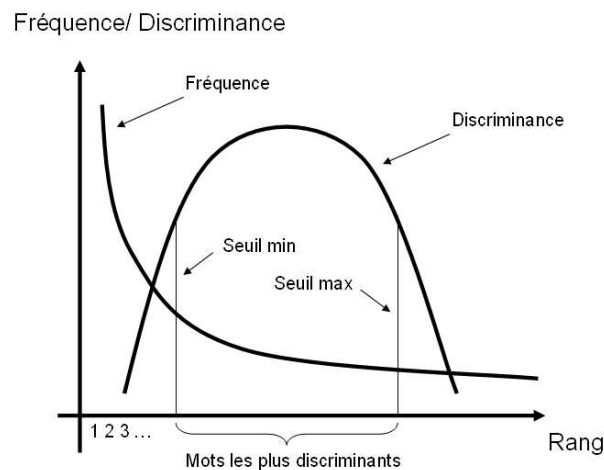


FIG. 2.3 – Loi de Zipf et conjecture de Luhn.

La loi de Zipf énonce que le produit de la fréquence d'apparition d'un descripteur

par son rang est approximativement constant : $frequence \times rang = constante$, ce qui traduit une fonction hyperbolique, comme le montre la figure 2.3. La conjecture de Luhn stipule qu'il y a une relation entre le rang d'un terme et sa discriminance.

Les descripteurs les moins discriminants sont ceux de rangs faibles – très fréquents, et de rangs élevés – très rares ; les descripteurs les plus discriminants sont ceux des rangs médians. Ce sont ces termes qu'il faut sélectionner et conserver dans le vocabulaire d'indexation.

Nous avons décrit différents aspects de la RI textuelle au travers du modèle vectoriel. Ce modèle est fondé sur les mots clés qui expriment par eux-mêmes la sémantique des documents qu'ils caractérisent. A l'inverse, la sémantique exprimée dans les images n'est ni explicite, ni directement manipulable par des systèmes informatiques. La recherche d'images nécessite des modèles et systèmes adaptés, que nous abordons dans le chapitre suivant.

Chapitre 3

... à la recherche d'images

La problématique de la Recherche d'Images (RIm) découle directement de celle de la RI. Il s'agit de retrouver des images pertinentes pour satisfaire les besoins d'un utilisateur. De manière analogue à un SRI textuel, un SRIm est modélisé par un modèle de représentation des documents, un modèle de requêtes, un modèle de connaissance le cas échéant, et un modèle de correspondance. Nous présentons dans ce chapitre un état de l'art en RIm en distinguant les approches signal et les approches symboliques. Nous présentons l'approche signal, qui est basée sur des caractéristiques de bas niveau des images, et montrons ses limites, notamment dans un contexte d'images personnelles. Nous présentons ensuite les approches symboliques, qui intègrent une sémantique associée aux pixels des images. Cette sémantique est intégrée par le biais de mots clés qui décrivent le contenu des images, ou en se basant sur des formalismes plus riches, comme celui des graphes conceptuels, qui constituent un cadre formel utile pour la représentation et la recherche d'images.

3.1 Signal et sens : le fossé sémantique

Du fait de la distance entre le signal et le sens, il est difficile d'extraire automatiquement le contenu sémantique des images [SWS⁺00]. Comme nous l'avons décrit précédemment, le fossé sémantique est l'écart qui sépare les pixels de l'image de l'interprétation du sens associé à ces pixels. Il découle de cet écart deux classes d'approches de RIm : les approches

basées sur le **signal** qui considèrent uniquement le niveau pixel des images, et les approches basées sur les **symboles** qui intègrent une interprétation des pixels de l'image.

3.2 Les approches signal

Les SRIm de type **signal** considèrent les pixels bruts de l'image, et leurs caractéristiques de bas niveau. Ces approches ont l'avantage de pouvoir gérer automatiquement de grandes collections d'images hétérogènes, dans un domaine large, c'est-à-dire qui présente une grande variabilité [SWS⁺00, VT00]. Nous présentons cette approche selon les modèles de documents, de requêtes, et de correspondance, en s'appuyant sur des exemples de systèmes existants. Par rapport à la description des modèles faite dans la partie 2.2.3, nous ne nous intéressons pas au modèle de connaissance car il n'apparaît pas dans ces systèmes.

3.2.1 Modèle de documents

Des *caractéristiques* – ou *traits* – de bas niveau, classiquement des informations de couleur, de texture et de forme [SWS⁺00], sont extraits automatiquement des images. Les couleurs sont représentées par des histogrammes qui représentent la distribution de couleur de l'image dans des espaces de couleurs tel que RGB (*Red, Green, Blue*), CMYK (*Cyan, Magenta, Yellow, black*), HSV (*Hue, Saturation, Value*) ou Luv (*Luminance et chrominance*). Les textures sont des propriétés définies sur des surfaces et non sur des pixels comme les couleurs. Elles modélisent des variations locales de l'intensité lumineuse dans plusieurs directions et à différentes échelles [CPW98, SS02]. Il existe de nombreuses méthodes pour extraire et représenter les textures. Les méthodes statistiques figurent parmi plus utilisées. Elles sont fondées sur des analyses des matrices de co-occurrences, desquelles sont extraites des caractéristiques comme l'énergie, l'entropie, le contraste, l'homogénéité ou la corrélation [CPW98]. Les formes sont exprimées par des fonctions mathématiques comme des transformées de Fourier ou des coefficients d'ondelettes. Ces différentes classes de traits permettent la construction d'autant d'index au travers d'un processus d'indexation qui peut

être entièrement automatique.

Descriptions globales des images

Lorsque les traits définis dans le modèle de document portent sur l'image entière, les index décrivent les caractéristiques de l'image de manière globale. C'est le cas des vecteurs de couleur moyenne de **QBIC**¹, un système conçu par Niblack et Flickner [NBE⁺93, FSN⁺95] pour IBM, et qui est historiquement un des premiers dans le domaine. Un vecteur de couleur moyenne est un vecteur à 3 dimensions représentant les composantes de couleur de l'image dans l'espace RGB de la totalité de l'image. C'est également le cas de **PicHunter** [CMM⁺00], système conçu par Cox au *NEC Research Institute*, qui représente les images avec des histogrammes globaux de couleur.

Cependant, considérer la totalité de l'image ne permet pas de conserver les informations locales de couleur et de texture. Afin de conserver ces informations, il est nécessaire de représenter individuellement les caractéristiques selon les parties de l'image. Pour cela, les images doivent être segmentées, c'est-à-dire subdivisées en **régions**, qui sont indexées individuellement.

Descriptions locales des images

Les régions sont des ensembles de pixels contigus qui forment une partition de l'image : tous les pixels de l'image appartiennent à une région, et les régions ne se chevauchent pas. Les régions sont des parties de l'image cohérentes selon une ou plusieurs caractéristiques de niveau signal, idéalement correspondant aux projections 2-D des objets réels (cas de segmentation *consistante*). Il est difficile dans la réalité de segmenter parfaitement une image de manière consistante – même manuellement, en raison d'ambiguïtés dans les limites des objets. Les systèmes de segmentation automatiques délimitent les contours des régions homogènes selon différentes méthodes.

¹QBIC = *Query By Image Content*.

Une méthode de segmentation consiste à délimiter les régions en utilisant un seuil de valeur pour le trait considéré (majoritairement la couleur). Ce seuil est choisi a priori ou dépendant d'un histogramme de ce trait pour l'image entière. Typiquement, le système **VisualSEEK** [SC96] développé à l'Université de Columbia par Smith et Chang segmente les images en régions homogènes de même couleur dominante ("*blob*" en anglais). D'autres systèmes intègrent également la texture dans le processus de segmentation, comme **SIMPLIcity**² [WJW01] conçu par Wang à Stanford. Ce système découpe dans un premier temps les images en blocs de taille fixe de 4×4 pixels, décrits individuellement par des vecteurs de traits de dimension $6 : 3$ pour la couleur moyenne du bloc dans l'espace de couleur Luv, et 3 pour l'énergie en bandes hautes fréquences des transformées en ondelettes. Ces blocs sont ensuite regroupés en classes par l'algorithme *k-means*, chaque classe correspondant à une région. Un autre exemple est le cas de **Blobworld** [CTB⁺99] développé par Carson à Berkeley, qui implante une segmentation des images fondée sur l'algorithme EM (*Expectation-Maximization*) combiné sur des traits de couleur et de texture. Un vecteur de trait à 6 dimensions (comme SIMPLIcity : 3 pour la couleur et 3 pour la texture) est calculé pour chaque pixel de l'image, et l'algorithme EM estime, par maximum de vraisemblance, les paramètres d'une mixture de K gaussiennes dans l'espace de traits, K étant le nombre de régions de l'image. Ainsi, le modèle fournit, pour chaque pixel, un estimé de la région à laquelle il appartient. Le système **NeTra** [MM99] de Ma et Manjunath segmente les images à l'aide d'une technique originale qui consiste à estimer la direction du changement de la couleur et de la texture à différents emplacements dans l'image et à différentes échelles, afin de détecter les bords des objets. Cette technique, appelée *flot de contour*, est décrite en détails dans [MM97].

Une fois les images segmentées, les descripteurs associés aux régions définies sont intégrés dans les index des images. De plus, des informations relatives aux *formes* des régions, à leurs tailles, ainsi qu'à leur disposition spatiale (position de leur barycentre) sur l'image

²SIMPLIcity = *Semantics-sensitive Integrated Matching for Picture Libraries*.

peuvent également être intégrées dans les index.

Pondération d'attributs de type signal

Le sens des valeurs numériques contenues dans les index est fondamentalement différent des poids des termes en RI (voir la partie 2.5). Nous pouvons faire une analogie entre les vecteurs ou histogrammes de caractéristiques et les vecteurs du modèle vectoriel. Les valeurs numériques associées aux traits de bas niveau représentent uniquement des quantités de ces traits, et ne symbolisent pas une *importance sémantique* dans l'image. Par exemple, la quantité de rouge d'une image ou d'une région exprimée dans un histogramme de couleurs représente uniquement la quantité relative de cette composante de couleur, et non l'importance de cette couleur pour décrire l'image ou la région. Une autre différence importante est que la notion d'*idf* n'est pas intégrée. Notons toutefois qu'une normalisation des dimensions de l'espace de couleurs RGB est effectuée dans QBIC : chaque composante d'un vecteur de couleur moyenne est divisée par l'écart-type des valeurs de cette composante pour toute la collection d'images. Ainsi le vecteur de couleur moyenne du document i est représenté sous la forme suivante :

$$\vec{v}_i = \left(\frac{R_i}{\sigma_R}, \frac{G_i}{\sigma_G}, \frac{B_i}{\sigma_B} \right)$$

où R_i , G_i et B_i sont respectivement les composantes R, G et B de l'image, et σ_R , σ_G et σ_B sont respectivement les écarts-types des composantes R, G et B pour toute la collection. Cette normalisation des dimensions est une réduction de l'écart-type qui vise à homogénéiser les distributions des valeurs pour les trois composantes. Notons également que SIMPLICITY [WJW01] se distingue des autres systèmes par la mise en œuvre d'une pondération inspirée du $tf \times idf$ du texte, décrite dans [WD01]. Des poids sont attribués aux régions en fonction de leur fréquence dans l'image et dans la collection. Cette pondération a été évaluée sur une tâche de classification d'images en 10 catégories sémantiques comme "*plage*", "*montagne*", ou "*bâtiments*". Les résultats montrent que la qualité de la classification est meilleure qu'avec de simples histogrammes de couleur. La différence fondamentale avec

ces histogrammes est la prise en compte de la distribution des régions dans la collection, ce qui rapproche SIMPLiCity de la RI sémantique.

Nous venons de présenter différents systèmes dans lesquels les images sont segmentées selon les couleurs et les textures. La représentation interne des documents dans les SRIm de type signal est essentiellement basée sur des caractéristiques de bas niveau : ce sont ces représentations que la fonction de correspondance confronte à la représentation interne d'une requête pour évaluer la pertinence système des documents.

3.2.2 Modèle de requêtes

Les requêtes sont représentées de la même manière que les documents, par des caractéristiques de bas niveau. L'interface d'interrogation doit permettre aux utilisateurs de spécifier une requête. Cela se fait généralement de manière non verbale, ce qui constitue un aspect important des SRIm. En effet, cela pose le problème de la facilité pour l'utilisateur de définir précisément ses besoins à travers l'interface d'interrogation. Selon le cas, l'utilisateur peut spécifier directement les attributs de bas niveau de l'image cible dans sa requête, interroger le système en esquissant un croquis, ou bien en présentant au système une image exemple de ce qu'il recherche.

Spécification directe de traits

L'interrogation du système par la spécification directe des caractéristiques visuelles des images recherchées est une approche de bas niveau : c'est la manière de formuler une requête la plus proche du système, et donc la plus éloignée de la représentation mentale de l'utilisateur. L'interface de QBIC permet de choisir les quantités relatives de rouge, vert et bleu, ou bien de sélectionner une couleur voulue depuis une palette de couleur. L'interface permet également de choisir une texture depuis un échantillon proposé. Ainsi, il est possible d'interroger QBIC à travers une requête du type "*images contenant approximativement 30%*

de rouge et 15% de bleu”, en précisant ces valeurs au système.

Ce mode d’interrogation est très limitant car les caractéristiques spécifiées portent sur l’image entière. Le fait que les caractéristiques soient globales ne permet pas à l’utilisateur de préciser leurs valeurs pour différentes zones de l’image ; il en découle un important problème de bruit dans les réponses.

Requête par croquis

A un niveau d’abstraction plus haut, l’utilisateur spécifie une requête en dessinant un croquis (*“sketch”* en anglais) à l’aide d’outils simples de dessin fournis par l’interface d’interrogation. Il peut ainsi spécifier sa requête en termes de rectangles et ovales de différentes couleurs et textures, et également indiquer leur disposition et leur taille sur l’image. Le système calcule alors un ensemble de traits à partir de ces spécifications.

Exemple 4 *Imaginons un utilisateur qui recherche des images de coucher de soleil sur la mer avec le système VisualSEEk. Il doit dessiner un rectangle bleu dans la partie basse pour la mer, un rectangle orange dans la partie haute pour le ciel, et un disque rouge/orange au centre pour le soleil couchant. Il peut également préciser les textures pour chaque région dessinée.*

Bien qu’illustratif, cet exemple de coucher de soleil est trivial, car il contient des grosses régions aux couleurs typiques, avec un petit niveau de détail. Lorsqu’il s’agit d’images plus complexes, le dessin d’un croquis devient une tâche longue et fastidieuse, et requiert des compétences artistiques chez l’utilisateur.

Requête par l’exemple

L’utilisateur doit fournir au système une ou plusieurs images en exemple de ce qu’il recherche. Les images peuvent ou non appartenir à la collection stockée dans le système. Ce dernier définit la requête comme une combinaison des caractéristiques signal des images

données en exemples. Les systèmes **AMORE**³ [MHH99] du *NEC Research Institute*, **MARS**⁴ [HMR96] de l'Université de l'Illinois, sont centrés sur ce type de requêtes.

La requête par l'exemple permet de s'abstraire du problème de la formulation de la requête. La requête étant constituée d'images, l'utilisateur ne se soucie pas de la manière de formuler son besoin d'information : il est implicite et le système doit le deviner en fonction des exemples fournis. Cette modalité d'interaction, bien qu'extrêmement simple pour l'utilisateur, pose les trois problèmes suivants :

1. L'information implicite est délicate à extraire. Tout le document donné en exemple n'est pas forcément pertinent, et l'utilisateur ne peut pas spécifier ce qui, dans l'image exemple, est pertinent pour la recherche ; la requête par l'exemple ne tient pas compte de cela. Pour cette raison, la requête par l'exemple associée au bouclage de pertinence est une technique fréquemment utilisée dans les SRIIm de type signal, comme PicHunter, qui implante un BP basé sur une approche probabiliste bayésienne. Cependant, Koenemann et Belkin ont montré dans [KB96] la nécessité d'un BP *transparent* (affichage de la requête générée par le système) et *pénétrable* (possibilité de modifier la requête générée avant qu'elle ne soit évaluée) pour améliorer la qualité des résultats. Or les approches signal ne fournissent généralement qu'un BP *opaque* [KB96].
2. La requête par l'exemple implique que l'utilisateur doive déjà avoir à disposition une image exemple satisfaisante de ce qu'il cherche. Or ce n'est pas toujours le cas. Les images exemples sont généralement obtenues à la suite d'une première interrogation à l'aide d'une autre modalité.
3. Le problème sans doute le plus important, et c'est également le cas pour les autres modalités d'interrogation, est que les requêtes sont exprimées par des caractéristiques de bas niveau, et le lien avec la sémantique n'est jamais explicitement établi. La fonction de correspondance se base uniquement sur ces représentations pour retrouver des

³AMORE = *Advanced Multimedia Oriented Retrieval Engine*.

⁴MARS = *Multimedia Analysis and Retrieval System*.

images.

3.2.3 Modèle de correspondance

La sélection d'images pertinentes pour une requête consiste à évaluer la similarité entre les représentations internes de la requête et des documents. Les modèles de documents et de requêtes étant basés sur des caractéristiques de bas niveau, cette correspondance est donc uniquement visuelle. Les similarités sont évaluées pour chaque caractéristique, puis combinées en une seule valeur de similarité globale. La combinaison privilégie éventuellement certaines caractéristiques jugées plus importantes selon les choix du concepteur du système ou de l'utilisateur. Diverses mesures de similarité entre vecteurs et histogrammes sont employées. Comme les vecteurs, les histogrammes définissent des points dans des espaces n -dimensionnels. Les mesures de similarité évaluent des distances entre ces points. Une distance est une fonction mathématique d de deux points x et y de l'espace, qui satisfait les propriétés suivantes :

- Non négativité : $d(x, y) \geq 0$
- Réflexivité : $d(x, y) = 0 \Leftrightarrow x = y$
- Symétrie : $d(x, y) = d(y, x)$
- Inégalité triangulaire : $d(x, y) + d(y, z) \geq d(x, z)$

Les distances de vecteurs sont basées sur une L_k -norme (ou distance de Minkowski) de la forme :

$$d(x, y) = \left(\sum_i |x_i - y_i|^k \right)^{\frac{1}{k}}$$

où $k \geq 1$ est le paramètre de la norme. Quand $k = 1$, la L_1 -norme, qui est appelée distance de Manhattan ou distance *City-Block* est la somme des différences. Quand $k = 2$, la L_2 -norme est la distance euclidienne. Par exemple, QBIC compare les vecteurs de couleur moyenne par une distance euclidienne. La distance euclidienne est également utilisée par Blobworld pour calculer la distance entre les régions. Des mesures plus spécialement dédiées aux histogrammes de couleurs et de textures intègrent des comparaisons inter-dimensions, comme

la distance quadratique :

$$d(x, y) = (x - y)^T A(x - y)$$

où $(x - y)^T$ dénote le vecteur transposé, et A est la matrice symétrique de similarité entre les dimensions. À l'opposé de la distance de Minkowski, la distance quadratique donne une similarité non nulle pour deux teintes de couleurs proches. Elle a donc l'avantage de représenter une comparaison perceptuellement plus juste. Les images sélectionnées par le système sont généralement présentées à l'utilisateur par ordre décroissant de pertinence système.

3.2.4 Discussion

Nous rappelons que le problème majeur que rencontrent les approches basées sur le signal vient de l'absence totale de sémantique dans la représentation des images et des requêtes. Par conséquent, l'évaluation de la similarité est uniquement d'ordre visuel. Cette similarité visuelle définit la pertinence système, et c'est donc à l'utilisateur de s'adapter au niveau de représentation du système.

Dans l'exemple précédent dans lequel un utilisateur est à la recherche de couchers de soleils, il est possible que le système lui retourne une image représentant une orange sur une table bleue sur fond orange, qui est visuellement similaire au croquis qu'il a dessiné, mais sémantiquement différente de l'image qu'il recherche. Cette limitation de l'expressivité des requêtes est due à la polysémie des représentations visuelles, c'est-à-dire à leur ambiguïté : deux représentations visuelles peuvent être similaires au niveau signal, tout en dénotant des sémantiques très différentes.

Pour répondre aux attentes des utilisateurs vis-à-vis de systèmes de gestion de collections de photos personnelles, les approches symboliques intègrent une sémantique dans la représentation des images et l'interrogation du système, en élevant ainsi leur niveau d'abstraction à celui des utilisateurs.

3.3 Les approches symboliques basées sur les mots clés

Les approches symboliques intègrent le contenu sémantique des images, le plus souvent le sens qu'elles dénotent. Un grand nombre de ces approches utilise des mots clés, comme c'est le cas pour les documents textuels. Des représentations plus riches sont également utilisées afin de permettre une description et une recherche plus fine des images.

Les images sont indexées par des mots clés qui représentent leur contenu sémantique. Comme pour les caractéristiques de bas niveau, les mots clés peuvent porter sur l'image entière ou bien sur les régions quand l'image est segmentée.

Mots clés globaux

Nous distinguons ici deux approches, selon que les mots clés proviennent ou non d'une analyse automatique de l'image. Pour indexer une image sans analyser son contenu, il est possible de faire appel à des informations textuelles externes à l'image elle-même. Le système **Chabot** [OS95] utilise une description textuelle existante des images de la collection, et intègre d'autres informations textuelles comme la date de prise de vue, la localisation de la scène, et le plan de vue. Il est possible d'utiliser également le *contexte* de l'image sur une page Web, c'est-à-dire le texte qui l'accompagne. C'est le cas du moteur de recherche d'images *Google Image*⁵. Ces approches sont très similaires à de la RI textuelle. Aslandogan, Thier, Yu Zou et Rishe [ATY⁺97] ont étudié l'impact de l'utilisation du thésaurus Wordnet dans un SRIm. Ils ont établi une mesure de similarité entre termes provenant d'une hiérarchie *EstUn*, en fonction du nombre d'arêtes séparant les mots dans la hiérarchie. La fonction de similarité entre un terme de requête t_q et un terme de document t_d est :

$$sim(t_q, t_d) = \frac{idf_{t_q}}{distance(t_q, t_d) + 1}$$

⁵Google Images est disponible à l'URL suivante : <http://images.google.fr/>.

où idf_{t_q} est la fréquence documentaire inverse de t_q (voir la partie 2.5) et :

$$distance(t_q, t_d) = \begin{cases} 0 & \text{si } t_q \text{ et } t_d \text{ sont identiques} \\ 1 & \text{si } t_q \text{ et } t_d \text{ sont synonymes} \\ \text{nombre d'arêtes} & \text{si } t_q \text{ et } t_d \text{ sont reliés dans la hiérarchie } EstUn \end{cases}$$

Cette approche est intéressante dans le sens où, de la même manière que la distance quadratique des histogrammes, elle permet de calculer une similarité inter-dimensions.

Le système PicHunter intègre également des annotations textuelles en plus d'une description de bas niveau des images. Des mots clés comme "*ciel*", "*arbre*" ou "*cheval*" sont associés aux images par le biais d'un vecteur booléen de dimension 138. L'interrogation étant essentiellement du type requête par l'exemple, ces descriptions textuelles sont transparentes pour l'utilisateur : elle sont utilisées de manière interne au système durant le BP. Les mots clés peuvent également provenir d'une analyse automatique de l'image. Vailaya, Figuieredo, Jain et Zhang [VFJZ99] ont utilisé des classifieurs bayésiens pour catégoriser des images de vacances de manière hiérarchique, dans des classes comme *intérieur/extérieur*, les images de la classe *extérieur* sont réparties en *ville/paysage*, et les images de la classe *paysage* sont réparties *coucher de soleil/forêt/montagne*. Ces mots clés globaux permettent essentiellement de catégoriser les images, mais ne permettent cependant pas une description précise des éléments de l'image.

Mots clés locaux

Les mots clés peuvent être associés aux régions des images lorsque ces dernières sont segmentées. Les systèmes d'indexation automatique des images avec des mots clés sont majoritairement basés sur des algorithmes adaptatifs. Ils extraient automatiquement des mots clés des régions de l'image à l'aide d'algorithmes d'apprentissage. Par exemple, les *Visual Keywords* (VK) de Lim [Lim00] sont des prototypes visuels extraits d'échantillons d'images, qui sont utilisés pour catégoriser des blocs d'images de 20×20 pixels. Des vecteurs de mots clés sont associés à ces blocs après un processus d'apprentissage par entraînement d'un réseau de neurones artificiels. Les VK se rapprochent des catégories visuelles

de Town et Sinclair [TS00, TS01], qui sont des classifications de régions d'image par des réseaux de neurones artificiels. Bissol, Mulhem et Chiaramella ont présenté dans [BMC03] une approche de recherche d'images personnalisable, similaire à celle de l'outil *FourEyes* de Photobook [PPS94] : les mots clés représentant des objets et des catégories d'images peuvent être définis dynamiquement en donnant au système des exemples positifs et négatifs. Lorsque les mots clés décrivent les objets, ils portent sur les régions de l'image et non sur l'image totale.

De la même façon qu'en RI textuelle, les mots clés peuvent être pondérés pour refléter leur importance. Ounis propose dans [Oun99] de pondérer les objets images en fonction de paramètres perceptifs qui traduisent la facilité d'un utilisateur à percevoir les objets images :

- la taille de l'objet, c'est-à-dire sa surface relative sur l'image,
- le contraste de l'objet, par rapport au reste de l'image,
- le niveau de composition de l'objet, qui indique si l'objet est une partie d'un autre objet,
- l'occlusion éventuelle de l'objet,
- la position de l'objet, représentée par le barycentre de la région,
- le nombre d'occurrences de l'objet, qui est analogue aux *tf* des mots clés du texte.

Ce papier définit l'importance d'un objet image comme une moyenne pondérée des valeurs de ces paramètres. Nos travaux s'inspirent de cette pondération, en reprenant et en étendant notamment les paramètres de taille, de position, et de nombre d'occurrences.

Si le formalisme des mots clés est satisfaisant pour représenter le contenu de documents textuels, les images nécessitent des représentations plus riches : Ounis a montré dans [OP97] l'intérêt de l'intégration des relations sémantiques entre les mots clés qui indexent les images. Les mots clés ne permettent pas la prise en compte explicite des positions des éléments sur les images, ni des relations spatiales et sémantiques entre ces éléments, ni les informations de hiérarchie entre les mots clés, comme le fait que "*chat*" est un cas

particulier de “*animal*”. Nous décrivons dans la partie suivante le formalisme des graphes conceptuels, qui permet de prendre en compte ces différents aspects de manière unifiée.

3.4 Les graphes conceptuels pour la recherche d’images

Le formalisme des graphes conceptuels a été introduit par Sowa [Sow84] comme un modèle de représentation de connaissances. L’annexe B fournit une description détaillée de ce formalisme.

3.4.1 Intérêts des graphes conceptuels pour la RI

Les graphes conceptuels permettent de représenter des connaissances de manière à la fois intuitive et précise. La représentation graphique présente l’avantage d’être facile à lire et à éditer à l’aide d’outils adaptés comme CoGITaNT⁶. Elle constitue une interface compréhensible entre les humains et les représentations internes manipulables par le système.



FIG. 3.1 – Deux images susceptibles d’être décrites par les mêmes mots clés.

La motivation principale de l’utilisation des graphes conceptuels pour la représentation

⁶CoGITaNT = *Conceptual Graphs Integrated Tools allowing Nested Typed graphs*. Il s’agit d’une bibliothèque développée depuis 1994 au sein de l’équipe “Représentation de connaissances par des graphes” du LIRMM, sous la direction de Chein et Mugnier.

et la recherche d'images est qu'ils permettent une description plus riche que les mots clés du contenu des images, tout en fournissant une représentation homogène des différents éléments d'un modèle de RI, c'est-à-dire les modèles de documents et de requêtes, le modèle de correspondance, et le modèle de connaissance. Les images 1 et 2 de la figure 3.1 sont décrites respectivement par les graphes $G1$ et $G2$ de la figure 3.2 :

- L'image 1 représente Matthieu et Jean, qui rament en position assise. Matthieu est à gauche de Jean. Un bateau, qui est sur la mer et en dessous du ciel, est situé entre Matthieu et Jean.
- L'image 2 représente Jean et Matthieu, en position assise. Jean est à gauche de Matthieu, qui est lui-même à gauche d'un bateau. Ce dernier est sur la mer et en dessous du ciel.

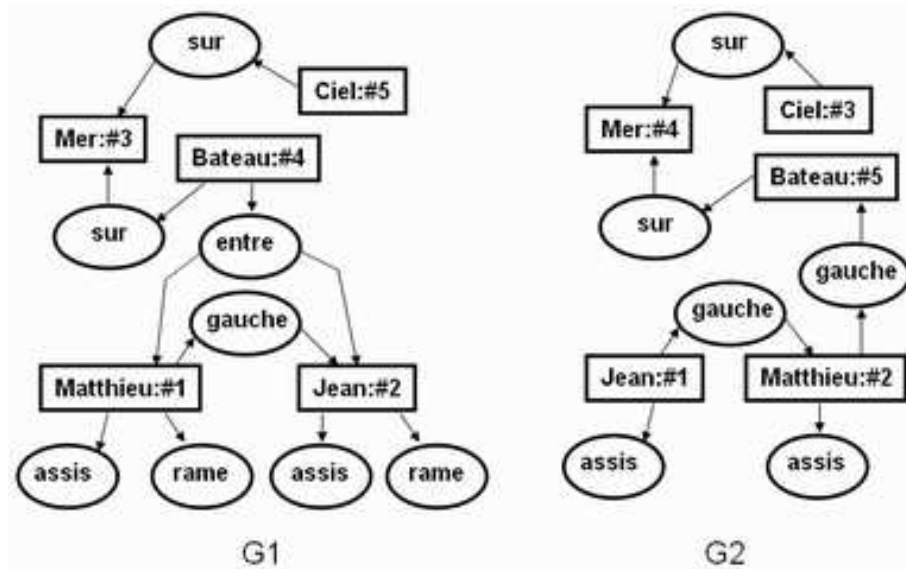


FIG. 3.2 – Deux graphes conceptuels décrivant respectivement les images 1 et 2 de la figure 3.1.

Comme le montre l'exemple ci-dessus, décrire une image à l'aide d'un graphe conceptuel permet d'intégrer les relations qui existent entre les objets visuels de l'image. Des caractéristiques de plus haut niveau des images peuvent également être exprimées, comme des

descriptions d’expressions faciales, des postures des personnes (assis, debout, etc.), des actions (courir, manger, etc.), ou des informations de disposition spatiale relative et absolue de ces objets. L’intégration des relations dans la description des documents et des requêtes a un impact majeur sur la qualité de la représentation interne des documents et sur l’expressivité du langage d’interrogation. En effet, l’utilisateur a la possibilité de formuler des requêtes complexes, et d’exprimer finement son besoin d’information. La précision de la recherche de tels systèmes, dits **orientés précision**, est ainsi théoriquement augmentée.

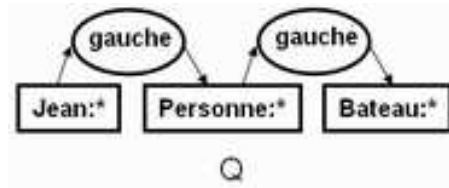


FIG. 3.3 – Un graphe conceptuel de requête.

Exemple 5 *Un utilisateur qui soumet au système la requête de la figure 3.3 qui décrit des “images contenant Jean à gauche d’une personne, cette personne étant elle-même à gauche d’un bateau” retrouve l’image de droite (D1) de la figure 3.1, sans nécessairement retrouver l’image de gauche (D2), qui contient également des personnes et des bateaux, toutefois disposés autrement sur l’image.*

Une sémantique dans la logique du premier ordre est explicitement établie par l’opérateur ϕ , qui associe une formule logique $\phi(G)$ à tout graphe conceptuel G (voir l’annexe B pour plus de détails). Cela confère aux graphes conceptuels un cadre formel d’étude qui rejoint celui du modèle logique de RI [vR86a, vR86b].

En plus de pouvoir représenter des connaissances, le formalisme définit des opérations qui permettent le raisonnement et l’inférence sur ces connaissances. Ces opérations, qui portent soit directement sur les graphes, soit sur les formules logiques associées, fournissent

un mécanisme de prise en compte de connaissances externes aux documents qui sont explicitées dans le support, comme les informations contenues dans les treillis de type de concepts et de relations. La figure 3.4 présente un exemple de treillis de types de concepts et de treillis de type de relations : les concepts et les relations y sont organisés selon un ordre partiel défini par le lien *EstUn*, noté \leq .

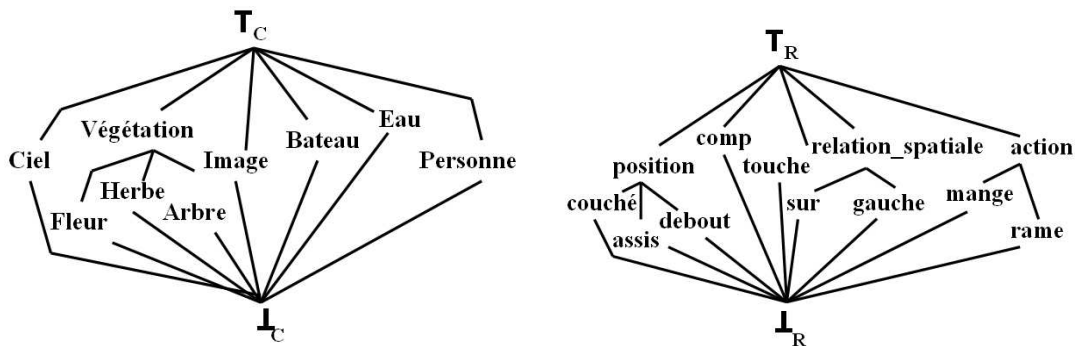


FIG. 3.4 – Les concepts et les relations sont organisés en treillis de types de concepts (gauche) et treillis de relations (droite).

Ces treillis contiennent un plus grand élément (\top_C pour les concepts et \top_R pour les relations), et un plus petit élément (\perp_C pour les concepts et \perp_R pour les relations). Les treillis décrivent les liens de généralité/spécificité entre les concepts (par exemple, le type de concept *Herbe* est inférieur au type de concept *Végétation*) et entre les relations (par exemple, le type de relation *gauche* est inférieur au type de relation *relation spatiale*).

L'opération de projection permet de sélectionner les graphes qui contiennent une information recherchée, compte tenu de ces connaissances externes. Elle est utilisée de la manière suivante en RI : l'ensemble Résultat des documents d'une collection D qui satisfont une requête q contient tous les documents tels qu'il existe au moins une projection du graphe de la requête dans les graphes associés à ces documents :

$$\text{Résultat} = \{d_i \in D \mid \exists p \text{ projection de } G_q \text{ dans } G_{d_i}\}$$

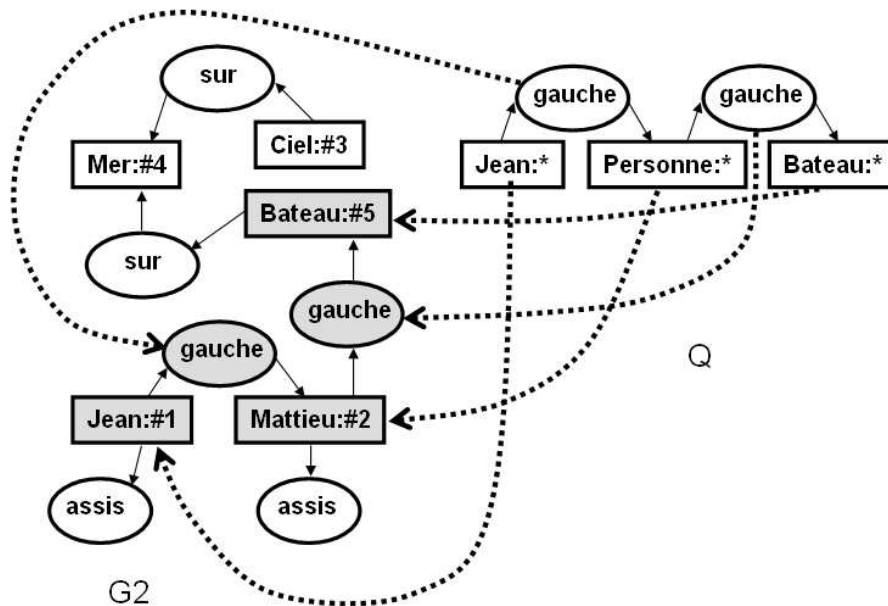


FIG. 3.5 – Projection de graphes conceptuels.

Exemple 6 Le graphe Q de la figure 3.3 se projette dans le graphe $G2$ de la figure 3.2, comme la figure 3.5 l'illustre, en montrant la correspondance des concepts et des relations dans le processus de projection. Les connaissances extérieures contiennent la relation $\text{Matthieu} \leq \text{Personne}$. Si le graphe $G2$ représente un document et le graphe Q représente une requête, alors l'existence d'une projection de Q dans $G2$ traduit la pertinence du document pour la requête, ce qui est satisfaisant pour la RI.

Le formalisme des graphes conceptuels permet donc de représenter les documents, les requêtes, les connaissances externes, et la correspondance au sein d'un modèle uniforme. Pour ces raisons, ce formalisme semble être un bon candidat pour la définition d'un langage d'indexation et de requête. Pour ces mêmes raisons, cette approche a déjà connu plusieurs tentatives. Les graphes conceptuels ont été largement utilisés pour la RI au Laboratoire CLIPS-IMAG par Chevallet [Che92], Kheirbek [Khe95], Mechkour [MMFB97, Mec95,

MBC95], Berrut et Fourel [BFM⁺97], Chiaramella [CMF96], mais aussi au LIRMM par Genest [Gen00], et au CCR⁷ par Montes-y-Gómez [MLG00]. Ils sont utilisés plus spécialement pour la recherche d'images par Mechkour [Mec95, MBC95], Chiaramella [CM97], Ounis [Oun98, OP98], au LI⁸ par Haj Thami Rachid [TADC03], à l'IPAL⁹ par Mulhem [MLLK03] et Debanne [DM04].

Le modèle EMIR²

Le modèle EMIR² [Mec95] est un modèle général de représentation d'images fondé sur les graphes conceptuels, qui a été développé par Mechkour au sein du Laboratoire CLIPS-IMAG. L'objectif de Mechkour était de modéliser le contenu des images en faisant abstraction de leur domaine et des systèmes qui les manipulent, en intégrant plusieurs niveaux d'interprétation de l'image, ainsi que des aspects liés à la subjectivité de l'interprétation et à l'imprécision de la description. Ce modèle intègre les notions d'incertitude et de degré de confiance accordé à la reconnaissance des objets de l'image, qui étendent le formalisme des graphes conceptuels en se basant sur les travaux de Morton [Mor87] et Wuwongse [WM93] concernant les concepts flous et les relations floues, et également ceux de Maher [Mah91] concernant la projection floue. Ce modèle distingue l'image logique et l'image physique. L'image physique est la matrice de pixels stockée dans le système informatique. L'image logique est inspirée des travaux de Gupta, Weymouth et Jain [GWJ91], qui ont réalisé le système VIMSYS dans le cadre du projet Infoscope à l'Université du Michigan. VIMSYS est un système à la frontière entre les SGBD et les SRIm, qui implante un modèle de représentations d'images composé de 4 vues représentant différents aspects de la même image. L'image logique de EMIR² correspond à l'interprétation de l'image. Elle est constituée de la combinaison des vues structurelle, spatiale, perceptive, symbolique.

⁷CCR = *Center for Computing Research*, National Polytechnic Institute, Mexico, Mexique.

⁸LI = *Laboratoire d'Informatique*, École Nationale Supérieure d'Informatique et d'Analyse des Systèmes (ENSIAS), Université Mohammed V, Rabat, Maroc.

⁹IPAL = *Image Processing and Application Laboratory*, laboratoire conjoint CNRS-National University of Singapore, Singapour.

- **La vue structurelle** est la décomposition d’une image en objets images jugés pertinents par l’indexeur. Elle est représentée par un graphe, et est utilisée pour la construction des autres vues.
- **La vue spatiale** contient des informations géométriques sur la forme et la position relative des objets spatiaux associés aux objets images. Elle regroupe les quatre espaces euclidien, métrique (relations proche/loin), vectoriel (relations nord, sud, est, ouest), et topologique.
- **La vue perceptive** regroupe un ensemble d’attributs visuels des objets, comme la couleur, la texture ou la brillance.
- **La vue symbolique** décrit le contenu sémantique de l’image.

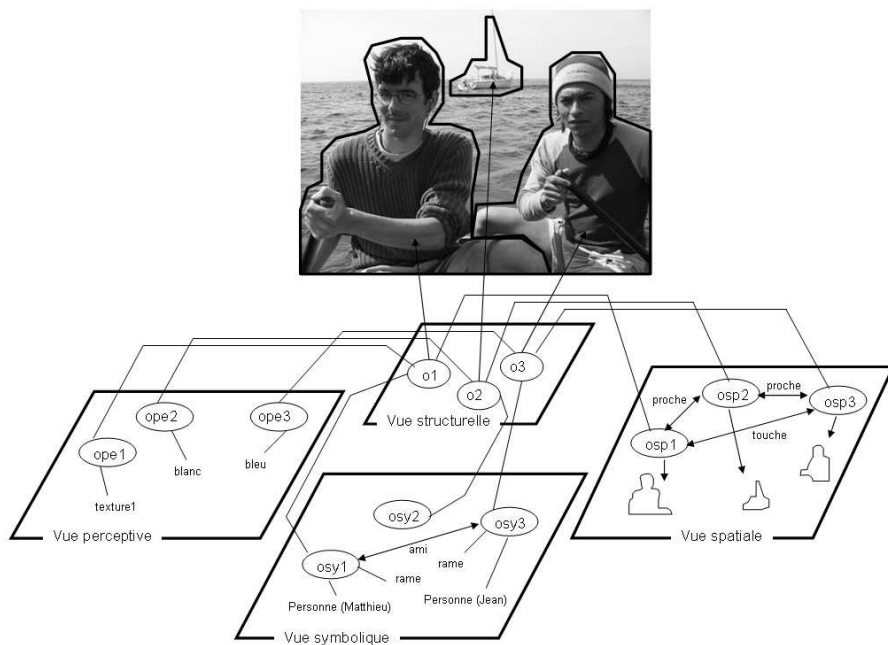


FIG. 3.6 – Vue globale d’une image dans le modèle EMIR².

Les vues spatiale, perceptive et symbolique sont des facettes du même objet image décrit par la vue structurelle. La description d’une image est dans ce modèle une abstraction qui

regroupe toutes ces vues partielles (voir la figure 3.6, sur laquelle se retrouvent ces quatre vues). Ce modèle est intéressant car il définit une structure homogène de représentation des différentes facettes d'une image.

3.4.2 Limites

L'opération de projection, bien que réalisant un processus de correspondance complexe, reste fondamentalement binaire : pour un document d et une requête q , une projection de q dans d existe ou n'existe pas. Il en résulte qu'une estimation de la pertinence fondée sur la seule projection de graphes est à valuation binaire. L'utilisation de l'opération de projection telle qu'elle est originalement définie pour la RI n'est donc pas totalement satisfaisante en raison de sa rigidité : la correspondance binaire ne permet pas d'établir un **classement** des documents, mais plutôt une **classification** en catégories *pertinents* et *non pertinents*. La classe *pertinents* contient les documents dans les graphes desquels il existe au moins une projection du graphe de la requête, la classe *non pertinents* contient tous les autres documents.

Cette correspondance binaire pose les mêmes problèmes que pour le modèle booléen de RI (voir l'annexe A). D'une part, la correspondance ne permet pas une estimation continue de la pertinence système, il est donc impossible d'ordonner les documents des résultats par ordre de pertinence estimée décroissante. D'autre part, cela génère du silence : seuls les documents totalement pertinents sont retrouvés par le système, et les documents répondant partiellement à la requête ne sont pas retrouvés, au même titre que les documents ne répondant pas du tout à la requête. Ce dernier aspect augmente considérablement le risque de réponse vide, ce qui est souvent très désorientant pour les utilisateurs non spécialistes.

Éventuellement, une catégorie supplémentaire de documents *partiellement pertinents* est intégrée à la classification des résultats, comme c'est le cas dans le système RELIEF [OP98]. Cette catégorie correspond aux graphes de documents dans lesquels le graphe de la requête se projette **partiellement** (ou de manière **non totale**), au sens de [Che92] : à défaut

du graphe de requête entier, seul un sous-graphe se projette dans le graphe du document. Toutefois, le problème de l'ordonnement des documents se révèle ici encore, car la mesure du degré de pertinence partielle n'est pas disponible, et les documents peu pertinents ne sont pas distingués des documents très pertinents.

Enfin, l'évaluation de la fonction de correspondance pose un problème de temps de réponse, lié à la complexité de l'opération. Mechkour a réalisé une implantation de ce modèle en un système opérationnel de RIm basé sur le formalisme des graphes conceptuels. La correspondance est effectuée vue par vue, à l'aide d'une extension de l'opération de projection de graphe, qui permet de calculer une valeur de correspondance. Rappelons que l'opération de projection est un isomorphisme de graphes, dont la complexité est exponentielle [CM92, MC96]. Mugnier et Chein ont montré dans [MC93] que la complexité pouvait être polynomiale dans le cas particulier où la structure des graphes est arborescente. Or les graphes de la vue structurelle sont les seuls à être arborescents, les graphes des autres vues sont quelconques. Par conséquent, comme le souligne la conclusion de [Mec95], le problème de temps de traitement des données s'est effectivement posé lors de la première implantation de EMIR² : l'expérimentation a été menée sur une collection de 48 documents et de 10 requêtes, et le temps de réponse du système était de l'ordre de 2 minutes par requête. Cela ne répond pas aux contraintes d'utilisabilité et de qualité d'un système définies dans [Nie94, BKB00], qui stipule que le temps de réponse doit être de l'ordre de 8 à 10 secondes.

Le système RELIEF

Dans le but de résoudre le problème de la complexité de l'opération de projection, tout en conservant le pouvoir expressif des graphes conceptuels dans le cadre de la RIm, Ounis et Pasca ont proposé dans le système RELIEF [OP97, OP98] l'idée originale de déplacer une partie des traitements des données au moment de l'indexation, afin de rendre plus rapide le processus de correspondance. La mise en œuvre de cette idée consiste à générer à l'indexation une structure de données contenant toutes les spécialisations de sous-graphes

de requêtes possibles. Le problème de projection est ainsi divisé en deux sous-problèmes, l'un étant traité durant l'indexation. Au moment de la recherche, il reste à identifier les sous-graphes composant la requête pour retrouver tous les index spécialisant la requête. Il en résulte que le temps de traitement de la requête devient polynomial.

Cependant, cette méthode a montré certains problèmes concernant la taille de l'index et le coût du processus d'indexation. Le temps requis pour la génération de l'index croît de manière exponentielle avec la taille de la collection. De plus, le système RELIEF ne permet pas de classer les documents retrouvés par ordre de pertinence décroissante. Seules trois classes de pertinences sont prévues, une classe de correspondance totale (cas où il existe une projection totale du graphe de la requête dans le graphe du document), une classe de correspondance partielle, et une classe de non correspondance, ce qui rend sommaire l'évaluation des réponses.

Le système DIESKAU

Le système DIESKAU¹⁰ [MLLK03] a été développé à l'IPAL dans le cadre du projet DIVA¹¹. Comme EMIR² et RELIEF, DIESKAU est également basé sur le formalisme des graphes conceptuels. Il met en œuvre une pondération des concepts, et implante une correspondance effectuée en deux temps :

1. Sélection des graphes de documents qui se projettent totalement ou partiellement dans le graphe de la requête.
2. Calcul d'une valeur de correspondance en fonction des poids des concepts en correspondance.

Ainsi, ce système permet une valuation continue de la correspondance, et les réponses peuvent être classées dans un ordre de pertinence système décroissant.

¹⁰DIESKAU = *Digital Image rEtrieval System based on Knowledge representAtion and image featUres.*

¹¹DIVA = *Digital Image and Video Album.*

3.5 Mise en perspective de notre travail

Nous avons présenté différentes approches pour la RIm. L' idée principale que l'on en retient est que les approches signal ne permettent pas aux utilisateurs de tels systèmes de rechercher des images selon la sémantique qu'elles expriment. Les approches symboliques permettent d'interroger le système de manière sémantique, et répondent davantage aux attentes des utilisateurs [RW03] dans ce cas de photographies personnelles et d'utilisateurs non experts. Même dans le cas d'utilisateurs non experts, il est nécessaire de prendre en compte le fait que les éléments visibles dans les images sont par leur nature même reliés (au moins par des relations spatiales), et l'utilisation d'index relationnels par des graphes conceptuels est une réponse à cette problématique. Cependant, la complexité algorithmique liée à la recherche de graphes conceptuels par l'utilisation de la projection de graphes ne favorise pas leur emploi en recherche d'information, et nous devons proposer un moyen de garder des index relationnels (même dégradés par rapport à la puissance d'expression originelle théorique des graphes) tout en tentant d'éviter l'écueil de la complexité du traitement de requêtes.

Deuxième partie

Un modèle vectoriel relationnel de recherche d'images

Chapitre 4

Description du modèle proposé

4.1 Rappel des objectifs

Ce travail est placé dans le contexte spécifique de gestion de collections d'images personnelles. Les utilisateurs cibles sont principalement le grand public, c'est-à-dire des utilisateurs non spécialistes de la recherche d'images.

Hypothèse 1 *Les utilisateurs cibles sont non spécialistes de la recherche d'images, et non experts dans l'utilisation de système de recherche d'images.*

Cette hypothèse oppose le public visé aux utilisateurs experts, qui ont une bonne connaissance des systèmes, des représentations internes, et de la façon dont sont traitées les données afin de fournir le résultat attendu. A l'opposé, un exemple d'utilisateur expert est un médecin qui recherche une radiographie d'une tumeur d'une certaine forme et d'une certaine taille. Il est capable d'en fournir une description très précise et détaillée, et attend du système une grande précision, qui s'accompagne éventuellement d'un temps de réponse plus long. Les utilisateurs d'un système de recherche d'images dans un contexte de collection d'images personnelles sont non experts, et la complexité des requêtes est plus petite que celle d'utilisateurs experts. Cette hypothèse a un impact sur leurs attentes face à de tels modèles et systèmes. Nous rappelons les objectifs fixés pour la définition du modèle, à la lumière de l'état de l'art présenté dans les deux chapitres précédents.

- Le modèle de représentation des images doit pouvoir exprimer leur contenu sémantique, afin que l'utilisateur puisse les retrouver par le biais de cette représentation. Le choix du langage d'indexation doit donc offrir un niveau de représentation en accord avec cette contrainte. Les modèles de documents et de requêtes doivent permettre de manipuler des représentations aussi précises que possible du contenu des documents et des requêtes. La représentation explicite des **relations sémantiques** entre les différents mots clés semble être une caractéristique importante pour un langage d'indexation et d'interrogation pour les images ; nous proposons de la prendre en compte en définissant un langage d'indexation qui supporte les relations.
- Le modèle de correspondance doit supporter le niveau de représentation du langage d'indexation, et doit évaluer une **correspondance partielle** entre les documents et les requêtes. Cette approche prend en compte l'inexpertise des usagers dans l'expression des requêtes, et la nécessité d'éviter autant que possible la production de réponses vides dues à des requêtes mal formulées, car trop éloignées de la représentation interne du contenu des images. Cet aspect du problème est important, et constitue une des raisons majeures pour lesquelles le modèle vectoriel, connu pour sa capacité à prendre en compte cette notion de correspondance partielle, constitue la base du modèle proposé.
- L'estimation de la pertinence système d'une image pour une requête de l'utilisateur doit correspondre à une valuation **continue** afin d'offrir la possibilité de trier les résultats par ordre de pertinence décroissante. Pour cela, il est nécessaire de permettre une valuation des éléments de l'index selon leur importance.
- Enfin, l'implantation de la fonction de correspondance doit respecter les contraintes de **temps de réponse** d'un système interactif [CMK66, Nie94], et fournir un résultat dans un temps de l'ordre de 8 à 10 secondes au maximum, même pour de grandes quantités de données.

La nature des images impose de représenter le contenu des documents et des requêtes de manière précise, plus élaborée qu'une simple liste de mots clés. Elle nécessite également la gestion de connaissances extérieures, de disposer d'une correspondance partielle documents - requêtes afin de limiter le silence, et de pouvoir ordonner les documents du plus pertinent au moins pertinent. Ces objectifs ont guidé nos choix de représentation, et le modèle proposé dans le cadre de ce travail vise essentiellement à répondre à ces contraintes.

4.2 Modèle de représentation des documents et des requêtes

Cette partie fournit une description de notre modèle de représentation d'images. Il est basé sur le modèle vectoriel de RI, et intègre les aspects relationnels des graphes conceptuels : les images sont représentées par des vecteurs dans un espace dont les dimensions sont formées par des relations mettant en jeu les objets de l'image.

4.2.1 Modèle d'image physique

Nous commençons par définir les éléments de base de notre modèle d'image physique, que nous allons manipuler par la suite. Ces définitions portent sur les images, les régions, et les objets image.

Définition 2 (Image) *Une image I est un ensemble de pixels connexes organisés dans une matrice rectangulaire, correspondant à une vue du monde réel, c'est-à-dire à une projection bidimensionnelle d'une scène selon un axe de prise de vue. Le nombre de pixels de I est son cardinal $||I||$, noté n_I .*

L'image est donc à la base un bloc de pixels, comme dans les approches signal dont nous avons fait mention dans l'état de l'art dans la partie 3.2. Nous définissons maintenant les notions de région, de segmentation en régions, d'objets images et de segmentation en objets images, qui permettent une description locale des éléments de l'image.

Définition 3 (Région) Une région $r \subseteq I$ est une projection bidimensionnelle selon un axe de prise de vue d'un objet physique (ou d'une partie d'un objet physique occlus) du monde réel, formant un ensemble non vide de pixels connexes d'une image : $r \in \mathbb{P}(I)$ où $\mathbb{P}(I)$ est l'ensemble des parties connexes non vides de I . Le nombre de pixels de r est son cardinal $\|r\|$, noté n_r .

L'ensemble des régions est en surjection avec l'ensemble des objets physiques visibles dans l'image : plusieurs régions peuvent détourner différentes parties d'un même objet physique, et tous les objets physiques visibles sont détournés par au moins une région.

Nous imposons une hypothèse sur la taille minimale des régions, similaire à celle largement utilisée dans le domaine de la cartographie [Weg99], qui définit un seuil de perception pour la surface des éléments représentés sur les cartes en deçà duquel les éléments ne sont plus considérés comme perceptibles :

Hypothèse 2 (Taille minimale d'une région) Une région doit avoir une taille minimale pour pouvoir être distinguée individuellement, et pour que l'objet physique qu'elle détourne puisse être identifié. Nous définissons un paramètre $\epsilon \geq 1$ qui est le nombre minimum de pixels d'une région : $\|r\| \geq \epsilon$.

En raison de la physiologie de l'œil, ce seuil est de l'ordre de 0.5 mm [Weg99], ce qui correspond approximativement à une surface de 4 pixels (un carré de 2×2 pixels par exemple) avec la résolution de référence pour les écrans de 72 ppp¹, et à une surface de 35 pixels avec une résolution standard pour une imprimante de qualité photographique de 300 ppp.

Définition 4 (Segmentation en régions) Une segmentation en régions d'une image I est un ensemble $S_I^R \subset \mathbb{P}(I)$ de régions formant une partition de l'image :

- $\forall r \in S_I^R : r \neq \emptyset$ (car $\|r\| \geq \epsilon$)
- $\forall r_i, r_j \in S_I^R : r_i \cap r_j = \emptyset$
- $\bigcup_{r \in S_I^R} r = I$

¹Point par pouce (ou dpi, *dot per inch* en anglais), 1 pouce = 2.54 cm.

Le nombre de régions de I est le cardinal $\|S_I^R\|$, noté $n_{S_I^R}$.

Une région est donc un élément d'une segmentation d'une image. Il existe plusieurs segmentations en régions possibles pour une image ; nous considérons dans notre modèle une segmentation en régions unique pour une image donnée. Les régions correspondent à des parties sémantiquement homogènes de l'image ; elles délimitent le contour des objets physiques ou des parties d'objets physiques qui apparaissent dans les images. Les objets physiques auxquels nous faisons référence sont les objets du monde réel qui sont susceptibles d'apparaître dans les images, comme des personnes, des arbres, des voitures, de la mer, des bateaux, etc. Un objet physique peut apparaître fractionné dans une image dans le cas d'une occlusion ; les régions délimitent alors les contours des différentes parties de cet objet.

Notre modèle représente les objets physiques apparaissant dans les images de manière plus flexible que les régions ne le permettent, à l'aide de la notion de plus haut niveau d'objet image.

Définition 5 (Objet image) *Un objet image est une projection bidimensionnelle selon un axe de prise de vue d'un ou de plusieurs objets physiques du monde réel, constitué d'une agrégation (union) de régions non nécessairement connexes. Le nombre de pixels de o est le cardinal $\|o\| \geq \epsilon$, noté n_o .*

Les objets images forment une segmentation de S_I^R (et donc également une segmentation de l'image), et sont des éléments d'une segmentation en objets images.

Définition 6 (Segmentation en objets images) *Une segmentation en objets images de S_I^R est un ensemble $S_I^O \subseteq \mathbb{P}(S_I^R)$ d'objets images formant une partition de S_I^R :*

- $\forall o \in S_I^O : o \neq \emptyset$ (car $\|o\| \geq \epsilon$)
- $\forall o_i, o_j \in S_I^O : o_i \cap o_j = \emptyset$
- $\bigcup_{o \in S_I^O} o = S_I^R$

Le nombre d'objets images de I est le cardinal $\|S_I^O\|$, noté $n_{S_I^O}$.

Les objets images forment également une partition de l'image, qui est moins fine que la partition définie par les régions :

$$\forall r \in S_I^r : \exists o \in S_I^O | r \in o$$

Autrement dit, toutes les régions appartiennent à un objet image. Il est possible que S_I^O soit identique à S_I^R , c'est-à-dire que tous les objets images sont constitués d'une seule région. Un objet image contient au moins une région, et au plus toutes les régions de l'image. Il a donc une certaine granularité, qui est décidée par l'indexeur².

4.2.2 Interprétation des objets images

Le lien entre les objets images et les objets physiques est établi par le biais d'une interprétation sémantique des objets images.

Définition 7 (Interprétation d'un objet image) *Un objet image o est associé à un concept³ $c = [t, \#o]$ dont le type t correspond à une interprétation sémantique de o , et le référent $\#o$ identifie l'objet image de manière unique. Les concepts sont en bijection avec les objets images.*

L'ensemble T_C des types de concepts est organisé dans un treillis $\langle T_C, \leq \rangle$, dont la relation \leq indique le lien de généralité/spécificité entre les types de concepts. Un objet image peut regrouper plusieurs objets physiques, ayant ou non la même interprétation sémantique. Dans le cas où les objets physiques ont la même interprétation sémantique, le concept associé à l'objet image a pour type à cette interprétation sémantique. Dans l'autre cas où les objets physiques n'ont pas la même interprétation sémantique, le concept associé à l'objet image a pour type le plus spécifique des types de concepts génériques.

Exemple 7 *Si un objet image regroupe plusieurs bateaux de différents types (des voiliers et des bateaux à moteur), et le treillis contient les relations Voilier \leq Bateau et Bateau à moteur \leq Bateau, alors c'est l'étiquette "Bateau" qui est associée à cet objet image.*

²Humain ou machine.

³Au sens des graphes conceptuels.

Les objet images peuvent englober des régions regroupées spatialement dans l'image, ou bien des régions qui représentent des objets physiques petits et nombreux, de telle sorte qu'ils sont perçus globalement comme un groupe qui constitue une entité à part entière dans l'image (une foule de personnes, une formation d'oiseaux en vol, une forêt d'arbres, un port de bateaux, etc.). Cette agrégation est une abstraction de la réalité que nous faisons le choix d'intégrer dans le modèle de représentation afin de faciliter la manipulation et la description des objets physiques.

Tous les objets images n'ont pas un contenu pertinent pour la représentation et la recherche des images : certains peuvent être jugés comme non pertinents dans la représentation par l'indexeur, soit parce que leur interprétation sémantique n'entre pas dans le cadre du domaine considéré (et ne correspond à aucune étiquette du treillis), soit parce que leur interprétation sémantique est indéterminée (zones sombres ou floues de l'image). Cette sélection des objets images pertinents correspond à l'utilisation d'un vocabulaire contrôlé pour le texte (un dictionnaire répertorie la liste des mots autorisés dans l'index, et le modèle considère uniquement ces derniers dans le langage d'indexation), ou au filtrage des mots outils du texte à l'aide d'un anti-dictionnaire (voir la partie 2.2.1 page 20). Ces objets images sont associés à des concepts de type \top_C : ce type de concept n'a aucun type générique, et n'est donc la réponse à aucune question. Autrement dit, la formule logique associée à ce type de concept n'implique aucune autre formule qu'elle-même. La figure 4.1 donne un exemple de segmentation d'une image en régions et en objets images étiquetés : des régions qui correspondent aux personnes, aux bateaux, aux bâtiments, au ciel et à l'eau sont détourés, et des concepts sont associés aux objets images. Un objet image qui ne correspond à aucune interprétation sémantique est associé à un concept de type \top_C .

4.2.3 Les graphes étoiles comme vocabulaire d'indexation

Les concepts sont mis en relations par des relations conceptuelles, qui sont organisées dans un treillis de type de relations $\langle T_R, \leq \rangle$, par le biais des **graphes étoiles**.



FIG. 4.1 – Exemple de segmentation en régions et en objets images.

Définition 8 (Graphe étoile) *Un graphe étoile est un graphe conceptuel élémentaire comprenant une et une seule relation r , et n_r concepts c_i attachés à cette relation ($n_r \geq 1$ est l'arité de r) :*

$$\langle r(c_1, c_2, \dots, c_{n_r}) \rangle$$

La figure 4.2 montre en exemple les graphes étoiles $\langle \text{rame}(\text{Jean} : \#2) \rangle$, $\langle \text{sur}(\text{Bateau} : \#4, \text{Mer} : \#3) \rangle$ et $\langle \text{entre}(\text{Bateau} : \#4, \text{Matthieu} : \#1, \text{Jean} : \#2) \rangle$. Ces graphes étoiles sont dits **spécifiques** car ils sont constitués de concepts spécifiques⁴. Un graphe étoile peut être **générique** lorsqu'il est constitué uniquement de concepts génériques. Notre définition des graphes étoiles diffère sensiblement de la définition de graphes étoiles de Chein et Mugnier [CM92, MC96], qui définissent un graphe étoile comme étant une relation attachée à des *types de concepts*. Dans notre définition, ce sont des concepts qui sont attachés à la relation. Les graphes étoiles peuvent être vus comme les briques de base servant à la construction de graphes conceptuels complets : il est possible de construire de tels graphes en reliant les graphes étoiles par une série d'opérations de jointure sur les concepts identiques.

Nous définissons notre vocabulaire d'indexation à partir de la **base canonique** (voir l'annexe B). Elle contient un ensemble de graphes étoiles en bijection avec les relations de

⁴Un concept spécifique que a un référent individuel, par opposition au référent universel : *.

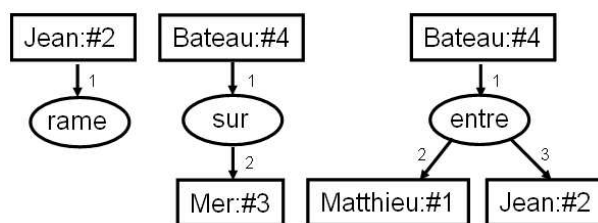


FIG. 4.2 – Graphes étoiles de relations unaires, binaire et ternaire, de gauche à droite.

T_R , définissant la signature de chaque relation, c’est-à-dire les types maximaux des concepts qui peuvent y être attachés. Tout graphe étoile peut s’obtenir par restriction de concepts et de relations à partir d’un graphe étoile de la base canonique, c’est-à-dire en remplaçant un des concepts par un concept de type inférieur, ou en remplaçant la relation par une relation de type inférieur.

Définition 9 (Vocabulaire d’indexation) *Le vocabulaire d’indexation est constitué de l’ensemble des graphes étoiles génériques obtenus par fermeture des graphes étoiles de la base canonique sur l’opération de restriction des types de concepts et de relations.*

Autrement dit, le vocabulaire d’indexation contient toutes les restrictions possibles des graphes étoiles de la base canonique par des spécialisations de types de concepts et de relations. Les graphes étoiles du vocabulaire d’indexation ne contiennent donc pas de référents, et dénotent des “classes sémantiques”. Deux graphes étoiles spécifiques $\langle gauche(Bateau :\#1, Personne :\#2) \rangle$ et $\langle gauche(Bateau :\#3, Personne :\#4) \rangle$ décrivant une image correspondent à la même “classe sémantique”, et sont donc considérés comme deux occurrences du terme d’indexation⁵ $\langle gauche(Bateau, Personne) \rangle$, au même titre que deux occurrences d’un même mot clé dans un texte. Un graphe étoile spécifique g d’une image est associé au terme d’indexation t le plus spécifique tel que $g \leq t$. De manière plus concrète, t est en fait la généralisation de g qui transforme chaque référent individuel de g en référent générique*.

⁵Pour faciliter la lecture, nous ne représentons pas le référent générique* dans les termes d’indexation.

Dans la suite, nous appellerons indifféremment “**graphe étoile**” un graphe étoile spécifique dont les concepts étiquettent des objets images et un graphe étoile générique dénotant un terme du vocabulaire d’indexation. La distinction explicite sera faite quand cela sera nécessaire.

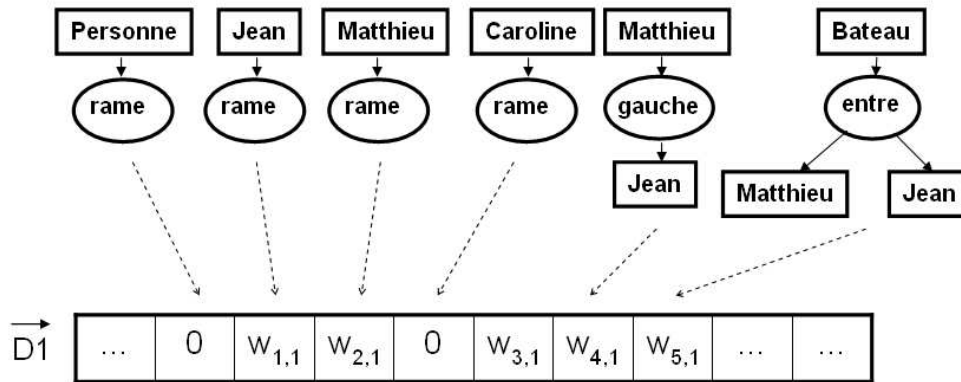


FIG. 4.3 – Représentation vectorielle de l’index d’un document.

Les graphes étoiles constituent des termes d’indexation suffisamment élaborés pour permettre de rendre explicites les relations entre les concepts. En ce sens, ils permettent une indexation relationnelle. Notre modèle représente les documents et les requêtes par des vecteurs dans l’espace des termes d’indexation. La figure 4.3 donne un exemple de vecteur qui représente une partie⁶ de l’index de l’image 1 de la figure 3.1 page 48. Des poids $w_{i,j}$ non nuls sont affectés aux dimensions correspondant aux termes qui décrivent l’image. La problématique de la pondération est développée dans le chapitre 5.

4.2.4 Compromis entre pouvoir expressif et complexité

Le modèle de documents et de requêtes définit les représentations internes du contenu des documents et des requêtes. Comme il a été vu précédemment, il est nécessaire que le langage d’indexation soit suffisamment riche pour indexer les documents de manière non

⁶L’espace entier n’est pas représenté.

ambiguë, et suffisamment simple afin de pouvoir être appliqué sur de grandes quantités de données. Si le langage d'indexation possède un pouvoir expressif trop faible, il peut arriver que des documents différents se voient affecter des index proches ou identiques. Le formalisme des mots clés a un pouvoir expressif relativement faible, car il ne permet pas d'exprimer de relation entre les mots clés.

Exemple 8 *Le formalisme des mots clés ne permet pas de distinguer les deux photographies de la figure 3.1 page 48 : bien qu'elles représentent des scènes différentes du point de vue de la disposition et de l'apparence des objets, elles sont susceptibles d'être décrites par le même ensemble de mots clés : { Matthieu, Jean, Mer, Ciel, Bateau }. Un utilisateur ne pourrait pas, à partir des seuls mots clés, exprimer une requête permettant de les distinguer, dans le but d'accéder à l'une ou à l'autre.*

Cet exemple montre l'utilité d'un langage d'indexation plus riche, qui prenne en compte la représentation explicite des relations qui existent entre les objets de l'image. Cependant, il est bien connu que la mise en œuvre de systèmes basés sur les mots clés est réalisable à l'aide d'algorithmes de faible complexité, même pour de grandes quantités de documents, comme c'est le cas pour les modèles classiques de RI. Le formalisme des mots clés bénéficie donc d'une faible complexité contrebalancée par une faible expressivité, comme l'illustre la figure 4.4.

A l'opposé, le formalisme des graphes conceptuels permet une indexation relationnelle des documents, qui permet de diminuer l'ambiguïté de l'indexation, en fournissant des structures sémantiques plus précises que les mots clés.

Exemple 9 *Les graphes conceptuels de la figure 3.2 page 49 décrivent plus précisément que les simples mots clés les images de la figure 3.1, et permettent ainsi de les distinguer. De même, les requêtes peuvent être spécifiées plus précisément, ce qui limite le **bruit** dans les résultats du système et qui, par conséquent, augmente la **précision** des réponses.*

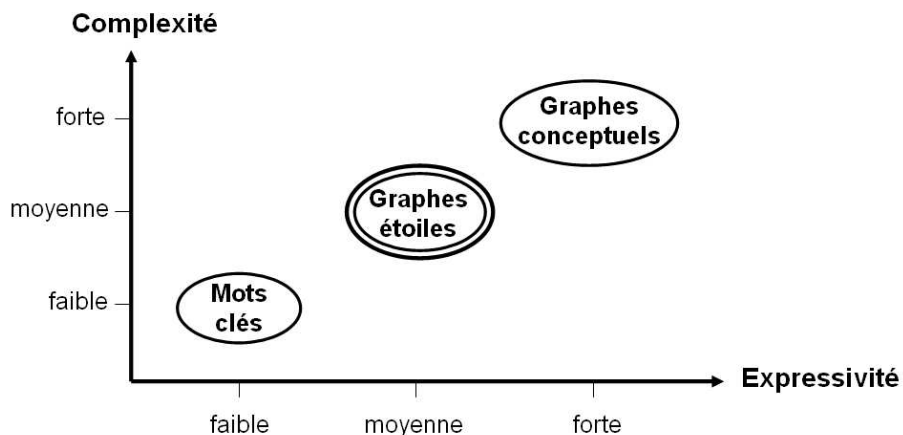


FIG. 4.4 – Relation entre l'expressivité du formalisme et la complexité des traitements pour les mots clés, les graphes étoiles et les graphes conceptuels.

Les graphes conceptuels permettent de différencier les photographies de l'exemple ci-dessus en fournissant une description plus expressive de leur contenu que les mots clés. Comme nous en avons déjà fait mention précédemment, ils posent néanmoins le problème de la complexité des algorithmes de traitement dans le cadre de la RI en raison de la NP-complétude de l'isomorphisme de graphes effectuée lors de l'opération de projection. Le formalisme des graphes conceptuels apporte donc une grande expressivité, contrebalancée par une grande complexité de traitement, comme le décrit la partie 3.4.2 page 55. Cette relation est illustrée à la figure 4.4.

A un niveau intermédiaire, les graphes étoiles, pris comme unités d'indexation, permettent une représentation à la fois plus expressive que les mots clés, et moins complexe que les graphes conceptuels considérés dans leur globalité. Pour les raisons de complexité et d'expressivité évoquées plus haut, l'utilisation des graphes étoiles en tant qu'éléments d'un langage d'indexation et de requêtes constitue un compromis entre **l'expressivité** du formalisme et **la complexité** des traitements associés (voir la figure 4.4). Les graphes étoiles sont plus simples que les graphes conceptuels qui indexent les documents. Le traitement indépendant des graphes étoiles comme unités d'indexation et de requêtes permet une cor-

respondance graphe étoile à graphe étoile à la fois plus souple (car fractionnée), et moins complexe qu’une correspondance qui mettrait en jeu des graphes complexes. Cette représentation à base de graphes étoiles conserve la sémantique du formalisme des graphes conceptuels, et est cohérente avec la décomposition en **arches** proposée par Ounis et Pasca dans le système RELIEF [OP98], avec les **tables de graphes** définies par Amati et Ounis [AO00], ainsi qu’avec les **triplets** de Bélaïd et Jaoua [AJ98]. Dans ces trois cas, les auteurs décomposent des graphes conceptuels en un ensemble de sous-graphes. Une arche est un triplet $[\text{Concept}_1] \rightarrow (\text{Relation}) \rightarrow [\text{Concept}_2]$, qui représente une relation binaire. Nous préférons utiliser le terme plus général de **graphe étoile**, car nous considérons le cas plus général des relations *n-aires*.

4.3 Processus d’indexation

Nous avons vu dans la partie précédente que le vocabulaire d’indexation de notre modèle est constitué de graphes étoiles. L’indexation d’un document consiste donc à sélectionner un sous-ensemble du vocabulaire d’indexation pour le décrire : le rôle de l’indexeur est, dans un premier temps, d’identifier les concepts dans une image, et dans un second temps, de les mettre en relation dans des graphes étoiles, qui constituent un ensemble de termes d’indexation de l’image. Dans un troisième temps, un processus automatique d’**expansion des documents** (voir la partie 2.2.1 page 20) complète la description du document en sélectionnant et en ajoutant des termes d’indexation aux documents [MCM02]. Nous détaillons ce processus d’expansion de documents, qui a pour objectif à la fois de faciliter la tâche de l’indexeur en complétant la description des documents, et de garantir une cohérence sémantique dans les index. En effet, il s’agit d’une explicitation des informations implicitement contenues dans les index.

L’expansion classique de requêtes textuelles est basée la plupart du temps sur les relations de synonymie qui existent entre les mots clés, et qui sont contenues dans un thésaurus comme WordNet [Mil90, SZ98]. Ainsi, lorsqu’une requête contient le mot “voiture”, le mot

“*automobile*” peut être ajouté à la requête afin de retrouver plus de documents pertinents (augmentation du rappel), car le mot “*voiture*” est synonyme du mot “*automobile*”. Dans le cas de la synonymie, la relation d’association des mots est symétrique. Dans notre cas, l’ordre partiel \leq sur les graphes étoiles ⁷ impose la distinction des directions d’expansion :

- spécifique \rightarrow générique, pour les documents (un document décrit par “*Arbre*” est également décrit par “*Végétation*”).
- générique \rightarrow spécifique, pour les requêtes (une requête qui porte sur “*Végétation*” porte également sur “*Arbre*”).

Un document indexé par un graphe étoile devrait être retrouvé quand l’utilisateur soumet au système une requête contenant un graphe étoile plus générique ; de même, une requête soumise par un utilisateur devrait retrouver les documents contenant des graphes étoiles spécifiques.

Exemple 10 *Le terme $\langle \text{gauche}(\text{Bateau}, \text{Personne}) \rangle$ d’une requête est en correspondance avec le terme $\langle \text{gauche}(\text{Bateau}, \text{Jean}) \rangle$ d’un document, en raison du fait que le graphe étoile $\langle \text{gauche}(\text{Bateau} : *, \text{Personne} : *) \rangle$ se projette dans le graphe étoile $\langle \text{gauche}(\text{Bateau} : \#1, \text{Jean} : \#2) \rangle$ en spécialisant le concept $[\text{Personne} : *]$ en $[\text{Jean} : \#2]$. Or ces termes, par construction de l’espace, correspondent à des dimensions différentes, et l’évaluation de la similarité dans le cadre d’un modèle vectoriel produirait une correspondance nulle.*

Cet exemple illustre la nécessité d’appliquer une expansion aux vecteurs (documents ou requête) afin que la correspondance entre des dimensions différentes mais liées par une relation de généralité/spécificité puisse être établie, et ainsi garantir la sémantique de la fonction de correspondance.

Nous choisissons d’appliquer l’expansion aux documents, car c’est une opération qui est effectuée *hors-ligne*, et qui n’a donc pas d’incidence sur la complexité de la fonction de correspondance. Le processus d’expansion de requêtes consiste donc à ajouter dans l’index

⁷Cet ordre est induit par les règles de dérivation sur les graphes.

d'une image les graphes étoiles qui sont plus génériques que les graphes étoiles sélectionnés par l'indexeur.

Définition 10 (Expansion de documents) Soit GE l'ensemble des graphes étoiles spécifiques décrivant un document, sélectionnés par l'indexeur. L'expansion de ce document est définie par l'ensemble de graphes étoiles spécifiques GE' tel que :

$$GE' = GE \cup \{ge_j | \exists ge_i \in GE : (ge_i \leq ge_j)\}$$

L'ensemble GE' est l'ensemble des termes d'indexation du document.

Des poids $w_{i,j}$ non nuls sont donc également affectés aux dimensions correspondant aux termes ajoutés. Le vecteur de la figure 4.3 est présenté **après** expansion à la figure 4.5 : le terme $\langle rame(Personne) \rangle$ a désormais un poids $w_{k,1}$ non nul. L'expansion peut être vue

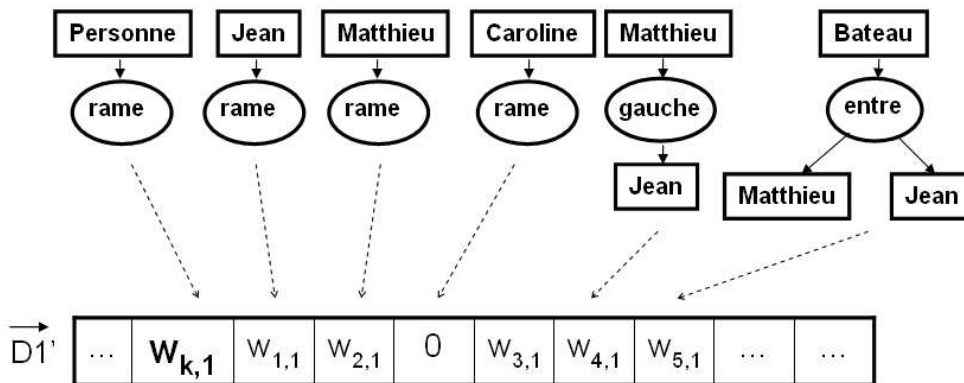


FIG. 4.5 – Vecteur $\vec{D1'}$ après expansion du document.

comme une intégration dans le modèle de représentation des documents des connaissances *a priori* contenues dans les treillis de types de concepts et de relation.

4.4 Fonction de correspondance

Le modèle de correspondance doit établir une mesure de pertinence système entre deux vecteurs de graphes étoiles représentant respectivement un document et une requête. La per-

tinence système est définie dans notre modèle par une mesure de similarité entre vecteurs :

$$\text{correspondance}(d, q) = \text{sim}(\vec{d}, \vec{q})$$

Les dimensions de l'espace vectoriel de notre modèle sont de nature symbolique, par opposition à des dimensions qui représenteraient des informations de niveau signal (voir la différence signal/symbole au chapitre 3). Elles ont donc une sémantique, comme c'est le cas pour le modèle vectoriel textuel (voir la partie 2.4 page 27). Suivant l'approche habituelle utilisée dans ce modèle, nous pouvons également baser la mesure de similarité de notre modèle sur le cosinus de l'angle formé par le vecteur d'un document d et le vecteur d'une requête q :

$$\text{sim}(\vec{d}, \vec{q}) = \cos(\vec{d}, \vec{q}) \quad (4.1)$$

Du fait de l'application de l'expansion de documents et de l'utilisation de la stratégie de pondération définie au chapitre 5, les documents sémantiquement proches de la requête obtiennent une forte valeur de similarité.

4.5 Conclusion

Nous avons présenté dans ce chapitre notre modèle relationnel de RI, adapté à la spécificité des images. Ce modèle est une extension du modèle vectoriel textuel classique, dont les dimensions sont des graphes étoiles. Nous l'avons décrit selon :

- un modèle de représentation de documents et des requêtes à base de vecteurs dans un espace de graphes étoiles,
- un modèle de connaissances incluant les treillis de types de concepts et de types de relations, utilisé pour l'expansion des documents lors du processus d'indexation.
- un modèle de correspondance similaire à celui utilisé dans les approches classiques.

Le modèle vectoriel est fondé sur une hypothèse implicite d'indépendance des dimensions. Or les dimensions générées par les graphes étoiles de notre modèle sont liées par

des relations de généralité/spécificité. Nous ne tenons pas compte actuellement de la non orthogonalité des dimensions.

Chapitre 5

Pondération des graphes étoiles

Le chapitre précédent a présenté le modèle vectoriel relationnel proposé. Ce modèle est une extension du modèle vectoriel de RI, dont les dimensions correspondent aux graphes étoiles de la collection ; les documents et les requêtes y sont donc représentés par des vecteurs de graphes étoiles. De même, c'est par des comparaisons de vecteurs que les documents du corpus peuvent être retrouvés en réponse à une requête. Une approche simple consisterait à attribuer une pondération booléenne aux graphes étoiles, indiquant leur présence ou leur absence dans l'index d'une image. Dans cette approche, les importances relatives des graphes étoiles dans les images ne seraient pas représentées, et le caractère discriminant de certains graphes étoiles de la collection ne serait pas exprimé (voir les approches classiques de pondération dans le modèle vectoriel textuel dans la partie 2.5 page 30). C'est la raison pour laquelle nous intégrons des pondérations dans notre modèle. Ce chapitre développe la problématique de la pondération des graphes étoiles qui décrivent les images, et propose une stratégie de pondération qui permet de refléter dans les vecteurs leur importance sémantique. L'objectif de ce chapitre est double :

1. Identifier des critères pertinents qui déterminent l'importance visuelle des objets images, en agissant dans une démarche d'observation et de capture de la réalité. En effet, les critères perceptifs choisis sont pragmatiques, et nous tentons de les identifier et de formuler des hypothèses.

2. Définir des modélisations pertinentes de ces paramètres.

Comme dans le modèle vectoriel classique (voir la partie 2.4 page 27), notre approche considère que le poids d'un graphe étoile doit refléter d'une part son importance dans le document, et d'autre part sa capacité à distinguer les documents les uns des autres dans la collection. Ce chapitre propose d'apporter des réponses aux deux questions suivantes concernant le graphe étoile de la figure 5.1 et la photographie qu'il accompagne :

1. "A quel point le graphe étoile $\langle gauche(\text{Bateau}, \text{Personne}) \rangle$ est-il un bon descripteur de l'image ?"
2. "Dans quelle mesure le graphe étoile $\langle gauche(\text{Bateau}, \text{Personne}) \rangle$ permet-il de discriminer les documents de la collection ?"



FIG. 5.1 – A quel point le graphe étoile est-il un bon descripteur de l'image ?

Pour répondre à la première question, il est nécessaire de s'interroger sur l'importance (locale) d'un graphe étoile en tant que descripteur d'image, c'est-à-dire sur sa **représentativité** de l'image.

Définition 11 (Importance d'un graphe étoile) *La pondération reflétant des critères de perception par les humains de "ce qui est important dans une image" ou bien, ce qui est*

*dualement équivalent en RI, de “pourquoi une image est perçue comme pertinente par rapport à un concept donné”, la notion d’importance d’un graphe étoile dans une image réfère à la pertinence de cette image pour ce graphe étoile. En d’autres termes, un graphe étoile est d’autant plus important qu’il est **représentatif** du contenu sémantique de l’image.*

Dans le cas de documents textuels, les approches les plus courantes considèrent le **nombre d’occurrences** d’un mot clé dans un document pour déterminer son importance dans ce document : il s’agit de la valeur **tf**, comme nous le présentons dans la partie 2.5 page 30. Dans ce contexte, la pertinence d’un document par rapport à un mot clé est directement liée à sa fréquence d’occurrence dans le texte. Il s’agit bien d’un critère de perception, qui a été abondamment vérifié expérimentalement. Intuitivement, dans le cas de graphes étoiles décrivant des images, il est raisonnable de se baser également sur leur nombre d’occurrences dans l’image, par analogie aux techniques de pondération des mots clés du texte.

Exemple 11 *Le graphe étoile <gauche(Bateau, Personne)> est d’autant plus représentatif de la photographie de la figure 5.1 qu’il y a de cas de “bateau à gauche d’une personne” dans l’image.*

Cependant, alors qu’un mot clé n’a pas de variabilité d’apparition autre que ses différentes formes lexicales, ce qui n’a pas de conséquence sur le comptage des occurrences et donc sur la pondération, un graphe étoile est susceptible de représenter des situations d’une grande variabilité, ce qui a un impact sur les importances individuelles de chaque occurrence. A la différence du texte, il est donc nécessaire d’analyser également les importances individuelles des occurrences de graphes étoiles. Afin de procéder à ces analyses, nous considérons, comme d’autres approches [OP98, ML02, MLLK03], que l’importance d’un graphe étoile dans un document dépend de l’importance de ses concepts composants, car le sens global du graphe étoile est porté par ses concepts. Une étude de l’importance des concepts doit donc permettre de définir l’importance des graphes étoiles. Une hypothèse séduisante est de considérer que les importances des concepts dépendent à leur tour des caractéristiques

physiques des régions de l'image étiquetées par ces concepts. Dans cette optique, il convient donc d'étudier l'importance des objets images.

Définition 12 (Importance d'un objet image) *L'importance d'un objet image réfère à la pertinence de l'image pour cet objet : un objet image est d'autant plus important qu'il est représentatif du contenu sémantique de l'image.*

Pour répondre à la seconde question, nous nous inspirons directement des techniques classiques utilisées pour le texte, en étudiant le pouvoir discriminant des graphes étoiles à partir de leur fréquence d'occurrence dans la collection. Ce point est détaillé dans la partie 5.3.2.

Avant de définir l'importance d'un graphe étoile, nous commençons donc par définir l'importance des objets images attachés aux concepts constitutifs des graphes étoiles. Pour cela, nous émettons des hypothèses relatives à l'importance des objets dans l'image, et donc aux aspects perceptifs évoqués plus haut.

5.1 Hypothèses relatives à l'importance des objets images

En nous inspirant des travaux de Ounis [Oun99] sur l'indexation des images à partir de critères d'importance visuelle, nous nous intéressons à la facilité, pour un observateur, de percevoir les différentes régions de l'image. Les hypothèses que nous énonçons expriment les relations entre cette notion d'importance des éléments d'une image et la facilité de perception de ces objets. Nous proposons d'isoler des critères qui font qu'un objet image est facilement visible et identifiable. Ces critères comprennent :

- la **géométrie des objets** : leurs formes et leurs tailles,
- la **topologie de l'image** : le nombre d'objets, leurs positions absolues et relatives, leurs occlusions éventuelles par d'autres objets ou par le bord de l'image, leur niveau

de composition, leur granularité¹, la présence éventuelle d'autres objets,

- les **caractéristiques de niveau signal** : la couleur, la brillance, et le contraste des objets.

Parmi ces critères, nous avons choisi de focaliser notre étude sur les deux premiers² : la **taille des objets images**, la **position des objets images**, le niveau de granularité des objets images dénoté par leur **éclatement** et la présence d'autres objets, dénotée par l'**hétérogénéité de l'image**.

5.1.1 La taille

Nous avons l'habitude de percevoir comme étant proches les objets de grande taille, et comme loin les objets de petite taille. D'après Messaris [Mes94], si l'observateur a une idée de la taille des objets dans le monde réel, alors la taille des objets d'une image lui donne une bonne indication de leur profondeur³. En raison de la perspective, il existe une relation entre la taille apparente des objets dans une image et leur profondeur [Mes94], une image bidimensionnelle étant une projection d'une scène tridimensionnelle. Le tableau 5.1 résume la relation entre la taille d'un objet dans une image et sa profondeur, en fonction de sa taille : lorsqu'un objet apparaît grand dans une image, cela signifie soit qu'il est proche – indépendamment de sa taille réelle, soit qu'il est grand dans la réalité. Lorsqu'il apparaît petit, cela signifie vraisemblablement qu'il est loin.

Un objet apparaît...	quand il est...	et de taille réelle...
grand	proche	grande
grand	proche	petite
petit	loin	grande
petit	loin	petite

TAB. 5.1 – Relation entre la taille apparente d'un objet et sa profondeur.

¹La granularité réfère à la "densité" des régions constituant un objet image, c'est-à-dire le nombre de régions par unité de surface.

²L'impact des caractéristiques signal pour l'indexation des images est un problème qui fait actuellement l'objet d'un autre travail de thèse dans l'équipe.

³Rappelons que la profondeur est la distance entre un objet et l'œil de celui qui le voit.

Une autre justification du paramètre de taille vient du domaine de l'Interaction Homme-Machine (IHM), dont les règles de conception ergonomique préconisent un affichage grand pour une perception plus facile des éléments importants, et un affichage réduit pour des éléments d'importance secondaire [SB97]. Par ailleurs, dans un cas limite fictif où un objet image aurait une taille nulle, il ne serait pas du tout représentatif du contenu de l'image ; dans l'autre cas limite où un objet image occuperait la totalité de l'image, il serait très représentatif du contenu de l'image.

Hypothèse 3 (Taille d'un objet image) *Un objet image de grande taille est plus représentatif du contenu sémantique d'une image qu'un objet image de petite taille.*

Cette hypothèse suppose que dans la figure 5.2, l'objet image bateau est plus important dans l'image de gauche que dans l'image de droite, en raison de sa plus grande taille.



FIG. 5.2 – Exemple de deux images entre lesquelles la taille de l'objet image varie.

5.1.2 La position

La physiologie de l'œil fait qu'il perçoit plus nettement les objets en vision centrale (grâce aux cônes situés au niveau de la fovéa⁴), qu'en vision périphérique (grâce aux bâtonnets). Ainsi, quand un observateur regarde une image dans sa globalité, c'est au centre

⁴La fovéa est la zone centrale de la rétine (la tâche jaune), dans laquelle la densité de cônes est la plus élevée.

de l'image qu'il perçoit les objets le plus nettement. Regarder le centre d'une image quelconque permet une meilleure appréhension globale de l'image que n'importe quel autre point. Par ailleurs, quand l'utilisateur regarde au centre de l'image, la totalité de son champ de vision net est dirigé vers l'image, ce qui n'est pas le cas pour les bords de l'image : ce qu'il y a à côté de l'image entre dans le champ de vision net, et parasite ainsi la perception du contenu de l'image. Le but des marges est justement d'interdire la présence d'information utile dans les zones non confortables de vision. Cela est consistant avec les hypothèses qu'ont émis Lim [Lim01] et Mulhem [ML02] sur l'importance des objets images selon leur positionnement au centre ou au bord d'une image.

Hypothèse 4 (Position d'un objet image) *Un objet image situé au centre d'une image est plus important qu'un objet image situé au bord de l'image.*

Cette hypothèse suppose que dans la figure 5.3, l'objet image bateau est plus important dans l'image de gauche que dans l'image de droite, en raison de sa position centrale.

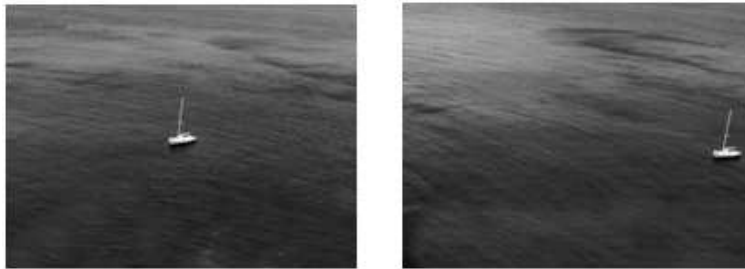


FIG. 5.3 – Exemple de deux images entre lesquelles la position de l'objet image varie.

5.1.3 L'éclatement

L'éclatement d'un objet image fait référence au nombre d'objets physiques englobés par cet objet image. Ce paramètre est une mesure du nombre de régions d'un objet image, pour

une taille donnée. Ce nombre augmente avec le nombre d'objets physiques et le nombre de parties occluses d'un objet physique.

Définition 13 (Éclatement d'un objet image) *L'éclatement d'un objet image est une mesure du nombre de régions qui le composent, et donc du nombre d'objets physiques ou de parties d'objets physiques qu'il englobe.*

L'éclatement soulève la question de l'importance d'un objet image représentant un objet physique par rapport à celle d'un objet image de même surface englobant plusieurs objets physiques, apparaissant donc plus petits dans l'image. L'éclatement d'un objet image a une incidence directe sur le niveau de détail perceptible des objets physiques qui le constituent : pour une surface donnée d'objet image, la taille des régions constituant cet objet image diminue quand leur nombre augmente. Intuitivement, il est raisonnable de penser qu'un gros objet est plus important que plusieurs petits objets en raison de la diminution du niveau de détail perceptible des objets physiques individuels.

Hypothèse 5 (Éclatement d'un objet image) *Un objet image peu éclaté est plus important dans une image qu'un objet image très éclaté.*

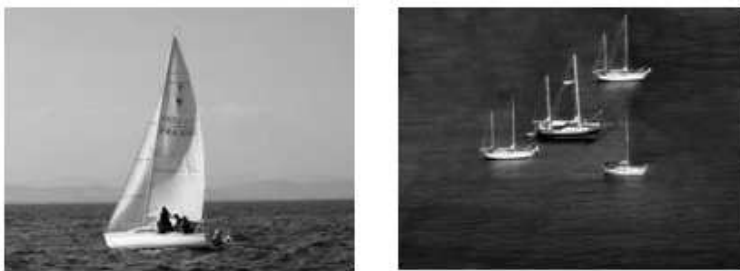


FIG. 5.4 – Exemple de deux images entre lesquelles l'éclatement de l'objet image varie.

Cette hypothèse suppose que dans la figure 5.4, l'objet image bateau est plus important dans l'image de gauche que dans l'image de droite, en raison de son aspect non éclaté (les quatre bateaux de l'image de droite forment un objet image unique).

5.1.4 L'hétérogénéité

L'hétérogénéité de l'image dénote la présence de plusieurs objets de nature différente dans une image.

Exemple 12 *Une image dans laquelle apparaissent un bateau, des personnes, du ciel et de la mer est plus hétérogène qu'une image dans laquelle apparaissent uniquement un bateau et de la mer.*

L'intégration de l'hétérogénéité parmi les paramètres relatifs à l'importance des objets images tire sa justification des sciences cognitives et de la manière dont le système cognitif traite les informations sensorielles. La mémoire du système cognitif d'un utilisateur comprend la mémoire à court terme et la mémoire à long terme. La mémoire à court terme détient les informations en cours de traitement, tandis que la mémoire à long terme est le lieu de stockage de la connaissance permanente [Pre94]. Lorsqu'un utilisateur regarde une image, les données sensorielles provenant de ses yeux sont stockées dans la mémoire à court terme. Les informations d'origine sensorielle y sont représentées sous forme symbolique. Contrairement au codage des stimuli en mémoire sensorielle, ces informations ne sont plus affectées par les caractéristiques physiques. Par exemple, la représentation d'une photo de bateau dans la mémoire à court terme traduit le fait qu'il s'agit d'un bateau. Les informations en provenance de la mémoire à long terme sont des mnèmes⁵ activés par le processeur cognitif. L'activation d'un mnème entraîne sa mise à disposition dans la mémoire à court terme. Cette activation se propage aux mnèmes associés ajoutant de nouveaux éléments dans la mémoire à court terme. La capacité de la mémoire à court terme est estimée à 7 ± 2 mnèmes, selon Miller [Mil56, Mil75]. Lorsque la mémoire à court terme est saturée, l'activation de nouveaux mnèmes efface de la mémoire ceux qui n'ont pas fait l'objet d'une réactivation, de la même façon que pour les registres d'un microprocesseur. L'œil ayant en fait un champ

⁵Un mnème ("*chunk*" en anglais) est une unité cognitive symbolique, une abstraction qui peut être associée à d'autres unités.

de vision nette très étroite, il balaie la surface d'une image par saccades oculaires successives. Ainsi, lorsqu'il y a beaucoup d'objets différents dans une image, l'effort cognitif de compréhension de l'image est plus important.

Ce constat fait l'objet du critère de **charge de travail** de conception ergonomique en IHM [SB97], qui implique l'ensemble des éléments de l'interface qui ont un rôle dans la réduction de la charge perceptive ou mnésique des utilisateurs et dans l'augmentation de l'efficacité du dialogue. Plus la charge de travail est élevée, plus grands sont les risques d'erreurs. De même, moins l'utilisateur est distrait par des informations non pertinentes, plus il peut effectuer sa tâche efficacement. Ce critère d'**efficacité** pour réaliser une tâche est donc à relier dans notre cas à la notion de **facilité** de perception d'un objet.

Par ailleurs, une étude menée au Laboratoire Cognition et Usages ⁶ [Tij03] a montré que la tâche de détection d'un objet cible dans une scène visuelle est influencée par la présence d'autres objets distracteurs (objets identiques perceptivement⁷, par exemple de la même couleur). L'augmentation du nombre d'objets distracteurs diminue les taux de réussite pour la tâche de détection, et augmente le temps de réponse. Cela montre qu'une information particulière est moins facilement visible.

Hypothèse 6 (Hétérogénéité de l'image) *Un objet image qui apparaît seul dans une image est plus important qu'un objet qui apparaît avec d'autres objets images.*

En d'autres termes, plus il y a d'objets images différents dans une image, moins un objet image particulier est représentatif du contenu sémantique de cette image. Le sens du paramètre d'hétérogénéité est similaire à celui d'un facteur de normalisation pour les mots clés d'un texte. Il tend en effet à favoriser les documents dans lesquels apparaissent moins d'informations différentes. Ainsi, les documents dans lesquels une information particulière est "diluée" parmi d'autres éléments sont moins représentatifs de cette information.

⁶Anciennement Laboratoire Cognition & Activités Finalisées.

⁷Il s'agissait, dans cette étude, de lettres et de chiffres de couleur.

Cette hypothèse suppose que dans la figure 5.5, l'objet image bateau est plus important dans l'image de gauche que dans l'image de droite, car l'image de gauche est moins hétérogène.



FIG. 5.5 – Exemple de deux images d'hétérogénéités différentes.

5.2 Modélisation de l'importance des objets images

Les hypothèses que nous venons d'énoncer nous paraissent pertinentes pour modéliser l'importance des objets images. Nous n'avons pas décrit tous les aspects liés à la perception, mais uniquement ceux qui nous paraissent les plus pertinents. Un problème important est à présent la prise en compte de ces hypothèses dans un modèle formel, capable en outre de les intégrer dans une vision globale de "*l'importance d'un objet dans une image*". Cette partie décrit comment nous modélisons ces hypothèses. Nous voulons exprimer cette importance au travers d'une notion de probabilité de pertinence d'un objet par rapport au contenu de l'image, c'est-à-dire la probabilité pour qu'une image soit jugée pertinente par rapport à cet objet. Notre choix de base pour la modélisation des paramètres repose sur une utilisation des probabilités et de la théorie de l'information de Shannon [Sha48], qui constituent un cadre formel approprié pour modéliser les paramètres. La modélisation revient donc à exprimer les critères introduits plus haut dans une approche probabiliste, ces quantités devant être elles-mêmes issues des caractéristiques géométriques et de position des objets images et des régions qui les composent.

La modélisation des critères *taille* et *position* met en jeu les surfaces relatives des objets images ainsi que leur probabilité de pertinence en fonction de leur emplacement dans l'image. La modélisation de l'éclatement et l'hétérogénéité est fondée sur une mesure de "désordre", exprimé par des calculs d'entropie sur les régions et les objets images.

5.2.1 Le paramètre taille

Le paramètre *taille* est modélisé à l'aide de la surface relative des objets images, qui est définie d'après la définition 5 d'un objet image (page 65).

Définition 14 (Surface relative d'un objet image) *La surface relative d'un objet image o d'une image I est :*

$$surface_I(o) = \frac{\sum_{r \subseteq o} (n_r)}{n_I} = \frac{n_o}{n_I}$$

où n_r est le nombre de pixels de chaque région r , n_o est le nombre de pixels de o et n_I est le nombre de pixels de I .

La surface relative d'un objet image o est isomorphe à une probabilité p_I , car elle satisfait les contraintes suivantes d'une distribution de probabilité :

- $\forall o \in S_I^O$, $surface_I(o) \geq 0$ car $n_o \geq \epsilon$ (voir l'hypothèse 2 page 64 sur la taille minimale des régions).
- $surface_I(o) + surface_I(\bar{o}) = 1$, où \bar{o} est l'image privée de o : $\bar{o} = I \setminus o$.
- $\sum_{o \in S_I^O} surface_I(o) = 1$, car les objets images de I forment une partition de I .

Les surfaces relatives des objets images définissent donc une distribution de probabilités : la surface relative d'un objet image correspond à la probabilité qu'un pixel tiré aléatoirement dans l'image appartienne à cet objet image.

La perception humaine des surfaces étant non linéaire, mais plutôt de nature logarithmique [RS90], nous utilisons pour notre paramètre *taille* une surface adaptée à cette perception, que nous nommons log-surface $_I$, et que nous définissons de la manière suivante :

$$\text{log-surface}(o) = \frac{\log(n_o)}{\log(n_I)} = \log_{n_I}(n_o)$$

où \log_{n_i} est le logarithme en base n_i . Notre paramètre taille est modélisé par cette log-surface I :

$$\text{taille}_I(o) = \log\text{-surface}_I(o) \quad (5.1)$$

L'équation 5.1 est cohérente avec l'hypothèse 3 : elle donne une valeur de pertinence plus grande aux objets images de surface plus grande, et inversement.

D'après l'hypothèse 2 page 64 sur la taille minimale des régions, le paramètre taille appartient à l'intervalle $[\log_{n_i}(\epsilon), \log_{n_i}(n_i)]$, qui est l'intervalle $[\log_{n_i}(\epsilon), 1]$. Remarquons que cet intervalle est $[0, 1]$ lorsque $\epsilon = 1$.

5.2.2 Le paramètre position

Afin de prendre en compte l'hypothèse 4 concernant la position des objets images, nous définissons une probabilité de pertinence d'un objet image, définie par une distribution non uniforme de probabilités, dans laquelle les pixels situés au centre de l'image sont associés à des probabilités de pertinence supérieures à celles des pixels situés au bord de l'image.

Cette distribution non uniforme de probabilités, nommée p' , doit satisfaire les contraintes énoncées plus haut, et doit être définie de telle sorte que les pixels du centre soient "*plus probables*" que les pixels du bord de l'image, selon l'hypothèse formulée sur ce paramètre. Elle peut être définie de différentes manières. La figure 5.6 donne quatre exemples de distributions, dont une distribution uniforme (a) qui n'intègre pas cet aspect de la position, présentée à titre illustratif. Pour simplifier l'explication de la figure 5.6, une image I est représentée (en coupe) par un segment en une dimension, deux objets images o_1 et o_2 de même taille sont représentés comme des parties de ce segment (l'un au centre de I et l'autre au bord), et la distribution de probabilités est représentée au-dessus de l'image par une courbe en 2D. Les probabilités associées aux objets images o_1 et o_2 pour chacun des cas ont pour valeurs les aires au-dessous de la courbe pour les intervalles correspondants. Ces distributions sont en 3 dimensions en réalité, symétriques par rapport à un axe central. Les quatre distributions suivantes sont présentées :

- (a) Distribution uniforme : $p(o_1) = p(o_2)$.
- (b) Distribution par paliers : $p'_1(o_1) > p'_1(o_2)$.
- (c) Distribution linéaire : $p'_2(o_1) > p'_2(o_2)$.
- (d) Distribution non linéaire : $p'_3(o_1) > p'_3(o_2)$.

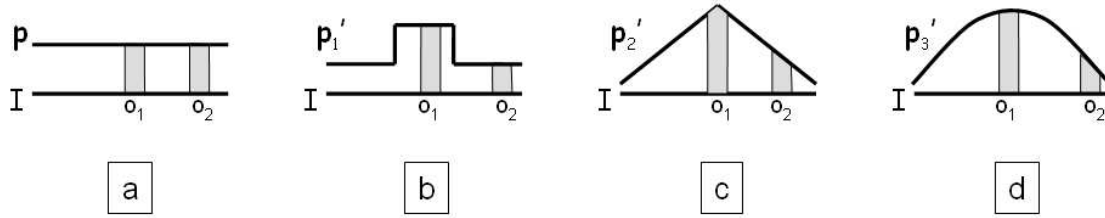


FIG. 5.6 – Distributions de probabilités (2D) : (a) uniforme, (b) par paliers, (c) linéaire, (d) non linéaire.

Les trois distributions (b), (c), et (d) sont satisfaisantes pour l'hypothèse 4, car elles permettent de donner plus d'importance aux objets images centraux. Le paramètre position est défini par la probabilité de pertinence de l'objet image selon cette distribution :

$$\text{position}_I(o) = p'(o) \quad (5.2)$$

où p' est l'une des trois distributions (b), (c), ou (d). L'équation 5.2 est cohérente avec l'hypothèse 4 : elle donne une valeur de pertinence plus grande aux objets images situés au centre de l'image, et inversement. Ce paramètre appartient à l'intervalle $[p'(x), p'(y)]$, x étant un objet image au bord de l'image, et y étant un objet image au centre de l'image, avec $p'(x) \in [0, 1]$, $p'(y) \in [0, 1]$ et $p'(x) < p'(y)$ (l'inégalité est stricte car nous ne considérons pas la distribution uniforme). Le plus grand intervalle pour ce paramètre est $[0, 1]$.

5.2.3 Le paramètre éclatement

L'éclatement réfère au nombre d'objets physiques englobés par un objet image. Pour modéliser ce paramètre, nous définissons une probabilité p_O d'une région r dans un objet

image o , qui correspond à la probabilité qu'un pixel tiré aléatoirement dans cet objet image appartienne à la région r :

$$\text{surface}_O(r) = \frac{n_r}{n_o} = p_O(r)$$

La probabilité $p_O(i)$ d'un pixel i est $1/n_o$. Cela nous permet de définir une entropie H_I^O d'un objet image o , qui est l'entropie définie au sens de Shannon [Sha48] sur la distribution des régions de o :

$$H_I^O(o) = \sum_{r \subseteq o} p_O(r) \times -\log(p_O(r))$$

Le terme “ $-\log(p_O(r))$ ” de la formule précédente représente la quantité d'information apportée par la région r dans l'objet image o . D'après l'hypothèse 2 sur la taille minimale des régions, $p_O(r) \in [\frac{\epsilon}{n_o}, 1]$, et par conséquent : $-\log(p_O(r)) \in [0, \log(\frac{n_o}{\epsilon})]$. La borne inférieure 0 est atteinte dans le cas où l'objet image est constitué d'une seule région, il englobe alors un seul objet physique, ce qui correspond à l'élément certain de la théorie des probabilités. Pour un nombre donné de régions dans l'objet image, l'entropie atteint un maximum en cas d'équirépartition de leur surface, ce qui correspond à une équiprobabilité des éléments. La borne supérieure $\log(\frac{n_o}{\epsilon})$ théorique correspond au cas où l'objet image serait constitué de $\frac{n_o}{\epsilon}$ régions de taille ϵ pixel(s).

Intuitivement, l'entropie est une mesure du “*désordre*” qui règne dans l'objet image. Cette définition de l'entropie permet d'unifier la prise en compte du nombre de régions qui composent l'objet image, et leur taille relative :

- La valeur d'entropie d'un objet image est faible quand le nombre de régions dans l'objet image est petit, ou quand le nombre de régions est grand et qu'une région est “*plus probable*” que les autres. Cela correspond, d'un point de vue perceptif, à un objet physique qui se distingue mieux du fait que la région qui le détoure est plus grande. La valeur minimum de l'entropie d'un objet image est 0 (une seule région de surface n_o).
- La valeur d'entropie d'un objet image augmente avec le nombre de régions ainsi

qu’avec le fait qu’elles sont de même taille. Cela correspond d’un point de vue perceptif à une situation dans laquelle les régions (correspondant aux objets physiques et aux parties d’objets physiques occlus) sont petites relativement à l’objet image. Sa valeur maximale est $n_o \times \frac{1}{n_o} \times \log\left(\frac{n_o}{\epsilon}\right) = \log\left(\frac{n_o}{\epsilon}\right)$.

En conclusion, l’entropie d’un objet image est minimale quand une seule région constitue l’objet image (autrement dit, l’objet image n’est pas éclaté) ; elle est maximale quand n_o régions constituent l’objet image (l’objet image est très éclaté). La valeur d’entropie d’un objet image o appartient donc à l’intervalle $[0, \log\left(\frac{n_o}{\epsilon}\right)]$. Pour les besoins de la modélisation, nous normalisons cette entropie en la divisant par sa valeur maximale, et nous définissons ainsi H_I^O -norm :

$$H_I^O\text{-norm}(o) = \frac{H_I^O(o)}{\log\left(\frac{n_o}{\epsilon}\right)}$$

Par construction, $H_I^O\text{-norm}(o)$ appartient à l’intervalle $[0, 1]$. Nous définissons maintenant le paramètre éclatement d’un objet image o comme étant la valeur complémentaire à 1 de son entropie normalisée $H_I^O\text{-norm}$:

$$\text{éclatement}_I(o) = 1 - H_I^O\text{-norm}(o) \tag{5.3}$$

L’équation 5.3 est cohérente avec l’hypothèse 5 : elle donne une valeur de pertinence plus grande aux objets images peu éclatés, et inversement. Ce paramètre appartient à l’intervalle $[0, 1]$.

5.2.4 Le paramètre hétérogénéité

L’hétérogénéité d’une image dénote la présence de plusieurs objets images de types différents. Comme le paramètre d’éclatement, ce paramètre est modélisé par la définition d’une entropie calculée dans le référentiel *image* (à la différence de l’entropie d’un objet image, qui était calculée dans le référentiel *objet image*), qui est une mesure du “*désordre*” de l’image. L’entropie H_I d’une image I est l’entropie définie sur la distribution des objets

images de I :

$$H_I = \sum_{o \in S_I^O} p_I(o) \times -\log(p_I(o)) \in [0, \log(\frac{n_I}{\epsilon})]$$

où $p_I(o)$ est la surface relative de l'objet (voir la définition 14).

Cette entropie unifie la prise en compte du nombre d'objets images et de leurs tailles relatives, et varie de la même manière que l'entropie d'un objet image :

- La valeur d'entropie d'une image est petite quand le nombre d'objets images dans l'image est petit, ou quand le nombre d'objets images est grand et qu'un objet image est "*plus probable*" que les autres.
- La valeur d'entropie d'une image augmente avec le nombre d'objets images ainsi qu'avec le fait qu'ils sont de même taille.

L'entropie d'une image est minimale quand un seul objet image constitue l'image (l'image n'est pas hétérogène) ; elle est maximale quand $\frac{n_I}{\epsilon}$ objets images constituent l'image (l'image est très hétérogène). Encore une fois, pour les besoins de la modélisation, nous normalisons cette entropie en la divisant par sa valeur maximale, et définissons ainsi H_I -norm :

$$H_I\text{-norm}(I) = \frac{H_I}{\log(\frac{n_I}{\epsilon})}$$

Nous définissons maintenant le paramètre hétérogénéité d'une image I comme étant la valeur complémentaire à 1 de son entropie normalisée H_I -norm :

$$\text{hétérogénéité}_I = 1 - H_I\text{-norm}(I) \tag{5.4}$$

L'équation 5.4, lorsqu'elle est utilisée dans la définition de l'importance d'un objet image, est cohérente avec l'hypothèse 6 : elle donne une valeur de pertinence plus grande objets images des images peu hétérogènes. Ce paramètre appartient à l'intervalle $[0, 1]$.

Remarquons que ce paramètre ne dépend pas d'un objet image en particulier, mais de la configuration de l'image. Elle est la même pour tous les objets d'une image.

5.2.5 Pondération des objets images

Nous venons de définir une modélisation pour les quatre paramètres considérés. Nous proposons maintenant de définir globalement l'importance d'un objet image en combinant ces paramètres. Nous pensons qu'il est souhaitable que la combinaison des paramètres donne une grande valeur d'importance aux objets images lorsque les quatre paramètres donnent une forte valeur de pertinence, et inversement. Nous pensons également que cette valeur d'importance ne devrait pas être nulle lorsque l'un des paramètres a une valeur nulle. Pour ces raisons intuitives (qu'il est nécessaire de vérifier), nous définissons l'importance d'un objet image par une combinaison additive :

$$\text{importance}_I(o) = \frac{1}{4}(\text{taille}_I(o) + \text{position}_I(o) + \text{éclatement}_I(o) + \text{hétérogénéité}_I) \quad (5.5)$$

Nous n'avons pas d'a priori sur les importances relatives des différents paramètres. Les paramètres peuvent éventuellement entrer en concurrence, comme par exemple dans le cas d'un gros objet image sur le bord d'une image, ou dans le cas d'un petit objet image au centre d'une image. Ces deux cas peuvent produire une valeur d'importance identique, ce qui n'est pas en contradiction avec les hypothèses sur la taille et la position.

5.3 Pondération de graphes étoiles

Nous rappelons que l'objectif de la pondération est de représenter pour un graphe étoile sa qualité en tant que descripteur de document, qui dépend d'une part du fait qu'il est représentatif du contenu d'un document donné, et d'autre part du fait qu'il permet de discriminer les documents de la collection.

5.3.1 Importance locale d'un graphe étoile dans un document

Nous avons modélisé précédemment les importances individuelles des objets images ; comme convenu, nous nous basons à présent sur cette modélisation pour définir l'importance d'un graphe étoile générique (terme d'indexation, voir la définition 9 page 69 du vo-

cabulaire d'indexation) pour un document. Pour cela, nous devons calculer les importances des graphes étoiles spécifiques de l'image associés ce terme d'indexation.

L'importance locale d'un graphe étoile spécifique est elle-même fondée sur une agrégation de l'importance de l'ensemble de ses concepts. La fonction qui combine les importances des concepts doit satisfaire les contraintes souhaitées suivantes :

- L'importance d'un graphe étoile spécifique doit augmenter avec l'importance des concepts qui le constituent.
- L'importance d'un graphe étoile spécifique ne doit pas dépendre de l'arité de la relation du graphe étoile.

La **moyenne** permet de définir l'importance d'un graphe étoile spécifique en satisfaisant ces contraintes. Nous proposons donc la formule suivante pour définir l'importance locale d'un graphe étoile spécifique s dans une image j :

$$\text{local-spécifique}_{s,j} = \frac{1}{\text{arité}_s} \sum_{c_k \in s} \text{importance}_j(c_k)$$

où l'importance du concept $c_k = [t, \#o]$ étiquetant l'objet image o est égale à l'importance de o :

$$\text{importance}_I(c_k) = \text{importance}_I(o)$$

L'importance locale d'un terme d'indexation i pour une image j est une fonction des importances des graphes étoiles spécifiques associés à ce terme, et est définie comme étant la somme de ces importances, normalisée en la divisant par la valeur d'importance locale maximale des graphes spécifiques du document, selon la formule couramment employée (voir l'équation 2.2 page 31) :

$$\text{local}_{i,j} = \frac{\sum_s \text{local-spécifique}_{s,j}}{\max_k \text{local-spécifique}_{k,j}}$$

où s dénote les graphes étoiles spécifiques associés à i et k dénote l'ensemble des graphes spécifiques de l'image j .

5.3.2 Importance globale d'un graphe étoile dans la collection

Comme dans les approches classiques du modèle vectoriel textuel (voir la partie 2.5), le calcul de l'importance globale d'un graphe étoile dans la collection vise à donner plus d'importance aux graphes étoiles les plus discriminants (voir la définition de la valeur d'*idf* dans la partie 2.5 page 30). Ce facteur pour un graphe étoile générique i est calculé selon la formule couramment utilisée (voir la formule 2.3 page 31) :

$$\text{global}_i = \log \left(\frac{N}{n_i} \right)$$

où N est le nombre d'images dans la collection, et n_i est le nombre d'images de la collection décrites par le graphe étoile i .

Remarquons que nous calculons ce facteur après expansion des documents (voir la définition 10 de l'expansion de documents page 75), donc dans des index qui contiennent toutes les généralisations des termes de l'index initial défini par l'indexeur. La signification du terme *global* est donc sensiblement différente de celle de l'*idf* [MCMO03, MOCM03]. Pour calculer l'*idf* d'un mot clé, il suffit de procéder à un simple comptage des documents dans lequel ce mot clé apparaît (voir la formule 2.2 page 31). Le calcul de la valeur global_i nécessite de comptabiliser non pas uniquement les documents indexés par i , mais également tous les documents indexés par un terme spécifique de i .

Exemple 13 *Pour calculer la valeur global de $\langle \text{gauche}(\text{Bateau}, \text{Personne}) \rangle$, il faut prendre en compte les documents indexés par $\langle \text{gauche}(\text{Bateau}, \text{Personne}) \rangle$, mais également les documents indexés par $\langle \text{gauche}(\text{Voilier}, \text{Personne}) \rangle$, $\langle \text{gauche}(\text{Bateau}, \text{Jean}) \rangle$, ou bien $\langle \text{gauche}(\text{Voilier}, \text{Matthieu}) \rangle$ (en supposant que le treillis contient les relations $\text{Voilier} \leq \text{Bateau}$, $\text{Jean} \leq \text{Personne}$, et $\text{Matthieu} \leq \text{Personne}$). En effet, tous ces documents sont **implicitement** indexés par $\langle \text{gauche}(\text{Bateau}, \text{Personne}) \rangle$, ce qui est explicité par l'opération d'expansion de documents.*

5.3.3 Poids d'un graphe étoile

Après avoir défini les importances locales et globales des graphes étoiles, nous pouvons maintenant définir la stratégie de pondération que nous proposons pour les graphes étoiles.

Poids d'un graphe étoile de document

En reprenant les approches de pondération classiques dans le modèle vectoriel textuel (voir la partie 2.5 page 30), nous définissons le poids d'un graphe étoile i dans un document j en se basant sur les deux sources suivantes : le terme $local_{i,j}$, qui est calculé pour le graphe étoile i dans le document j , et le terme $global_i$, qui est calculé pour le graphe étoile i , indépendamment du document dans lequel il apparaît :

$$w_{i,j} = local_{i,j} \times global_i \quad (5.6)$$

Le poids d'un graphe étoile combine donc son **importance locale** dans le document j et son **importance globale** dans la collection.

Poids d'un graphe étoile de requête

Dans le cadre de notre modèle vectoriel étendu, les requêtes, comme les documents, sont représentées par des vecteurs de graphes étoiles. La pondération des graphes étoiles des requêtes est fondée sur leur **importance globale** dans la collection, comme il est d'usage de le faire [BYRN99] (voir la formule 2.3 page 31). Par conséquent :

$$w_{i,q} = global_i \quad (5.7)$$

5.4 Conclusion

Nous avons approfondi dans ce chapitre la problématique de la pondération des dimensions de l'espace vectoriel de notre modèle, qui correspondent à des graphes étoiles. Ces poids sont des combinaisons d'importances locales et globales. L'importance locale d'un

terme d'indexation étant une fonction des importances locales des concepts qui le composent, nous nous sommes intéressés à la modélisation de l'importance des concepts. Pour cela, nous avons émis comme hypothèses que l'importance des objets images est une fonction de leur taille, leur position, leur éclatement, et de l'hétérogénéité de l'image. Nous avons modélisé ces paramètres en nous appuyant sur la théorie des probabilités et de l'information, et avons défini une valeur d'importance pour un objet image basée sur cette modélisation. Cela nous a permis de proposer une stratégie de pondération pour les termes d'indexation de notre modèle vectoriel étendu.

Ce chapitre termine la présentation de notre modèle. Le chapitre suivant présente une implantation de notre modèle et les expérimentations que nous avons menées. Ces expérimentations concernent d'une part la validation de nos hypothèses, et d'autre part l'évaluation des performances du système de recherche d'image implanté selon notre modèle.

Troisième partie

Mise en œuvre : Expérimentation et validation

Chapitre 6

Validation de la pondération

La présentation du modèle vectoriel relationnel de recherche d'information a fait l'objet du chapitre 4. Une stratégie de pondération originale dédiée aux termes d'indexation de notre modèle (les graphes étoiles) a été présentée au chapitre 5. Cette stratégie de pondération nous a amenés à introduire des hypothèses relatives à l'importance des objets images, que le présent chapitre propose de valider expérimentalement. Rappelons que ces hypothèses portent sur les critères suivants :

- la **taille des objets images**,
- la **position des objets images**,
- l'**éclatement des objets images**,
- l'**hétérogénéité de l'image**.

Nous tentons également de déterminer expérimentalement dans quelle mesure ces critères peuvent se combiner entre eux, afin d'apporter une validation à la formule 5.5 proposée pour combiner ces critères. En effet, la seule connaissance de la validité des critères ne permet pas de conclure sur la validité de leur combinaison.

La pondération reflétant des critères perceptifs humains, l'établissement de ces critères et leur modélisation doivent être validés à l'aide d'expérimentations confrontant des personnes et des images dans des situations prédéterminées de recherche d'images.

6.1 Description de l'expérimentation

Nous avons mis en place une évaluation confrontant des utilisateurs et des photographies dans une situation de sélection de photographies relatives à un sujet donné, permettant de tester la validité des 4 hypothèses. Un questionnaire est rempli par les utilisateurs à la fin de l'évaluation afin de cerner explicitement les critères qui ont déterminé leur comportement. Nous décrivons dans cette partie la collection d'images que nous avons constituée, donnons des détails sur les participants, et décrivons le déroulement de cette évaluation.

6.1.1 La collection d'images

Deux modalités de variations sont envisagées pour valider chacun des critères (voir le tableau 6.1). La justification du choix de deux modalités est de nature perceptive : il est

Critère	Modalité 1	Modalité 2
taille	Gros	Petit
position	Au centre	Au bord
éclatement	Un objet physique	Plusieurs objets physiques
hétérogénéité	Un type d'objet	Plusieurs types d'objets

TAB. 6.1 – Modalités de variation des 4 critères.

nécessaire que les différences entre les modalités soient suffisamment nettes pour être perçues sans ambiguïté par les participants. Les 4 critères variant selon 2 modalités, nous avons $2^4 = 16$ cas de situations logiques d'apparition d'objets images, présentés à la figure 6.1. Dans cette figure, l'objet image considéré est le disque noir (ou le groupe de disques noirs, qui constitue un objet image éclaté). Par exemple, la situation 5 de cette figure correspond à un objet image **gros** et **non éclaté**, situé au **centre** d'une image **hétérogène**. La situation 12 correspond à un objet image **petit** et **éclaté**, situé au **bord** d'une image **non hétérogène**.

La nature intrinsèque des objets étant sujette à des variations pouvant induire des différences de comportement chez les participants, 4 types d'objets (catégories sémantiques) sont considérés :

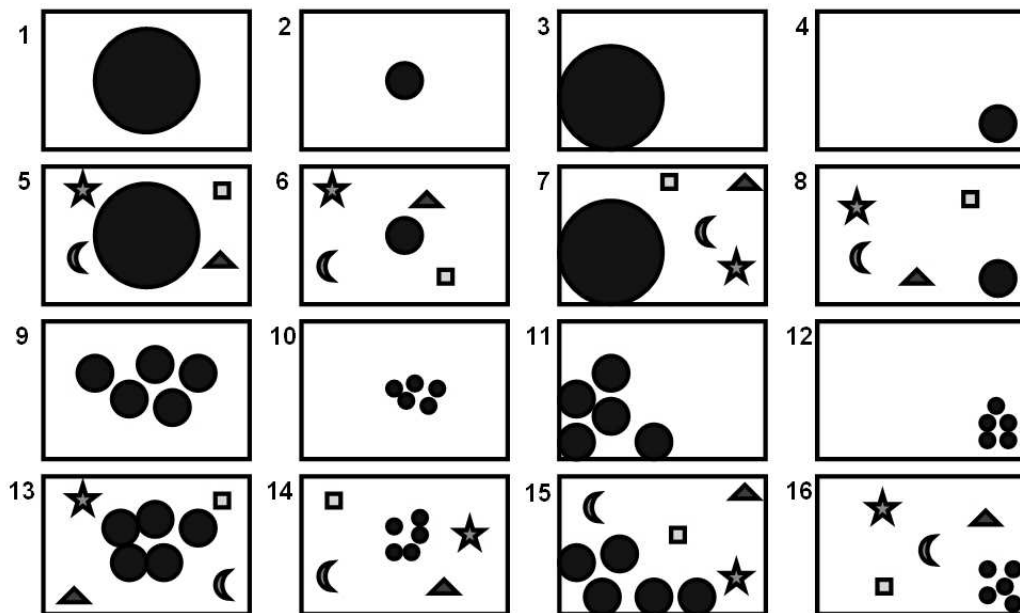


FIG. 6.1 – 16 situations logiques d'apparition d'un objet image (symbolisé ici par les disques noirs), selon les 2 modalités de chacun des 4 critères.

1. des objets inanimés (des bateaux),
2. des animaux (des oiseaux),
3. des êtres humains (des visages d'enfants),
4. une personne nommée (le visage de Jean).

Nous avons effectué des tests sur ces 4 types d'objets afin de nous affranchir de ces variations, et vérifier si les comportements des participants sont ou non influencés par le type des objets. Ces types d'objets sont susceptibles d'apparaître dans des collections de photographies personnelles. Les 3 premiers types d'objets sont non identifiés, à la différence du quatrième, qui est un visage identifié. Pour ce dernier cas, la situation dans laquelle plusieurs instances de l'objet image apparaissent ne se produit pas. Le critère *éclatement* ne s'applique donc pas, par conséquent il n'y a que $2^3 = 8$ situations logiques d'apparition pour ce type d'objet, qui correspondent aux 8 premiers cas de la figure 6.1.

Une collection d'images contenant chacune des situations d'apparition de ces 4 types d'objets a été constituée pour les besoins de l'évaluation. Le plus grand soin a été apporté à la sélection des images afin que les réponses des participants soient le moins possible influencées par leurs goûts personnels :

- La **qualité esthétique** des images est relativement constante. Les photographies proviennent pour la plupart d'une source unique (même photographe, non professionnel).
- La **qualité physique** des images est normalisée en ajustant le contraste et la luminosité. De plus, seules les images nettes et claires ont été retenues.
- La **définition et l'orientation** des images sont toutes identiques (400×300 pixels, format paysage).
- Le **degré de typicalité**¹ des objets est constant dans la mesure du possible : les objets images représentent des objets physiques de même degré de typicalité. Par exemple, pour le cas des bateaux, nous avons sélectionné uniquement des photographies de voiliers de couleur blanche, et les photographies de bateaux à moteurs ou de bateaux de couleurs différentes n'ont pas été retenues dans la collection.
- Les visages sont tous **vus de face**.
- La **parité fille-garçon** a été respectée pour les visages d'enfants.
- Par ailleurs, pour chacun des types d'objets et pour chacune des situations d'apparition, plusieurs exemples d'images sont disponibles, et sont choisis aléatoirement lors de la présentation des images.

Au total, la collection contient 175 images. Les images sont présentées aux participants par paires (voir le déroulement dans la partie 6.1.3). Pour nous assurer d'une couverture totale des critères et de leur combinaison, nous considérons toutes les paires non ordonnées de 2 images parmi les 16 situations, qui sont au nombre de :

$$C_{16}^2 = \frac{15 \times 16}{2} = 120$$

¹Nous entendons par "degré de typicalité" le rapprochement entre un exemplaire d'une catégorie sémantique et la représentation la mieux partagée parmi une population, c'est-à-dire le *prototype* qui illustre le mieux la catégorie considérée.

pour les trois premiers types d'objets, et :

$$C_8^2 = \frac{7 \times 8}{2} = 28$$

pour le quatrième type d'objet. Au total, $3 \times 120 + 28 = 388$ paires d'images sont présentées aux participants.

6.1.2 Les participants

Nous avons procédé à une pré-évaluation à l'aide de 5 participants, afin de nous assurer du bon déroulement de l'évaluation. Cette pré-évaluation nous a permis de régler des problèmes mineurs, d'améliorer l'interface d'évaluation, et d'ajuster le questionnaire afin de cibler plus précisément les questions. L'évaluation a ensuite eu lieu à l'aide de 30 participant(e)s (14 hommes et 16 femmes) âgé(e)s de 24 à 50 ans.

Le choix du nombre de participants a été dicté par deux raisons. D'une part, le nombre 30 correspond à une taille d'échantillon à partir de laquelle ce dernier est considéré comme grand [Bai98], ce qui permet de supposer la population normale et d'utiliser des outils statistiques. D'autre part, nous nous sommes assurés au cours de l'évaluation que les observations sur l'échantillon sont suffisamment proches des observations sur la population entière, en utilisant une méthode pour vérifier la "suffisance" des données appelée **"split-half analysis of consistency"** [MB86]. Cette méthode consiste à séparer aléatoirement en deux moitiés X et Y les données collectées, et à comparer les observations que l'on peut tirer de chacune des moitiés indépendamment, en calculant leur taux de corrélation $\rho_{X,Y}$:

$$\rho_{X,Y} = \frac{\text{covariance}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

où σ_X et σ_Y sont les écarts-types des séries de données X et Y , et $\text{covariance}(X, Y)$ est leur covariance. Si les mêmes conclusions peuvent être tirées des deux moitiés, alors une taille représentative de l'échantillon est atteinte. Lorsque la taille de l'échantillon est suffisante, agrandir cette taille en collectant plus d'information ne change que peu les observations que

l'on peut en tirer. Si les observations diffèrent, il est nécessaire de collecter plus d'information. Martin et Bateson [MB86] précisent qu'à partir d'une corrélation de 0.7, les données peuvent être considérées comme étant fiables. Le taux de corrélation des données que nous avons collectées lors de notre évaluation en fonction du nombre de participants, calculé selon cette méthode, est donné à la figure 6.2. Il vaut 0.962 pour les 30 participants : la corrélation entre les données est donc très forte, ce qui indique que le nombre de participants est suffisant.

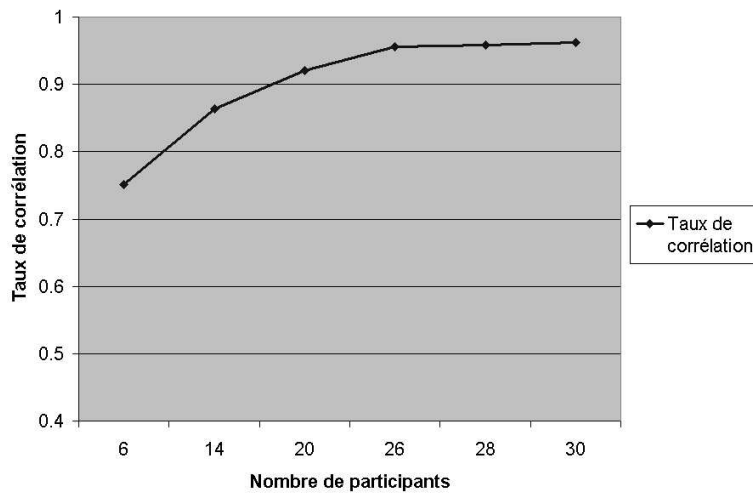


FIG. 6.2 – Taux de corrélation des données en fonction du nombre de participants.

6.1.3 Déroutement

La tâche des participants est de choisir la photographie **la plus pertinente** du sujet recherché, parmi deux photographies présentées par le système. La notion de pertinence étant jugée trop spécialisée (car propre au domaine de la RI), nous avons choisi de demander aux participants d'évaluer la **représentativité** d'une image d'un sujet donné.

Le système présente aux participants des séries de 2 images parmi les 16 situations de la figure 6.1, pour un type d'objet donné (par exemple : *bateau*). Il leur est demandé de choisir

l'image qui, selon eux, est la plus représentative de ce type d'objet. Toutes les combinaisons de situations sont présentées aux participants, pour chacun des types d'objets.

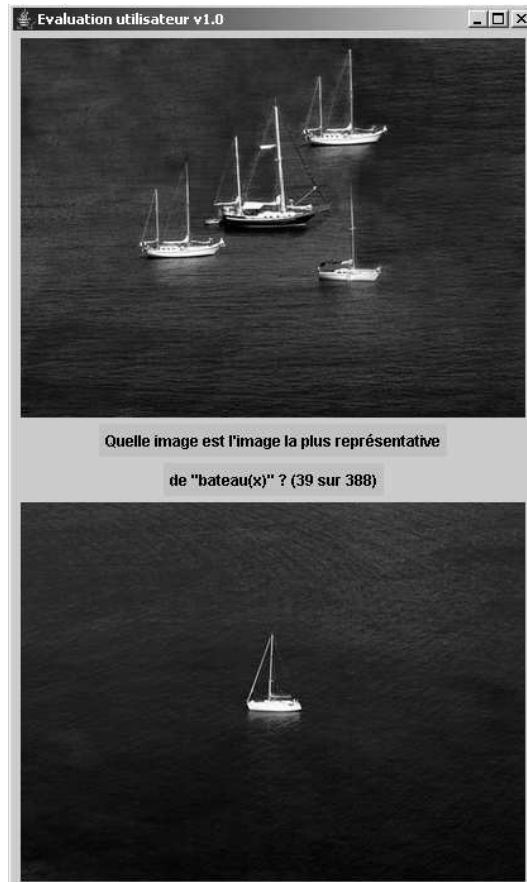


FIG. 6.3 – Interface d'évaluation.

Dans le but de minimiser la lassitude des participants et de maintenir leur attention, les paires d'images des différents types d'objets leur sont présentées dans un ordre varié (séries randomisées). Un prototype d'évaluation a été développé à cet effet, dont l'interface est présentée à la figure 6.3. Elle est composée de l'affichage des deux images, d'un libellé précisant le sujet recherché correspondant aux images affichées, ainsi que d'un compteur indiquant l'état d'avancement de l'évaluation. Dans cette figure, il est demandé aux participants de sélectionner une des deux images de l'objet *bateau*, correspondant aux situations 9

(de la figure 6.1) pour l'image du haut et 2 pour l'image du bas. C'est volontairement que nous n'avons pas permis aux participants d'exprimer le fait que deux images sont également représentatives d'un sujet, afin de les contraindre à faire des choix dans les situations où ils n'en n'auraient pas faits par facilité.

A l'issue de l'évaluation, les sujets remplissent un questionnaire destiné à cerner explicitement les critères qui ont guidé leur choix. Ce questionnaire a été réalisé en prenant soin de poser des questions ouvertes : une première partie du questionnaire demande aux participants d'exprimer librement ces critères. Ensuite, des affirmations neutres et indirectes sont proposées aux participants, et ils indiquent leur accord par un choix d'une valeur qualitative parmi 5 (de "*Pas du tout d'accord*" à "*Totalement d'accord*"). Enfin, il est demandé aux participants de classer les 4 critères par ordre d'importance selon eux. Au total, l'évaluation dure approximativement 30 minutes.

6.2 Interprétation des données collectées

Nous décrivons maintenant l'analyse et l'interprétation des données collectées. L'ensemble des choix des participants constitue une série de distributions empiriques, que nous proposons de modéliser à l'aide de lois de probabilités. Nous commençons par effectuer cette modélisation, et décrivons les résultats pour chaque critère. Nous analysons ensuite la combinaison de ces critères, et présentons les résultats du dépouillement du questionnaire.

6.2.1 Modélisation des données

Chacune des 120 combinaisons de situation met en jeu (i.e. fait varier) 1, 2, 3 ou 4 critères. Par exemple, la combinaison des situations 1 et 2 (de la figure 6.1) met en jeu le critère *taille* uniquement, les autres critères restant constants par ailleurs ; la combinaison des situations 6 et 16 met en jeu les critères *position* et *éclatement*, les autres restant constants. Pour analyser les données collectées, nous proposons d'agréger les 120 combinaisons en catégories selon les critères mis en jeu. Ces catégories correspondent aux éléments de l'ensemble

des parties de l'ensemble $\mathbb{C} = \{T, P, E, H\}$ où T, P, E et H (pour *taille*, *position*, *éclatement* et *hétérogénéité* respectivement) représentent un critère mis en jeu dans une combinaison. La figure 6.4 présente cet ensemble des parties de \mathbb{C} organisé sous la forme d'un treillis de Gallois, chaque nœud indiquant les critères qui varient. Par exemple, la catégorie $\{T, P, H\}$ regroupe les combinaisons mettant en jeu à la fois les critères *taille*, *position* et *hétérogénéité*, comme la combinaison (1,8).

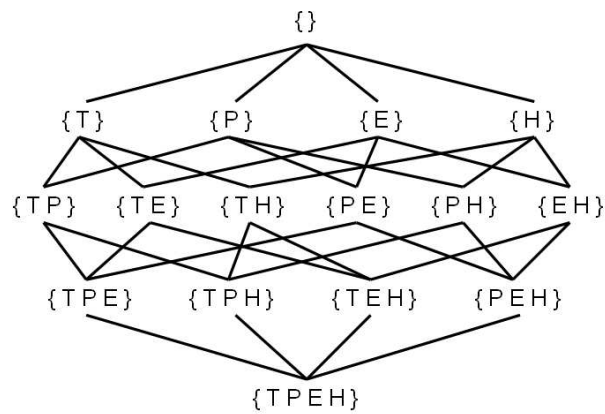


FIG. 6.4 – Treillis de Gallois associé aux critères mis en jeu dans les combinaisons de situations.

Une analyse des 120 combinaisons de situations permet de les catégoriser, et ainsi d'affecter 8 combinaisons à chaque catégorie. Par exemple, la catégorie $\{T, P, H\}$ est constituée des combinaisons (1,8), (2,7), (3,6), (4,5), (9,16), (10,15), (11,14) et (12,13) : toutes ces combinaisons font varier la *taille*, la *position* et l'*hétérogénéité*, l'*éclatement* restant constant.

Nous proposons de modéliser les choix des utilisateurs à l'aide de la **loi binomiale**, qui repose sur un type d'expérience particulier appelé expérience – ou épreuve – de Bernoulli. Cette loi caractérise un phénomène observé qui donne lieu à deux résultats possibles : le succès ou l'insuccès. Une variable aléatoire discrète Y_e (où e est un élément du treillis de

Gallois²), appelée variable de Bernoulli, peut prendre les valeurs 1 ou 0 avec les probabilités respectives p_e et $1 - p_e$. Dans le cadre de notre évaluation, l'expérience consiste à présenter une paire d'images à un participant. Le choix d'une image constitue le résultat de l'épreuve, qui confirme ou infirme une ou plusieurs de nos hypothèses :

$$Y_e = \begin{cases} 1 & \text{si le choix confirme l'hypothèse des critères de } e \text{ (cas du succès)} \\ 0 & \text{si le choix infirme l'hypothèse des critères de } e \text{ (cas de l'insuccès)} \end{cases}$$

Si nous associons aux valeurs possibles de la variable aléatoire Y_e la probabilité qui leur correspond, nous obtenons la loi de probabilité Y_e . L'espérance mathématique – ou moyenne – de la variable aléatoire binaire Y_e est p_e , et correspond à la probabilité d'observation de la valeur 1. La probabilité p_e est estimée sur l'échantillon selon la formule classique :

$$p_e = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

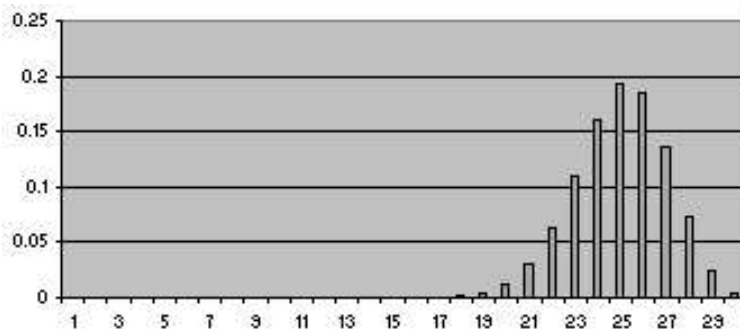


FIG. 6.5 – Exemple de distribution binomiale $B(30,0.8)$.

La figure 6.5 montre un exemple de distribution binomiale $B(n, p)$, pour $n = 30$ (nombre de tirages) et $p = 0.8$ (probabilité associée à chaque tirage). Par exemple, la valeur associée au “bin” 27 indique la probabilité d'obtenir 27 succès en effectuant 30 tirages, la probabilité de succès étant de 0.8 pour chaque tirage.

²Nous ne considérons pas l'ensemble vide, qui correspondrait aux combinaisons qui ne font varier aucun critère.

Nous nous intéressons dans un premier temps aux 4 catégories qui ne font varier qu'un critère, ce qui nous permettra de valider nos hypothèses. Nous nous intéresserons ensuite aux 11 autres catégories, qui correspondent aux combinaisons de critères.

Nous voulons estimer les probabilités des variables aléatoires binaires Y_e . Les moyennes et les écarts-types des données fusionnées pour les 30 participants sont présentés au tableau 6.2 (voir les distributions empiriques aux figures 6.6 à 6.9). Par exemple, la valeur 0.76 pour le critère *taille* du type d'objet *bateau* représente le rapport du nombre de choix confirmant l'hypothèse sur la taille, sur le nombre total de combinaisons mettant en jeu la taille uniquement, pour ce type d'objet. Autrement dit, cette valeur correspond à la probabilité estimée qu'un participant trouve une image de gros bateau plus représentative de *bateau* qu'une image de petit bateau. Nous voyons dans ce tableau que le critère *éclatement* semble

Critère	Bateau	Oiseau	Visage d'enfant	Visage nommé
taille	0.76 (0.15)	0.74 (0.18)	0.77 (0.16)	0.97 (0.09)
position	0.84 (0.13)	0.82 (0.12)	0.77 (0.15)	0.70 (0.27)
éclatement	0.38 (0.20)	0.53 (0.25)	0.60 (0.25)	N/A (N/A)
hétérogénéité	0.88 (0.12)	0.65 (0.19)	0.81 (0.12)	0.86 (0.22)

TAB. 6.2 – Moyennes (et écarts-types) des critères pour les 30 participants.

moins important que les autres : les probabilités estimées pour ce critère sont proches de 0.5, par opposition aux autres critères, dont les probabilités estimées sont plus proches de 0.7 ou 0.8, et tendent à confirmer nos hypothèses initiales.

Variable aléatoire	Probabilité (é.-t.)	Test T de Student (H_0)
$Y_{\{T\}}$	0.81 (0.14)	6.57E-13
$Y_{\{P\}}$	0.78 (0.16)	1.84E-10
$Y_{\{E\}}$	0.50 (0.23)	0.947954
$Y_{\{H\}}$	0.78 (0.14)	6.74E-12

TAB. 6.3 – Probabilités estimées (et écarts-types) pour tous les types d'objets, et probabilités associées au test T de Student (H_0).

Les probabilités estimées pour les variables aléatoires $Y_e, e \in \{\{T\}, \{P\}, \{E\}, \{H\}\}$ sont présentées dans le tableau 6.3. Elles sont comprises entre 0.78 et 0.81, excepté pour

le critère *éclatement*, dont la probabilité estimée est de 0.5. Afin de valider statistiquement nos résultats, nous avons recours à un test d'hypothèse unilatéral, l'hypothèse nulle étant $H_0 : Y_e > 0.5$ pour chacun des critères, et l'hypothèse alternative étant $H_1 : Y_e \leq 0.5$. Les probabilités associées au test T de Student [Bai98] sont également présentées dans le tableau 6.3 ; elles correspondent à la probabilité de rejeter à tort l'hypothèse H_0 , alors qu'elle est vraie. Les tests pour les 3 critères *taille*, *position* et *hétérogénéité* sont **très hautement significatifs** avec des valeurs de significativité très inférieures au seuil de 1%. Nous pouvons donc conclure que nos hypothèses 3, 4 et 6 correspondant respectivement aux critères *taille*, *position* et *hétérogénéité* sont statistiquement valides. Le test T de Student pour le critère *éclatement* donne une probabilité de 0.95, ce qui nous permet pas de conclure, avec un seuil de significativité acceptable, que l'hypothèse 5 correspondant au critère *éclatement* est statistiquement valide.

6.2.2 La taille

L'histogramme de la figure 6.6 présente la distribution empirique des choix des participants pour le critère *taille*. Par exemple, le "bin" 7 correspond au nombre de personnes qui ont sélectionné 7 fois sur 8 l'image contenant le gros objet.

Nous constatons que les distributions sont relativement homogènes selon les types d'objets, mis à part le cas du visage nommé, pour lequel le pic dans la distribution indique qu'une grande majorité de participants a sélectionné l'image contenant un visage de grande taille (un gros plan ou un portrait).

6.2.3 La position

L'histogramme de la figure 6.7 présente la distribution empirique des choix des participants pour le critère *position*. Cette distribution est moins homogène selon les types d'objets que la précédente, ce qui signifie que le critère *position* n'a pas le même impact pour tous les types d'objets. Nous constatons que la distribution pour le visage nommé est plus étalée que

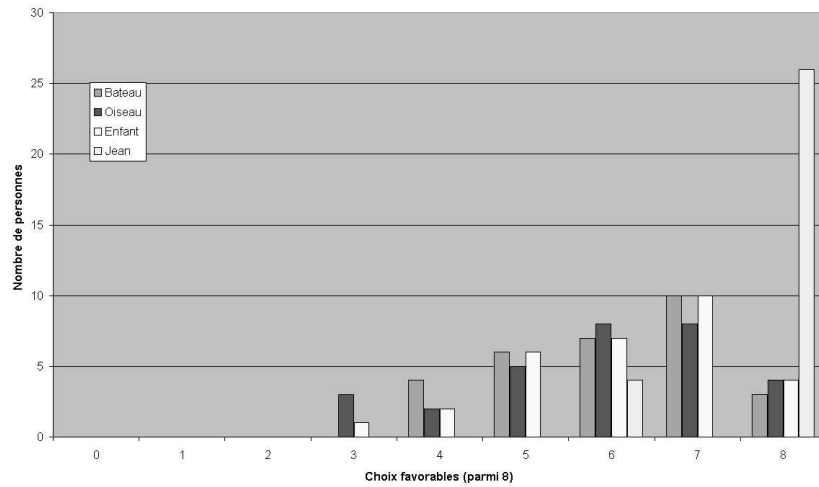


FIG. 6.6 – Distribution des choix selon le type d'objet pour le critère *taille*.

les autres (ce qui est confirmé par l'écart-type de 0.27 plus grand pour la position du visage nommé au tableau 6.2). En effet, 5 participants n'ont sélectionné que 2 fois sur 8 l'image contenant le gros visage nommé. Le critère de position est donc moins important dans ce cas.

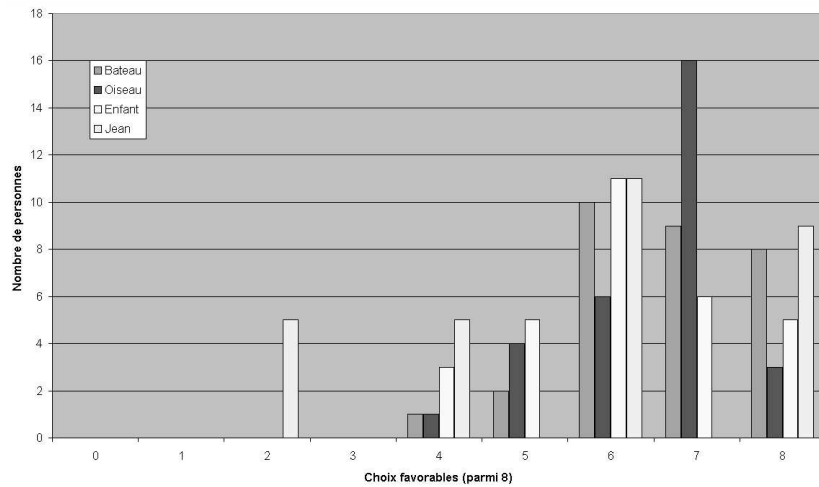


FIG. 6.7 – Distribution des choix selon le type d'objet pour le critère *position*.

6.2.4 L'éclatement

Ce critère ne semble pas pertinent pour modéliser l'importance des objets images. Les statistiques extraites des données reflètent le fait qu'en moyenne, les participants ont indiqué **une fois sur deux** que l'image contenant l'objet éclaté est plus représentative que l'image contenant l'objet non éclaté. Cela traduit le fait que dans les deux situations, les images sont également représentatives du sujet.

L'histogramme de la figure 6.8 présente la distribution empirique des choix des participants pour le critère *éclatement*. Comme l'indique la probabilité estimée pour ce critère, les valeurs de l'histogramme sont distribuées autour de 0.5, c'est-à-dire 1 choix sur 2 confirme l'hypothèse. De plus, c'est le cas pour tous les types d'objets (sauf pour le visage nommé, pour lequel le critère *éclatement* ne s'applique pas).

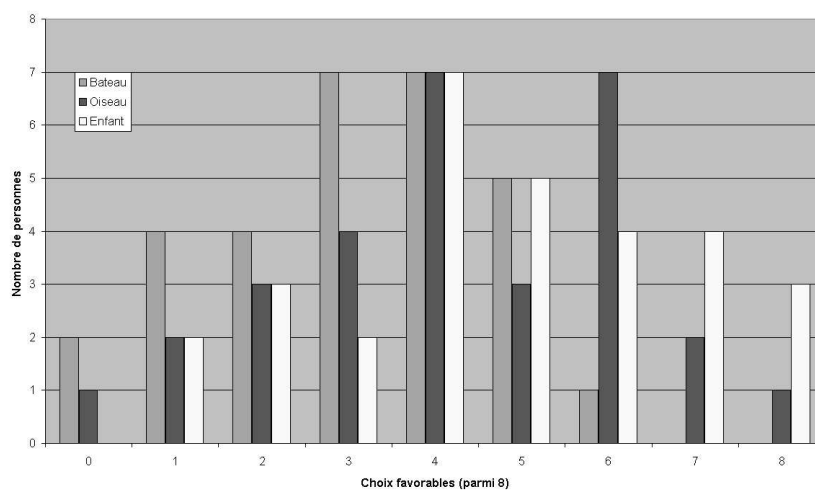


FIG. 6.8 – Distribution des choix selon le type d'objet pour le critère *éclatement*.

6.2.5 L'hétérogénéité

L'histogramme de la figure 6.9 présente la distribution empirique des choix des participants pour le critère *hétérogénéité*. Cette distribution est relativement homogène pour

les types d'objets, bien que la distribution pour les oiseaux soit légèrement décalée vers le centre, ce qui peut signifier que le critère *hétérogénéité* a moins d'impact pour ce type d'objet.

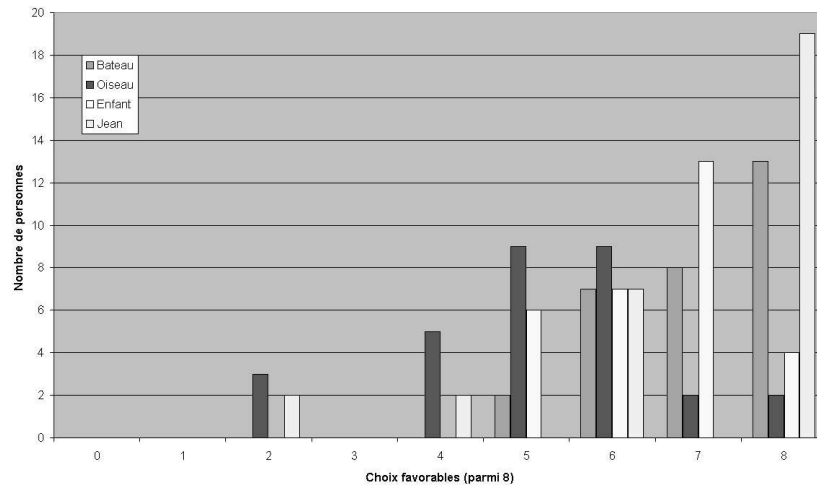


FIG. 6.9 – Distribution des choix selon le type d'objet pour le critère *hétérogénéité*.

6.2.6 Combinaison de critères

Pour aller plus loin dans l'analyse, nous nous interrogeons sur les dépendances entre les critères et la manière dont ils se combinent.

Nous nous intéressons pour cela aux 11 autres catégories du treillis de Gallois, qui correspondent aux combinaisons de critères. Lorsque 2 critères sont mis en jeu dans une combinaison, le choix d'une combinaison peut soit confirmer les 2 hypothèses associées à ces critères *en même temps*, soit confirmer une seule des 2 hypothèses, *tout en infirmant* l'autre. Par exemple, le choix de la situation 1 dans la combinaison (1,6) confirme les hypothèses sur les critères *taille* et *hétérogénéité*, tandis que le choix de la situation 5 dans la combinaison (2,5) confirme l'hypothèse sur le critère *taille*, et infirme l'hypothèse sur le critère *hétérogénéité*). Le tableau 6.4 présente ces résultats, avec la convention de notation suivante : le critère dont l'hypothèse est infirmée par le test est noté entre parenthèses. Par exemple, la

variable $Y_{\{T(E)\}}$ modélise la probabilité associée aux choix qui confirment l’hypothèse sur le critère *taille* et infirment l’hypothèse sur le critère *éclatement*. Soulignons que la probabilité de $Y_{\{T(E)\}}$ est le complément à 1 de $Y_{\{E(T)\}}$.

V.A.	Probabilité	V.A.	Probabilité
$Y_{\{TP\}}$	0.91	$Y_{\{T(P)\}}$	0.54
$Y_{\{TE\}}$	0.74	$Y_{\{T(E)\}}$	0.70
$Y_{\{TH\}}$	0.91	$Y_{\{T(H)\}}$	0.47
$Y_{\{PE\}}$	0.72	$Y_{\{P(E)\}}$	0.73
$Y_{\{PH\}}$	0.94	$Y_{\{P(H)\}}$	0.51
$Y_{\{EH\}}$	0.76	$Y_{\{(E)H\}}$	0.73

TAB. 6.4 – Probabilités estimées pour les variables mettant en jeu 2 critères.

Dans le tableau 6.4, nous comparons les résultats par ligne. Nous constatons dans la première ligne que la taille et la position se combinent en se renforçant : la probabilité associée à $Y_{\{TP\}}$ est de 0.91, et est donc plus élevée que les probabilités associées à $Y_{\{T\}}$ (0.76) et à $Y_{\{P\}}$ (0.81) individuellement (voir le tableau 6.3). Dans la première ligne, nous constatons également que lorsque ces critères varient dans des sens qui confirment le critère *taille* tout en infirmant le critère *position*, alors la probabilité devient 0.54. Cela traduit le fait que les participants ont sélectionné une fois sur deux (en moyenne) une situation correspondant à un gros objet situé sur le bord de l’image, en la jugeant plus pertinente qu’une situation correspondant à un petit objet au centre de l’image (choix de la situation 3 de la figure 6.1 dans la combinaison (2,3), par exemple). Cela indique que les critères *taille* et *position* sont d’importances comparables.

Ce comportement des variables est similaire pour toutes les paires de critères pertinents, qui correspondent aux variables $Y_{\{TP\}}$, $Y_{\{TH\}}$ et $Y_{\{PH\}}$. En ce qui concerne les variables intégrant le critère *éclatement* ($Y_{\{TE\}}$, $Y_{\{PE\}}$ et $Y_{\{EH\}}$), il est intéressant de noter que la probabilité associée ne change que peu entre les variables qui confirment ou infirment l’hypothèse sur ce critère.

De la même manière que nous avons analysé les variations de 2 critères dans une combi-

raison, nous avons analysé les variations de 3 et 4 critères, qui révèlent une confirmation de ce comportement. Nous présentons dans le tableau 6.5 les résultats de l'analyse de $Y_{\{TPEH\}}$. La deuxième colonne de ce tableau présente les probabilités associées à cette variable pour le trois cas suivants :

- le choix confirme toutes les hypothèses (la première ligne),
- le choix confirme 3 hypothèses et infirme l'une d'elles (les quatres lignes suivantes),
- le choix confirme 2 hypothèses et en infirme 2 (les trois dernières lignes).

V.A.	Probabilité
$Y_{\{TPEH\}}$	0.93
$Y_{\{TPE(H)\}}$	0.80
$Y_{\{TP(E)H\}}$	0.91
$Y_{\{T(P)EH\}}$	0.86
$Y_{\{(T)PEH\}}$	0.83
$Y_{\{TP(EH)\}}$	0.77
$Y_{\{(T)P(E)H\}}$	0.69
$Y_{\{T(PE)H\}}$	0.71

TAB. 6.5 – Probabilités estimées pour la variable $Y_{\{TPEH\}}$ mettant en jeu les 4 critères.

Nous voyons dans ce tableau que, comme dans le tableau 6.4, les combinaisons des critères pertinents renforce leur probabilité, et le critère *éclatement* influence peu ces probabilités.

6.3 Questionnaire

Le dépouillement des questionnaires permet de confirmer les critères qui ont guidé les choix des participants. Le tableau 6.7 présente les affirmations proposées aux participants, ainsi que les rangs moyens (et les écarts-types) des choix qu'ils ont effectués, parmi 5 (1 pour "Pas du tout d'accord" et 5 pour "Totalemt d'accord"). Les affirmations sont groupées par critères, et nous nous intéressons pour chaque critère :

- à déterminer si il a influencé les choix des participants,
- à connaître la modalité la plus pertinente de ce critère pour les participants,

– à savoir si le type d’objet a une influence sur le critère considéré.

Le tableau 6.7 présente également les probabilités associées à un test bilatéral T de Student, l’hypothèse nulle étant $H_0 : \text{RangMoyen} \neq 3$ où RangMoyen est le rang moyen pour chaque question, et l’hypothèse alternative étant $H_1 : \text{RangMoyen} = 3$. Nous voulons vérifier que les rangs moyens sont significativement différents de la valeur médiane. Les critères *taille*, *position* et *hétérogénéité* ont significativement (seuil inférieur à 1%) influencé les choix des participants, dans le sens de nos hypothèses. De plus, d’après les participants, ces critères s’appliquent pour tous les types d’objets. Ces statistiques confirment également que les participants ne sont pas sensibles au critère *éclatement*, et cela pour tous les types d’objets (seuil inférieur à 5%).

Affirmation	Rang moyen (é.-t.)	Test T de Student (H_0)
La taille a influencé vos choix.	4.67 (0.68)	8.55E-15
Un gros objet est plus représentatif.	4.00 (0.93)	1.25E-05
Un petit objet est plus représentatif.	1.62 (0.92)	1.01E-07
Réponses indépendantes du type d’objet.	4.04 (1.12)	7.18E-05
La position a influencé vos choix.	4.54 (0.82)	1.58E-10
Un objet au centre est plus représentatif.	4.37 (0.87)	4.15E-08
Un objet au bord est plus représentatif.	1.71 (0.86)	8.52E-08
Réponses indépendantes du type d’objet.	4.33 (0.76)	6.33E-09
L’éclatement a influencé vos choix.	3.29 (1.19)	0.129175
Un objet non éclaté est plus représentatif.	3.17 (0.92)	0.191192
Un objet éclaté est plus représentatif.	3.33 (1.09)	0.073871
Réponses indépendantes du type d’objet.	3.58 (1.38)	0.024928
L’hétérogénéité a influencé vos choix.	4.42 (1.08)	2.77E-08
Une image non hétérogène est plus repr.	4.62 (0.58)	6.22E-13
Une image hétérogène est plus repr.	1.75 (0.85)	1.16E-07
Réponses indépendantes du type d’objet.	4.21 (1.06)	5.71E-06

TAB. 6.6 – Rangs moyens des réponses au questionnaire.

Les participants ont classé les 4 critères que nous leur avons proposés par ordre d’importance (1 pour le plus important, 4 pour le moins important). Les rangs moyens de ces classements sont présentés au tableau 6.7. Le critère *taille* est celui qui est classé comme étant le plus important. Ce résultat est en accord avec le rang moyen le plus élevé pour la

taille au tableau 6.6 : 4.67. Viennent ensuite les critères *position* et *hétérogénéité*.

Critère	Rang moyen (écart-type)
taille	1.94 (0.98)
position	2.12 (1.02)
éclatement	3.44 (0.63)
hétérogénéité	2.50 (1.26)

TAB. 6.7 – Rangs moyens (et écarts-types) des classements des 4 critères par ordre d’importance.

Le critère *taille* est perçu comme le plus important, devant les critères *hétérogénéité* et *position*, dans l’ordre. Les participants ont donc une bonne perception des critères qui influencent leurs choix (le critère *éclatement* est classé comme étant le moins pertinent). Cependant, le classement des 3 autres critères est différent de celui obtenu selon les probabilités du tableau 6.3, ce qui prouve l’intérêt de notre protocole d’évaluation, car la simple analyse du questionnaire ne reflète pas la réalité.

6.4 Nouvelle formule de pondération des objets images

Nous avons montré que les hypothèses concernant la *taille*, la *position* et l’*hétérogénéité* des objets images sont statistiquement valides, tandis que l’hypothèse sur le critère *éclatement* ne l’est pas. Le critère *éclatement* n’a pas significativement influencé les choix des participants, de plus sa combinaison avec les autres critères ne change pas les conclusions que nous pouvons tirer des données. Au vu de ces résultats, nous proposons de ne pas conserver ce paramètre dans la modélisation de l’importance des objets images donnée par notre formule 5.5, et ajustons la définition de la valeur d’importance pour les objets images de la manière suivante :

$$\text{importance}_I(o) = \frac{1}{3}(\text{taille}_I(o) + \text{position}_I(o) + \text{hétérogénéité}_I) \quad (6.1)$$

6.5 Discussion

Nous avons modélisé les données à l'aide la loi binomiale. Cependant, il a pu arriver qu'au cours de l'évaluation, le comportement des participants change, en raison d'une évolution de leurs critères de représentativité des objets. Leurs choix sont donc susceptibles de ne pas être indépendants les uns des autres, ce qui n'est pas pris en compte dans la modélisation à l'aide de la loi binomiale³. Nos hypothèses sont néanmoins soutenues pas des tests statistiques de significativité des données collectées.

6.6 Conclusion

Nous avons présenté dans ce chapitre une expérimentation dont l'objectif est de tester la validité statistique de nos hypothèses. Pour cela, nous avons constitué une collection d'images, mené une campagne d'évaluation auprès de 30 participants, et analysé les données collectées.

Nous tirons de ces analyses la conclusion que 3 parmi les 4 critères sont validés : la taille, la position et l'hétérogénéité ; le critère éclatement n'est pas validé par nos test. Par ailleurs, nous avons constaté que les critères validés se combinent en se renforçant. Cela apporte une confirmation notre intuition concernant la combinaison additive des critères, et valide l'utilisation de l'addition dans les formules 6.1 et 5.5.

³Il aurait peut-être été souhaitable de faire remplir en plus aux participants un questionnaire de pré-évaluation, afin de comparer les réponses avant et après l'évaluation, et mesure le biais induit par l'évaluation dans les réponses fournies.

Chapitre 7

Evaluation du système

Le chapitre 6 a présenté une validation expérimentale de notre modèle de pondération des objets images, qui constitue une part importante du modèle. Rappelons que notre langage d'indexation est composé des types de graphes étoiles, qui constituent les dimensions d'un espace vectoriel, dans lequel sont représentés les documents et les requêtes. Le modèle proposé a été implanté en utilisant le système SMART [Sal71] comme noyau ; le présent chapitre traite de l'évaluation de ce système. Comme il en est fait mention dans la partie 2.3 page 25, les SRI sont comparés sur la base de courbes de rappel/précision calculées sur des collections de tests. Nous commençons par décrire la collection que nous avons utilisés, et nous décrivons ensuite les évaluations menées sur cette collection. L'objectif de ces évaluations est de comparer notre approche relationnelle avec une approche sans relation (basée sur des mots clés) et avec une approche orientée précision basée sur la projection de graphes. Nous avons également évalué l'impact de notre fonction de pondération dans le système.

7.1 Collection de tests

Nous avons utilisé une collection de tests contenant 498 photographies personnelles et 38 requêtes. Les photographies de cette collection sont décrites par des graphes conceptuels selon le modèle de représentation d'images EMIR² de Mechkour [Mec95], générés après une indexation automatique des photographies par des *Visual Keywords* suivant les travaux

de Lim [Lim00]. Cette indexation utilise un apprentissage à base de réseaux de neurones, et produit une liste de mots clés par image, chaque mot clé étant associé à une région de l'image. Les régions sont en bijection avec les concepts des graphes conceptuels. Le treillis



FIG. 7.1 – Aperçu des images de la collection de tests utilisée.

de types de concepts, dont la construction est manuelle, est composée de 104 concepts, et est de profondeur 6. Il contient les mots clés issus du processus d'indexation de la collection (comme *Homme*, *Bâtiment*, *Ciel*, ou *Piscine*), et explicite les liens de généralité/spécificité entre eux. Le treillis de type de relations est également construit manuellement. Il contient 46 types de relations, et est de profondeur 3. Il contient essentiellement des relations spatiales, de composition, et de positions absolues des objets dans l'image. Les graphes conceptuels décrivant les images contiennent 29 relations en moyenne. La figure 7.1 donne un aperçu des photographies de cette collection.

Le système a été évalué sur 38 requêtes (sous la forme de graphes conceptuels), pour lesquelles nous avons établi manuellement la liste des documents pertinents. Les graphes des requêtes sont composés de 4 relations en moyenne. La figure 7.2 présente une requête de la collection décrivant “*images d'une personne, avec des bâtiments situés de part et d'autre de la personne*”. Parmi ces requêtes, 29 sont des requêtes ne mettant pas en jeu des relations explicites entre les concepts (par exemple, “*images de bâtiments*”), et 9 requêtes

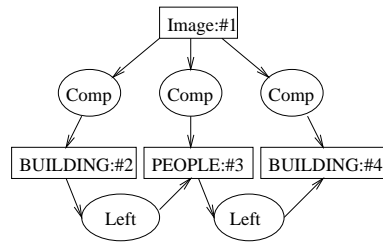


FIG. 7.2 – Un exemple de requête de la collection de tests.

utilisent des relations explicites entre les concepts, comme la requête de la figure 7.2.

Afin d’obtenir des index de documents composés de graphes étoiles selon notre modèle, nous décomposons les graphes conceptuels. La figure 7.3 montre la décomposition d’un graphe conceptuel en graphes étoiles, et l’ensemble des graphes étoiles résultants. Après

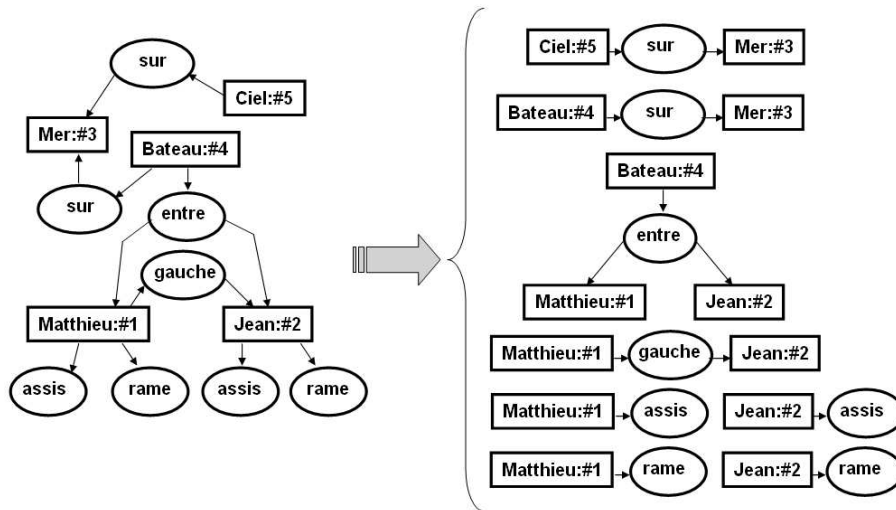


FIG. 7.3 – Décomposition d’un graphe conceptuel en graphes étoiles.

décomposition des graphes et expansion des index des documents, les vecteurs des documents comprennent 454 dimensions non nulles en moyenne. Après l’expansion des documents, la collection comprend 15411 termes d’indexation au total. Les vecteurs des requêtes

contiennent en moyennes 4 dimensions non nulles. Les documents et les requêtes sont représentées par des vecteurs de termes d'indexation, qui sont pondérés dans les documents selon l'équation 5.6, et dans les requêtes selon l'équation 5.7. La valeur d'importance des concepts des graphes étoiles des documents est calculée en utilisant le critère *taille* uniquement de la formule 6.1. La correspondance est évaluée par un cosinus selon la formule 4.1.

7.2 Evaluations comparatives

L'objectif de notre évaluation est de comparer les résultats de notre système vectoriel relationnel :

- d'une part à des systèmes basés sur des mots clés (comme les systèmes décrits dans la partie 3.3 page 46) afin d'évaluer l'impact de l'utilisation des relations dans un SRIm,
- d'autre part à des systèmes basés sur l'utilisation de la projection de graphes conceptuels comme fonction de correspondance.

7.2.1 Comparaison avec les mots clés

Nous comparons ici les résultats de notre système avec les résultats d'un système vectoriel basé sur les mots clés uniquement, c'est-à-dire sans les relations (les mots clés sont les concepts des graphes conceptuels). Une expansion de documents est appliquée aux vecteurs du système basé sur les mots clés afin de rendre la comparaison pertinente. La figure 7.4 présente les résultats sous la forme de courbes de la précision des systèmes pour les 38 requêtes en fonction du rappel. Nous constatons que la précision de notre système est largement supérieure à la précision du système basé sur les mots clés, dans les petites valeurs de rappel (ce qui correspond à la précision mesurée sur les documents les mieux classés).

Les courbes pour les deux systèmes ont des précisions comparables pour des grandes valeurs de rappel, qui se rejoignent lorsque le rappel vaut 1. Cela signifie que l'ensemble des documents retrouvés par les deux systèmes est identique (le système basé sur les mots clés retrouve tous les documents contenant des bâtiments et des personnes pour la requête

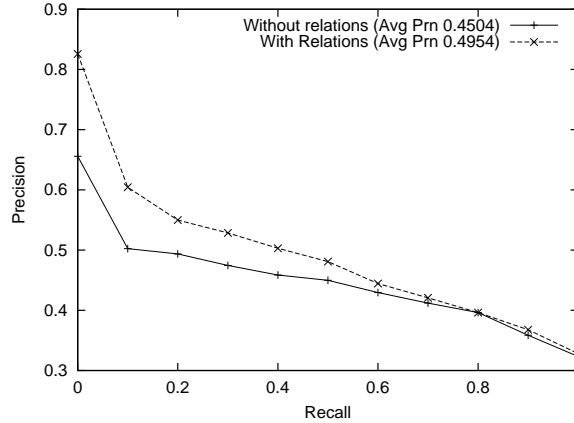


FIG. 7.4 – Courbes de rappel/précision pour notre système et le système basé sur les mots clés.

de la figure 7.2) .

La courbe de notre système est au-dessus de la courbe du système basé sur les mots clés pour toutes les valeurs de rappel, et le gain relatif est de 9.99% en précision moyenne. Cela confirme le fait que l’indexation relationnelle permet une représentation plus précise des documents et des requêtes.

7.2.2 Comparaison avec la projection de graphes

Nous proposons maintenant de comparer les résultats de notre système avec un système utilisant la projection totale de graphes comme mesure de similarité entre les documents et les requêtes. Nous utilisons pour cela un prototype qui implante cette projection de graphes [MT01]. Ce prototype utilise une pondération des concepts fondée sur la surface des régions, et permet de classer les documents retrouvés grâce à une valeur de pertinence système calculée en faisant la somme des poids associés aux sous-graphes sur lesquelles la requête se projette.

Pour cette comparaison, nous présentons les résultats obtenus pour toutes les requêtes, et également les résultats obtenus d’une part pour le sous-ensemble des 29 requêtes ne mettant

pas en jeu de relations explicites entre les concepts, et d'autre part pour le sous-ensemble des 9 requêtes mettant en jeu des relations explicites entre les concepts.

La figure 7.5 présente les courbes de rappel/précision pour l'évaluation concernant l'ensemble de requêtes. La précision pour les faibles valeurs de rappel est légèrement plus faible

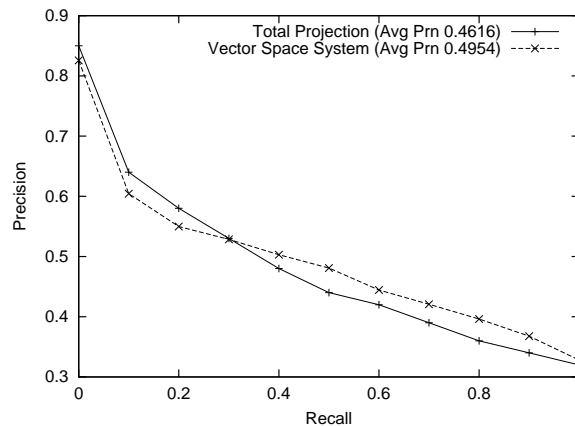


FIG. 7.5 – Courbes de rappel/précision pour notre système et le système basé sur la projection de graphes.

pour notre système que pour le système implantant la projection totale de graphes. Cela peut s'expliquer par le fait que le système basé sur la projection (qui est un système orienté précision), représente des informations de jointure entre les graphes étoiles qui ne sont pas représentées dans les vecteurs de notre système. Par exemple, la requête de la figure 7.2 met en relation trois concepts [Building :#2], [People :#3] et [Building :#4] par le biais de deux relations (left). La décomposition de ce graphe conceptuel ne permet pas de conserver l'information que le concept [People :#3] est le même dans les deux graphes étoiles résultants. La courbe de notre système passe "au-dessus" de la courbe du système basé sur la projection pour une valeur de rappel de 0.3. La similarité de vecteur effectuée par notre système évalue une correspondance *partielle* document-requête. Il en résulte une plus grande souplesse que la correspondance totale, qui est binaire – bien qu'elle soit pondérée. Ainsi, notre système retrouve plus de documents pertinents pour des grandes valeurs de rappel, qui sont

donc en correspondance partielle avec la requête. Ces documents ne sont pas retrouvés par le système basé sur la projection, ce qui se traduit par un gain relatif de 7.32% de précision moyenne pour notre système.

Les courbes de rappel/précision de la figure 7.6 présentent les résultats de l'évaluation pour les sous-ensembles de requêtes.

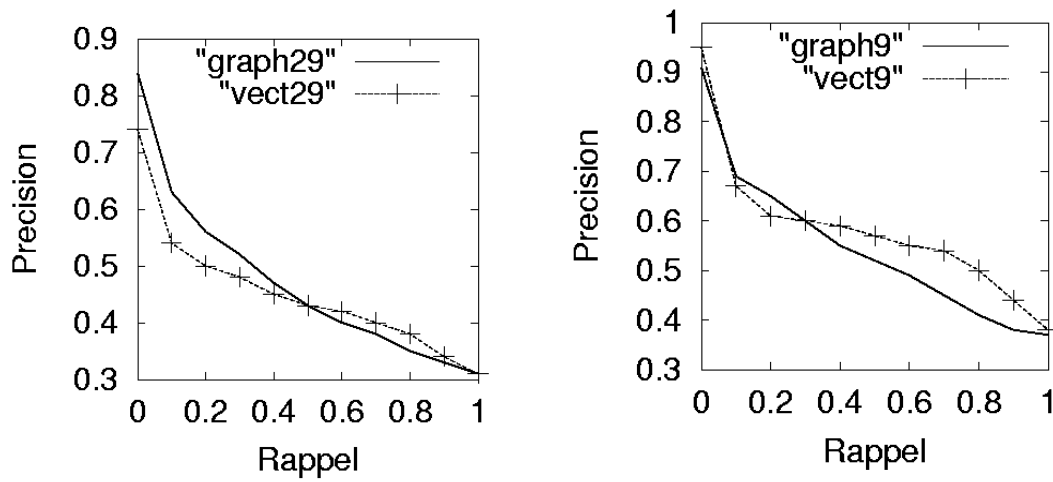


FIG. 7.6 – Courbes de rappel/précision pour notre système et le système basé sur la projection de graphes. A gauche : 29 requêtes sans relation, à droite : 9 requêtes avec relations.

Pour l'évaluation avec les requêtes sans relation (courbe de gauche de la figure 7.6), l'approche vectorielle obtient de moins bons résultats pour des petites valeurs de rappel (rappel plus petit que 0.5). Par ailleurs, la précision moyenne est de 0.4454 pour le système basé sur la projection, et 0.4257 pour notre système. Les requêtes ne comportant pas de relation dans ce cas, tous les documents pertinents sont en correspondance totale avec les requêtes. Le système basé sur la projection n'est alors pas "pénalisé" par sa rigidité, et se comporte comme un système utilisant des mots clés additionnés d'inférence.

La courbe de droite de la figure 7.6 porte sur les requêtes mettant en jeu des relations explicites entre concepts. Il est pertinent de noter que c'est dans ce cas que notre système obtient les meilleurs résultats, avec une précision moyenne de 0.5477, contre une précision

moyenne de 0.5139 pour la projection. La courbe de notre système est au-dessus de celle du système basé sur la projection pour toutes les valeurs de rappel, exceptées les valeurs entre 0.1 et 0.3. Ce résultat n'a cependant pas la même force que les précédents, en raison du petit nombre de requêtes.

Soulignons que le temps de traitement pour une requête est inférieur à 1 seconde¹ dans cette évaluation pour notre système, alors que le système basé sur la projection traite une requête en 6 secondes en moyenne. Notons également que ce temps de traitement pour notre modèle ne devrait pas augmenter sensiblement avec la taille et la complexité des graphes conceptuels (en utilisant le cosinus), contrairement au système qui implante la projection totale, en raison de la complexité des algorithmes de traitement de ces graphes.

7.2.3 Impact de la stratégie de pondération

Ce troisième volet de l'évaluation a pour objectif d'évaluer l'impact de l'intégration de notre stratégie de pondération dans le système. Pour cela, nous comparons les résultats à une pondération booléenne, et nous évaluons les éléments de la pondération séparément. Nous avons également évalué la différence entre l'*idf* classique (calculé *avant* l'expansion de documents) et notre définition du terme *global*. Le tableau 7.1 présente les résultats de cette évaluation selon la stratégie de pondération utilisée pour les documents et les requêtes, sous la forme de la valeur de précision moyenne et du gain relatif en pourcentage par rapport à la pondération booléenne.

Document	Requête	Précision moyenne	Gain relatif (%-age)
booléen	booléen	0.4850	0.0
<i>local</i>	booléen	0.4893	+ 0.89
<i>local</i> × <i>idf</i>	<i>idf</i>	0.4924	+ 1.53
<i>local</i> × <i>global</i>	<i>global</i>	0.4954	+ 2.14

TAB. 7.1 – Précisions moyennes et gains relatifs pour 3 stratégies de pondération différentes.

Nous constatons que l'intégration de la valeur *local* améliore sensiblement la précision

¹ Sur un processeur Pentium III 733MHz.

moyenne du système. Dans le cadre de cette évaluation, seul le critère *taille* a été intégré au système. Les résultats des tests d'hypothèses présentés au chapitre précédent indiquent que l'intégration des critères *position* et *hétérogénéité* doivent améliorer encore le gain de précision relative, en fournissant un classement plus pertinent des documents retrouvés.

L'intégration de l'*idf* améliore la précision du système, ce qui est un résultat attendu en RI. L'utilisation de la valeur *global* permet également d'obtenir une meilleure précision. Du fait de l'application de l'expansion de documents, les graphes étoiles les plus génériques (i.e. composés de concepts et de relations peu profonds dans les treillis) apparaissent dans davantage de documents, tandis que les graphes étoiles les plus spécifiques (i.e. composés de concepts et de relations placés profonds dans les treillis) apparaissent dans moins de documents. Par conséquent, le terme *global* permet de favoriser les documents décrits par des graphes étoiles spécifiques. Ainsi, le fait que le terme *global* donne une précision moyenne plus élevée que l'*idf* nous permet d'affirmer que les graphes étoiles les plus spécifiques permettent de mieux discriminer les documents de cette collection.

7.3 Conclusion

Nous avons présenté dans ce chapitre les résultats de différentes évaluations d'un système implantant le modèle que nous avons proposé. Nous avons montré que l'intégration des relations dans la représentation des documents et des requêtes par le biais des graphes étoiles permet d'améliorer notablement la précision du système comparativement, à l'utilisation de simples mots clés, du fait d'un plus grand pouvoir expressif. Nous avons montré également que notre approche vectorielle soutient la comparaison avec la projection de graphe dans ce contexte de RI. De plus, notre approche permet de diminuer considérablement le temps de traitement des requêtes. Enfin, nous avons montré qu'une version simplifiée de notre stratégie de pondération permet d'augmenter la qualité des résultats que nous avons obtenus sur cette collection. Ce chapitre, qui apporte une validation de notre approche générale, termine la présentation de la mise en œuvre de notre modèle.

Quatrième partie
Conclusion générale

Chapitre 8

Conclusion

8.1 Synthèse et contributions

Le travail présenté dans le cadre de cette thèse s'intéresse à la recherche d'information adaptée au média image, l'objectif étant de permettre l'indexation de documents images et la recherche de ces documents par des utilisateurs. De manière similaire aux travaux de recherche d'information sur des documents textuels, définir un modèle de recherche d'information pour les images nécessite la mise en place d'un certain nombre d'éléments : modèle de documents et de requêtes, modèle de correspondance partielle, formalisation des connaissances entrant en jeu. Un élément fondamental d'un modèle de recherche d'information est lié à la pondération utilisée dans le modèle de documents et de requêtes.

Le cadre de ce travail se focalise sur des photographies, sur des utilisateurs non experts, et sur l'utilisation de descriptions symboliques du contenu des images et des requêtes. Au cours de l'état de l'art que nous avons réalisé, nous avons étudié la modélisation des systèmes de recherche d'information et nous avons détaillé le modèle vectoriel classique de recherche d'information sur des textes. Notre travail portant sur les images, nous avons également exposé des travaux existants sur la recherche d'images, en mettant en relief la difficulté d'appliquer aux images les techniques classiques de recherche d'information. Cette difficulté provient du fossé sémantique, qui sépare les pixels constituant une image de l'interprétation du sens associé à ces pixels. Nous avons en particulier distingué les approches

existantes en RIm selon qu'elles considèrent ou non le sens associé aux pixels. Dans notre cadre de travail sur la recherche de photographies personnelles, nous avons constaté que les approches qui utilisent de la sémantique associée aux pixels sont bien adaptées. Plus particulièrement, les travaux existants sur la recherche d'images ayant pour support le formalisme des graphes conceptuel a eu notre préférence, grâce à la possibilité de prendre en compte des relations à l'intérieur des index de ces images.

Nous proposons une extension du modèle vectoriel de recherche d'information pour les images. Il comprend un modèle de représentation de documents basé sur un modèle d'image physique. Notre définition repose sur la notion d'objet image, un objet image étant une abstraction d'objets physiques du monde réel apparaissant dans les images. Les documents – et les requêtes – sont représentés par des vecteurs d'un espace dont les dimensions sont des graphes étoiles. Un graphe étoile est un graphe conceptuel élémentaire constitué d'une seule relation à laquelle des concepts (associés aux objets images) sont connectés. Cela nous permet de représenter les relations entre les objets images. Le modèle proposé comprend également un modèle de connaissances, constitué de treillis de types de concepts et de relations. Ce modèle de connaissances est utilisé lors de l'opération d'expansion de documents que nous avons définie pour le processus d'indexation. Enfin, le modèle comprend une fonction de correspondance basée sur l'utilisation classique du cosinus.

Cette utilisation de graphes conceptuels élémentaires comme descripteurs d'images dans le cadre d'un modèle classique de recherche d'information donne à nos travaux un aspect novateur. Le cœur de notre proposition réside en effet dans l'idée de fusionner le modèle vectoriel classique de RI avec le formalisme de représentation de connaissances des graphes conceptuels pour la définition d'un modèle de RIm.

L'approche que nous définissons fait des graphes étoiles un excellent compromis entre le manque d'expressivité des mots clés couramment rencontrés en RI textuelle et la complexité algorithmique des traitements associés aux graphes conceptuels. D'autre part, l'utilisation du modèle vectoriel nous permet de traiter le problème de la correspondance binaire liée à

l'utilisation de l'opération de projection de graphes.

Comme nous l'avons décrit plus haut, la pondération des index de documents est un point clé en RI. Un résultat important de ce travail de thèse est la définition d'un modèle de pondération pour les graphes étoiles. La pondération des graphes étoiles est fondée sur la notion d'**importance** des concepts qui les constituent, cette notion étant elle-même reliée aux caractéristiques géométriques et de position des objets images. Nous avons émis quatre hypothèses relatives à cette notion d'importance (concernant la taille, la position et l'éclatement des objets images, ainsi que l'hétérogénéité des images), et proposé une fonction de pondération des objets images fondée sur des critères correspondant à ces hypothèses. Nous avons apporté un grand soin à la validation expérimentale du modèle de pondération proposé. L'expérimentation réalisée auprès de 30 participants a établi la validité des critères de taille, position et hétérogénéité, et a invalidé le critère d'éclatement.

L'extension du modèle vectoriel de RI a conduit à la réalisation d'un système opérationnel de recherche d'images. Nous avons analysé l'impact de la représentation relationnelle proposée, comparativement à une approche basée sur des mots clés. Nous avons également montré que la correspondance de vecteurs étendus de graphes étoiles donne des résultats comparables (en précision relative) à la correspondance de graphes par le biais de la projection, tout en réduisant le temps de traitement des requêtes. Nous avons également montré que l'intégration partielle de notre stratégie de pondération dans le système permet d'améliorer la qualité des résultats.

8.2 Perspectives

Ce travail, comme tout travail de thèse, pose de nombreuses questions et donne lieu à de nombreuses perspectives, parmi lesquelles :

- A court terme, il est important d'évaluer l'intégration complète de nos travaux, en utilisant les graphes étoiles et notre modèle de pondération. Une collection de plusieurs milliers de photographies personnelles entièrement annotées est en cours de

réalisation dans l'équipe et nous allons l'utiliser pour cette évaluation.

- Au niveau de la modélisation par l'intermédiaire d'un espace vectoriel, notre principe d'expansion des données est susceptible de générer de nombreuses dimensions non pertinentes, et de produire des effets de bords dus à la grande dépendance entre les dimensions de l'espace. Il convient dès lors de tenter d'estimer théoriquement et expérimentalement l'impact de ces points afin de répondre à ces questions. Nous pourrions par exemple nous poser la question de l'emploi de décomposition en valeurs singulières (SVD) de l'espace vectoriel pour réduire l'espace "utile" des données, comme cela a été fait pour le modèle vectoriel sur des documents textuels.
- Pour les aspects de pondération des objets images, nos propositions sont prometteuses, mais nous pouvons nous poser la question de la complétude des critères choisis, ainsi que des dépendances entre ces critères. Par exemple, la taille d'un objet n'est pas totalement indépendante de sa position : un gros objet a forcément des parties proches du centre de l'image. L'explicitation de ces dépendances de manière formelle pourrait amener à combiner ces critères en amont, c'est-à-dire avant d'évaluer la pondération définitive. Nous allons nous intéresser à ces aspects à l'avenir à un niveau expérimental en nous penchant de nouveau sur les évaluations de la pondération, de manière à avoir des points de départ nous permettant d'intuiter des points clés de la pondération.
- Idéalement, le travail proposé dans cette thèse a pour but d'être intégré dans un système totalement automatique d'indexation et de recherche symbolique d'images photographiques. Cette intégration passe par la mise en commun des travaux effectués dans l'équipe, et sera source de futures investigations dans le domaine de la Recherche d'Images.

Cinquième partie
Annexes et bibliographie

Annexe A

Autres modèles de recherche d'information

Cette annexe décrit les modèles booléen, probabiliste et logique de RI.

Trois modèles de RI textuelle sont dits “classiques” [vR79]. Il s’agit du modèle **booléen**, du modèle **vectoriel** [Sal71], et du modèle **probabiliste** [RJ76]. Un autre modèle, le modèle **logique** [vR86a], est fondé sur une interprétation logique du contenu des documents.

Ces modèles ont en commun le vocabulaire d’indexation. Seule la manière de comparer les documents à la requête change. Ces modèles se basent sur le formalisme des mots clés. Le vocabulaire d’indexation $V = \{t_i\}, i \in \{1, \dots, n\}$ est constitué de n mots ou racines de mots qui apparaissent dans les documents.

A.1 Modèle booléen

Le modèle booléen est basé sur la théorie des ensembles. Dans ce modèle, les documents et les requêtes sont représentés par des ensembles de mots clés. Ce modèle doit son nom à l’utilisation des opérateurs *et*, *ou* et *non* pour la représentation des documents et des requêtes. Chaque document d_j est représenté par un ensemble de termes, et c’est la conjonction de ces termes qui constitue l’index des documents. Une représentation en logique des prédicats est :

$$d_j = t_1 \wedge t_2 \wedge \dots \wedge t_k \wedge \neg t_{k+1} \wedge \dots \wedge \neg t_n$$

ou bien

$$d_j = t_1 \wedge t_2 \wedge \dots \wedge t_k$$

Ce modèle est donc basé sur l'hypothèse du monde fermé : la présence (resp. l'absence) d'une proposition atomique dans la modélisation du document indique que le terme en question représente (resp. ne représente pas) le document. La requête est une expression booléenne dont les termes sont reliés par des opérateurs de conjonction, de disjonction ou de négation, qui peut se ramener à une forme normale disjonctive (ou DNF, "Disjonctive Normal Form") :

$$q = t_1 \wedge t_2 \vee (\neg t_3 \wedge t_4)$$

La fonction de correspondance vérifie si l'index de chaque document d_j implique l'expression logique de la requête q . Les documents retrouvés sont ceux dont l'index vérifie $d_j \Rightarrow q$, \Rightarrow étant l'implication de la logique propositionnelle. Le résultat de cette fonction est donc binaire.

Ce modèle a l'avantage d'être très simple à mettre en œuvre, et de posséder une fonction de correspondance de calcul rapide. C'est ce qui lui a valu sa popularité et son utilisation dans de nombreux systèmes, notamment pour les moteurs de recherche sur Internet. Cependant, ce modèle possède plusieurs inconvénients :

- l'indexation des documents. Il n'y a pas de possibilité d'exprimer l'importance plus forte ou moins forte d'un terme dans un document.
- la fonction de correspondance est également binaire, c'est-à-dire qu'un document est soit pertinent, soit non pertinent pour une requête. Cela a deux conséquences sur les résultats. D'une part le corpus est partitionné en un sous-ensemble des documents pertinents, et un sous-ensemble des documents non pertinents : l'ordonnancement des réponses n'est pas possible. D'autre part, un document ne satisfaisant que partiellement la requête n'est pas retrouvé.
- la formulation d'une requête avec des opérations booléennes est difficile et parfois trompeuse pour l'utilisateur, car la sémantique des opérations booléennes n'est pas

similaire à celle de la langue naturelle. Par exemple, le *ou* dans la langue naturelle a un sens exclusif alors que ce n'est pas le cas dans le langage logique. De plus, il n'est pas intuitif de tenir compte des parenthèses, ni des priorités des opérateurs dans la formulation de la requête.

Le principal inconvénient du modèle booléen est son indexation binaire. Des extensions, par exemple l'utilisation de la *p-norme* dans le cadre du modèle booléen étendu [SFW83], ont été proposées pour prendre en compte ces faiblesses, et améliorer ainsi les performances du modèle. Par exemple, une manière de rendre l'indexation plus souple est d'associer un poids à chaque terme d'indexation pour dénoter son importance dans le document. C'est un modèle fréquemment utilisé sur le Web.

A.2 Modèle probabiliste

Ce modèle est fondé sur le calcul de la probabilité de pertinence d'un document pour une requête. On se rapproche ici de la notion de classification probabiliste. L'idée est de retrouver des documents qui ont en même temps une forte probabilité d'être pertinents, et une faible probabilité d'être non pertinents.

Les documents – et les requêtes – sont représentés par des vecteurs de booléens dans un espace à n dimensions.

$$d_i = (w_{1,j}, w_{2,j}, \dots, w_{n,j})$$

$$q = (w_{1,q}, w_{2,q}, \dots, w_{n,q})$$

avec $w_{k,j} \in \{0, 1\}$ et $w_{k,q} \in \{0, 1\}$. La valeur de $w_{k,j}$ (resp. $w_{k,q}$) représente le fait que le terme t_k apparaît ou non dans le document d_j (resp. la requête q).

La fonction de correspondance évalue la pertinence du document d_j pour la requête q :

$$\text{correspondance}(d_j, q) = \frac{P(\text{rel} | d_j)}{P(\text{nrel} | d_j)} \quad (\text{A.1})$$

avec $P(rel|d_j)$ (resp. $P(nrel|d_j)$) la probabilité que la réponse soit pertinente (resp. non pertinente) pour la requête q , étant donné le document d_j comme réponse. En utilisant le théorème de Bayes, van Rijsbergen [vR79] estime cette probabilité :

$$correspondance(d_j, q) = \frac{P_r(d_j) \cdot p(rel)}{P_r(d_j) \cdot p(rel) + P_n(d_j) \cdot p(nrel)}$$

dans laquelle $P_r(d_j)$ et $P_n(d_j)$ représentent respectivement la probabilité pour d_j d'être un document pertinent ou non pertinent pour la requête q , $p(rel)$ et $p(nrel)$ sont respectivement les probabilités de pertinence et de non pertinence d'un document quelconque du corpus, lesquelles sont supposées fixées pour un corpus donné. Cela nous ramène à l'estimation de $P_r(d_j)$ et $P_n(d_j)$. Si on considère que les termes sont indépendants :

$$p_r(d_j) = \prod_{t_i \in d_j} P_r(t_i)$$

$$p_n(d_j) = \prod_{t_i \in d_j} P_n(t_i)$$

avec $P_r(t_i)$ (resp. $P_n(t_i)$) la probabilité d'apparition du terme t_i dans un document pertinent (resp. non pertinent). On peut calculer ces probabilités par échantillonnage, en choisissant deux ensembles de documents considérés comme étant pertinents et non pertinents à partir du nombre d'occurrences des termes dans ces ensembles.

Le modèle probabiliste se démarque des deux autres modèles classiques car il utilise l'outil des probabilités pour représenter explicitement le concept fondamental de la RI qu'est la pertinence. Il est difficile de mettre en oeuvre un tel modèle, en raison du problème de calcul des probabilités initiales. La méthode décrite ici consiste à se baser sur un échantillon de documents pour lesquels on connaît la pertinence, et à répercuter les probabilités ainsi obtenues au reste du corpus. En pratique, la plupart des expérimentations se basent sur des méthodes *tf * idf* (voir la description du modèle vectoriel à la section 2.4) [JWR98], en supposant que la répartition des termes est décrite par une distribution de probabilité.

A.3 Modèle logique

En 1986, van Rijsbergen [vR86a, vR86b] a modélisé la pertinence d'un document par rapport à une requête au moyen d'un principe d'incertitude logique. Soient deux formules logiques d et q (les représentations du document et de la requête), une fonction de comparaison entre d et q mesure l'incertitude qui existe dans l'implication $d \Rightarrow q$ relative à un ensemble de données K . Cette fonction détermine l'extension minimale à apporter à K pour établir la preuve de $d \Rightarrow q$.

Un document est considéré comme un ensemble de "phrases" interprétées dans une certaine sémantique définie. Il en est de même pour la requête, qui est généralement formée d'une seule phrase.

Ce principe, dans sa forme la plus générale, exprime la mesure de la correspondance entre le document et la requête comme suit : si une ou plusieurs phrases impliquent la requête, cette dernière est considérée comme satisfaite. Les données K correspondent à l'ensemble de la connaissance dont on dispose au moment de l'évaluation de l'implication.

Si l'implication ne peut pas être satisfaite par rapport à cet ensemble d'informations, il est nécessaire d'ajouter de nouvelles informations de façon à ce que l'implication devienne satisfaite par rapport à l'ensemble d'informations étendu. L'extension minimale signifie qu'au cours de la mesure d'incertitude, seules les informations nécessaires à la satisfaction de $d \Rightarrow q$ sont ajoutées.

Ce modèle est théorique, et pour le mettre en œuvre il faut en définir les éléments clés : les formules logiques d et q , l'implication \Rightarrow , et l'ensemble de données K . Dans le modèle de base proposé par van Rijsbergen, l'implication représentée par \Rightarrow peut être toute implication définie et elle ne se limite pas à l'implication logique $d \Rightarrow q$. On peut voir le modèle logique comme un méta-modèle de recherche d'informations, dont les instances sont opérationnelles. Par exemple, le formalisme des graphes conceptuels [Sow84] a été proposé dans [Che92, Mec95, Oun98] comme formalisme opérationnel pour le modèle logique.

A.4 Discussion sur ces modèles

Le formalisme des mots clés a l'avantage d'être simple et facile à mettre en oeuvre. Il pose cependant un problème d'ambiguïté des termes d'indexation. En effet la polysémie et la synonymie des mots clés limitent la précision de l'indexation. La polysémie est la propriété d'un mot qui a plusieurs sens. Le mot "pompe", par exemple, est un terme polysémique. En effet, il peut désigner un appareil (pompe à vélo) comme il peut exprimer le mot chaussure en argot. La synonymie désigne le fait que plusieurs mots peuvent avoir le même sens. Par exemple, "bicyclette" et "vélo" sont des synonymes.

L'indexation sémantique latente (ou LSI, "Latent Semantic Indexing") est une extension du modèle vectoriel. C'est une méthode d'indexation automatique dont l'objectif est de représenter le contenu sémantique (ou conceptuel) des documents. Le LSI est fondé sur une décomposition en valeurs propres pour générer la matrice d'association terme-document, et construire un espace "sémantique" dans lequel les termes et les documents proches sont placés les uns à côté des autres. Le but est de palier les problèmes de synonymie et de polysémie, pour améliorer le rappel et la précision de la recherche. Des termes qui n'apparaissent pas dans un document peuvent quand même être proches du document, si cela est consistant avec les schémas principaux d'association dans les données. L'interrogation s'effectue en utilisant une requête pour identifier un point dans l'espace sémantique, afin de retourner les documents dans le voisinage de ce point.

Annexe B

Le formalisme des graphes conceptuels

Cette annexe décrit les éléments du formalisme des graphes conceptuels qui servent de base à nos travaux. Elle contient donc une description non exhaustive du formalisme, qui a été introduit dans les années 80 par Sowa [Sow84] comme un modèle de représentation de connaissances. Il a été ensuite largement utilisé et étendu, notamment par Chein et Mugnier [CM92, MC96] et Amati et Ounis [AO00] qui ont effectué un travail de formalisation ultérieur à celui de Sowa. Le formalisme a été étendu par Morton [Mor87], Maher [Mah91] et Wuwongse [WM93] pour prendre en compte l'incertitude par le biais des ensembles flous. Il a été utilisé en RI par Chevallet [Che92], qui s'est basé sur ce formalisme pour la recherche de composants logiciels, et par Genest [Gen00] qui a étendu le formalisme pour la RI. Il a été utilisé plus spécifiquement en RIm par Mechkour [Mec95, MMFB97] qui a fondé l'implantation des modèles PRIME et EMIR² sur ce formalisme, Ounis et Pasca [OP98] et Mulhem et Lim [ML02] qui ont mené des travaux dans la représentation et la recherche d'images basés sur le formalisme.

Les graphes conceptuels sont des graphes orientés, connexes, et bipartis, composés de deux sortes de nœuds : les **concepts** et les **relations conceptuelles**. Les concepts représentent les entités, attributs, états et événements, et les relations conceptuelles (ou relations) dénotent les interconnexions entre les concepts. Les graphes conceptuels sont définis par rapport à un **support**. Avant de définir formellement un graphe conceptuel, nous présentons cette notion de support, qui constitue le vocabulaire de base du formalisme.

B.1 Support

Les graphes conceptuels sont définis par rapport à un support.

Définition 15 (Support) *Un support est un quintuplet :*

$$S = (T_C, T_R, \sigma, M, \tau)$$

dans lequel :

- T_C est l'ensemble des types de concepts, qui correspondent aux classes sémantiques.
- T_R est l'ensemble des types de relations, disjoint de T_C .
- σ est l'application qui associe une signature à toute relation de T_R .
- M est l'ensemble des référents (ou marqueurs individuels), qui représentent les individus considérés dans la base de connaissance.
- τ est un prédicat de conformité sur $M \times T_C$.

Ces cinq éléments sont détaillés ci-après.

B.1.1 Les concepts et les relations

L'ensemble T_C des concepts est partiellement ordonné par une relation *EstUn*, notée \leq , qui induit un treillis sur T_C . Son plus grand élément est le type de concept universel \top_C et son plus petit élément est le type de concept absurde \perp_C , de telle sorte que $\forall t_C \in T_C, \top_C \leq t_C \leq \perp_C$.

Exemple 14 *Le type de concept [Fleur] est inférieur au type de concept [Végétation], selon le treillis de type de concepts de la figure B.1 (à gauche).*

L'ensemble T_R est également partiellement ordonné. Il est partitionné en sous-ensembles de relations de même arité¹ : $T_R = \bigcup_{i \geq 1} T_{R_i}$, où T_{R_i} est l'ensemble des relations d'arité i . Le plus grand élément de T_R est le type de relation universel \top_R , et son plus petit élément est le type de relation absurde \perp_R , de telle sorte que $\forall t_R \in T_R, \top_R \leq t_R \leq \perp_R$.

¹L'arité (ou la valence, ou encore le degré) d'une relation est le nombre de concepts connectés à cette relation.

Exemple 15 Le type de relation [rame] est inférieur au type de relation [action₁], selon le treillis de type de relations de la figure B.1 (à droite).

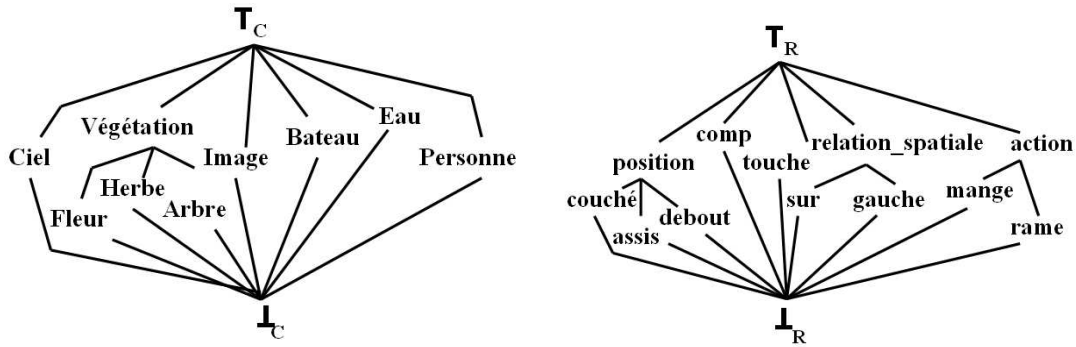


FIG. B.1 – Les concepts et les relations sont organisés en treillis de types de concepts (à gauche) et treillis de relations (à droite).

B.1.2 Les signatures et les graphes étoiles

La signature $\sigma(r)$ de $r \in T_{R_k}$ est le k -uplet (tc_1, \dots, tc_k) de $(T_C)^k$ qui indique le type maximal des concepts associés à r , et vérifie la contrainte :

$$\forall r_1, r_2 \in T_{R_k} : r_1 \leq r_2 \Rightarrow \sigma(r_1) \leq \sigma(r_2)$$

où l'ordre sur les signatures est l'ordre produit sur $(T_C)^k$. En notant $\sigma_i(r)$ le $i^{\text{ème}}$ concept de $\sigma(r)$, on écrit $type(\sigma_i(r_1)) \leq type(\sigma_i(r_2))$.

Une signature $\sigma(r) = (tc_1, \dots, tc_k)$ est représentée par un graphe conceptuel appelé **graphe étoile**. Un graphe étoile est un graphe conceptuel élémentaire comprenant une et une seule relation r , et n_r concepts c_i attachés à cette relation ($n_r \geq 1$ est l'arité de r) :

$$\langle r(c_1, c_2, \dots, c_{n_r}) \rangle$$

L'ensemble des graphes étoiles associés à l'ensemble des types de relations T_R d'un support est appelé la **base canonique** de ce support. La base canonique contient un sous-ensemble de graphes conceptuels en bijection avec les relations de T_R . Elle est unique lorsqu'elle est minimale, c'est-à-dire qu'aucun graphe de la base canonique n'est exprimable à l'aide d'une combinaison d'autres graphes de la base.

Les graphes conceptuels dérivés des graphes de la base canonique par une ou plusieurs des quatre opérations canoniques sont appelés **graphes canoniques**. Ainsi la base canonique limite l'expression des graphes conceptuels aux seuls graphes *bien formés*, c'est-à-dire qui ont un sens. La fermeture de la base canonique avec les quatre opérations canoniques produit tous les graphes conceptuels bien formés du support. La figure B.2 montre la situation des ensemble des graphes conceptuels.

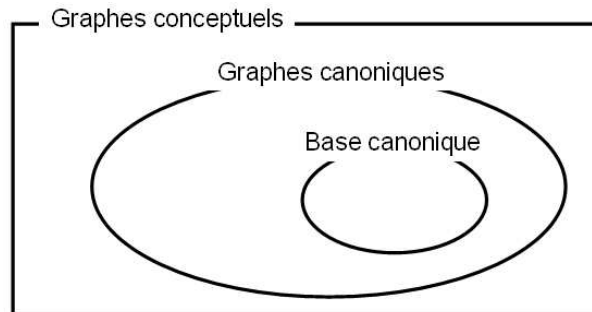


FIG. B.2 – Situation de la base canonique et des graphes canoniques et conceptuels.

L'ensemble des graphes étoiles est un *ordre* muni de la relation \leq . Les relations étant structurées dans un treillis, il en est de même pour les graphes étoiles.

B.1.3 Les référents

L'ensemble M contient notamment le référent universel $*$ qui représente un individu non spécifié et le référent absurde \emptyset . M est partiellement ordonné par \sqsubseteq de la manière suivante :

tous les référents de $M - \{*, \emptyset\}$ sont incomparables deux à deux, et nous avons :

$$\forall m \in M : * \sqsubseteq m \sqsubseteq \emptyset$$

On appelle **concept individuel** un concept dont le référent appartient à $M - \{*, \emptyset\}$, et **concept générique** un concept dont le référent est le référent universel $*$ (ou référent générique).

Le prédicat τ satisfait les contraintes suivantes : $\forall m \in M - \{*, \emptyset\}$, et $\forall t, t' \in T_C$:

- $\tau(m, \top_C)$ et $\neg\tau(m, \perp_C)$
- $\neg\tau(\emptyset, t)$
- $\tau(*, t)$ si $t \neq \perp_C$, $\neg\tau(*, \perp_C)$
- $(t' \leq t) \wedge \tau(m, t') \Rightarrow \tau(m, t)$

Le support contient les contraintes syntaxiques et fournit des informations relatives au domaine de l'application, par rapport auxquelles les graphes conceptuels sont définis.

B.2 Graphes conceptuels

Définition 16 *Un graphe conceptuel est un S -graphe, c'est-à-dire un graphe défini sur un support S . Un S -graphe est noté $G = (C_G, R_G, A_G, l_G)$, où :*

- C_G est l'ensemble des sommets concepts du graphe et R_G est l'ensemble des sommets relations, avec $C_G \cap R_G = \emptyset$ et $C_G \neq \emptyset$. Appelons N_G l'ensemble des sommets $C_G \cup R_G$.
- A_G est l'ensemble des arêtes. Le graphe étant biparti, les arêtes ont une extrémité dans C_G et une extrémité dans R_G .
- l_G est une application qui associe une étiquette à tout nœud du graphe. Si $c \in C_G$, alors $l(c) \in T_C \times M - \{\emptyset\}$ est le couple (*type(c), référent(c)*) tel que $\tau(\text{référent}(c), \text{type}(c))$. Il existe une relation d'ordre partiel sur les étiquettes de C_G , qui est l'ordre produit sur $T_C \times M - \{\emptyset\}$. Si $r \in R_G$, alors $l(r) \in T_R$ est le type du sommet relation. Par ailleurs, l_G satisfait les contraintes de σ et τ .

Le formalisme des graphes conceptuels est lié à logique du premier ordre, par le biais de l'opérateur Φ [Sow84], qui associe une formule logique à un graphe conceptuel (ou à un support). Une formule logique $\Phi(G)$ est associée à un graphe G , de telle sorte que :

- à chaque concept est associé un prédicat unaire de même nom que le type du concept, et le paramètre est soit une constante correspondant au référent du concept s'il est défini, soit une variable quantifiée existentiellement si le référent est le référent existentiel.
- à chaque relation de degré n est associé un prédicat n -aire de même nom que le type de la relation, et dont les paramètres sont les paramètres (variables ou constantes) des concepts liés par la relation.
- la formule $\Phi(G)$ associée à G est la conjonction des prédicats associés aux concepts et relations de G .

Il est possible de manipuler les éléments du formalisme de manière graphique ou bien par le biais des formules logiques associées.

B.3 Opérations canoniques sur les graphes conceptuels

Des opérations de construction permettent de dériver des graphes à partir d'autres graphes. Ces opérations sont la copie, la restriction, la simplification et la jointure.

- la **copie**, qui duplique un graphe conceptuel.
- la **restriction**, qui consiste à *spécialiser* dans le graphe conceptuel original soit une des relations en la remplaçant par une relation de type plus petit, soit un des concepts, par l'une des deux transformations suivantes :
 - Attribution d'un référent (conforme à son type) à un concept générique.
 - Remplacement du type t d'un concept par un type $t' \leq t$, à la condition que le référent du concept soit conforme au nouveau type.
- la **simplification**, qui consiste à éliminer les éventuelles relations dupliquées.
- la **jointure** de deux graphes conceptuels contenant un concept identique en commun,

qui est obtenue en éliminant ce concept d'un des deux graphes, et en raccrochant tous les arcs restant pendants à l'autre graphe.

B.4 Projection de graphes

Un graphe conceptuel H est dit **inférieur** à un autre graphe conceptuel G (noté $H \leq G$) si H résulte d'au moins une opération de copie, restriction, jointure, ou simplification appliquée à G . Dans ce cas, H étant **canoniquement dérivable** de G , H est une spécialisation de G , et G est une généralisation de H (ou G subsume H). Cette relation de spécialisation/généralisation sur les graphes définit une relation d'ordre partiel sur les graphes.

L'opération de projection est l'opération centrale du formalisme qui permet d'identifier dans un graphe des sous-graphes de même structure, elle correspond à un morphisme de graphe. Il existe une projection d'un graphe G dans un graphe H si et seulement si $H \leq G$ [Sow84, CM92]. L'existence d'une projection de G dans H est équivalente [Sow84, CM92] à :

$$H \Rightarrow G$$

où \Rightarrow désigne l'implication logique.

Formellement, une projection d'un graphe $G = (C_G, R_G, A_G, l_G)$ dans un graphe $H = (C_H, R_H, A_H, l_H)$ est une application $\Pi : N_G \rightarrow N_H$, telle que :

1. les arêtes et les étiquettes des arêtes de G sont conservées dans H .
2. les étiquettes des sommets de G sont éventuellement spécialisées dans H .

Autrement dit, la projection plaque un graphe sur un autre tout en vérifiant la spécialisation des étiquettes sur les concepts et les relations, et en préservant l'ordre des arcs.

L'opération de projection est utilisée en recherche d'information pour identifier les documents pertinents pour une requête, en sélectionnant les graphes de documents dans lesquels le graphe de la requête se projette.

Bibliographie

- [AJ98] H. Bélaïd Ajroud and A. Jaoua. Abstraction of objects by conceptual clustering. *Journal of Information Sciences*, 1(109) :79–94, 1998.
- [AO00] G. Amati and I. Ounis. Conceptual graphs and first order logic. *The Computer Journal*, 43(1) :1–12, 2000.
- [ATY⁺97] Y. Alp Aslandogan, C. Thier, C.T. Yu, J. Zou, and N. Rishe. Using semantic contents and wordnet in image retrieval. In *Proceedings of SIGIR-97, 20th ACM International Conference on Research and Development in Information Retrieval*, pages 286–295, Philadelphia, USA, 1997.
- [Bai98] G. Baillargeon. *Probabilités, statistique et techniques de régression*. SMG, Trois-Rivières, Québec, 1998.
- [BFM⁺97] C. Berrut, F. Fourel, M. Mechkour, P. Mulhem, and Y. Chiaramella. *Indexing, navigation and retrieval of multimedia structured documents : the PRIME information retrieval system*. Chapter of Deliverable D11, FERMI ESPRIT BRA 8134, 1997.
- [BKB00] A. Bouch, A. Kuchinsky, and N. Bhatti. Quality is in the eye of the beholder : Meeting users’ requirements for internet quality of service. In *Conference on Human Factors in Computing Systems - CHI’00*, pages 297–304, 2000.
- [BMC03] S. Bissol, P. Mulhem, and Y. Chiaramella. Dynamic learning of indexing concepts for home image retrieval. In *Proceedings of CBMI’03 - Third Inter-*

national Workshop on Content-Based Multimedia Indexing, pages 335–341, 2003.

- [BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. In *In journal of Computer Networks and ISDN Systems (30)*, pages 107–117, Brisbane, 1998.
- [BYRN99] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press Series/Addison-Wesley, 1999.
- [Che92] J.P. Chevallet. *Un Modèle Logique de Recherche d'Informations applique au formalisme des Graphes Conceptuels*. Thèse de doctorat, Univerité Joseph Fourier, Grenoble, 1992.
- [Chi97] Y. Chiaramella. Browsing and querying : Two complementary approaches for multimedia information retrieval. In *Hypertext – Information Retrieval – Multimedia (HIM)*, pages 9–26, 1997.
- [CM92] M. Chein and M.-L. Mugnier. Conceptual graphs : Fundamental notions. *Revue d'intelligence artificielle*, 6(4) :365–406, 1992.
- [CM97] Y. Chiaramella and M. Mechkour. *Indexing an image test collection*. Technical Report, FERMI BRA 8134, 1997.
- [CMF96] Y. Chiaramella, P. Mulhem, and F. Fourel. A model for multimedia information retrieval. In *Technical report, FERMI ESPRIT BRA 8134, University of Glasgow*, 1996.
- [CMK66] C.W. Cleverdon, J. Mills, and M. Keen. *Factors Determining the Performance of Indexing Systems*. ASLIB Cranfield Research Project, 1966.
- [CMM⁺00] I.J. Cox, M.L. Miller, T.P. Minka, T.V. Papathornas, and P.N. Yianilos. The bayesian image retrieval system, pichunter : Theory, implementation, and psychophysical experiments. *IEEE Transactions on Image Processing*, 9(1), 2000.

- [CPW98] C.H. Chen, L.F. Pau, and P.S.P. Wang. *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*. World Scientific Publishing, 1998.
- [CTB⁺99] Chad Carson, Megan Thomas, Serge Belongie, Joseph M. Hellerstein, and Jitendra Malik. Blobworld : A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*. Springer, 1999.
- [DM04] E. Debanne and P. Mulhem. Query by example for symbolic still image retrieval. In *Proceedings of CORIA-2004, Conference en Recherche Information et Applications*, pages 363–376, Toulouse, France, 2004.
- [FSN⁺95] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, and B. Dom et al. Query by image and video content : the QBIC system. *IEEE Computer*, 28(9) :23–32, 1995.
- [Gen00] D. Genest. *Extension du modèle des graphes conceptuels pour la recherche d'information*. Thèse de Doctorat, Université Montpellier II, 2000.
- [Gui96] O. Guinaldo. Conceptual graphs isomorphism : Algorithm and use. In *ICCS'96 Conference*, volume 1115, pages 160–174, 1996.
- [GWJ91] A. Gupta, T.E. Weymouth, and R. Jain. Semantic queries with pictures : The vimsys model. In Guy M. Lohman, Amílcar Sernadas, and Rafael Camps, editors, *17th International Conference on Very Large Data Bases, September 3-6, 1991, Barcelona, Catalonia, Spain, Proceedings*, pages 69–79. Morgan Kaufmann, 1991.
- [HMR96] Thomas S. Huang, Sharad Mehrotra, and Kannan Ramchandran. Multimedia analysis and retrieval system (mars) project. *Proceedings of the 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval, University of Illinois at Urbana-Champaign*, 1996.

- [JS88] S.F. Weiss J.B. Smith. Hypertext. *Communications of the ACM*, 31(7) :816–819, 1988.
- [JWR98] K. Sparck Jones, S. Walker, and S.E. Robertson. *A probabilistic model of information retrieval : development and comparative experiments*. Technical Report TR446, University of Cambridge Computer Laboratory, 1998.
- [KB96] J. Koenemann and N.J. Belkin. A case for interaction : A study of interactive information retrieval behavior and effectiveness. In *Conference on Human Factors in Computing Systems - CHI'96*, pages 205–212, 1996.
- [Khe95] A. Kheirbek. *Modèle d'intégration d'un système de recherche d'informations et d'un système hypermédia basé sur le formalisme des graphes conceptuels. Application au système RIME*. Thèse de Doctorat, Université Joseph Fourier, 1995.
- [Lim00] J.-H. Lim. Photograph retrieval and classification by visual keywords and thesaurus. *New Generation Computing*, 18 :147–156, 2000.
- [Lim01] J.-H. Lim. Building visual vocabulary for image indexation and query formulation. *Pattern Analysis and Applications (Special Issue on Image Indexation)*, 4(2/3) :125–139, 2001.
- [Lov68] J. Lovins. Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics*, 11 :22–31, 1968.
- [Luh57] H.P. Luhn. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, 1(4) :309–317, 1957.
- [LV03] P. Lyman and HR. Varian. *How much information ?* Projet du School of Information Management and Systems, University of California at Berkeley. URL : <http://www.sims.berkeley.edu/research/projects/how-much-info/>, 2003.

- [Mah91] P. E. Maher. Conceptual graphs - a framework for uncertainty management. In *NAFIPS'91*, pages 106–110, 1991.
- [MB86] P. Martin and P. Bateson. *Measuring Behaviour : An Introductory Guide*. Cambridge University Press, 1986.
- [MBC95] M. Mechkour, C. Berrut, and Y. Chiamarella. Using conceptual graph framework for image retrieval. In *International conference on MultiMedia Modeling - MMM'95*, pages 127–142, Singapore, 1995.
- [MC93] M.-L. Mugnier and M. Chein. Characterization and algorithmic recognition of canonical conceptual graphs. In *Proceedings of International Conference on Conceptual Structures - ICCS'93, Lecture Notes in AI 699*, pages 294–311, 1993.
- [MC96] M.-L. Mugnier and M. Chein. Représenter des connaissances et raisonner avec des graphes. *Revue d'intelligence artificielle*, 10(1) :7–56, 1996.
- [MCM02] J. Martinet, Y. Chiamarella, and P. Mulhem. Un modèle vectoriel étendu de recherche d'information adapté aux images. In *INFORSID'02*, pages 337–348, Nantes, 2002.
- [MCMO03] J. Martinet, Y. Chiamarella, P. Mulhem, and I. Ounis. Photograph indexing and retrieval using star-graphs. In *In Proceedings of CBMI'03 - Third International Workshop on Content-Based Multimedia Indexing*, pages 335–341, Rennes, 2003.
- [Mec95] M. Mechkour. *Un modèle étendu de représentation et de correspondance d'images pour la recherche d'informations*. Thèse de Doctorat, Université Joseph Fourier, Grenoble, 1995.
- [Mes94] P. Messaris. *Visual Literacy : Image, Mind, and Reality*. Westview Press, 1994.
- [MHH99] S. Mukherjea, K. Hirata, and Y. Hara. Amore : A world wide web image retrieval engine. *The WWW Journal*, 2(3) :115–132, 1999.

- [Mil56] G.A. Miller. The magical number seven, plus or minus two : Some limits on our capacity for processing information. *The Psychological Review*, 63 :81–97, 1956.
- [Mil75] G.A. Miller. *The Psychology of Communication*. Basic Books, New York, second edition, 1975.
- [Mil90] G. Miller. Wordnet : an on-line lexical database. *International Journal of Lexicography*, 4, 1990.
- [ML02] P. Mulhem and J.H. Lim. Symbolic photograph content-based retrieval. In *CIKM'2002 - ACM Conference on Information and Knowledge Management*, pages 94–101, Virginia, USA, 2002.
- [MLG00] M. Montes-y-Gomez, A. Lopez-Lopez, and A.F. Gelbukh. Information retrieval with conceptual graph matching. In *Database and Expert Systems Applications*, pages 312–321, 2000.
- [MLLK03] P. Mulhem, J.H. Lim, W.K. Leow, and M. Kankanhalli. Advances in digital home photo albums. In S. Deb (Ed.), editor, *In Multimedia Systems and Content-Based Image Retrieval*. Idea Group Publishing, 2003.
- [MM97] W.Y. Ma and B.S. Manjunath. Edge flow : a framework of boundary detection and image segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 744–749, 1997.
- [MM99] Wei-Ying Ma and B. S. Manjunath. Netra : A toolbox for navigating large image databases. *Multimedia Systems*, 7(3) :184–198, 1999.
- [MMFB97] M. Mechkour, P. Mulhem, F. Fourel, and C. Berrut. PRIME-GC : A medical information retrieval prototype on the WEB. In *7th International Workshop on Research Issues in Data Engineering (RIDE '97) High Performance Database Management for Large-Scale Applications*, page 2, 1997.

- [MOCM03] J. Martinet, I. Ounis, Y. Chiaramella, and P. Mulhem. A weighting scheme for star-graphs. In *In Proceedings of ECIR'03 - 25th BCS-IRSG European Conference on Information Retrieval Research*, pages 546–554, Pisa, 2003.
- [Mor87] S. Morton. *Conceptual Graphs and Fuzziness in Artificial Intelligence*. PhD thesis, University of Bristol, 1987, 1987.
- [MP04] J. Martinez and P.Valduriez. Vers un passage à l'échelle pour un SGBD d'images. In *Proceedings of CORIA-2004, Conférence en Recherche Information et Applications*, pages –, Toulouse, France, 2004.
- [MT01] P. Mulhem and T. Tan. Impact of labeling processes on symbolic image. In *Symposium on Image Processing, Honolulu, USA*, 2001.
- [MTT99] R. Mandala, T. Tokunaga, and H. Tanaka. Combining multiple evidence from different types of thesaurus for query expansion. In *Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval*, pages 191–197, Berkley, USA, 1999.
- [Nam00] F. Namer. FLEMM : un analyseur flexionnel du français à base de règles. *TAL*, 41(2) :523–549, 2000.
- [NBE⁺93] W. Niblack, R. Barber, W. Equitz, M. Flicker, E. Glasman, D. Petkovic, P Yan-ker, C Faloutsos, , and G. Taubin. The QBIC project : Querying images by content using color, texture, and shape. *Research Report, Report Rj 9203(81511), IBM Almanden Research Center, San Jose, CA.*, 1993.
- [Nie94] J. Nielsen. *Usability Engineering*. Morgan Kaufmann, San Francisco, 1994.
- [OP97] I. Ounis and M. Pasca. The relief retrieval system. In *KDEX'97*, 1997.
- [OP98] I. Ounis and M. Pasca. Relief : Combing expressiveness and rapidity into a single system. In *SIGIR'98*, pages 266–274, 1998.
- [OS95] V. E. Ogle and M. Stonebraker. Chabot : Retrieval from a relational database of images. *IEEE Computer*, 28(9) :40–48, September 1995.

- [Oun98] I. Ounis. *Un modèle d'indexation relationnel pour les graphes conceptuels fondé sur une interprétation logique*. Thèse de doctorat, Université Joseph Fourier, Grenoble, 1998.
- [Oun99] I. Ounis. A flexible weighting scheme for multimedia documents. In *Database and Expert Systems Applications*, pages 392–405, 1999.
- [Por80] M. Porter. An algorithm for suffix stripping. *Program*, 14(3) :130–137, 1980.
- [PPS94] A. Pentland, R. Picard, and S. Sclaroff. Photobook : Content-based manipulation of image databases. In *In SPIE Storage and Retrieval for Image and Video Databases II, number 2185*, 1994.
- [Pre94] J. Preece. *Human-Computer Interaction*. Addison-Wesley, 1994.
- [QF93] Yonggang Qiu and Hans-Peter Frei. Concept-based query expansion. In *Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval*, pages 160–169, Pittsburgh, USA, 1993.
- [RJ76] S. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27 :129–146, 1976.
- [Roc71] J.J. Rocchio. *Relevance feedback in information retrieval*. Prentice Hall, 1971.
- [Rod99] K. Rodden. How do people organise their photographs ? In *In Proceedings of 25th BCS-IRSG Colloquium on IR*, 1999.
- [RS90] A.S. Rojet and E.L. Schwartz. Design considerations for a space-variant visual sensor with complex-logarithmic geometry. *10th International Conference on Pattern Recognition*, 2 :278–285, 1990.
- [RW86] V.V. Raghavan and S.K.M. Wong. A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Science*, 37(5) :279–287, 1986.

- [RW03] K. Rodden and K.R. Wood. How do people manage their digital photographs ? In *In Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI'03*, pages 409–416, Florida, USA, 2003.
- [Sal71] G. Salton. *The SMART Retrieval System*. Prentice Hall, 1971.
- [SB88] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. In *Information Processing and Management*, pages 513–523, 1988.
- [SB90] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Journal of the ASIS*, 41 :4 :288–297, 1990.
- [SB97] D.L. Scapin and J.M.C. Bastien. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, 16 :220–231, 1997.
- [SC96] J.R. Smith and S.-F. Chang. Visualseek : A fully automated content-based image query system. In *ACM Multimedia*, pages 87–98, 1996.
- [SCI02] D.J. Schiano, C.P. Chen, and E. Isaacs. How teens take, view, share and store photos. In *CSCW - Interactive Poster*, 2002.
- [SFW83] G. Salton, E.A. Fox, and H. Wu. Extended boolean information retrieval. *Communications of the ACM*, 26(11) :1022–1036, 1983.
- [Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 :379–423, 623–656, 1948.
- [SM83] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [Sma94] M. Smaïl. *Raisonnement base de cas pour une recherche évolutive d'information*. Université de Nancy I, 1994.
- [Sow84] J. F. Sowa. *Conceptual Structures*. Addison-Wesley, Reading, MA, 1984.

- [SP99] Amit Singhal and Fernando C. N. Pereira. Document expansion for speech retrieval. In *Research and Development in Information Retrieval*, pages 34–41, 1999.
- [SS02] S. Singh and M. Singh. Evaluation of segmentation and texture algorithms combinations for scene analysis. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics - SMC'02*, 2002.
- [SWS⁺00] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(12) :1349–1380, December 2000.
- [SWY75] G. Salton, A. Wong, and C. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11) :613–620, 1975.
- [SZ98] Dong-Ho Shin and Byoung-Tak Zhang. A two-stage retrieval model for the TREC-7 ad hoc task. In *Text REtrieval Conference*, pages 439–445, 1998.
- [TADC03] R. Oulad Haj Thami, T. Filali Ansary, M. Daoudi, and H. Chaarani. Generic : un système pour la recherche d'images basée sur les attributs visuels et les connaissances. In *Proceedings of CORESA'03 - Compression et REprésentation des Signaux Audiovisuels*, 2003.
- [Tij03] C. Tijus. *Rapport d'activités*. Laboratoire Cognition & Usages, 2003.
- [TS00] C. Town and D. Sinclair. Content based image retrieval using semantic visual categories. In *Technical Report TR2000-14, AT&T Laboratories Cambridge*, Cambridge, UK, 2000.
- [TS01] Christopher Town and David Sinclair. Ontological query language for content based image retrieval. In *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 75–81, 2001.
- [VFJZ99] A. Vailaya, M. Figueiredo, A. Jain, and H.-J. Zhang. Content-based hierarchical classification of vacation images. In *In Proceedings of IEEE Multimedia*

Systems'99 - International Conference on Multimedia Computing and Systems - ICMCS'99, pages 518–523, Florence, Italy, 1999.

- [vR79] C.J. van Rijsbergen. *Information Retrieval, 2nd edition*. Dept. of Computer Science, University of Glasgow, 1979.
- [vR86a] C.J. van Rijsbergen. A new theoretical framework for information retrieval. In *SIGIR'86*, pages 194–200, 1986.
- [vR86b] C.J. van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29(6) :481–485, 1986.
- [VT00] R.C. Veltkamp and M. Tanase. *Content-Based Image Retrieval Systems : A Survey*. Technical Report UU-CS-2000-34, 2000.
- [WC91] J.A. Waterworth and M.H. Chignell. A model for information exploration. In *Hypermedia*, volume 3(1), pages 35–58, 1991.
- [WD01] J. Z. Wang and Y. Du. Rf x ipf : A weighting scheme for multimedia information retrieval. In *ICIAP*, pages 380–385, 2001.
- [Weg99] G. Weger. *Cartographie Volume 1 : Sémiologie Graphique et conception cartographique*. Institut Géographique National - École Nationale des Sciences Géographiques, Marne La Vallée, 1999.
- [WJW01] J.Z. Wang, J.L., and G. Wiederhold. SIMPLIcity : Semantics-sensitive Integrated Matching for picture Libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9) :947–963, 2001.
- [WM93] V. Wuwongse and M. Manzano. Fuzzy conceptual graphs. In *ICCS'93*, pages 430–449, 1993.
- [ZD69] P. Zunde and M.E. Dexter. Indexing consistency and quality. *American Documentation*, 20(3) :259–267, 1969.
- [Zip32] G.K. Zipf. *Selective Studies and the Principle of Relative Frequency in Language*. Harvard University Press, 1932.

[Zip49] G.K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949.