



HAL
open science

Fonctions de Croyance et Indexation MultimodaleApplication à l'Identification de Personnes dans des Albums

Said Kharbouche

► **To cite this version:**

Said Kharbouche. Fonctions de Croyance et Indexation MultimodaleApplication à l'Identification de Personnes dans des Albums. Interface homme-machine [cs.HC]. Université de Rouen, 2006. Français. NNT: . tel-00232806

HAL Id: tel-00232806

<https://theses.hal.science/tel-00232806>

Submitted on 1 Feb 2008

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.

N° : ...

FONCTIONS DE CROYANCE ET INDEXATION MULTIMODALE

APPLICATION À L'IDENTIFICATION DE PERSONNES DANS DES ALBUMS

Thèse de doctorat

présentée en vue de l'obtention du titre de

Docteur de l'Université de Rouen

Discipline : Traitement du Signal et des Images

présentée par

Saïd KHARBOUCHE

Soutenue le 8 décembre 2006 devant le jury composé de :

O. COLOT	Professeur, Université des Sciences et Technologies de Lille	Rapporteur
G. QUENOT	Chargé de recherches HDR au CLIPS-IMAG, Grenoble	Rapporteur
C. GARCIA	Chargé de recherches à France Telecom, Rennes	Examinateur
M. PLU	Chargé de recherches à France Telecom, Lannion	Examinateur
A. BENSRAIR	Professeur à l'INSA de Rouen	Examinateur
P. MICHE	Professeur à l'Université de Rouen	Directeur de Thèse
P. VANNOORENBERGHE	Maître de conférences à l'Université de Toulouse III	Co-directeur de Thèse
C. LECOMTE	Maître de conférences à l'Université de Rouen	Encadreur de Thèse

UNIVERSITÉ DE ROUEN

LABORATOIRE D'INFORMATIQUE, DE TRAITEMENT DE L'INFORMATION ET DES SYSTÈMES

Sommaire

Introduction	15
1 Système d'indexation et de la recherche multimédia	19
1.1 Système d'indexation multimédia	20
1.1.1 Phase d'indexation	21
1.1.2 Phase de recherche	24
1.1.3 Contexte du travail	25
1.2 Indexer par le contenu ou le contexte?	25
1.2.1 Indexation par le contenu	26
1.2.2 Indexation par le contexte	29
1.2.3 Indexation par le contenu et le contexte	32
1.3 De la nécessité d'une architecture de fusion hybride	33
2 Fusion d'informations pour l'indexation multimédia	35
2.1 Méthodes classiques de fusion d'informations	35
2.1.1 Système de vote majoritaire	36
2.1.2 Combinaison linéaire	37
2.1.3 Théorie des probabilités	38
2.1.4 Théorie des possibilités	39
2.1.5 Fusion par des techniques d'apprentissage	39
2.2 Fonctions de croyance	41

2.2.1	Fonctions de croyance	41
2.2.2	Fusion d'informations	43
2.2.3	Quelques outils spécifiques	44
2.2.4	Affaiblissement	44
2.2.5	Éléments focaux : Noyau	44
2.2.6	Prise de décision	45
2.2.7	Notation	45
2.3	Applications des fonctions de croyance à l'indexation	45
2.3.1	Système Diogenes	46
2.3.2	Boucle de pertinence automatique	51
2.4	Conclusion	54
3	FUSINDEX : Un système de FUSion d'INDEX	57
3.1	Description du système	58
3.1.1	Moteurs d'indexation	58
3.1.2	Position du problème et solution envisagée	59
3.2	Description des différents moteurs d'indexation	61
3.2.1	Détecteur de visage : <i>DV</i>	62
3.2.2	Reconnaisseur de visage : <i>RV</i>	63
3.2.3	Reconnaisseur de genre Homme/Femme : <i>H/F</i>	64
3.2.4	Reconnaisseur de Prénom audio : <i>RPa</i>	65
3.2.5	Reconnaisseur de Prénom texte : <i>RPt</i>	66
3.2.6	Informations contextuelles : Horloge <i>Hr</i> , Lieux <i>L</i>	66
3.3	Hierarchisation des informations et indexation de la base	67

3.3.1	Organisation hiérarchique	67
3.3.2	Base indexée	68
3.4	Génération des métadonnées par fusion d'informations	70
3.4.1	Sources d'informations disponibles	70
3.4.2	Génération des métadonnées	70
3.4.3	Évaluation des degrés de pertinence associés aux métadonnées	72
3.4.4	Prise en compte de la fiabilité et Fusion Contenu et Contexte	74
3.5	Conclusion	76
4	FUSINDEX et l'indexation par le contenu	79
4.1	Introduction	79
4.2	Modélisation du contenu à partir de l'image : I	81
4.2.1	Détecteur de visage : DV	81
4.2.2	Reconnaisseur de visage : RV	82
4.2.3	Reconnaisseur de genre homme/femme : H/F	85
4.2.4	Source d'information IZ pour la gestion des zones détectées	91
4.2.5	Croyances associées aux métadonnées issues de l'image I	93
4.3	Modélisation du contenu à partir des commentaires : Com	94
4.3.1	Définition d'une matrice de recouvrement	95
4.3.2	Scores	96
4.3.3	Fusion inter-commentaires	97
4.3.4	Détection des nouveaux prénoms	98
4.3.5	Croyances associées aux métadonnées issues des commentaires Com	99
4.4	Exemple d'indexation par le contenu d'un document multimédia	99

4.4.1	Analyse du contenu de l'image : I	100
4.4.2	Analyse des commentaires du document : Com	104
4.4.3	Fusion entre image I et commentaires Com	107
4.5	Conclusion	108
5	FUSINDEX et l'indexation par le contexte	109
5.1	Introduction	110
5.2	Etude du contexte spatio-temporel	112
5.2.1	Concept de proximité temporelle	112
5.2.2	Concept de proximité géographique	112
5.3	Renforcement et contestation	114
5.3.1	Extraction du contexte pour l'indexation d'un document	114
5.3.2	Incertitude associée au contexte	115
5.3.3	Calcul des coefficients de fiabilité	116
5.4	Exemple d'indexation par le contexte d'un document multimédia	118
5.4.1	Extraction du contenu du document à indexer I_1	118
5.4.2	Extraction du contexte des documents I_2 et I_3	120
5.4.3	Informations contextuelles associées aux documents	123
5.4.4	Processus de fusion du contenu et du contexte	124
5.5	Conclusion	125
6	Évaluation des performances du système	127
6.1	Exemples	127
6.1.1	Exemple 1	128
6.1.2	Exemple 2	131

6.2	Corpus et méthodologie d'évaluation	134
6.2.1	Le corpus	134
6.2.2	Conditions d'évaluations	136
6.2.3	Les dates et les coordonnées géographiques	138
6.3	Evaluations	139
6.3.1	Evaluation 1 : Fusion du contenu	140
6.3.2	Evaluation 2 : Utilisation d'un reconnaiseur H/F simulé	140
6.3.3	Evaluation 3 : Fusion contenu et contexte	141
6.3.4	Evaluation 4 : Recherche d'images	142
6.4	Conclusion	143
	Conclusion et perspectives	147
	Bibliographie	151

Liste des tableaux

3.1	Exemple de fonctions de masse associées aux sources I , Com et $Cont$ pour une personne P_i du carnet d'adresse B	73
3.2	Exemple de fonctions de masse associées aux sources I , Com et $Cont$, fusion $m^\Psi[., P_i]$ et probabilité pignistique $BetP_{m^\Psi[., P_i]}$ avec $\alpha_I = \alpha_{Com} = \alpha_{Cont} = 1$	74
3.3	Exemple de fonctions de masse associées aux sources I , Com et $Cont$, fusion $m^\Psi[., P_i]$ et probabilité pignistique $BetP_{m^\Psi[., P_i]}$ avec $\alpha_I = \alpha_{Com} = 0.1$ et $\alpha_{Cont} = 1$	74
3.4	Exemple de fonctions de masse $m^\Psi[., P_i]$ associées aux différents prénoms du carnet d'adresses $B = \{Homer, Marge, Bart, Lisa, Maggie\}$	75
3.5	Exemple de probabilités pignistiques $BetP_{m^\Psi[., P_i]}$ associées aux fonctions de masse du tableau 3.4 pour les prénoms du carnet d'adresses B	75
4.1	Intervalle de distance et bonnes réponses associées.	85
4.2	Distances obtenues par le reconnaiseur de visages RV	101
4.3	Fonctions de masse $m_{\alpha_{RV}}^\Omega[RV, Z_j](.)$ obtenues par le reconnaiseur de visages RV	101
4.4	Probabilités pignistiques associées aux fonctions de masse $m^\Omega[I', Z_j](P_i)$	103
4.5	Crédibilités associées aux fonctions de masse $m^\Omega[I', Z_j](P_i)$	103
4.6	Probabilités pignistiques associées aux fonctions de masse $m^\Omega[I, Z_j](P_i)$	104
4.7	Fonctions de masse $m^\Psi[Com', P_i](.)$ pour les éléments focaux Per, \overline{Per} et Ψ	106
4.8	Fonctions de masse $m^\Psi[Com, P_i](.)$ pour les éléments focaux Per et Ψ	107
4.9	Fonctions de masse $m^\Psi[I \oplus Com, P_i](.)$ pour les éléments focaux Per et Ψ	107
4.10	Probabilité pignistique $BetP_{m^\Psi[I \oplus Com, P_i]}(Per)$	107
5.1	Fonction de masse $m^\Psi[I \oplus Com, P_i](.)$ associée au document I_1	119

5.2	Paramètres liés aux notions de proximité temporelle et Géographique.	123
5.3	Fonction de masse contextuelle $m_{\alpha_{2,i}}^{\Psi}[I_2, P_i]$ associée au document I_2	123
5.4	Fonction de masse contextuelle $m_{\alpha_{3,i}}^{\Psi}[I_3, P_i]$ associée au document I_3	124
5.5	Fonction de masse $m_{\alpha_{Cont}}^{\Psi}[Cont, P_i]$ associée au document I_1	124
5.6	Fonction de masse finale $m^{\Psi}[I \oplus Com \oplus Cont, P_i]$ associée au document I_1 . . .	124
6.1	Distances obtenues par le reconnaisseur de visages RV pour l'exemple 1.	128
6.2	Résultats obtenus par le système après la fusion des informations.	131
6.3	Distances obtenues par le reconnaisseur de visages RV pour l'exemple 2.	131
6.4	Résultats obtenus par le système après la fusion des informations $I \oplus Com$. . .	133
6.5	Les 25 personnes connues du carnet d'adresses.	135

Liste des figures

1.1	Structure d'un système d'indexation et de recherche de documents multimédia.	21
2.1	Fusion de trois sources image, voix et texte pour la classification.	40
2.2	Page requête et page des résultats du système Diogenes	46
2.3	Architecture du système Diogenes	47
2.4	Etapas de la boucle de pertinence du système Diogenes	50
2.5	Image requête et images retournées par un tri simple (a) et par l'approche proposée (b).	54
2.6	Précision/Rappel dans la catégorie Football (48 images) pour un tri simple et pour l'approche proposée.	55
3.1	Illustration de l'interface Someone-Photo mettant en évidence les trois catégories : Albums, Personnes et Lieux.	59
3.2	FUSINDEX : Outils d'indexation et illustration de la multimodalité.	60
3.3	Architecture technologique de l'interface Someone-Photo.	61
3.4	Outil de détection des visages <i>DV</i> sur l'image. A chaque image analysée, les informations fournies par l'outil sont le nombre de visages détectés, les positions et tailles des fenêtres englobantes ainsi qu'un score de détection associé à chaque zone.	62
3.5	Outil de reconnaissance des visages <i>RV</i> sur l'image. A chaque visage détecté, l'information disponible est rassemblée dans un vecteur de distances permettant d'établir un tri au niveau des personnes <i>P</i> reconnues.	63
3.6	Outil de reconnaissance de genre H/F à partir de l'image. A chaque image et pour chaque visage détecté, l'outil fournit une information sur le genre H/F de la personne détectée.	64

3.7	Outil de reconnaissance de prénoms à partir des commentaires audio. A chaque prénom détecté est associé un score $RPa(.)$	65
3.8	Outil de reconnaissance de prénoms à partir des commentaires textuels. A chaque prénom détecté est associé un score $RPt(.)$	66
3.9	Hierarchie des sources d'information utilisées pour engendrer les métadonnées.	68
3.10	La base des personnes de la famille Simpson contient 5 personnes connues.	71
3.11	Document à indexer composé d'une image avec la détection de deux zones, de deux commentaires textuels et trois commentaires vocaux.	71
4.1	Evaluation du détecteur de visages DV en fonction du score s_{DV} fourni par le moteur, allocation de masse associée au détecteur DV	81
4.2	Taux de bonne reconnaissance en fonction du rang des visages bien reconnus par le reconnaisseur de visages RV	83
4.3	Evolution du rang de classement du reconnaisseur de visage en fonction de la distance fournie par le moteur d'indexation RV	84
4.4	Taux de bonne reconnaissance calculé pour le moteur RV dans chaque intervalle de distance et courbe de régression choisie pour approximer ce taux.	86
4.5	Bonne reconnaissance en fonction de la différence entre les distances respectives 'homme' et 'femme' $ d_{h,j} - d_{f,j} $	87
4.6	Étapes pour construire le jeu de masses lié au reconnaisseur HF/P_i	88
4.7	Calcul et mise à jour de la probabilité F_i sur le sexe féminin de la personne P_i	90
4.8	Personnages du carnet d'adresses utilisés pour l'indexation	100
4.9	Document à indexer : Image et Commentaires.	100
5.1	Les documents $I_k, I_{k'}$ et $I_{k''}$ dans un espace spatio-temporel permettant de définir les concepts de proximité en terme de distance et de temps.	110
5.2	Concept de proximité temporelle et fonction g_T Choisie.	113
5.3	Concept de proximité géographique et fonction g_{Lp} Choisie.	114

5.4	Concept d'éloignement géographique et fonction g_{Le} Choisie.	115
5.5	Allure de la courbe représentant le coefficient $\alpha_{k',i}$ en fonction de $maxBel_{k',i}$	118
5.6	Document à indexer I_1 : image associée, aucun commentaire vocal ou textuel.	119
5.7	Document I_2 de la base indexée : Image associée, pas de commentaire vocal ou textuel.	120
5.8	Document I_3 de la base indexée : Image et commentaires Associés.	121
6.1	Document de l'exemple 1 à indexer : image sans Commentaire.	129
6.2	Document de l'exemple 2 à indexer : image avec commentaire textuel.	131
6.3	$DatesCoordonnesA$: Histogramme de la distribution des dates.	139
6.4	$DatesCoordonnesB$: Histogramme de la distribution des dates.	139
6.5	Performances en phase d'évaluation 1 : Fusion du contenu.	141
6.6	Performances en phase d'évaluation 2 : Fusion du contenu avec H/F' (H/F simulé).	142
6.7	Performances en phase d'évaluation 3 : contenu et contexte.	143
6.8	Performance en phase d'évaluation 4 : courbes rappel/precision de RV , Com , $Fusindex$ et système de vote.	144
6.9	Performance en phase d'évaluation 5 : courbes rappel/precision de $Fusindex$ et système de vote avec et sans H/F simulé.	145

Introduction

De nos jours, il est devenu facile pour une personne de prendre de nombreuses photos, de les stocker et de les envoyer à d'autres personnes. En effet, d'une part, les ventes d'appareils photo numériques de tous genres se sont multipliées et ont été intégrés aux téléphones portables. Ces appareils glissés dans la poche permettent de prendre de nombreuses photos en tous lieux et à tout moment. D'autre part, le développement de l'accès aux outils informatiques et aux réseaux de télécommunications multimédia permettent de stocker, copier, partager, envoyer toutes ces photos pour un coût dérisoire [1, 2]. Ainsi, chacun se retrouve à posséder un nombre colossal de photos personnelles qu'il a lui-même prises ou qui lui ont été envoyées par sa communauté qui, elle aussi, ne fait que s'agrandir¹. La gestion de ces bases importantes d'images devient alors un réel problème des nombreux usagers de ces technologies devenues accessibles à tous. Le but est d'organiser, structurer, classer ces collections afin que leur parcours et la sélection de photos particulières soient efficaces et agréables.

Cette organisation, est également appelée indexation de nos jours dans le jargon scientifique et nous définirons plus précisément ce terme dans la suite. Pour le grand public, il existe de multiples façons d'organiser des albums photos, mais les plus courantes sont de les organiser selon les personnages, selon les dates, selon les lieux ou selon les événements par exemple [3, 4, 5, 6]. Dans cette thèse, nous nous intéresserons à l'indexation selon les personnes car l'indexation par dates et lieux qui facilite le tri des photos selon certains critères est un problème presque entièrement résolu à l'heure actuelle. Plus généralement, l'indexation de collections de photos peut être complètement manuelle, entièrement automatique ou semi-automatique. Dans le cas d'une indexation manuelle, on laisse à l'utilisateur le loisir d'annoter personnellement ces photos (une par une ou par groupe) ce qui rend le travail coûteux et fastidieux si on possède une grande collection d'images. Une indexation complètement automatique laisse le système mis en œuvre se charger d'indexer les photos sans aucune intervention humaine. Malheureusement, ces systèmes d'indexation automatique exploitent des informations bas-niveau (les couleurs, les

¹A titre d'illustration, on peut lire sur le web (<http://blog.forret.com/2006/10/a-picture-a-day-flickers-storage-growth>) que pour le service de partage de photos en ligne Flickr, il y a 920.000 photos partagées par jour.

textures présentes dans l'image, les dates d'acquisition, ...) pour identifier des concepts de haut-niveau dont la sémantique est très difficilement appréhendable. Ce type d'indexation reste encore une tâche difficile à réaliser [7, 8, 9] pour les albums de photos personnelles qui se caractérisent par leurs grandes variabilités et des catégories illimitées. Lorsque l'indexation est semi-automatique, le système se comporte comme un système d'indexation automatique dont l'utilisateur cherche à valider le processus étape par étape. Le rôle de tels systèmes est d'aider l'utilisateur à annoter ces images afin de lui permettre une indexation manuelle plus rapide et efficace de ses albums. Cela consiste, pour chaque photo à indexer, à proposer un ensemble de descripteurs triés selon leurs degrés de pertinence ou leur fiabilité.

Cette thèse se situe dans la cadre de l'organisation semi-automatique d'albums photo et s'intègre dans un cadre applicatif particulier d'un prototype de service développé par la division recherche et développement de France Telecom². Dans ce cadre applicatif, les photos peuvent être partagées entre plusieurs personnes et peuvent être commentées vocalement et/ou textuellement par ces différents utilisateurs. Le processus d'indexation développé dans cette thèse ne se limite pas seulement à l'indexation d'une collection d'images mais traite également leurs commentaires associés ce qui rend ces contenus multimédia. D'autres informations peuvent également être associées aux photos comme les dates et les lieux d'acquisition de l'image (qui sont connus avec une grande précision grâce notamment au développement de moyens de géo-localisation des appareils multimédia) et peuvent être exploitées pour l'organisation de la base [10].

Ainsi, le travail envisagé dans le cadre de cette thèse se focalise sur des documents multimédias avec différentes modalités : image, texte, son et données. L'un des objectifs à atteindre concerne la fusion des informations issues de ces différentes modalités dans le but d'identifier les personnages figurant dans les images qui permettent ainsi d'indexer les documents. Chacun des documents de la collection est représenté par ses contenus relatifs aux différents médias mais est aussi considéré dans son contexte. Pour analyser chaque contenu d'un document, nous utilisons des outils d'indexation qui leur sont spécifiques. Le contexte d'une image est exploité à partir de descripteurs déjà calculés sur des documents de la base en exploitant les dates et lieux d'acquisition des images associées. La contribution essentielle de ce travail concerne donc **l'indexation de documents multimédia par leur contenu et leur contexte**.

Une des particularités des outils automatiques d'analyse de ces contenus est l'incertitude et l'imprécision de leurs résultats. Pour gérer au mieux ces informations incertaines dans le système d'indexation développé, nous avons résolument opté pour la théorie des fonctions de croyance [11]. Cette théorie a également été choisie pour ses méthodes adaptatives de fusion d'informations et d'identification du conflit entre des sources de fiabilité variable. Elle constitue de

²Ce travail a été réalisé dans le cadre d'un contrat de recherche externe avec France Telecom (num. 46129144)

plus un cadre suffisamment générique pour englober les autres théories des mesures de confiance. Parmi celles-ci, les mesures classiques de probabilité peuvent être vues comme des fonctions de croyance particulières [12] permettant ainsi d'établir un lien entre information statistique et incertitude. L'objectif de cette thèse est de montrer comment les fonctions de croyance peuvent être utilisées pour l'indexation multimodale d'albums personnels et plus particulièrement pour la reconnaissance de personnes dans ces photos.

Ce document s'organise de la façon suivante. Le premier chapitre donne une vision panoramique des systèmes d'indexation de documents multimédia en proposant au lecteur les grands principes qui les régissent. Il tente également de donner quelques pistes sur l'utilité d'indexer par le contenu **et** par le contexte et sur la nécessité d'une architecture de fusion hybride alliant ces deux types d'information. Un état de l'art sur les différentes techniques de fusion d'informations dans un contexte d'utilisation en indexation de documents multimédia est ensuite proposé dans le chapitre 2. La théorie des fonctions de croyance, vue comme un outil de représentation et de fusion d'informations incertaines y est détaillée d'un point de vue mathématique. Enfin, deux applications utilisant les fonctions de croyance pour l'indexation sont présentées et permettent de mettre en évidence les avantages de ce formalisme. Le troisième chapitre donne une vue globale sur les principaux éléments du système d'indexation proposé dans cette thèse. Nous y détaillons les différents outils d'indexation utilisés. Ils permettent d'exploiter les informations issues des documents et nécessitent d'être organisés sous forme hiérarchique. Les modèles et les techniques employés pour l'indexation d'un document via son contenu et son contexte sont ensuite présentés respectivement dans les chapitres 4 et 5 du manuscrit. Les fonctions de croyance occupent une place majeure dans la représentation des incertitudes liées aux informations extraites mais permettent également de les fusionner offrant ainsi une gestion multimodale du document en vue de son indexation. Dans chacun de ces chapitres, un exemple didactique permet de clarifier les mécanismes développés au sein du système et d'avoir un aperçu sur ses principales capacités. Afin de juger de ses performances, nous présentons dans le chapitre 6, les résultats des évaluations menées sur le système dans le cadre d'une application dédiée à l'identification de personnes dans une collection de plus de 450 documents. Les expérimentations permettent de démontrer la capacité du système proposé à fusionner les informations fournies par les différents outils d'indexation et d'accroître ainsi la précision en terme d'indexation. Enfin, de manière assez classique, le manuscrit se termine par une conclusion et la présentation de quelques perspectives qui semblent de bonnes pistes de recherche pour les travaux futurs.

Chapitre 1

Systeme d'indexation et de la recherche multimedia

Ce chapitre donne un aperçu de différents systèmes d'indexation multimedia qui ont déjà été mis en œuvre dans le cadre d'applications spécifiques en présentant leurs architectures ou en décrivant ceux qui sont en cours de développement. Comme nous l'avons précisé dans l'introduction, la problématique de ce travail de thèse consiste à faciliter l'indexation et l'organisation d'albums de photos numériques. Pour cette raison, la majorité des travaux évoqués dans cet état de l'art se focalise sur cette problématique.

Le chapitre s'organise de la façon suivante. Dans un premier temps, la structure de base d'un système d'indexation multimedia est présentée. Nous détaillons les différents éléments de la phase d'indexation des documents mais également de la phase de recherche dans la base. Nous différencions dans ce premier paragraphe les informations extraites du contenu du document des informations extraites de son contexte. Bien que la structure du document puisse constituer une information importante pour la phase d'indexation (comme par exemple dans le cas d'indexation de documents XML), nous nous focalisons uniquement dans le cadre de notre application sur le contenu et le contexte. Ces deux types d'informations peuvent être utilisés de façon à faciliter l'indexation à partir d'une base de documents existante préalablement validée : la base d'apprentissage. Dans un deuxième temps, nous essayons au travers de la littérature de répondre à la question sur la nécessité d'indexer par le contenu ou par le contexte. Pour cela, on dresse un état de l'art sur les travaux traitant de l'indexation basée sur ces deux catégories d'informations. Enfin, on aborde la problématique de l'indexation hybride qui consiste à indexer la base de documents par l'utilisation conjointe d'informations sur le contenu et sur le contexte. Nous donnerons quelques éléments de réponse à la question posée dans la conclusion de ce chapitre.

1.1 Système d'indexation multimédia

Avant de présenter les concepts liés à l'indexation de documents, nous définissons ce qu'on entend par document multimédia. De nos jours, on utilise le mot multimédia pour désigner toute application utilisant ou servant à travailler sur au moins un média spécifique : image, son ou texte. Ainsi, nous appellerons document multimédia un document numérique qui contient au moins un média visuel (image, vidéo), un média sonore (musique, parole, ...) ou encore un média textuel (texte, HTML, ...). La vidéo numérique est à elle seule un bon exemple de document multimédia puisqu'elle se compose de séquences d'images, de sons et voire même de textes sous forme de transcriptions. Les pages WEB constituent un autre exemple de document numérique multimédia composées à la fois de textes, de sons, d'images et de vidéos. La photographie prise par un appareil numérique peut être considérée elle aussi comme un document multimédia si toutefois on lui associe d'autres informations comme des données textuelles sous forme de métadonnée. Dans le cadre de ce travail de recherche, nous nous focalisons sur l'indexation de ce dernier type de documents multimédia. Ainsi, les documents multimédia d'étude sont composés de photographies auxquelles peuvent être associées des données textuelles et des données sonores. Le but de ce travail consiste à faciliter l'indexation de tels albums.

La figure 1.1 illustre les principaux composants d'un système d'indexation de documents multimédia [13]. Dans la pratique, le système peut se diviser en deux phases, une phase d'indexation (hors-ligne) et une phase de recherche (en-ligne). Le but principal de la phase d'indexation consiste à structurer la base, souvent de grande taille, des documents multimédia riches en contenu. La structuration de cette base devra faciliter la phase de recherche et de navigation dans la base, phase rendue beaucoup plus commode de par son indexation.

On peut associer à chacun des documents multimédia un certain nombre d'informations qui permettent de caractériser le contexte dans lequel le document a été produit. Prenons l'exemple d'une photo qui a été prise à un instant t_1 dans un lieu précis l_1 . Ces données t_1, l_1 extrinsèques au document lui-même peuvent être vues comme des informations contextuelles au sens où elles permettent de caractériser le contexte spatio-temporel où a été prise la photo. Imaginons maintenant qu'une photo de la base ait été prise à un instant t_2 proche temporellement de t_1 dans un lieu l_2 proche géographiquement¹ de l_1 . Que peut-on dire de l'index à associer à la deuxième photo par rapport à l'index de la première ? Nous tenterons de répondre à cette question comme une contribution de ce travail. D'une certaine manière, les informations contextuelles permettent de caractériser l'environnement dans lequel le document a été produit. A titre d'exemple, l'origine du document (propriétaire, expéditeur, ...), le destinataire du document, les dates et lieux de

¹Nous définirons les notions de 'proche géographiquement' et 'proche temporellement' dans le chapitre 4.

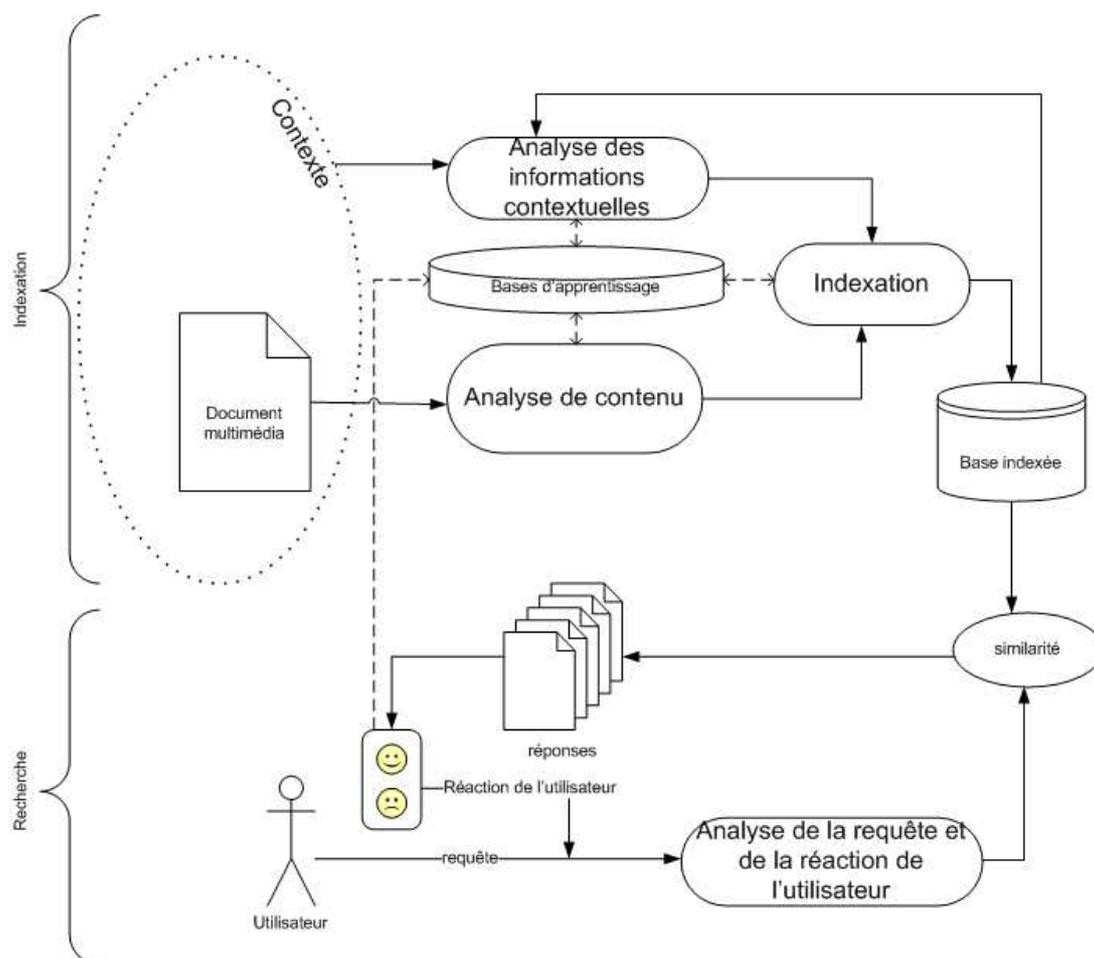


Figure 1.1 : Structure d'un système d'indexation et de recherche de documents multimédia.

création, ... sont autant d'informations qu'il s'agira de prendre en considération lors de la phase d'indexation. A partir de ces considérations, nous détaillons dans les paragraphes suivants les phases d'indexation et de recherche dans la base.

1.1.1 Phase d'indexation

La phase d'indexation, consiste à ne présenter le document dans la base indexée qu'avec un ensemble très limité de données, appelées *métadonnées*², par rapport à la quantité réelle qu'il contient. Ces données extraites des documents doivent être les plus pertinentes et discriminantes possibles au regard de la phase de recherche. Ainsi, la précision de cette extraction se détermine souvent au détriment de la rapidité du processus d'indexation. Néanmoins, ce dernier critère n'est pas primordial dans cette phase puisque, pour la majorité des systèmes, elle s'effectue en amont

²Le mot métadonnée sera préféré aux mots index ou descripteur.

(hors-ligne) de la phase de recherche où le temps n'a que peu d'importance. Dans cette phase d'indexation, le document passe par deux étapes. La première consiste à extraire et analyser les informations contenues dans le document et/ou dans son contexte. La deuxième étape consiste à placer dans la base indexée un vecteur descriptif du document (index ou métadonnées). Ces étapes sont décrites dans les paragraphes suivants.

Analyse de contenu

Si on s'intéresse uniquement au contenu, la phase d'extraction consiste à analyser le document multimédia afin de calculer un ensemble de métadonnées qui indexe d'une manière optimale le contenu de ce document. Ces métadonnées sont choisies en fonction du type d'indexation visé. Ce processus se focalise uniquement sur le contenu du document multimédia sans traiter le contexte en appliquant les techniques de traitement de l'image, de la parole ou du texte selon la nature des données numériques présentes dans le document à indexer. Là encore, ce processus dépend largement de l'objectif visé en terme d'indexation. Comme nous allons voir dans ce chapitre, il existe plusieurs systèmes qui exploitent le contexte dans le but de décrire le contenu (passer du contexte au contenu).

Analyse d'informations contextuelles

Comme illustré dans la figure 1.1, l'étape de traitement des informations contextuelles consiste à récupérer et à analyser ces informations afin de ne conserver que celles qui sont utiles vis-à-vis de l'indexation. L'origine de ces informations peut être diverse. Il peut s'agir des données accompagnant le document à indexer (date de création, lieu, taille, ...) ou d'autres informations provenant de sources complètement indépendantes du document lui-même. Dans ce dernier cas, on peut citer comme exemple des informations extraites du contenu d'un autre document (déjà indexé) partageant certaines caractéristiques communes avec le document à indexer. Nous verrons que ces dernières données pourront s'avérer utiles dans le processus d'indexation.

Indexation

Dans l'étape d'indexation proprement dite, le système est chargé de calculer le vecteur³ descriptif final du document à indexer, l'index. Il peut s'agir d'un vecteur de métadonnées qui matérialise la réalisation de concepts saisissables comme par exemple la présence d'un paysage

³On parle généralement de vecteur lorsque les informations possèdent des dimensions continues et où elles peuvent être projetées dans un espace multidimensionnel.

(mer, arbres, montagnes), la présence d'un personnage (identité, prénom), etc. Il peut être question également de notions insaisissables comme la distribution et la proportion des couleurs ou des textures dans l'image, le spectre dans un signal audio, etc. Ces caractéristiques peuvent être automatiquement générées ou prédéfinies par le concepteur du système. Dans notre problématique particulière, le calcul de cet ensemble de métadonnées s'appuie sur les informations résultantes de l'analyseur de contenu et de l'analyseur de contexte.

Les valeurs du vecteur descriptif représentent des degrés d'appartenance à un concept donné mais, selon la nature des informations extraites, elles peuvent prendre :

- des valeurs binaires (appartenance ou non appartenance au concept) ;
- des valeurs de scores qui permettent de quantifier d'une certaine manière l'appartenance à un concept (par exemple le cas d'une mesure de probabilité) ;
- des mesures de distance matérialisant le fait que le document à indexer s'éloigne du concept étudié (la présence d'arbres dans une photo est plus proche du concept 'nature' que du concept 'ville' bien que l'on trouve des arbres dans les villes) ;
- des valeurs ordinales qui permettent d'ordonner les concepts auxquels le document peut appartenir suivant une relation d'ordre prédéfinie (le document appartient plus au concept 'nature' qu'au concept 'ville').

Chacune de ces valeurs peut être représentée sur un axe à partir duquel il est possible de trier, d'organiser les documents de la base selon leur degré d'appartenance au concept. Le choix de ces axes reste toutefois une des problématiques importantes de la phase d'indexation qui est souvent guidée par l'analyse de documents déjà indexés rassemblés dans la base d'apprentissage.

Base d'apprentissage

Les systèmes d'indexation multimédia les plus sophistiqués tirent profit de la mise en place d'un mécanisme d'apprentissage automatique. Ce processus a pour but de s'adapter aux besoins, qui sont généralement guidés par l'utilisateur, en terme d'indexation. Il permet d'analyser la pertinence des différents descripteurs implémentés dans les trois étapes de la phase d'indexation (analyse de contenu, analyse du contexte et indexation). Cette phase d'apprentissage est généralement basée sur des techniques de reconnaissance de forme [14, 15] qui font appel aux méthodes de discrimination supervisée comme les Séparateurs à Vaste Marge (SVM)[16], les chaînes de Markov cachés (HMM), les réseaux neuronaux, ... [17]. L'utilisation de tels algorithmes d'apprentissage nécessitent de pouvoir disposer d'une base de documents suffisamment renseignée afin de faciliter la mise à jour des paramètres associés aux descripteurs. Ces bases d'apprentissage peuvent être alimentées soit directement par le concepteur du système (annotations manuelles des documents sur les concepts prédéfinis) ou via une boucle, appelée boucle de pertinence qui

permet à l'utilisateur d'annoter la pertinence des réponses du système lors de la recherche [18]. Pour certains systèmes, le mécanisme d'apprentissage peut également proposer l'indexation de certains documents de façon automatique et permettre d'augmenter la base des documents en apprentissage sans intervention de l'utilisateur.

1.1.2 Phase de recherche

De manière concrète, l'utilité première d'un système d'indexation multimédia est de permettre à un utilisateur de rechercher dans une base de documents multimédia ceux qui correspondent à sa requête. Cette phase de recherche, aussi appelée en-ligne, s'exécute en temps réel ce qui nécessite d'optimiser la complexité des algorithmes et les masses de données à traiter afin de diminuer le temps de réponse. Ce temps de réponse est l'un des critères importants qui conditionne la qualité du système d'indexation multimédia. Dans la pratique, la requête de l'utilisateur est tout d'abord analysée pour permettre le calcul d'une mesure de similarité entre le document recherché et les documents de la base. Éventuellement, cette requête peut être affinée par la boucle de pertinence et les retours de l'utilisateur [18, 19, 20].

Analyse de la requête

La requête est vue par le système comme un document multimédia. En effet, elle doit être décomposée dans le même espace des descripteurs extraits des documents indexés. Ceci est rendu nécessaire par le fait qu'il faut qu'elle soit comparable aux documents de la base. Ainsi, la requête se compose des mêmes éléments que les document multimédia, c'est-à-dire : texte, son et image. Par ailleurs, on peut également analyser une requête fruit de plusieurs sous-requêtes (multi-requêtes) comme par exemple l'utilisation conjointe de mots-clés et d'une image. L'analyse de la requête consiste donc à calculer son vecteur descriptif. Si elle prend la même forme que les documents indexés (par exemple recherche d'images par leurs contenus), elle empruntera les mêmes chemins d'analyse que les documents à indexer (c'est-à-dire analyse du contenu, analyse du contexte et indexation) pour le calcul de son vecteur descriptif [21].

Mesure de similarité

L'étape suivante consiste à comparer le vecteur extrait de la requête aux vecteurs des documents indexés pris individuellement ou par groupe. Ce processus consiste à trier les documents indexés vis-à-vis de la requête. Ceci est réalisé par l'intermédiaire d'une fonction de similarité qui peut posséder certaines propriétés des mesures de distance : symétrie, séparation et inégalité

triangulaire. Cette fonction de similarité peut être une fonction booléenne (pertinent/non pertinent), une fonction floue admettant des valeurs comprises entre 0 et 1, une fonction exponentielle, etc. Le choix de cette fonction est crucial car l'ordre des réponses proposées à l'utilisateur varie d'une fonction à l'autre. Cette variabilité influence la précision des réponses du système [22].

Analyse de la réaction de l'utilisateur

L'analyse de la réaction de l'utilisateur est le processus qui consiste à ajuster le vecteur descriptif de la requête en fonction des réactions de l'utilisateur face aux résultats proposés [22]. Cette analyse peut conduire à affiner la requête dans un cycle itératif (boucle de pertinence) ou à modifier les paramètres de la phase d'indexation. En plus des informations extraites de la requête, l'analyse de la réaction a également comme but d'enrichir les informations sur l'objectif de l'utilisateur en profitant des informations issues de la pertinence de certains documents de la base et, a permis d'ordonner les pertinences par ordre d'importance. Finalement, les informations extraites de la réaction de l'utilisateur permettent d'indexer certains documents eu égard aux concepts prédéfinis et ainsi d'enrichir la base d'apprentissage.

1.1.3 Contexte du travail

Dans le cadre de notre problématique d'indexation d'albums de photos numériques, notre objectif consiste à produire des métadonnées puis à les proposer à l'utilisateur. En d'autres termes, il s'agit essentiellement d'un problème d'indexation où l'on cherche à discriminer des personnes dans les photos prises par les téléphones cellulaires. Cette tâche permet de faciliter pour l'utilisateur l'annotation de ses documents selon les personnages présents dans les photos [6]. Nous ne nous intéresserons pas, par conséquent, à la phase de recherche qui reste un problème à part entière. Par contre, les réactions de l'utilisateur pourront être l'objet d'un bouclage pour adapter le système à son utilisateur.

1.2 Indexer par le contenu ou le contexte ?

Au travers des différents travaux traitant de l'indexation de documents multimédia et de certaines applications, nous tentons de répondre à la question : Indexer par le contenu ou par le contexte ? Dans ce but, on dresse un état de l'art en insistant sur les réalisations proches de notre travail en termes d'objectifs.

1.2.1 Indexation par le contenu

Les systèmes d'indexation par le contenu se focalisent uniquement sur l'extraction et l'analyse des informations du contenu du document multimédia. Historiquement, il s'agit des premières informations bas niveau qui ont été utilisées pour l'indexation. Par la suite et de façon à combler le **fossé** entre le contenu sémantique du document et les index calculés, des informations de plus haut niveau ont été étudiées. Le contenu peut ainsi être représenté par un ensemble de descripteurs bas et haut niveau. Ce type d'indexation peut être mis en œuvre pour des systèmes destinés à indexer des documents numériques dont les contextes respectifs sont inconnus ou indisponibles. Nous détaillons quelques exemples de systèmes basés sur le contenu.

Les systèmes d'indexation et de recherche d'images décrits dans [22] sont basés uniquement sur le contenu visuel extrait des images. Pour chaque image à indexer, le système calcule un index à l'aide de descripteurs extraits sur les trois attributs suivants : la couleur, la texture et la forme. Ces descripteurs caractérisent l'information en valorisant leurs proportions et leurs distributions relatives et absolues dans l'image. Ils font l'objet de plusieurs travaux de recherche dans le domaine du traitement d'images et de la reconnaissance de forme [23] (le choix de l'espace colorimétrique, les ondelettes, la quantification, le suivi de contours, ...). Les algorithmes utilisés dans ce cadre permettent de réduire le bruit et la redondance des informations avant de déterminer les points et les zones d'intérêts dans l'image. A partir de ces informations, ils peuvent calculer le vecteur discriminant qui résume au mieux l'image dans la base. En phase de recherche, ces systèmes utilisent une image présentée par l'utilisateur comme requête qui permet de rechercher et de présenter les images similaires de la base indexée. Une série d'échanges entre l'utilisateur et le système (le bouclage de pertinence) permet d'affiner la requête [21, 20]. Malheureusement, comme cité plus haut, ce type de système se limite aux concepts de contenu bas niveau car jugé incapable de saisir tous les concepts de haut niveau sémantique visés par l'utilisateur.

L'image requête est riche en informations, ce qui rend difficile la détermination de l'objectif et l'intention de l'utilisateur. Comme exemple, une image requête contenant un coucher de soleil sur la mer peut engendrer une ambiguïté sur le but de l'utilisateur : soleil, mer ou les deux. Pour faciliter le traitement de concepts de plus haut niveau, certains systèmes exploitent de manière conjointe les textes qui peuvent être codés avec l'image ou autour de l'image (format JPEG2000, PNG, pages WEB, images médicales) [24, 25, 26, 27]. Les textes descriptifs dans les vidéos peuvent également être pris en compte. D'autres systèmes exploitent les textes figurant dans l'image par l'utilisation de processus de reconnaissance des caractères dans l'image (par exemple www.riya.com). Ces derniers systèmes exploitent à la fois les concepts bas et haut niveaux de l'utilisateur et établissent les liens entre ces deux niveaux améliorant ainsi les performances. Cependant, ces techniques sont limitées par la difficulté d'annoter manuellement ou de façon semi-

automatique de grandes quantités de documents, surtout pour les documents qui ne contiennent pas de données textuelles.

Dans [24], les auteurs proposent une méthode qui combine des descripteurs visuels et des mots-clés. L'index visuel se détermine à partir de l'histogramme des couleurs, de la forme et de la texture en utilisant les transformées de "Hough" et de "Fourier". En ce qui concerne les mots-clés, la méthode proposée décompose le mot-clé dans une base de données lexicales externe. La combinaison entre ces deux informations complémentaires se fait à l'aide de SVM dont les données d'apprentissage sont extraites via le retour de pertinence. Les expériences menées avec cette approche sur une base de documents représentant des concepts liés aux activités humaines (art, archéologie, architecture, ...) montrent que la combinaison entre l'image et les mots-clés améliore significativement la précision du système.

Le système **Retimm** proposé dans [25] a été conçu pour indexer et rechercher les images par leurs contenus et les mots-clés extraits des commentaires et des titres. Ce système accepte les requêtes multiples (images et/ou mots clés) pour lesquelles le vecteur descriptif de l'image est calculé à partir :

- des distributions des couleurs (quantification sur 256 niveaux) et des textures (utilisation de filtres de Gabor) dans l'image ;
- d'un ensemble de mots-clés prédéfinis.

Le calcul de similarité ne s'effectue pas par une fonction de similarité ordinaire mais par un système de vote basé sur l'algorithme des k plus proches voisins dans la base d'apprentissage. Le calcul d'un score à valeur binaire 0 ou 1, comptabilisé pour chacune des caractéristiques du vecteur descriptif, permet d'éviter le calcul de la distance dans l'espace multidimensionnel. Le score final associé à chaque image de la base correspond à la somme des votes pour toutes les dimensions. La rapidité du système en phase de recherche est un des avantages mis en évidence par les auteurs.

Une autre approche présentée dans [28] consiste à définir automatiquement les concepts qui peuvent être associés à des segments de vidéo numérique. Ces concepts sont construits de façon hiérarchique à partir de l'outil d'annotation **VideoAnnEx** d'IBM proposé dans [9]. Cette hiérarchie de concepts permet de définir des liens de similarité ou de dissemblance entre concepts. L'apprentissage de chaque concept à partir du contenu de l'image (couleur, texture, forme) est basé sur des méthodes utilisant des SVM. La probabilité qu'une vidéo appartienne à un concept donné est ajustée par l'intermédiaire de deux facteurs. Un facteur de renforcement permet d'augmenter cette probabilité si un concept similaire est également probable (notion de propagation ascendante dans la hiérarchie). Inversement, un facteur de confusion diminue la probabilité si un concept dissemblable est détecté (notion de propagation horizontale dans la hiérarchie).

Dans [8], les auteurs décrivent un système d'annotation automatique de segments vidéo qui combine les trois modalités : image, audio et texte. En premier lieu, ils définissent deux niveaux de concepts, le haut niveau et le niveau intermédiaire. Le niveau intermédiaire pour l'audio contient : silence, explosion, musique, etc qui sont extraits à partir de classificateurs basés sur des modèles de Markov cachés. Pour les informations visuelles, ce niveau englobe des concepts comme le feu, le ciel, en extérieur, en intérieur, etc extraits à partir de SVM. Les concepts haut niveau sont détectés par les outils de reconnaissance de la parole qui traduisent la parole en texte. A partir de ce texte, le système détecte des mots clés. La fusion se fait à l'aide de SVM, où les trois vecteurs descriptifs des images, de l'audio et du texte sont regroupés dans un seul vecteur. Dans cet article, les auteurs montrent que la fusion à l'aide de SVM est plus performante qu'une fusion basée sur un réseau Bayésien.

Dans [29], les auteurs proposent une méthodologie de fusion d'informations issues de différentes sources de médias (image, texte, audio). Le choix de la stratégie de combinaison des différentes modalités est justifiée par deux raisons. D'une part, la combinaison qui consiste à calculer le produit des probabilités issues de chacune des modalités exige que les sources d'informations soient complètement indépendantes entre elles. D'autre part, bien que la combinaison linéaire (fusion par pondération) soit très tolérante au bruit, elle reste une méthode de combinaison linéaire incapable d'explorer les dépendances entre les modalités. Ainsi, avant l'étape de fusion, une étape préliminaire consiste chronologiquement :

- à réduire le bruit via une analyse en composantes principales ;
- à déterminer l'espace des axes indépendants via une analyse en composantes indépendantes ;
- à former les différentes modalités via une catégorisation en modalités indépendantes.

A chaque modalité est associé un classifieur supervisé du type SVM qui permet de construire une fonction discriminante. La fusion des différentes modalités est réalisée elle aussi à l'aide d'un séparateur à vaste marge qui détermine la fonction discriminante finale. Cet algorithme de fusion est appelé SKF (Super-Kernel Fusion). Les expérimentations menées sur deux bases d'images et de vidéos permettent de constater une amélioration de l'ordre de 2% à 5% sur la précision.

Parmi les systèmes d'indexation multimédia basés sur le contenu, le système **MARVEL**⁴ [30], a été mis en œuvre par **IBM** pour gérer de grandes quantités de documents multimédia notamment le vidéo. Le système proposé a pour objectif de limiter l'annotation manuelle au profit d'une annotation presque totalement automatique. Toutefois, cela exige une annotation manuelle entre 1% et 5% des documents pour la base d'apprentissage. Les sources media sont le texte, la vidéo et l'audio. Le texte est soit issu de l'audio par des processus de reconnaissance de la parole, soit extrait du texte codé dans la norme MPEG-7. Les informations issues de la vidéo sont les

⁴www.research.ibm.com/marvel.

couleurs, les textures, les formes et les mouvements des objets dans la scène. Via la classification basée sur des *SVM*, ce système fusionne les informations issues des documents déjà annotés manuellement afin de construire les détecteurs de concepts sémantiques. Dans la phase d'indexation (ou d'annotation), chaque détecteur de concept donne des scores normalisés (compris dans l'intervalle $[0, 1]$) où chaque score peut être renforcé ou dégradé par les autres détecteurs de concepts. Ce système accepte des requêtes basées sur des descriptions visuelles, sur des concepts ou sur des mots-clés.

Un système pour fusionner les informations issues de l'image et de l'audio est également présenté dans [31]. Pour l'identification du locuteur, le système combine linéairement les deux scores résultants respectivement d'un classificateur traitant la vidéo et d'un deuxième classificateur traitant l'audio. Les coefficients utilisés pour cette combinaison sont calculés à partir des fiabilités des deux sources. Une autre application de ce système consiste à détecter les changements de locuteur. Là encore, le mode de fonctionnement consiste à combiner linéairement les deux probabilités issues des deux classificateurs vidéo et audio. La dernière application se situe dans le domaine de l'interface homme/machine, et consiste à déterminer automatiquement les ordres vocaux destinés à la machine via une caméra et un microphone. Les scores pour chaque source (vidéo et audio) sont issus d'un mélange gaussien et la fusion se fait aussi à l'aide d'une méthode de combinaison linéaire.

1.2.2 Indexation par le contexte

Les systèmes d'indexation multimédia basés sur le contexte sont souvent des systèmes qui ont été mis en œuvre pour traiter les photos et les vidéos prises par des appareils cellulaires (téléphone mobile, PDA, ...) [32]. Contrairement à l'indexation basée sur le contenu, seules les informations issues du contexte dans lequel se trouve le document lors de sa création sont prises en considération. Dans de nombreux cas, les informations contextuelles semblent suffisantes pour satisfaire l'attente de l'utilisateur désireux de rechercher ou d'indexer des documents. Les systèmes peuvent être différenciés entre eux par le type d'informations contextuelles qu'ils utilisent pour l'indexation.

Ainsi, dans [3], les auteurs utilisent l'instant et le lieu où l'image a été capturée. Ces informations sont utilisées pour ajouter un certain nombre de métadonnées associées à chacune des images. Plusieurs catégories de métadonnées peuvent être prises en compte : le nom du lieu (en décomposant le pays, la ville, la rue, le parc, ...), la saison, la période dans la journée (matin tôt, matin, après-midi, soir, nuit,...), la météo (pluie, ciel dégagé, beau ...), la température et le décalage horaire. Ainsi les photos sont organisées et indexées selon les métadonnées produites

par le système.

Sur la même idée, on trouve dans [33] un système appelé **PhotoCompas** pour organiser les photos selon deux structures hiérarchiques : une hiérarchie des événements et une hiérarchie des lieux. Ces deux structures hiérarchiques sont construites à l'aide de règles de discrimination exploitant les informations de date et de lieu. Les structures interfèrent entre elles dans la mesure où un lieu peut contenir plusieurs événements alors qu'un événement ne contient qu'un seul lieu. **PhotoCompas** propose également un service de nomination automatique des différents événements et lieux. Il est à noter que le système ne traite que les images issues d'un seul appareil photo numérique (identifié par un codage interne dans l'image), ce qui signifie que l'on ne peut pas avoir deux images avec les mêmes dates et dans deux lieux différents.

A partir d'informations contextuelles (dates et lieux) et d'une série d'images complètement ou partiellement annotées, le système décrit dans [4] permet de retrouver l'identité des personnes qui se trouvent dans une image. Ainsi, pour chaque image, il propose un ensemble de personnes candidates à partir d'un ensemble d'estimateurs : un estimateur de temps, un estimateur d'événement, un estimateur de lieu, un estimateur de lieux voisins, un estimateur de temps proches et un estimateur de co-présence. Pour une image à annoter, chaque estimateur attribue un score, qui prend la forme de probabilités estimées à partir de statistiques sur les images déjà annotées, à chaque personne d'un groupe prédéfini par l'utilisateur. Finalement, un estimateur global, présentant les candidats à l'utilisateur, réalise la fusion des autres estimateurs via deux méthodes distinctes. La première, appelée **Padding**, consiste à sélectionner dans chacune des listes ordonnées générées par les différents estimateurs les personnes candidates suivant leurs fiabilités respectives. La deuxième méthode fusionne les scores des différents estimateurs par une combinaison linéaire pondérée. Les expériences réalisées sur ce système montrent que la combinaison linéaire donne les meilleures précisions.

Le système d'indexation de photos personnelles de **Microsoft** décrit dans [5] est accessible via le web et les appareils cellulaires. Il organise les images selon les lieux géographiques :

- par les noms de lieux standard (France, Paris, ...);
- par les lieux dépendant de l'utilisateur (mon bureau, ma maison, ...);
- par le code postal, la longitude et/ou la latitude ou par les coordonnées euclidiennes.

Grâce à un algorithme particulier [34], l'image est indexée par une valeur unique au lieu de deux (latitude et longitude) mais avec une marge d'erreur de 25 mètres. L'information contextuelle relative au lieu est récupérée grâce à l'entête de l'image (équipement GPS ou manuellement introduite). L'objectif de ce système consiste à faciliter la navigation dans la base des images indexées qui sont placées sur une carte interactive aux endroits où elles ont été capturées.

Plog, développé par **HP**, est un système accessible depuis le web et les téléphones cellulaires

avec des interfaces spécifiques à ces différents terminaux [1]. Sa conception permet de faciliter l'accès et le partage des photos prises par des appareils cellulaires dotés de caméras numériques. L'indexation des photos se fait par l'analyse des dates de création qui sont regroupées par des algorithmes de classification tels que les K -means ou la classification hiérarchique descendante. Le même type d'information est utilisé dans [35] où les groupes d'images sont organisés selon un calendrier ou une structure hiérarchique. Le regroupement des images s'inspire d'une classification hiérarchique où le nombre de groupes est mal défini. Chaque groupe est ensuite représenté par des échantillons d'images sélectionnés en fonction de leur écart temporel, leur différence de contraste et la taille du groupe.

L'une des applications les plus abouties dans ce domaine est sans doute le système mis en œuvre pour faciliter le partage des photos prises par des téléphones cellulaires décrits dans [2, 36]. Pour chaque image capturée, on s'intéresse à la création d'une liste, la plus pertinente possible, des éventuels destinataires de cette image. Cette liste est estimée à partir du traitement de trois types d'informations contextuelles : des données temporelles, des données spatiales et des informations de nature sociale. Les informations temporelles (date de capture) et spatiales (identification de la cellule radio) sont issues des équipements de réseaux mobiles alors que les informations sociales (identité du propriétaire, co-présence d'un autre utilisateur au même endroit, etc) sont issues de serveur web MMM2 qui collectent et traitent toutes les données disponibles. Par les mêmes auteurs [10], une adaptation pour passer du contexte au contenu a été étudiée en identifiant le nom du lieu où l'image a été prise. Ceci est rendu possible grâce aux images déjà annotées de la même cellule radio par d'autres utilisateurs et aux informations contextuelles extraites. Pour fusionner toutes ces informations, ce système utilise la combinaison linéaire avec des coefficients préfixés sans traiter les informations visuelles présentes dans l'image. Dans [37], on utilise des caméras équipées de GPS et d'une interface Bluetooth. L'équipement GPS permet d'identifier les coordonnées exactes de l'appareil photo contrairement aux coordonnées géographiques qui laissent une marge d'erreur de plusieurs mètres à plusieurs kilomètres selon la taille de la zone de couverture radio. L'interface Bluetooth permet d'identifier la présence simultanée d'autres utilisateurs [38]. Le système exploite également la dégradation ou la disparition des signaux GPS à l'intérieur de bâtiments pour distinguer les images prises en intérieur ou à l'extérieur. Les données sont fusionnées par un algorithme SFA (Sparse Factor Analysis) [39] qui est un processus de classification partiellement supervisée basé sur l'algorithme EM. L'utilisation d'un tel algorithme permet de gérer les informations manquantes ou inaccessibles comme c'est le cas pour des systèmes de capture d'images qui ne sont pas équipés d'interface Bluetooth.

1.2.3 Indexation par le contenu et le contexte

Malgré la cohérence qui peut exister entre le contexte et le contenu d'un document multimédia, on ne trouve, dans la littérature, que peu de travaux qui manipulent simultanément ces deux types d'informations qui peuvent être complémentaires. Dans [4], les auteurs n'exploitent que des informations extraites du contexte (les informations spatio-temporelles associées à l'image) pour identifier les identités des personnes figurant dans l'image à indexer. Autrement dit, **on remonte au contenu du document via son contexte**. Les auteurs rejettent même parfois l'utilisation d'outils d'indexation basés sur le contenu (détecteurs de visages par exemple) pour identifier le nombre de personnes présentes dans l'image. Ils justifient ce choix par le manque de fiabilité et de sensibilité de ces dispositifs qui apporteront de nombreuses fausses détections ou des non-détections, dégradant ainsi la performance du système.

Dans [37], des descripteurs du contenu sont utilisés pour identifier les lieux et les visages dans les images prises par des téléphones cellulaires. Ceux-ci permettent d'enrichir le formalisme introduit dans [10, 2, 36] qui ne traitent que des informations contextuelles. Pour identifier les lieux, le système se base sur les informations contextuelles issues du système MMM2 [2]. L'analyse de l'image est effectuée par la méthode CVA (Cognitive Visual Attention) et la fusion se fait par une combinaison linéaire pondérée dont l'ajustement est réalisé par une base d'apprentissage. Pour identifier les visages, plusieurs algorithmes de reconnaissance de visages sont mis en compétition. La fusion entre le contenu et le contexte du document se base sur le même algorithme SFA [39] décrit précédemment. Les expérimentations menées dans ce travail montrent que la fusion entre le contexte et le contenu améliore remarquablement la précision du système.

Le système *Acoi* [40] a été conçu pour indexer et rechercher plusieurs types de documents multimédia notamment les images et les pages WEB. Il utilise le contexte (expéditeurs, destinataires, URL, titre, type de documents, ...) et le contenu (texte, analyse d'image, détecteurs de visage, ...) via des détecteurs spéciaux. Ces détecteurs se divisent en deux catégories :

- des détecteurs capables de calculer des caractéristiques bas niveau (couleur, texture, ...);
- des détecteurs capables d'établir des liens entre les caractéristiques bas niveau et les concepts haut niveau.

Selon les auteurs, l'un des points forts du système est son adaptativité permettant d'intégrer de nouveaux descripteurs sans aucune difficulté.

Le système de *Xerox* décrit dans [7] a comme objectif l'annotation de segments vidéos acquis lors de réunions ou de conférences à partir d'un ensemble de mots-clés constituant les concepts haut niveau. Le choix des mots-clés restreint néanmoins les capacités du système. D'un point de vue du contenu, la segmentation de la vidéo et de l'audio se fait à l'aide de méthodes dérivées

des chaînes de Markov cachées (HMM) dont l'apprentissage est basé sur un algorithme de ré-estimation itératif. La transcription textuelle extraite à partir de la vidéo est traitée pour faciliter l'apprentissage audio par l'outil de reconnaissance de la parole. Dans ce système, les informations contextuelles sont extraites des annotations manuelles prises lors des réunions ou après la visualisation de la vidéo mais ne permettent pas d'identifier le début ou la fin des segments vidéos.

1.3 De la nécessité d'une architecture de fusion hybride

Dans le domaine du multimédia, la plupart des outils disponibles à ce jour sont habituellement destinés à une seule modalité. Par exemple, les logiciels de recherche d'informations sur le WEB n'utilisent pas l'image, les systèmes de recherche d'images ne mélangent pas les descriptions visuelles et textuelles, l'analyse de vidéo est habituellement faite séparément sur le son et l'image. L'une des raisons de cet écueil est que les outils d'analyse du contenu de ces media sont dérivés de champs scientifiques différents et parfois très séparés (analyse de la parole et traitement du signal, analyse de photos et traitement d'images). Ces approches ont prouvé leur inaptitude à intégrer les informations provenant de plusieurs modalités. Pourtant, une indexation efficace nécessite une approche multimodale dans laquelle la collaboration de différentes modalités devrait être utilisée. L'intégration des trois modalités est courante dans le cadre de l'interprétation sémantique des informations de la vidéo par des indexeurs humains. Plusieurs études ont montré que les performances des analyses et de la compréhension automatique du multimédia (notamment en terme de robustesse) peuvent être grandement améliorées par la combinaison de différentes modalités.

Au travers des travaux mentionnés dans les paragraphes précédents, nous avons pu constater que la majorité des données disponibles pour l'indexation multimédia sont issues d'outils mis au point dans le domaine du traitement du signal et des images ainsi que dans le domaine de la reconnaissance des formes et de l'intelligence artificielle [41]. Ces données extraites de différents moteurs d'indexation peuvent être vues comme des sources d'information qu'il s'agit de fusionner dans le but :

- d'améliorer la qualité de l'indexation en utilisant la complémentarité des sources d'informations disponibles. Nous avons évoqué plus haut l'augmentation de la précision de tels systèmes qui combinent les différents médias comparativement aux systèmes qui n'exploitent qu'un seul média.
- d'augmenter la précision des réponses en quantifiant la fiabilité de chacune des sources. En effet, une information non fiable est capable de dégrader l'apport des autres sources

d'informations et peut diminuer les performances du système.

- de lever les ambiguïtés pouvant intervenir entre les différentes sources.

Une telle vue de la problématique de fusion nécessite d'être capable de représenter les imprécisions et les incertitudes de chacune des informations dans un même formalisme. Si tel est le cas, la fusion pour les documents multimédias pourrait être rendue possible sous deux optiques différentes :

Fusion multimodale qui consiste à combiner les informations issues des différentes sources de media (image, audio, texte) et toutes autres informations associées aux documents.

Fusion contenu/contexte qui consiste à fusionner les informations issues du contenu avec les informations contextuelles. Cette fusion peut être multimodale dans la mesure où elle peut concerner des informations de différents média.

Dans le cadre de ce travail de thèse, nous nous sommes focalisés sur un système capable de fusionner différentes sources d'informations pour l'aide à l'indexation de documents multimédia. Nous avons insisté sur la nécessité de posséder un formalisme mathématique capable de représenter l'information disponible le plus fidèlement possible. Dans le chapitre suivant, quelques cadres théoriques classiques utilisés en fusion d'informations seront présentés. Nous détaillerons plus particulièrement le cadre basé sur les fonctions de croyance qui permet de généraliser plusieurs approches. Nous illustrerons l'application d'un tel formalisme dans le cadre d'applications à l'indexation de documents multimédia.

Chapitre 2

Fusion d'informations pour l'indexation multimédia

La fusion d'informations est connue de nos jours comme un des outils de choix pour le traitement et l'amélioration de la qualité de la décision dans plusieurs domaines [12] comme la finance, la médecine, la reconnaissance des formes, etc. Dans notre problématique d'indexation multimédia, l'utilisation d'un tel outil devient encore plus déterminant et primordial à cause de la quantité et de la nature différente des informations traitées (visuelles, vocales, textuelles) et surtout à cause de leurs imperfections. On entend par imperfection de l'information des notions liées à son imprécision, son incertitude associée ou parfois même à son manque (incomplétude). De plus, la fusion doit permettre de prendre en compte les types de relations existant entre les différentes sources (discordances, conflit, complémentarité, raffinement, etc). Elle se doit également de prendre en considération la façon dont la source 'observe' le document visé (au sens de son contenu et/ou de son contexte).

Dans ce chapitre, nous nous attachons à donner un aperçu des différentes techniques de fusion d'informations dans un contexte d'utilisation en indexation de documents multimédia. La théorie des fonctions de croyance comme un outil de fusion d'informations est ensuite détaillée d'un point de vue mathématique. Enfin, deux applications utilisant les fonctions de croyance pour l'indexation sont présentées et permettent de mettre en évidence les avantages de ce formalisme.

2.1 Méthodes classiques de fusion d'informations

Dans les systèmes d'indexation et de recherche de documents multimédias, les techniques de fusion d'informations sont utilisées pour prendre en compte la multimodalité. Généralement, les méthodes employées sont basées sur des outils probabilistes [42, 43], des techniques de combi-

raison linéaire [31, 21] ou encore des outils de classification (*clustering* en anglais) [37, 8]. Avant de détailler un certain nombre de ces outils de modélisation et de fusion d'informations, nous posons la problématique de l'indexation et définissons quelques variables qui seront utilisées dans la suite de ce chapitre.

Soit Ω un référentiel ou un ensemble fini composé de q hypothèses qui sont toutes exclusives et exhaustives du problème posé :

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_q\}. \quad (2.1)$$

Dans la pratique, Ω représente l'ensemble des concepts auxquels un document peut appartenir. Ainsi, pour un problème de catégorisation, cet ensemble peut être constitué des différentes catégories à discerner. Soient S_1, S_2, \dots, S_n un ensemble de n sources d'informations. Dans notre problématique d'indexation, on va supposer que le document à indexer, que nous noterons $x \subseteq X$ où X désigne la base des documents à indexer, est observé¹ par chacune des sources d'information S_j (avec $j = \{1, 2, \dots, n\}$). Ainsi, le but d'un système d'indexation consiste à associer au document x le concept $\omega \in \Omega$ auquel il appartient. La caractérisation de x par S_j conduit à un vecteur descriptif, noté $f_j(x)$ de dimension finie d_j , défini par :

$$\begin{aligned} f_j : X &\rightarrow \mathbb{R}^{d_j} \\ x &\rightarrow f_j(x). \end{aligned}$$

La nature du vecteur f_j peut être diverse selon la source d'information S_j considérée. Si on observe le contenu d'une image x , le vecteur $f_j(x)$ peut être constitué d'un ensemble d'observations sur les couleurs ou les textures présentes dans l'image. Le vecteur peut également se limiter à une valeur de score quantifiant l'appartenance de x au concept $\omega \in \Omega$.

Le but des outils de modélisation et de fusion qui seront présentés dans la suite de ce chapitre consiste à améliorer la qualité de l'indexation (quantifier de manière plus pertinente le concept $\omega \in \Omega$ auquel x appartient) dans un contexte multimodal. Nous présenterons des techniques classiques comme le vote majoritaire, la combinaison linéaire et la théorie des probabilités mais aussi des techniques moins classiques dérivées du domaine de l'Intelligence Artificielle.

2.1.1 Système de vote majoritaire

Le vote majoritaire est un système qui prend uniquement en Considération l'ordre des hypothèses solutions attribuées par les sources. Ces listes ordonnées sont ensuite transformées en scores pour chacune des hypothèses lors de l'étape de fusion. Un des algorithmes les plus connus

¹au sens de son contenu et de son contexte.

est l'algorithme de *Borda* qui a été proposé par *Jean Charles de Borda* en 1770 [44]. Dans cet algorithme, chaque source S_j attribue à chaque hypothèse ω un nombre de points en fonction de son rang dans la liste ordonnée. Pour un exemple à deux sources S_1 et S_2 , une hypothèse ω reçoit $(q - i)$ points si elle est rangée au $i^{\text{ème}}$ rang par la source S_1 et $(q - j)$ points si elle est rangée au $j^{\text{ème}}$ rang par la source S_2 . Ceci lui confère un score de $(2q - (i + j))$ lors de l'étape de fusion qui consiste à sommer les points attribués par chacune des sources. Avec q le nombre d'hypothèses et, $i, j \in \{1, 2, \dots, q\}$.

Pour l'indexation de documents multimédia, cette approche a été utilisée dans [25] pour fusionner les deux modalités : l'image et le texte. Les avantages de cette technique de fusion sont sa facilité de mise en œuvre, son adaptabilité aux nombres de sources et surtout sa rapidité. Néanmoins, travaillant sur des listes, elle ne permet pas de prendre en compte la pertinence de chacune des sources. La méthode n'est pas non plus capable de gérer les conflits qui peuvent intervenir entre sources, ce qui rend son application un peu limitée.

2.1.2 Combinaison linéaire

La combinaison linéaire est une technique qui s'attache à combiner les scores finaux attribués aux différentes hypothèses de Ω par chacune des sources S_j . Pour un document inconnu x et une hypothèse ω , chaque source S_j calcule un vecteur descriptif $f_j(x)$ qui est transformé en un score défini par :

$$s_j : \mathbb{R}^{d_j} \times \Omega \rightarrow \mathbb{R}$$

$$(f_j(x), \omega) \rightarrow s_j(f_j(x), \omega).$$

A l'étape de fusion, le score final $s(x, \omega)$ qui sera attribué à l'hypothèse ω pour le document x sera calculé à l'aide d'une combinaison linéaire par l'équation suivante :

$$s(x, \omega) = \sum_{j=1}^n \alpha_j s_j(f_j(x), \omega). \quad (2.2)$$

où α_j représente un coefficient de pondération relatif à la source S_j . Les scores peuvent être calculés et optimisés à l'aide de systèmes d'apprentissage ou être mis à jour après chaque document indexé par l'utilisateur [21, 20]. Cette technique de fusion, simple et rapide, est utilisée dans plusieurs systèmes d'indexation multimédia [21, 31, 4]. Cependant, elle reste une méthode critiquable [29] car elle est très tolérante au bruit et ne gère pas de manière explicite les dépendances entre les sources.

2.1.3 Théorie des probabilités

La théorie des probabilités est sans doute la théorie la plus classique et la plus ancienne pour traiter l'incertitude et pour fusionner les informations. Soit A un élément quelconque de Ω . La probabilité que l'événement A soit vrai (ou pertinent) peut être quantifiée par la quantité $p(A)$. On définit pour tout A de Ω une fonction p , appelée distribution de probabilité par :

$$\begin{aligned} p : \Omega &\rightarrow [0, 1] \\ A &\rightarrow p(A). \end{aligned}$$

vérifiant les axiomes suivants :

$$\begin{aligned} p(\Omega) &= 1 \\ p(\emptyset) &= 0 \\ p(A \cup B) &= p(A) + p(B) \text{ si } B \subseteq \Omega \text{ et } A \cap B = \emptyset. \end{aligned}$$

Le couple (Ω, p) est appelé espace probabilisé. Ce dernier axiome permet de mettre en évidence que la mesure de probabilité est une mesure additive et donc la relation suivante est vérifiée :

$$\sum_{i=1}^q p(\omega_i) = 1. \quad (2.3)$$

L'un des principaux outils de la théorie des probabilités est sans aucun doute le théorème de Bayes qui permet la fusion d'informations entre sources. Pour une source d'information donnée, on suppose que chaque concept d'indexation ω_i peut être modélisé par une fonction de vraisemblance définie de Ω dans $[0, +\infty]$ notée $L[v](\omega_i)$. Ainsi, pour une valeur v_1 mesurée par la source S_1 pour le document x , on définit la probabilité conditionnelle que l'on ait la mesure v_1 si ω_i est vrai par $p[\omega_i](v_1) = L[v_1](\omega_i)$. La probabilité *a posteriori* du concept ω_i ayant mesuré les valeurs v_1 par la source S_1 et v_2 par la source S_2 se calcule par la formule suivante :

$$p[v_1, v_2](\omega_i) = \frac{p[\omega_i](v_1) p[\omega_i](v_2)}{\sum_{\omega_j \in \Omega} p[\omega_j](v_1) p[\omega_j](v_2)}. \quad (2.4)$$

qui permet de fusionner les informations apportées par les deux sources.

La modélisation probabiliste n'apparaît pas toujours suffisante ni très bien adaptée pour représenter l'incertitude et modéliser les méconnaissances [45, 12]. De plus cette théorie ne permet pas la gestion des conflits entre les sources puisque aucune autre information que la probabilité *a posteriori* n'est disponible.

2.1.4 Théorie des possibilités

La théorie des possibilités est issue des travaux sur les sous-ensembles flous de Zadeh [46, 47, 48] qui a proposé la notion d'appartenance partielle d'un individu à un concept. *Dubois* et *Prade* ont mis en place la théorie des possibilités [49] qui permet de modéliser l'imprécision et l'incertitude associées à un élément d'information. Pour ce faire, la théorie fait appel à deux mesures, une mesure de possibilité $\Pi(\cdot)$ et une mesure de nécessité $N(\cdot)$ toutes deux définies de Ω vers $[0, 1]$. La mesure $\Pi(\omega)$ quantifie le fait que le concept ω soit possible tandis que $N(\omega)$ quantifie le degré de certitude associé au concept. Par le biais des relations liant la possibilité et la nécessité, l'incertitude et l'imprécision liés à une information sont modélisées de façon à ce qu'un concept puisse être possible sans être certain.

Si on possède deux distributions de possibilités Π_1 et Π_2 associées à deux sources d'information S_1 et S_2 , alors il existe de nombreux opérateurs qui permettent de combiner ces deux distributions :

- les T -normes qui s'attachent à combiner des sources concordantes ;
- les T -conormes qui combinent des sources discordantes ;
- les opérateurs adaptatifs qui permettent de combiner les sources en fonction de leurs degrés de conflit respectifs.

Le résultat conduit à une distribution de possibilité qui résume les informations disponibles dans S_1 et S_2 . En ce qui concerne la décision, elle est généralement prise par des critères qui maximisent la possibilité (ou la nécessité) sur l'ensemble des concepts.

L'un des avantages de cette technique de combinaison est sa facilité de mise en œuvre et la gestion du conflit entre sources. Elle permet ainsi, en représentant l'imprécision et l'incertitude par deux fonctions, de traiter des problèmes non modélisables par des approches probabilistes classiques [49]. Nous verrons dans les paragraphes suivants que la théorie des possibilités et les méthodes probabilistes restent des cas particuliers de la théorie des fonctions de croyance. Dans le contexte de l'indexation multimédia, cette technique a été utilisée pour l'indexation d'images basées sur le contenu [50, 51].

2.1.5 Fusion par des techniques d'apprentissage

Parmi les techniques classiques d'apprentissage statistique, on a souvent tendance à distinguer les méthodes de discrimination des méthodes de classification. Dans un contexte d'indexation, les algorithmes de discrimination cherchent à estimer une fonction $\omega_i = \mathcal{F}(f_j)$ (où f_j est le vecteur descriptif associé à la source S_j et ω_i le concept d'indexation) à partir d'un historique de mesures

(f_j, ω_i) . Les méthodes de classification consistent quant à elles à regrouper les données f_j entre elles pour former des agglomérats dans l'espace multidimensionnel auxquels seront associés des concepts. Pour fusionner les données dans un contexte multimodal, la technique consiste dans un premier temps à regrouper tous les vecteurs descriptifs en un seul vecteur de grande dimension (voir figure 2.1). Le vecteur ainsi construit est utilisé comme données d'entrées d'un algorithme

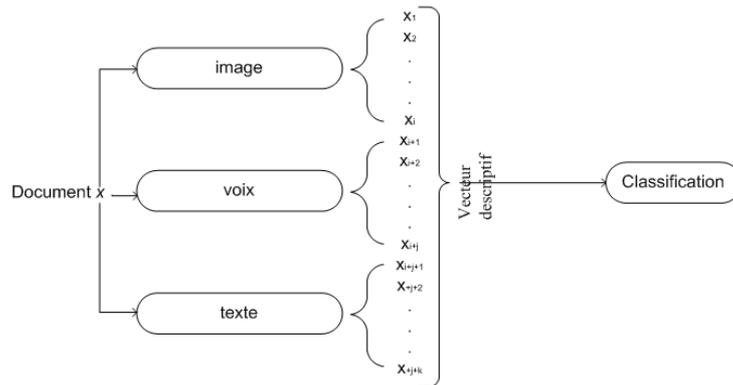


Figure 2.1 : Fusion de trois sources image, voix et texte pour la classification.

de discrimination du style séparateur à vaste marge (SVM : Support Vector Machines) ou de classification comme l'algorithme EM (Expectation-Maximization). Des applications de ce type de techniques sont détaillées dans les paragraphes suivants.

Méthode SVM

Les séparateurs à vaste marge [16] ont été introduits en 1992 pour gérer les problèmes de discrimination ou de régression non linéaire. La technique se base sur deux points fondamentaux :

- La transformation non linéaire des entrées, de l'espace d'entrée vers un espace de redistribution de grande dimension (espace de *Hilbert*) muni d'un produit scalaire.
- Le choix d'une séparation linéaire optimale dans cet espace de grande dimension. Au lieu de chercher un hyperplan dans l'espace des entrées, on passe d'abord dans un espace de représentations intermédiaires qui facilite la recherche d'une séparation linéaire.

Les SVM sont appliqués avec une grande efficacité dans les systèmes de reconnaissance des formes et à l'indexation multimédia [8]. Un des points faibles de cette technique est le fait qu'elle ne supporte pas des données incomplètes.

Méthode EM

L'algorithme EM (Expectation-Maximization) qui a été proposé par *Dempster* en 1977 est un algorithme itératif qui permet l'estimation des paramètres d'une loi de probabilité [52]. La technique, qui s'appuie parfois sur des données manquantes, consiste à déterminer les paramètres de la loi (souvent un mélange de lois gaussiennes) par un estimateur du maximum de vraisemblance. Elle a été utilisée dans l'algorithme SFA [39] pour la fusion d'informations incomplètes. Elle a l'avantage de traiter des données manquantes comme c'est le cas pour les images prises par des caméraphones [37]. Malheureusement, là encore, l'hypothèse d'une loi de probabilité ne permet pas de gérer l'incertitude des événements. De plus, la méthode se contraint à un modèle rigide qui ne permet pas de s'adapter facilement à l'intégration de nouvelles sources d'information. En outre, elle considère que chacune des sources possède la même fiabilité.

2.2 Fonctions de croyance

Les travaux de *Dempster* sur les bornes inférieure et supérieure d'une famille de distributions de probabilités [45] ont permis à *Shafer* d'asseoir les bases de la théorie des fonctions de croyance [11]. Ce formalisme théorique peut revêtir plusieurs interprétations et appellations (théorie des fonctions de croyance ou théorie de Dempster-Shafer). *G. Shafer* a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines. L'utilité de ces fonctions, comme une alternative aux probabilités subjectives, a été démontrée plus tard de manière axiomatique par Ph. Smets [53] au travers du modèle des croyances transférables. Ce modèle fournit une interprétation non probabiliste cohérente et permet de clarifier le concept sous-jacent à la théorie. Dans cette section, nous abordons quelques éléments mathématiques liés aux fonctions de croyance [12]. Le concept de fonction de croyance est tout d'abord présenté, permettant ainsi au lecteur de comprendre comment sont représentées les informations dans une structure de croyance. Quelques outils associés à ce formalisme sont ensuite détaillés, y compris la combinaison et la prise de décision.

2.2.1 Fonctions de croyance

Soit un ensemble fini $\Omega = \{\omega_1, \omega_2, \dots, \omega_q\}$, généralement appelé cadre de discernement. Une fonction de croyance *bel* (nommée fonction de crédibilité) est une fonction de 2^Ω vers $[0,1]$ définie par :

$$bel(\emptyset) = 0. \tag{2.5}$$

et $\forall l \geq 1, \forall i = 1, \dots, l \ A_i \subseteq \Omega$ on a :

$$bel\left(\bigcup_{i=1,\dots,l} A_i\right) \geq \sum_{I \subseteq \{1,\dots,l\}, I \neq \emptyset} (-1)^{|I|+1} bel\left(\bigcap_{i=1,\dots,l} A_i\right). \quad (2.6)$$

Le cas particulier où $l = 2$ et $A_1 \cap A_2 = \emptyset$ illustre le fait que la croyance attribuée à l'union de deux sous-ensembles disjoints de Ω est supérieure ou égale à la somme des croyances attribuées à chacun des sous-ensembles. Si l'inégalité est transformée en égalité dans l'équation (2.6), la fonction résultante bel , appelée généralement fonction de croyance bayésienne, devient une fonction de probabilité. Ainsi, le cadre classique des probabilités n'est qu'un cas particulier des fonctions de croyance qui peut être interprétée comme une généralisation de la théorie des probabilités pour la modélisation des imprécisions et incertitudes.

Une fonction de croyance peut également être définie mathématiquement par une fonction (ou allocation) de masse, notée m définie de 2^Ω dans $[0, 1]$, qui vérifie :

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (2.7)$$

De par sa définition, $m(\cdot)$ (contrairement à une distribution de probabilité) ne conserve donc pas la monotonie par rapport à l'inclusion :

$$A \subset B \not\Rightarrow m(A) < m(B). \quad (2.8)$$

ce qui fait qu'une fonction de croyance est suradditive. Chaque sous-ensemble $A \subseteq \Omega$ tel que $m(A) > 0$ est appelé élément focal de m . Ainsi, la masse $m(A)$ représente le degré de croyance attribué à la proposition A qui n'a pas pu, compte tenu de l'état de la connaissance, être affectée à un sous-ensemble plus spécifique que A . Une fonction telle que $m(\emptyset) = 0$ est dite normale. Dans le modèle des croyances transférables [53], la condition $\sum_{\emptyset \neq A \subseteq \Omega} m(A) = 1$ n'est pas supposée et $m(\emptyset) > 0$ est acceptée. Cette spécificité permet d'introduire la notion de monde ouvert en supposant que la croyance ne peut être placée sur un sous-ensemble de Ω . Dans ce contexte, l'ensemble vide \emptyset peut être interprété comme une hypothèse non clairement définie dans l'ensemble de discernement par opposition au monde fermé où l'ensemble des hypothèses est exhaustif. Étant donnée une fonction de masse m , on peut définir bel la fonction de crédibilité et pl la fonction de plausibilité respectivement comme suivant :

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B), \quad \forall A \subseteq \Omega, \quad (2.9)$$

et

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B), \quad \forall A \subseteq \Omega. \quad (2.10)$$

Les fonctions m , bel et pl représentent trois facettes de la même information.

Une des principales difficultés consiste à modéliser la connaissance disponible en initialisant de manière adéquate les fonctions de croyance. Supposons, par exemple, qu'il faille modéliser la connaissance concernant la culpabilité de trois suspects $\Omega' = \{Philippe, Arthur, Laurence\}$ exprimée par un témoin dans une affaire de meurtre. Sans connaissance particulière sur la culpabilité des trois suspects, la fonction de croyance est caractérisée par l'allocation de masse vide $m(\Omega') = 1$ qui représente l'ignorance totale. Supposons maintenant que le témoin a aperçu la chevelure blonde du meurtrier qui ne peut correspondre qu'à *Philippe* ou *Arthur*. L'allocation de masse prend alors la forme $m(\{Philippe, Arthur\}) = 1$. Si le témoin a vu le meurtrier, il peut formaliser sa connaissance sous la forme d'une fonction attribuant la totalité de la masse sur un des singletons de Ω' . L'ignorance partielle peut évidemment être codée elle aussi sous la forme d'une allocation de masse. Modéliser le plus fidèlement l'information disponible confère aux fonctions de croyance l'avantage de ne pas introduire d'information a priori contrairement aux fonctions de probabilités. En effet, le concept d'équi-probabilité n'est pas identique au concept d'ignorance totale mais constitue déjà une forme précise de connaissance.

2.2.2 Fusion d'informations

Supposons que nous disposions de deux allocations de masse m_1 et m_2 définies sur le même référentiel Ω . Ces deux fonctions peuvent être agrégées par un opérateur de combinaison conjonctif noté \odot . Le résultat de cette opération conduit à une fonction de croyance unique à laquelle correspond une fonction de masse, notée m_{\odot} , qui peut être définie par :

$$m_{\odot}(A) = (m_1 \odot m_2)(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C). \quad (2.11)$$

Cette règle conjonctive est parfois dénommée règle de combinaison de *Dempster* non normalisée. Si nécessaire, l'hypothèse de normalisation $m_{\odot}(\emptyset) = 0$ peut être retrouvée en divisant chaque masse par un coefficient adéquat. L'opérateur résultant, qui est connu sous le nom de règle de Dempster et noté \oplus , est défini par :

$$(m_1 \oplus m_2)(A) = \frac{(m_1 \odot m_2)(A)}{1 - m_{\odot}(\emptyset)} \quad \forall A \subseteq \Omega. \quad (2.12)$$

La quantité $m_{\odot}(\emptyset)$, qui représente le degré de conflit entre les fonctions m_1 et m_2 , peut être calculée par :

$$(m_1 \odot m_2)(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \quad (2.13)$$

L'utilisation de la règle de *Dempster* est possible si les fonctions de masse m_1 et m_2 ne sont pas en conflit total c'est-à-dire s'il existe deux éléments focaux B et C respectivement de m_1 et m_2 tels que $B \cap C \neq \emptyset$. Cette règle permet ainsi de combiner des informations incertaines extraites sous forme de fonctions de croyance.

2.2.3 Quelques outils spécifiques

Dans le modèle des croyances transférables [53], la fonction de masse a une interprétation naturelle puisque la masse $m(A)$ représente le degré de croyance attribué à la proposition $A \subseteq \Omega$ qui n'a pas pu être assignée à un sous-ensemble plus spécifique. Si maintenant, on apprend que la croyance allouée au sous-ensemble $B \subseteq \Omega$ est totale alors la masse $m(A)$ initialement allouée à A est transférée au sous-ensemble $A \cap B$ d'où le nom du modèle introduit par *Ph. Smets* dans [53]. La fonction de masse résultante est celle obtenue par l'application de la règle dite de conditionnement qui n'est qu'un cas particulier de la règle de combinaison de *Dempster*. Ainsi, le résultat de la combinaison d'une fonction de masse m conditionnellement à B produit une fonction de croyance, notée $bel(.|B)$ qui peut se calculer $\forall A \subseteq \Omega$ par :

$$bel(A|B) = bel(A \cup \overline{B}) - bel(\overline{B}). \quad (2.14)$$

2.2.4 Affaiblissement

Lorsque la source d'information dont la fonction de croyance est extraite n'est pas totalement fiable, il est possible d'introduire une opération d'affaiblissement. Dans ce cas, un coefficient α qui représente une sorte de méta-connaissance concernant la fiabilité de la source, permet de transférer une partie de la croyance vers l'ensemble Ω . Ainsi, une fonction de masse affaiblie, notée m_α , peut se déduire de m par :

$$m_\alpha(A) = \alpha m(A) \quad \forall A \subsetneq \Omega \quad (2.15)$$

$$m_\alpha(\Omega) = 1 - \alpha + \alpha m(\Omega). \quad (2.16)$$

Ainsi, une fonction de croyance est capable de représenter plusieurs types de connaissance et constitue par ce biais un cadre riche et flexible pour la représentation d'informations incertaines.

Ce coefficient α peut être calculé à l'aide d'un algorithme d'apprentissage statistique comme dans [54].

2.2.5 Éléments focaux : Noyau

Les éléments focaux sont les éléments A de 2^Ω dont la masse $m(A)$ est non nulle et, le noyau $N[m]$ représente l'ensemble de ces éléments focaux :

$$N[m] = \{A/A \subseteq \Omega \text{ et } m(A) > 0\}. \quad (2.17)$$

2.2.6 Prise de décision

Au niveau pignistique, une fonction de croyance unique, sorte de résumé exhaustif de l'information disponible au niveau crédal, est utilisée pour la prise de décision. En basant son raisonnement sur des arguments de rationalité développés dans le modèle des croyances transférables, Ph. Smets propose de transformer une fonction de masse m en une fonction de probabilité $BetP_m$ définie sur Ω (appelée fonction de probabilité pignistique) qui se formalise pour tout ω_k par :

$$BetP_m(\omega_k) = \frac{1}{1 - m(\emptyset)} \sum_{A \in 2^\Omega / \omega_k \in A} \frac{m(A)}{|A|}. \quad (2.18)$$

où $|A|$ représente la cardinalité du sous-ensemble $A \subseteq \Omega$ et $k \in \{1, 2, \dots, q\}$. Dans cette transformation, la masse de croyance $m(A)$ est uniformément distribuée parmi les éléments de A .

2.2.7 Notation

On suppose que la connaissance détenue par une source S (corpus d'évidence) sur la valeur ω_0 prise par une variable d'intérêt notée x à valeurs dans Ω peut être quantifiée par une fonction de croyance $m_\alpha^\Omega[S, x]$ définie de 2^Ω dans $[0, 1]$ comme suivant :

$$m_\alpha^\Omega[S, x](A). \quad (2.19)$$

Également, la fonction de crédibilité :

$$bel_\alpha^\Omega[S, x](A). \quad (2.20)$$

et la fonction de plausibilité :

$$pl_\alpha^\Omega[S, x](A). \quad (2.21)$$

Où α est le coefficient de fiabilité et Ω le cadre de discernement.

2.3 Applications des fonctions de croyance à l'indexation

De nombreux travaux s'attachent à comparer le cadre formel des fonctions de croyance à plusieurs autres méthodes de fusion notamment aux méthodes probabilistes (dans le cadre Bayésien), à la théorie des possibilités et à la combinaison linéaire. Dans [55], les auteurs mettent en compétition ces trois méthodes de fusion qui sont appliquées dans un problème de vérification de signature. Les expérimentations menées dans ces travaux ont montré l'efficacité de la combinaison de *Dempster* par rapport aux autres méthodes. Pour la fusion d'informations multimodales

(audio et vidéo), les fonctions de croyance ont été jugées plus performantes [56] que la combinaison linéaire, les méthodes Bayésiennes et la théorie des possibilités. Ce paragraphe détaille deux applications particulières des fonctions de croyance à l'indexation. La première application décrit un système appelé **Diogenes** qui consiste à rechercher des images sur le WEB en fusionnant le texte avec l'image. La deuxième application consiste à rechercher des images par leurs contenus via une boucle de pertinence.

2.3.1 Système Diogenes

Le système **Diogenes** [57, 58] a été mis en place pour chercher sur le WEB des photos de personnes à partir de leurs noms. Il requiert le nom de la personne recherchée et retourne une liste d'images de cette personne triées selon leurs degrés de similarité vis-à-vis du point de vue du système. La figure 2.2 montre le résultat d'une requête sur *Abraham Lincoln*. Le système

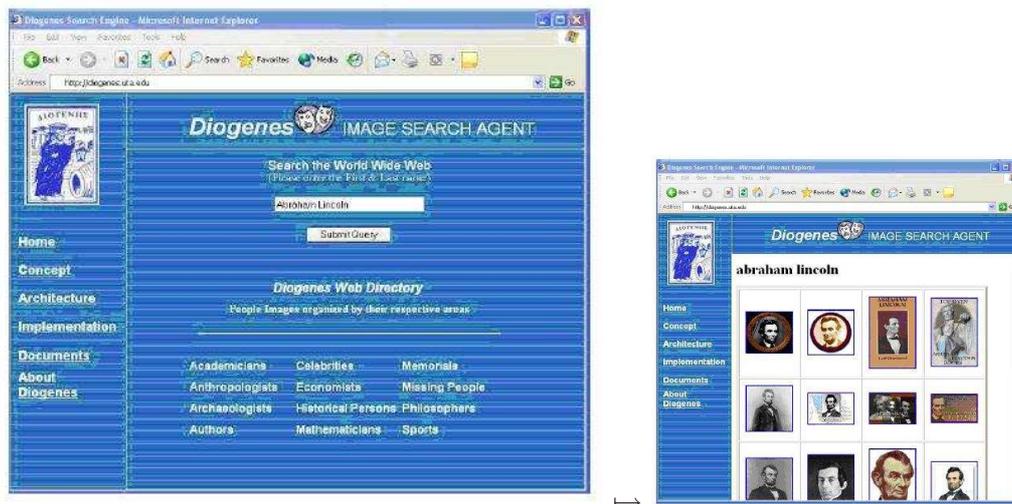


Figure 2.2 : Page requête et page des résultats du système **Diogenes**.

utilise deux modules principaux (outils d'indexation) pour analyser les pages WEB. Le premier est l'analyseur de contenu de l'image *FR* (Face Recognizer) qui se compose lui-même de deux modules : le détecteur de visage et le reconnaisseur de visage. Le deuxième module est l'analyseur de texte *TA* (Text Analyzer) qui analyse les textes HTML.

Architecture du système Diogenes

L'architecture du système **Diogenes** est illustrée à la figure 2.3. Il se compose des éléments suivants :

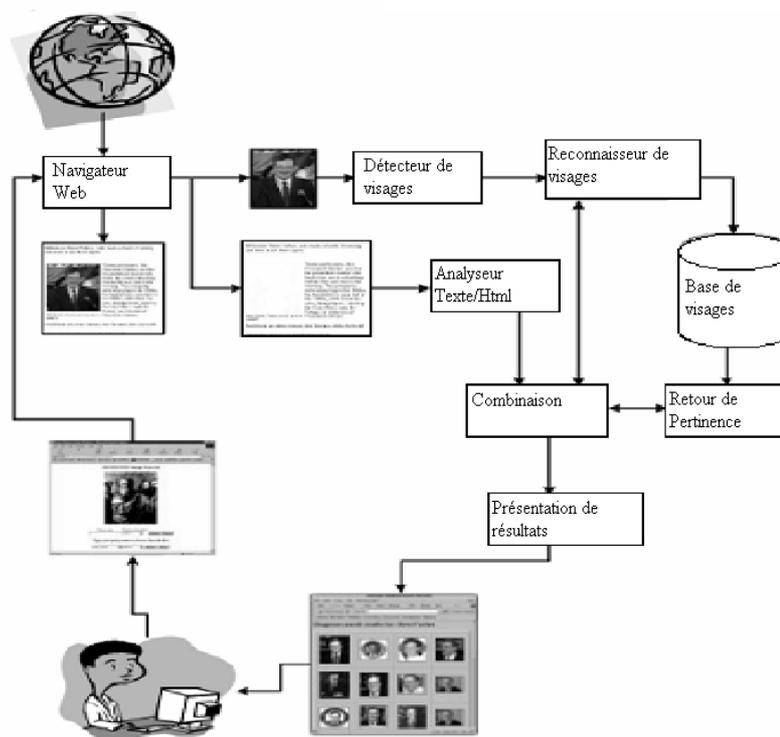


Figure 2.3 : Architecture du système **Diogenes**.

1. une interface homme/machine où l'utilisateur saisit sa requête. Le système détermine tous les liens électroniques (URL) vers les sites WEB qui sont en rapport avec le nom saisi (requête). Ce module obtient les résultats à travers les moteurs de recherche connus comme Google, Altavista, ... ;
2. un navigateur WEB qui visite chaque URL en séparant les images et les textes avant de les sauvegarder temporairement pour l'analyse ;
3. un détecteur de visage (Face Detector-*FD*-) dans les images qui est basé sur l'utilisation de réseaux de neurones. Il sert à vérifier l'existence d'un visage humain dans l'image en déterminant les coordonnées du rectangle qui englobe ce visage ;
4. un système de reconnaissance de visage (Face Recognition-*FR*-) qui compare le visage détecté aux visages qui se trouvent dans la base des visages connus ;
5. un analyseur de texte HTML (Text/html Analyzer-*TA*-) qui a pour but d'analyser les textes de la page WEB (titres, commentaires associés aux images, ...) afin de valoriser la présence du nom de la personne requête dans cette page ;
6. une boucle de pertinence (Feedback Processor) qui alimente la base de visages au moment de la requête ;

7. et d'un module de fusion qui combine les données fournies par les deux modules analyseur de texte TA et reconnaisseur de visage FR .

Sources d'Information

Les fonctions de croyance sont utilisées dans le système pour modéliser les méconnaissances et combiner les différentes sources d'information à l'aide de la règle de combinaison de *Dempster* dans un cadre de discernement $\Psi = \{Per, \overline{Per}\}$ où Per désigne le fait que l'image à indexer est pertinente vis-à-vis de la requête. Nous détaillons les deux sources d'information utilisées dans le cadre de cette application : l'analyse de l'image et l'analyseur de texte HTML.

L'analyse de l'image nous fournit deux informations : la présence d'un visage dans l'image, et le nom de la personne. La première information est donnée par le détecteur de visage (FD) et la deuxième est donnée par le reconnaisseur de visage (FR) si toutefois la présence d'un visage a été détectée. Seule la croyance associée au fait qu'une personne P est présente dans l'image I déduite du reconnaisseur de visage est quantifiée. Ainsi, à partir d'une requête sur une personne P , la masse allouée au fait qu'il est pertinent, que P soit dans l'image I est calculée par :

$$m^{\Psi}[FR, I](Per) = C_{FR} * d_{FR}(I, P).$$

où $d_{FR}(I, P)$ représente une mesure de similarité entre la personne P et le visage qui se trouve dans l'image I . Dans cette dernière équation, C_{FR} est une constante multiplicatrice calculée de cette façon :

$$C_{FR} = \frac{1 - m^{\Psi}[FR, I](\Psi)}{\sum_P d_{FR}(I, P)}.$$

où $m^{\Psi}[FR, I](\Psi)$ représente l'incertitude globale qui est ajustée à une valeur constante calculée à partir de la base des visages connus. Le système estime que si l'image contient plusieurs visages, l'identification de l'un d'entre eux est suffisant pour indexer le document. Ainsi, à chaque image est associée une fonction de masse $m^{\Psi}[FR, I](.)$ qui quantifie la croyance pour que l'image I réponde au concept d'indexation (personne P). Ceci constitue une première source d'information.

La deuxième source d'information est entièrement élaborée à partir de l'analyseur de texte HTML (TA). Ce dernier outil valorise l'association existante entre le nom et/ou prénom de la personne et les textes de la page WEB. De façon à évaluer la pertinence de l'image via ce moteur d'indexation, le système calcule des scores (proportionnels aux fréquences d'apparition) associés à différents types de texte extraits de la page où se trouve l'image :

- s_{freq} : textes de la page web où se trouve l'image ;
- s_{tag} : tags html communs entre l'image et le nom ;
- s_{path} : URL de l'image ;

– s_{alt} : texte alternatif à l'image.

Afin de mesurer le degré d'association entre un nom donné et l'image, ces différents scores sont combinés de façon linéaire en tenant compte de tous les mots présents dans la requête (nom et prénom). Le degré d'association entre un mot W et l'image I extraite d'une page web donnée est donc calculé par :

$$d(W, I) = (w_{freq} \cdot s_{freq}) + (w_{tag} \cdot s_{tag}) + (w_{path} \cdot s_{path}) + (w_{alt} \cdot s_{alt}).$$

où w_{freq} , w_{tag} , w_{path} , et w_{alt} représentent les coefficients de pondération qui sont optimisés par le système via une procédure d'apprentissage. Pour une personne P , le calcul est réalisé par :

$$d(P, I) = d(W, I) = 0.25 \times d(I, W_1) + 0.75 \times d(I, W_2).$$

où W_1 et W_2 représentent respectivement le prénom et la nom de famille de la personne P . La fonction de masse est calculée de la façon suivante :

$$\begin{aligned} m^\Psi[TA, I](Per) &= C_{TA} \cdot \frac{d(I, P)}{d_{max}}; \\ m^\Psi[TA, I](\Psi) &\propto \frac{1}{d_{max}}. \end{aligned}$$

avec d_{max} une constante globale de normalisation et C_{TA} une constante multiplicatrice.

Combinaison des sources d'information

Pour combiner les deux sources d'informations FR et TA , le système **Diogenes** utilise la règle de combinaison de *Dempster* pour agréger les fonctions de masse $m^\Psi[FR, I](\cdot)$ et $m^\Psi[TA, I](\cdot)$ ce qui conduit à la fonction de masse $m_{\oplus}^\Psi[\cdot, I]$ définie par :

$$m_{\oplus}^\Psi[\cdot, I] = m^\Psi[FR \oplus TA, I] = m^\Psi[FR, I] \oplus m^\Psi[TA, I].$$

Puisque tous les éléments focaux sont des singletons ou l'ensemble Ψ , la formule de combinaison pour la personne P peut se calculer facilement. Cette opération est réalisée pour toutes les images de la base. Le résultat attendu du système est la liste d'images triées selon leurs degrés d'association $m_{\oplus}^\Psi[\cdot, I](Per)$ entre la requête P et l'image I .

Construction de la base des visages

Lorsqu'une requête est soumise au système, la base des visages peut ne contenir aucun visage connu de la personne P . Pour construire une base de visages pour cette personne P , le système suit une procédure qui peut être illustrée par la figure 2.4. Les différentes étapes de cette procédure sont :

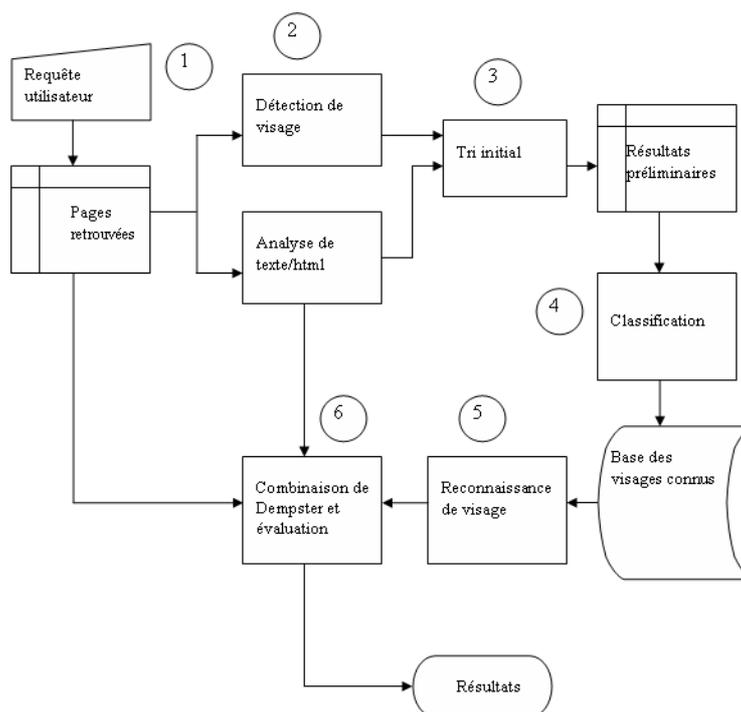


Figure 2.4 : Etapes de la boucle de pertinence du système **Diogenes**.

1. requête utilisateur : le nom de la personne P ;
2. liste des pages WEB retournées par des moteurs pour élimination des pages qui ne contiennent pas de visages par le FD ;
3. tri initial basé sur l'analyseur de texte puis sélection des images qui se trouvent en tête de liste comme résultats préliminaires ;
4. classification hiérarchique des images, la base d'apprentissage contenant les visages de la personne P correspondant au cluster le plus grand ;
5. reconnaissance des visages dans les images de la base d'apprentissage ;
6. combinaison des informations entre l'analyseur de texte et le reconnaisseur de visages.

Cette stratégie permet de constituer une base initiale de visages qui permet au système de traiter la même requête en ne conservant aucune des informations obtenues dans la phase d'apprentissage. Seules les images ayant été retrouvées dans l'étape de construction de la base de visages connus sont utilisées dans la phase de recherche.

Résultats

Les expérimentations menées sur **Diogenes** permettent de mettre en évidence une précision remarquable du système par rapport au moteurs de recherche classique comme Google et Altavista. Ces expérimentations montrent également l'utilité des fonctions de croyance pour la fusion d'informations. En effet, les auteurs ont comparé en terme de précision de recherche les résultats obtenus par l'analyseur de texte *TA* uniquement, par le reconnaiseur de visages *FR*, par la combinaison linéaire entre *TA* et *FR* et par la combinaison de *Dempster*. A partir de ces résultats, on constate que la méthode de fusion basée sur les fonctions de croyance donne les meilleures précisions.

2.3.2 Boucle de pertinence automatique

Dans ces paragraphes, nous présentons une application des fonctions de croyance à la mise en œuvre d'une boucle de pertinence automatique utilisée pour la recherche d'images par le contenu [59]. La phase de recherche de tels systèmes permet de trier par ordre de pertinence les images de la base dont la mesure de similarité est la plus proche de l'image requête. Dans le cadre de cette application, nous avons exploité cette observation afin d'accumuler plus d'informations sur le concept d'indexation associé à la requête.

Soit I_0 une image requête qui sera considérée comme une source d'information S_0 à laquelle nous attribuerons la fiabilité α_0 . En effet, cette source d'information va permettre le calcul des similarités entre l'image requête et les images de la base. Dans la phase de présentation des résultats, le système retourne une série d'images $\{I_1, I_2, \dots, I_n\}$. Nous noterons $\{S_0, S_1, \dots, S_n\}$ l'ensemble des sources d'informations auxquelles seront associées des fiabilités $\{\alpha_0, \alpha_1, \dots, \alpha_n\}$. L'idée consiste à exploiter un ensemble de sources d'information $\{S_0, S_1, \dots, S_i\}$ pour décider de la pertinence de l'image I_{i+1} en fonction de la mesure de similarité entre images et de la fiabilité des différentes sources. Dans ce processus, chacune des sources participe dans la décision pour sélectionner la prochaine image similaire à l'image requête en combinant toutes les informations issues des images précédentes.

Soit U l'ensemble des images dans la base indexée et $R \subseteq U$ l'ensemble des images retournées par le système. On notera $(U - R)$ l'ensemble des images candidates. Dans le cadre de discernement $\Psi = \{Per, \overline{Per}\}$, chaque source S_i examine chaque image I_j dans l'ensemble $(U - R)$ en lui attribuant un degré de pertinence qui conduit à la définition d'une fonction de masse $m^\Psi[S_i, I_j](\cdot)$. Le degré de croyance alloué à l'hypothèse *Per* (le fait que l'image soit pertinente au sens de la requête) par la source S_i doit être proportionnel à la distance entre les deux images

I_i et I_j . Cette distance, qui sera notée $d(I_i, I_j)$ peut être calculée à partir de leurs vecteurs descriptifs respectifs f_i et f_j . Plus les deux images I_i et I_j sont semblables, plus leur mesure de similarité est faible ce qui nous permet d'attribuer la masse $m^\Psi[S_i, I_j](.)$ en utilisant les équations suivantes :

$$\begin{aligned} m^\Psi[S_i, I_j](Per) &= 1 - d(I_i, I_j); \\ m^\Psi[S_i, I_j](\overline{Per}) &= d(I_i, I_j); \\ m^\Psi[S_i, I_j](\Psi) &= 0. \end{aligned}$$

Pour prendre en compte le fait que toutes les images n'apportent pas la même quantité d'information, la fonction de masse est affaiblie à partir de coefficients de fiabilité α_i . Il s'agit ainsi d'augmenter l'incertitude associée en fonction du rang auquel l'image a été retrouvée. Pour cela, on utilise l'opération d'affaiblissement qui a été donnée à la formule (2.15), ce qui conduit à :

$$\begin{aligned} m_{\alpha_i}^\Psi[S_i, I_j](Per) &= \alpha_i.(1 - d(I_i, I_j)); \\ m_{\alpha_i}^\Psi[S_i, I_j](\overline{Per}) &= \alpha_i.d(I_i, I_j); \\ m_{\alpha_i}^\Psi[S_i, I_j](\Psi) &= 1 - \alpha_i. \end{aligned}$$

Dans ces équations (une pour chaque $i \in \{1, 2, \dots, L\}$, avec L le nombre d'images dans la base), le coefficient α_i représente la fiabilité associée à la source d'information S_i dérivée de la $i^{\text{ème}}$ image retournée I_i . Chaque α_i prend ses valeurs dans l'intervalle $[0, 1]$ tel que $\alpha_i = 1$ si la source est complètement fiable et $\alpha_i = 0$ si elle n'est pas fiable. Puisque l'image I_0 représente l'image requête, nous aurons $\alpha_0 = 1$.

Dans le cadre de la boucle de pertinence, il nous faut combiner les n sources d'informations en utilisant la règle de la combinaison de *Dempster* (équation 2.12) pour chaque image de la base des images candidates. Ainsi, pour chaque image I_j de $U - R$, la fonction de masse finale est calculée par :

$$m^\Psi[., I_j] = \bigoplus_{i=0}^{n-1} m_{\alpha_i}^\Psi[S_i, I_j].$$

Cette phase de combinaison permet au système de quantifier le degré de croyance de chaque image I_j de la base candidate ayant observé l'ensemble des images déjà retournées. Le tri associé à l'image I_n suivante est réalisé par le calcul de la probabilité pignistique à partir des $n - 1$ sources. Cette image de la base candidate est celle à qui on attribue la plus grande valeur de la probabilité pignistique pour l'hypothèse *Per* :

$$BetP_{m^\Psi[., I_n]}(Per) \geq BetP_{m^\Psi[., I_j]}(Per).$$

tel que $I_j \in U - R$. L'image retournée I_n sera incluse dans l'ensemble des sources R d'images retournées. L'opération de recherche peut être arrêtée en mettant un seuil sur la cardinalité de R ou/et en mettant un autre seuil sur la probabilité pignistique pour l'hypothèse *Per*.

Nous avons évalué la boucle de pertinence proposée dans le cadre de l'indexation de la base COLUMBIA² qui contient 1115 images classées en 20 catégories. Cette base généraliste comporte essentiellement des images couleur de paysages naturels ou de lieux habités. Concernant le calcul des vecteurs descriptifs, nous nous sommes limité à l'attribut couleur grâce à son pouvoir discriminant par rapport aux autres attributs. Nous travaillons dans l'espace HSV [60] qui s'avère plus dédié à l'évaluation de la similarité perceptuelle entre deux couleurs. On réalise sur les images de la base une quantification statique et uniforme sur les trois axes de l'espace HSV. Le vecteur descriptif utilisé dans cette évaluation est basé sur des histogrammes de couleur de neuf sous-images qui la constituent [61]. Nous réduisons le nombre de couleurs dans l'espace HSV par une quantification uniforme et statique de 162 couleurs [62]. Avec cette méthode, une image est représentée par un vecteur descriptif de 9×162 dimensions. La mesure de similarité (distance) entre les deux images I_i et I_j est une mesure vectorielle utilisée dans [21] :

$$d(I_i, I_j) = \frac{1}{9} \sum_{k=1}^9 \frac{\sum_{c=1}^{162} \min(hist_{i,k}(c), hist_{j,k}(c))}{\min(\sum_{c=1}^{162} hist_{i,k}(c), \sum_{c=1}^{162} hist_{j,k}(c))}.$$

avec $hist_{i,k}(c)$ (resp. $hist_{j,k}(c)$) représente la proportion de la couleur "c" dans le $k^{\text{ème}}$ rectangle de l'image I_i (resp. I_j).

Les coefficients de fiabilité des différentes sources d'informations sont obtenus par une fonction simple de calcul de paramètre α_i . Ainsi, pour chaque source d'information S_i dérivée de la $i^{\text{ème}}$ image retournée I_i , on calcule $\alpha_i = \alpha^i$. Ainsi la fiabilité de la source S_0 de l'image requête est toujours égale à 1. Dans cette expérimentation on a fixé $\alpha = 0,9$.

Les performances de la méthode sont illustrées à la figure 2.5 qui propose de gauche à droite l'image requête et 4 images retournées par le système. La première ligne (a) correspond à une liste des images triées uniquement en fonction de leur degré de similarité tandis que la seconde ligne (b) présente les images effectivement retournées par notre approche. On peut s'apercevoir que les images (2) et (4) initialement proposées par le système ont été rejetées par notre algorithme permettant de ne proposer que des images pertinentes au sens de la requête. Cet avantage met en évidence le fait que fusionner des informations dont les incertitudes et les fiabilités ont été correctement modélisées permet d'accroître la robustesse du système. En terme de performances, nous illustrons à la figure 2.6 une courbe de précision/rappel calculée pour les 48 images de la catégorie *football*. Là encore, on remarque que l'approche proposée est plus satisfaisante qu'un tri simple des images. Cet algorithme peut être intégré facilement dans n'importe quel système d'indexation et de recherche d'images. Toutefois, les performances de la boucle automatique dépendent fortement des premières images retournées et les temps de calcul sont relativement longs. Néanmoins, cette approche nous semble l'une des rares applications concrètes de la théorie

²<http://www.cs.washington.edu/research/imagedatabase/groundtruth/>.

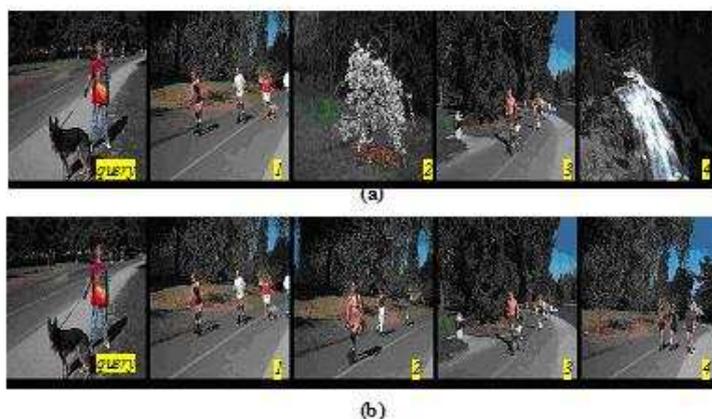


Figure 2.5 : Image requête et images retournées par un tri simple (a) et par l'approche proposée (b).

des fonctions de croyance à l'indexation de documents.

2.4 Conclusion

Dans ce chapitre, plusieurs techniques de fusion d'informations ont tout d'abord été présentées dans un contexte d'utilisation en indexation de documents multimédia. Dans ce cadre, la fusion est généralement employée pour la prise en compte de la multimodalité des informations (image, audio, texte). On ne peut pas réellement conclure qu'une des techniques présentées est idéale au sens où elle répond à toutes les problématiques de l'indexation. D'ailleurs, dans certains cas, il pourrait paraître judicieux d'utiliser de manière simultanée plusieurs de ces techniques à la fois en exploitant certaines de leurs propriétés.

Dans le cadre de cette thèse, nous avons choisi de travailler avec la théorie des fonctions de croyance. En effet, ce cadre générique permet d'englober les formalismes de plusieurs théories classiques comme les probabilités et la théorie des possibilités. Les fonctions de croyance offrent de nombreux mécanismes et outils pour gérer l'incertitude, modéliser la méconnaissance et prendre en compte des informations partielles. Sa souplesse et sa capacité à intégrer des notions sur la fiabilité des différentes sources d'information contribue à conclure sur son utilisation dans le cadre de l'indexation de documents multimédia. Les quelques applications des fonctions de croyance à l'indexation permettent d'être optimiste sur l'adaptation de cette technique de fusion dans le cadre d'un système d'indexation. Dans le chapitre suivant, nous proposons la contribution essentielle de ce travail de thèse où les fonctions de croyance sont utilisées dans un système de

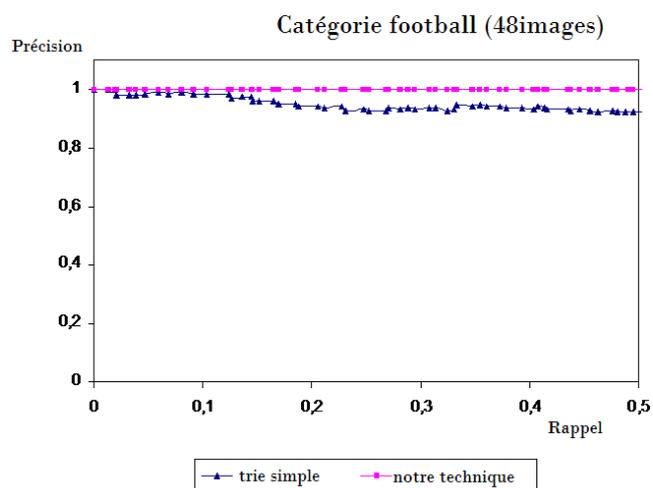


Figure 2.6 : Précision/Rappel dans la catégorie Football (48 images) pour un tri simple et pour l'approche proposée.

reconnaissance de personnes dans des albums photos.

Chapitre 3

FUSINDEX : Un système de FUSion d'INDEX

La principale problématique du programme de recherche dans lequel s'inscrit ce travail de thèse concerne une offre de service semi-automatique d'indexation pluri-média d'albums de photographies. L'un des objectifs du projet consiste à produire des fonctions semi-automatiques d'indexation d'images et de vidéos personnelles¹ auto-produites en combinant des techniques d'indexation automatique et d'annotation manuelle d'images. Le système recherché devra, dans un premier temps, disposer d'outils d'indexation automatique basés sur l'exploitation de différents médias tels que l'image, la parole et le texte. Des outils d'annotation manuelle doivent également être proposés comme service d'indexation. Dans le premier chapitre, nous avons mis en évidence le fait que la qualité des résultats spécifiques à une technique d'indexation particulière n'exploitant qu'un seul média est asymptotique et tend à atteindre un plafond en terme de performance. La solution envisagée repose sur la fusion d'informations multimédia en tirant profit du contenu des documents mais également de leur contexte associé [63]. Par conséquent, il est envisagé dans le cadre de cette thèse, la mise en œuvre d'un prototype opérationnel d'indexation basée sur l'utilisation conjointe de plusieurs médias. Ce système a été baptisé **FUSINDEX** pour FUSion d'INDEX. Dans ce chapitre, nous nous attachons à décrire le système proposé dans la section 3.1, puis nous détaillerons les différents outils d'indexation utilisés (section 3.2). Le système exploite des informations du contenu et du contexte des documents qui nécessitent d'être organisées sous forme hiérarchique (section 3.3) puis fusionnés pour produire des index pertinents. Nous verrons comment ces index sont générés à la section 3.4 où nous expliquons également comment sont évalués leurs degrés de pertinence respectifs.

¹Dans ce document, nous nous focaliserons sur l'indexation de bases d'images statiques.

3.1 Description du système

Le système **FUSINDEX** s'intègre dans un service semi-automatique d'indexation pluri-média d'albums de photographies personnelles baptisé **ImagINUM** pour **Images Indexation by a Network of Users and Machines**. Comme son nom l'indique, la philosophie de ce service est résolument une approche orientée *réseau* qui consiste à partager les informations multimédia disponibles entre utilisateurs et entre machines. Ainsi, l'outil propose aux utilisateurs référencés de déposer des photos sur leur compte, de les visualiser et de les partager². Il a également la possibilité d'ajouter des commentaires textuels et vocaux à ses photos formant ainsi une base de documents multimédia qu'il s'agit d'organiser et d'indexer pour en faciliter l'accès.

La navigation dans la base est rendue possible en classant les documents sous la forme d'une structure arborescente où chaque feuille de l'arbre est associée à une catégorie spécifique de documents appelée *topic*. Un *topic* est une classe dans laquelle est rangé un ensemble de documents relatifs à un même **concept d'indexation**. Ces concepts peuvent être proposés par l'utilisateur ou automatiquement générés par le système. Dans la version dédiée à la gestion d'albums de photos personnelles, une interface Homme/Machine, baptisée Someone-Photo et illustrée à la figure 3.1, a été développée pour faciliter l'accès et la navigation dans la base [64]. Dans l'architecture de l'interface SOMEONE, nous nous limitons volontairement aux concepts suivants : Album, Personne et Lieu. L'utilisateur associe donc à chacune de ses photos les *topics* qu'il souhaite [65] et qu'il organise à sa façon (cf. figure 3.1).

3.1.1 Moteurs d'indexation

L'architecture du projet **ImagINUM** est centrée sur un ensemble de fonctionnalités en terme d'indexation qui exploitent la multimodalité des documents multimédia. Nous appellerons moteur d'indexation, un outil qui permet d'obtenir des index à partir d'un média donné. Ainsi, certains moteurs peuvent fournir des informations à partir de l'image, des commentaires textuels ou des commentaires vocaux. La section 3.2 s'attachera à décrire chacun de ces moteurs dans la version qui nous a été fournie. Nous verrons un peu plus loin dans le chapitre suivant comment ils ont été adaptés pour répondre à nos attentes. Les liens entre ces moteurs et le système **FUSINDEX** sont illustrés à la figure 3.2. Les spécificités technologiques associées à chacun de ces moteurs sont à trouver dans les domaines du traitement des images, de la parole et de l'information textuelle. Comme on peut le remarquer sur la figure 3.2, ce panel d'outils disponibles est également enrichi d'une fonctionnalité d'annotation manuelle des documents qui permet une indexation personnalisée de la base. On peut donc catégoriser les moteurs d'indexa-

²Voir par exemple [http : \ photos . orange . fr](http://photos.orange.fr) .

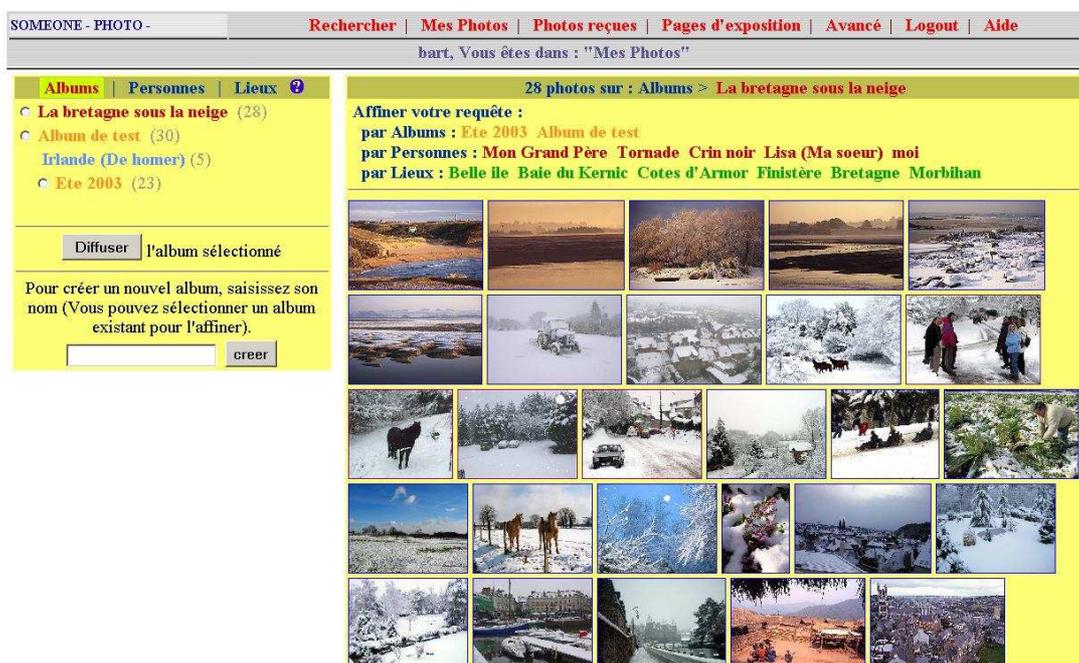


Figure 3.1 : Illustration de l'interface Someone-Photo mettant en évidence les trois catégories : Albums, Personnes et Lieux.

tion en quatre classes : les outils d'indexation basés sur les informations extraites de l'image, les outils d'indexation basés sur les informations extraites de la parole, les outils d'indexation basés sur les informations extraites du texte et les outils d'indexation basés sur les annotations personnelles.

3.1.2 Position du problème et solution envisagée

La problématique à laquelle souhaite répondre le système **FUSINDEX** consiste à proposer un maximum de descripteurs pertinents pour l'utilisateur dans un contexte d'offre semi-automatique d'indexation pluri-médias des documents de la base. On appelle descripteur pertinent, un descripteur qui a été accepté par l'utilisateur et qui s'avère utile pour la recherche d'images dans la base. Le rôle de ces descripteurs, qui peuvent varier fortement d'un utilisateur à l'autre, consiste à indexer de façon autonome et automatique les documents présentés au système. Ainsi, chaque moteur propose un résultat qu'il s'agit d'exploiter au mieux pour indexer le document. La figure 3.3 illustre les principaux composants de l'interface Someone-Photo. Elle met en évidence la constitution de la base des documents multimédia (albums photos, commentaires textuels et vocaux et son interface WEB de dépôt dans la base) et le listing des descripteurs (topics) permettant l'indexation de ces documents.

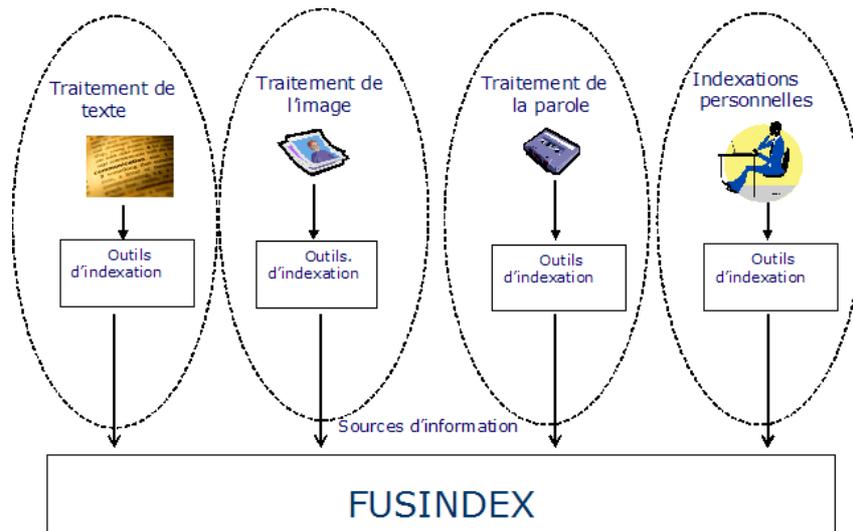


Figure 3.2 : FUSINDEX : Outils d'indexation et illustration de la multimodalité.

De façon à agréger l'ensemble de ces résultats d'indexation, nous proposons une approche qui s'inspire de travaux récents sur la fusion d'informations. On considère chaque élément d'indexation comme une source d'information qu'il s'agit de modéliser sous un même formalisme afin de les fusionner. Les différentes sources d'information à disposition proviennent en grande partie d'informations obtenues via le traitement d'images, de la parole (audio) et du texte. Nous verrons qu'elles peuvent être enrichies par des informations contextuelles qui permettent d'indexer les documents par leur contenu et par leur contexte. En terme d'indexation, la fonction d'évaluation doit tenir compte du nombre de descripteurs proposés et acceptés. Il est également envisagé de pondérer l'impact de chaque descripteur sur la fonction d'évaluation en fonction de sa précision et de son utilité. Pour répondre à ces attentes, nous avons résolument choisi le cadre théorique des fonctions de croyance, qui propose des fonctionnalités avantageuses en terme de modélisation d'information imprécise et incertaine, d'analyse du conflit potentiel entre les sources, d'évaluation de la fiabilité, de fusion d'informations, ...

Dans le cadre de cette thèse, nous nous intéressons aux descripteurs basés sur l'identité des personnes qui se trouvent dans le document à indexer. Chacun des moteurs fournit donc une information sur la (ou les) personne(s) qui sont censée(s) indexer le document dans le *topic* considéré. Soit B la base des personnes connues par le système définie par $B = \{P_1, P_2, \dots, P_n\}$. Cette base qui contient n personnes est également appelée "le carnet d'adresses" relatif à chaque utilisateur du système. En d'autres termes, le problème qui est posé consiste à identifier les personnes présentes dans l'image, les prénoms des personnes prononcés dans le commentaire audio, les prénoms des personnes apparaissant dans le commentaire textuel en s'appuyant sur

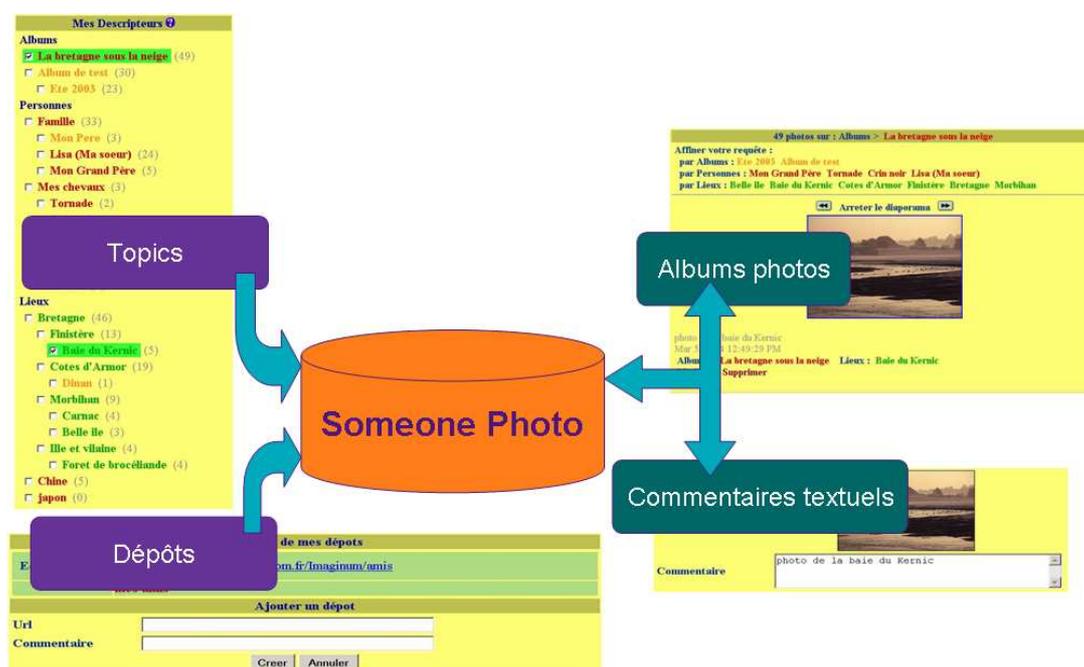


Figure 3.3 : Architecture technologique de l'interface Someone-Photo.

les différents moteurs d'indexation. Il s'agit donc d'une indexation selon les personnes qui est à rapprocher avec le système développé dans [4].

3.2 Description des différents moteurs d'indexation

De façon à exploiter la complémentarité des différents médias, le système **FUSINDEX** exploite les informations recueillies par différents moteurs d'indexation. Les informations utiles pour l'indexation sont fournies de manière automatique par des outils d'analyse automatique de l'image [66], de l'audio [67] et du texte [68] associés à chaque document. Les outils disponibles dans le cadre de notre application sont :

- un détecteur de visage, noté DV ;
- un reconnaiseur de visage, noté RV ;
- un reconnaiseur de genre Homme/Femme basé sur l'image, noté H/F ;
- un reconnaiseur de Prénom basé sur l'audio, noté RPa ;
- un reconnaiseur de Prénom basé sur le texte, noté RPt ;
- une horloge Hr ;
- un descripteur de lieux, noté L .

Les trois premiers d'entre eux DV , RV et H/F analysent l'image tandis que les outils RPa et RPt analysent respectivement les commentaires vocaux et textuels. Ces différents moteurs s'attachent à extraire des informations pertinentes sur le contenu de chaque document dans la base. Les deux derniers moteurs Hr et L traitent les données contextuelles attachées au document à indexer (date et coordonnées géographiques). Dans la description qui en est faite dans les paragraphes suivants, on se limite volontairement aux entrées et sorties de ces outils sans préciser les mécanismes et les méthodes exploitées pour interpréter l'image et ses données associées.

Nous nous limitons également ici à la description de ces seuls outils. Ceux-ci seront utilisés dans le processus de fusion d'index proposé. Il est évident que tout nouvel outil disponible pourra être intégré dans ce processus de fusion améliorant d'autant plus la fiabilité du système.

3.2.1 Détecteur de visage : DV

Le détecteur de visage DV est un outil d'analyse d'images qui consiste à localiser les visages dans l'image associée au document à indexer. Il est basé sur une architecture neuronale convolutionnelle qui a été spécialement conçue pour détecter de façon robuste des visages dans des scènes complexes [69]. Les algorithmes proposés s'accommodent des rotations du visage dans le plan image de près de 60 degrés. Les informations fournies par ce moteur sont illustrées à la figure 3.4. Pour chaque image traitée, les informations suivantes sont disponibles : le nombre z

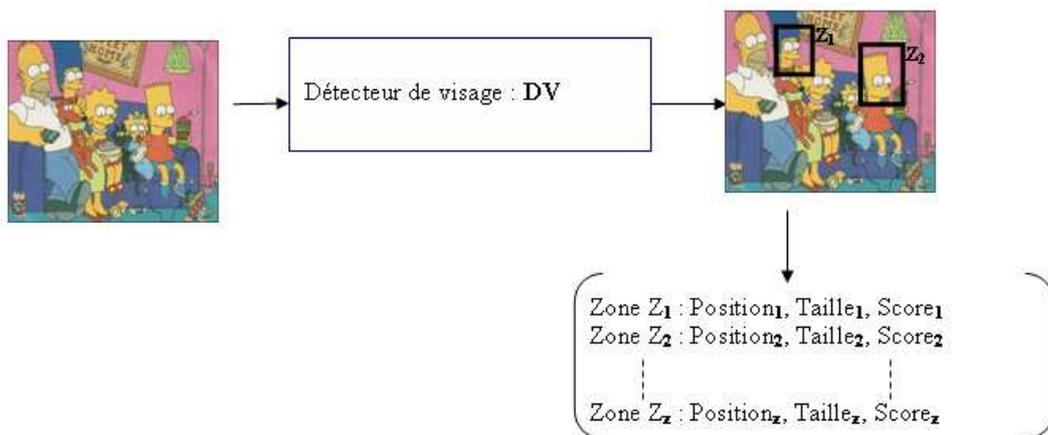


Figure 3.4 : Outil de détection des visages DV sur l'image. A chaque image analysée, les informations fournies par l'outil sont le nombre de visages détectés, les positions et tailles des fenêtres englobantes ainsi qu'un score de détection associé à chaque zone.

de visages détectés dans l'image, les positions et les tailles des zones Z_j pour $j = 1, \dots, z$ rectan-

gulaires englobant les visages dans l'image (un rectangle par visage) ainsi qu'un score positif non normalisé (compris entre une valeur minimale et une valeur maximale) associé à chaque visage détecté. Les positions et tailles des zones Z_j sont utilisées par le reconnaisseur de visage.

3.2.2 Reconnaisseur de visage : RV

L'outil de reconnaissance de visages RV a pour but de reconnaître les visages préalablement détectés par le détecteur DV en se basant sur une base de personnes connues préalablement répertoriées dans le carnet d'adresses B . Ceci présuppose donc la constitution d'une base de visages des personnes connues où chaque personne P_i peut être apparentée à un ensemble de visages. Nous reviendrons plus tard dans le document sur la constitution de cette base d'apprentissage. Le moteur RV est basé sur de nouvelles techniques d'extraction de signatures (un ensemble d'informations bas niveau extraites de l'image) basées sur l'analyse discriminante linéaire qui prend en compte la structure bidimensionnelle des images de visages [66]. L'outil obtient de bonnes performances en terme de reconnaissance et permet de rendre plus robuste la prise de décision vis-à-vis des changements de pose, des variations d'expression faciale et d'occultations partielles des visages [70]. A partir des informations validées pour les zones détectées dans l'image et de la base d'apprentissage, le reconnaisseur de visages RV quantifie la similarité de chaque zone détectée avec les visages des personnes connues sous la forme d'une mesure de distance. Ce processus d'indexation est illustré à la figure 3.5 qui fait apparaître la validation des zones par l'outil DV . A

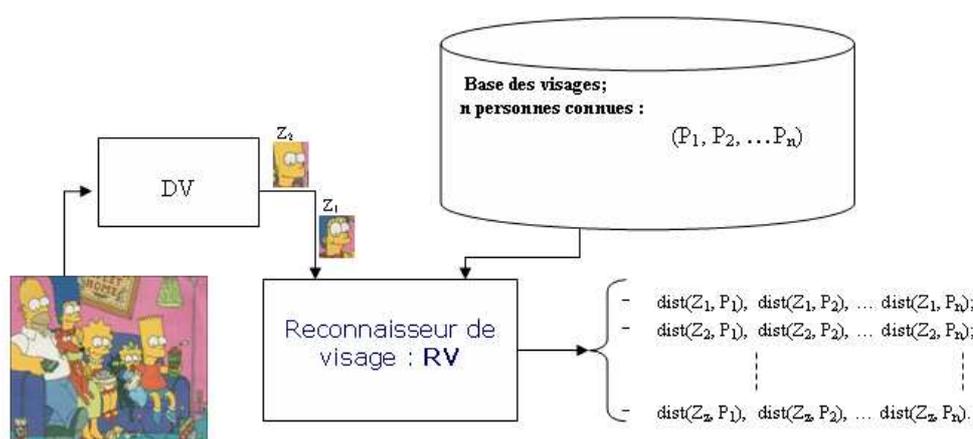


Figure 3.5 : Outil de reconnaissance des visages RV sur l'image. A chaque visage détecté, l'information disponible est rassemblée dans un vecteur de distances permettant d'établir un tri au niveau des personnes P reconnues.

partir des différentes zones Z_1, Z_2, \dots, Z_z détectées et du carnet d'adresses $B = \{P_1, P_2, \dots, P_n\}$,

le système RV fournit automatiquement $z \times n$ distances :

$$\begin{pmatrix} dist(Z_1, P_1), & dist(Z_1, P_2), & \dots & dist(Z_1, P_n) \\ dist(Z_2, P_1), & dist(Z_2, P_2), & \dots & dist(Z_2, P_n) \\ \dots & \dots & dist(Z_j, P_i), & \dots \\ dist(Z_z, P_1), & dist(Z_z, P_2), & \dots & dist(Z_z, P_n) \end{pmatrix} \quad (3.1)$$

où $dist(Z_j, P_i)$ quantifie la distance entre la zone Z_j et la personne P_i . Plus cette distance est petite, plus la zone Z_j peut être assimilée à un visage connu de la personne P_i .

3.2.3 Reconnaiseur de genre Homme/Femme : H/F

L'outil de reconnaissance de genre Homme/Femme, noté H/F est basé sur le même principe que l'outil de reconnaissance de visages RV . Son objectif consiste à analyser les zones préalablement détectées par le DV et à émettre une hypothèse sur le sexe de la personne détectée. Le principe de ce moteur d'indexation est illustré à la figure 3.6. Il consiste à identifier le sexe de

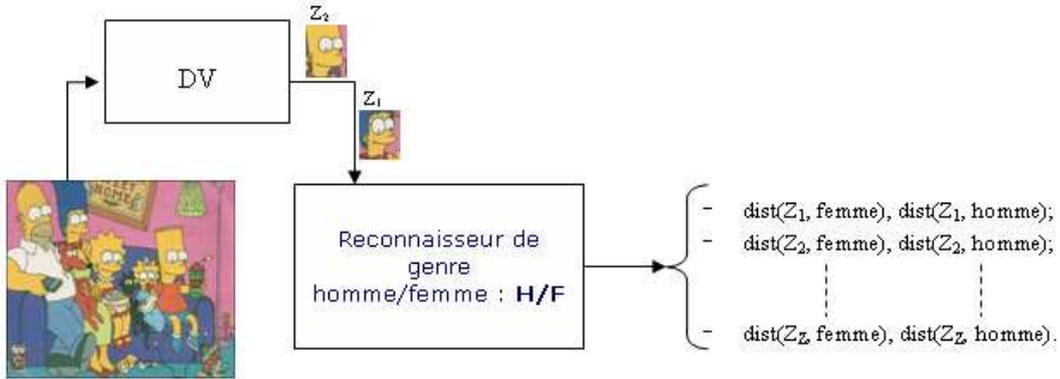


Figure 3.6 : Outil de reconnaissance de genre H/F à partir de l'image. A chaque image et pour chaque visage détecté, l'outil fournit une information sur le genre H/F de la personne détectée.

la personne à partir d'informations extraites de son visage en se basant sur les visages connus catégorisés en deux classes : Homme et Femme. La constitution de cette base se fait simplement en triant dans la base des visages connus les visages masculins et les visages féminins. Ainsi, si on suppose que le DV détecte les zones Z_1, Z_2, \dots, Z_z , le moteur H/F fournira automatiquement les $z \times 2$ distances suivantes :

$$\begin{pmatrix} dist(Z_1, H), & dist(Z_1, F) \\ dist(Z_2, H), & dist(Z_2, F) \\ \dots, & \dots \\ dist(Z_z, H), & dist(Z_z, F) \end{pmatrix}. \quad (3.2)$$

où $dist(Z_j, H)$ (respectivement $dist(Z_j, F)$) représente la distance entre la zone Z_j et le fait que ce visage soit un visage masculin (respectivement féminin). Les trois outils qui viennent d'être décrits DV , RV et H/F permettent d'extraire des informations sur le contenu du document à partir de l'image.

3.2.4 Reconnaiseur de Prénom audio : RPa

Parmi les moteurs capables d'extraire des informations pertinentes sur le contenu d'un document, un reconnaiseur de prénoms dans les commentaires audio est utilisé dans le système **FUSINDEX**. Cet outil d'indexation a pour but de reconnaître les prénoms à partir de l'analyse d'un commentaire audio associé à chaque document de la base. On cherche à repérer, par l'analyse de la parole avec un outil dédié, les différents prénoms prononcés dans le commentaire audio du document. L'outil en question est le fruit du consortium de recherche NEOLOGOS³ pour la construction d'une base de données vocale [67]. L'optimisation de cette base en terme de réduction du nombre de locuteurs tout en conservant leur diversité (genre H/F, accent régional et âge des locuteurs⁴) constitue l'originalité de la technologie employée. Le reconnaiseur de prénoms RPa , dont le mécanisme d'indexation est présenté à la figure 3.7, se base sur cette technologie. Le

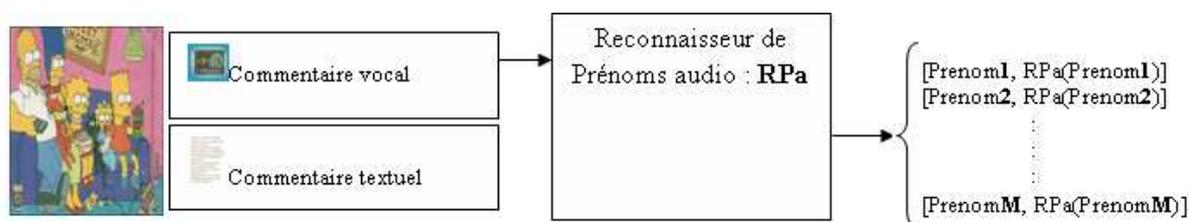


Figure 3.7 : Outil de reconnaissance de prénoms à partir des commentaires audio. A chaque prénom détecté est associé un score $RPa(.)$.

résultat proposé par ce moteur est une liste des prénoms prononcés dans le commentaire avec des degrés de confiance associés à chaque prénom. Il s'agit d'une mesure de probabilité *a posteriori* pour que le prénom soit effectivement prononcé dans le commentaire. Pour information, l'outil utilisé détecte également les surnoms comme Papa, Maman, ...

³Programme de recherche Technolanguage www.technolanguage.net.

⁴www.donne-ta-voix.fr.

3.2.5 Reconnaisseur de Prénom texte : *RPt*

Le reconnaisseur de prénom *RPt* est un moteur qui consiste à reconnaître, à partir d'un commentaire textuel associé à chaque document, les différents prénoms présents à partir de l'analyse de texte. Le résultat est une liste des prénoms présents dans le commentaire avec des degrés de confiance associés à chaque prénom. Il s'agit là encore d'une probabilité *a posteriori* que la réponse soit correcte. Il est basé sur un analyseur syntaxique développé spécifiquement pour faciliter l'indexation d'albums de photos personnelles [68]. La figure 3.8 présente le fonctionnement du moteur d'indexation basé sur cet outil d'analyse textuelle. De la même façon que le *RPa*,

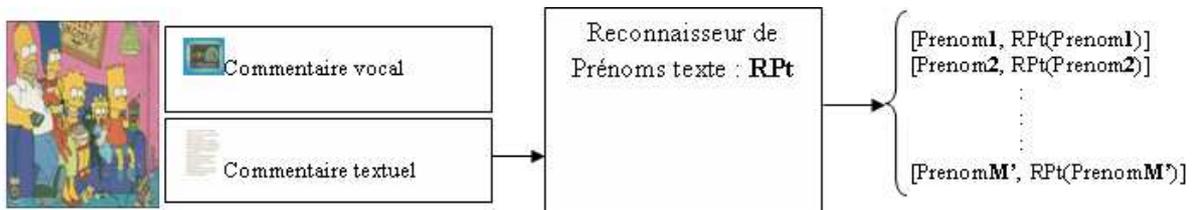


Figure 3.8 : Outil de reconnaissance de prénoms à partir des commentaires textuels. A chaque prénom détecté est associé un score $RPt(.)$.

le moteur *RPt* détecte aussi les surnoms. L'attention du lecteur est attirée sur le fait qu'il ne s'agit pas d'un simple analyseur de texte mais d'un outil capable de donner un score sur le fait que le prénom d'une personne est présent dans le commentaire (exemple : 'Chapelle Rucellai à Florence').

3.2.6 Informations contextuelles : Horloge *Hr*, Lieux *L*

Nous avons insisté au paragraphe 1.3 sur l'utilité d'une gestion adéquate des informations contextuelles dans l'indexation de documents multimédia. Dans le cadre de ce travail, nous avons exploité les informations de date et de lieu. L'exploration par la date ou le lieu des images d'une collection personnelle est proposée par de nombreux logiciels (Microsoft World Wide Media eXchange, Hello, ShoZu). Néanmoins, aucun ne propose d'organisation automatique des documents, celles-ci devant être organisées manuellement [71]. Nous pensons que la prise en compte de ces informations de façon systématique est bénéfique, notamment dans le cas où le nombre de documents devient assez grand.

La date d'acquisition de l'image fournit une information supplémentaire riche sur le contexte dans lequel a été capturée l'image. Cette information est corrélée à une variable notée *Hr* caractérisant l'horloge comme échelle de temps. Dans l'architecture informatique du système **Ima-**

gINUM, la date est récupérable, soit par l'appareil photo numérique, soit par introduction manuelle de l'utilisateur. Dans le cas d'utilisation d'appareil du style "cameraphone", la date peut être transmise directement par le réseau de communication GSM. Cette information prend une forme standard comme par exemple : année/mois/jour/heure/minute/seconde.

Les coordonnées géographiques du lieu dans lequel l'image a été capturée constituent elles aussi des informations à part entière qu'il s'agit d'exploiter parmi les informations contextuelles. Le lieu, noté par une variable L , est soit introduit de façon manuelle par l'utilisateur soit directement récupéré via un équipement 'GPS' attaché au "cameraphone" dans lequel l'identificateur des cellules radio peut être utilisé. Ces informations prennent souvent la forme de deux valeurs réelles référencées dans un système de coordonnées géographiques du style latitude et longitude.

En terme de contexte, l'idée défendue dans ce travail consiste à supposer que deux documents qui ont été pris à des intervalles de temps très courts et à des endroits géographiques très proches doivent être assimilés à des concepts d'indexation similaires. La technique s'appuie donc sur la comparaison en terme de temps et de distance (proche temporellement et proche géographiquement) du document à indexer et de chaque document déjà annoté dans la base. Nous détaillerons la prise en compte formelle de ces deux concepts associés aux dates et aux lieux dans le chapitre 5 de cette thèse. Notons ici que seules ces informations temporelle et géographique ont été prises en compte mais d'autres informations contextuelles (destinataire, expéditeur, à caractère social, ...) pourraient être envisagées dans le cadre du système proposé.

3.3 Hiérarchisation des informations et indexation de la base

A partir des informations de contenu et des informations contextuelles extraites par les différents moteurs présentés à la section 3.2, l'objectif en terme d'indexation est de représenter chaque document par un ensemble fini de métadonnées ou de concepts. Comme il a été mentionné, une indexation selon les personnes est envisagée ce qui correspond à fournir pour métadonnée un (ou plusieurs le cas échéant) prénom(s) candidat(s) dans la base B des prénoms, pour indexer le document. Dans cette section, nous présentons la façon dont les informations sont hiérarchisées pour atteindre cet objectif et le résultat attendu quant à l'indexation de la base des documents.

3.3.1 Organisation hiérarchique

Le système proposé organise les différentes sources d'information sous la forme d'une hiérarchie dont le niveau le plus haut correspond à la distinction entre le contenu du document et son

contexte. La production des métadonnées peut être schématisée par la figure 3.9 qui illustre la hiérarchie utilisée. En ce qui concerne le contenu, les outils d'indexation peuvent être regrou-

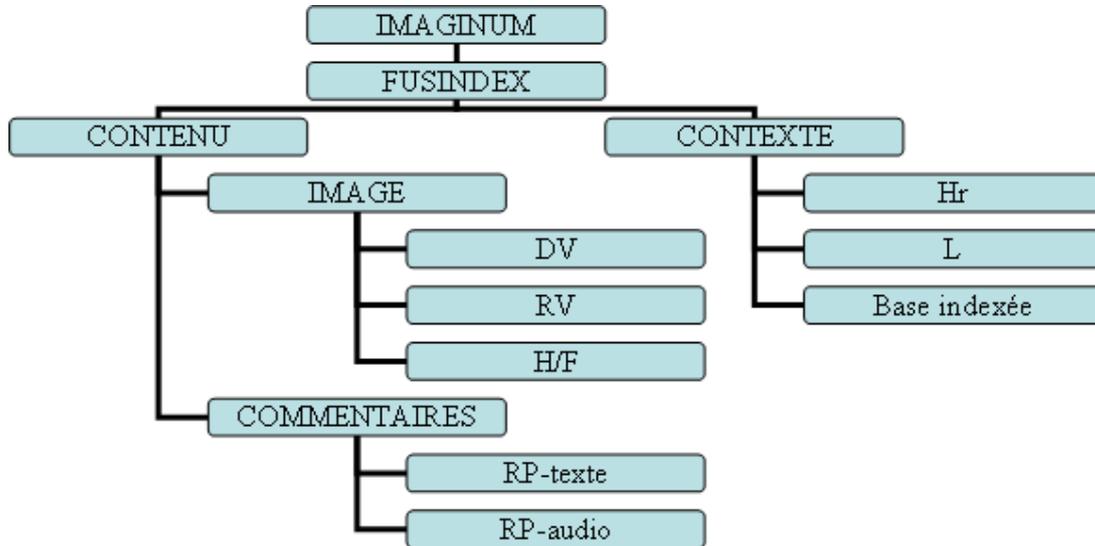


Figure 3.9 : Hiérarchie des sources d'information utilisées pour engendrer les métadonnées.

pés en deux sources d'information : les informations de type image, que nous noterons source I , et commentaires, notés Com , qui incluent les commentaires vocaux et textuels. La source I regroupe les outils qui analysent le contenu de l'image : le détecteur de visage DV , le reconnaissseur de visage RV et le reconnaissseur de genre homme/femme H/F . La source d'information Com se décline en deux catégories : les commentaires textuels qui sont traités par le RPt et les commentaires vocaux qui sont traités par le RPa . En ce qui concerne le contexte, le système **FUSINDEX** s'appuie sur des informations provenant d'images déjà indexées dans la base. Cette source d'information, notée $Cont$, regroupe les informations recueillies par les images voisines (avec leurs sources I et Com associées) à partir des moteurs d'indexation Hr et L .

3.3.2 Base indexée

Soit un document I_k appartenant à un ensemble fini de K documents définissant la base E des documents à indexer : $E = \{I_1, I_2, \dots, I_k, \dots, I_K\}$. Mathématiquement, le document I_k peut être défini dans la base indexée par un vecteur descriptif noté T_k calculé par **FUSINDEX**. Ainsi, le système est chargé d'évaluer pour chaque document I_k de la base une fonction définie par :

$$\begin{aligned}
 FUSINDEX : E &\rightarrow [0, 1]^n \\
 I_k &\rightarrow T(I_k) = T_k = (\tau_{k,1}, \tau_{k,2}, \dots, \tau_{k,i}, \dots, \tau_{k,n})^t.
 \end{aligned} \tag{3.3}$$

où $\tau_{k,i}$ est une valeur comprise entre 0 et 1 qui quantifie l'appartenance de la personne P_i (i variant de 1 à n) dans le document I_k (k variant de 1 à K). Autrement dit, les métadonnées qui seront fournies par le système correspondent aux identités des personnes connues dans la base B . Dans le chapitre suivant, nous aborderons une stratégie pour introduire les métadonnées liées aux nouvelles personnes détectées dans le document.

La base indexée peut donc se résumer à une matrice T de taille $K \times n$ valeurs définie par :

$$T = \begin{pmatrix} \tau_{1,1} & \cdots & \tau_{1,i} & \cdots & \tau_{1,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tau_{k,1} & \cdots & \tau_{k,i} & \cdots & \tau_{k,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tau_{K,1} & \cdots & \tau_{K,i} & \cdots & \tau_{K,n} \end{pmatrix}. \quad (3.4)$$

La matrice T (ou le vecteur T_k) est dite *hard* si la variable $\tau_{k,i}$ prend une valeur binaire dans l'ensemble des valeurs possibles de $\{0, 1\}$ (vrai ou faux). La présence d'une personne P_i dans le document I_k est alors représentée par :

$$\tau_{k,i} = 1 \Rightarrow P_i \in I_k. \quad (3.5)$$

qui permet d'indexer la base E relativement aux n concepts prédéfinis. La matrice T (ou le vecteur T_k) est dite floue si $\tau_{k,i}$ varie dans l'intervalle $[0, 1]$ et représente le degré d'appartenance au concept $P_i \in I_k$. Dans le cas flou, les métadonnées présentées à l'utilisateur seront triées et listées selon leurs valeurs respectives de $\tau_{k,i}$ par ordre décroissant de la plus grande à la plus petite : $\text{rang}(P_i) > \text{rang}(P_j)$ si $(\tau_{k,i} > \tau_{k,j})$ pour $(i, j) \in \{1, 2, \dots, n\}$.

Dans le système **FUSINDEX**, la variable $\tau_{k,i}$ peut représenter :

- soit un intervalle dont les bornes inférieure et supérieure de la probabilité [45] correspondent à la crédibilité et à la plausibilité de l'événement " P_i se trouve dans I_k ". Cet intervalle quantifie le degré de croyance dans le fait que le document I_k peut être associé au concept P_i . Ainsi, l'événement "la personne P_i est dans l'image associé au document I_k " sera représentée par :

$$(\text{bel}_{k,i}, \text{pl}_{k,i}). \quad (3.6)$$

- soit la probabilité pignistique associée à cet intervalle. On quantifie ainsi la probabilité que l'événement "la personne P_i est dans l'image associée au document I_k " soit vraie qui est définie par :

$$\tau_{k,i} = \text{Bet}P_{k,i}. \quad (3.7)$$

La probabilité $\text{Bet}P_{k,i}$, associée au document I_k est définie sur l'espace probabilisé qui correspond à l'ensemble des personnes du carnet d'adresses B . Il est à noter qu'on peut passer de cette mesure

à un vecteur T_k hard (information binaire, cf. équation (3.5)) en extrayant le maximum de cette probabilité sur l'ensemble des P_i .

3.4 Génération des métadonnées par fusion d'informations

Pour chaque document I_k à indexer dans la base E , la pertinence d'une métadonnée liée à une personne connue est une fonction dépendante des informations. Par exemple, plus le degré de certitude pour qu'une personne figure dans l'image est grand, plus la métadonnée générée sera pertinente au sens de l'indexation. Dans ces paragraphes, nous détaillons le processus utilisé pour produire et quantifier la pertinence des métadonnées. Il est basé sur la représentation des incertitudes par des fonctions de croyance et la fusion de celles-ci dans un cadre de discernement adapté.

3.4.1 Sources d'informations disponibles

Dans l'architecture du système tel que nous le décrivons (voir le paragraphe 3.2 sur les moteurs d'indexation disponibles), on dispose de trois sources d'information capables de fournir des métadonnées en terme d'indexation :

- Image : *I*. Les outils d'indexation qui traitent l'image *DV*, *RV* et *H/F* peuvent indiquer la présence d'une personne connue dans l'image à indexer.
- Commentaires : *Com*. Les outils d'indexation qui traitent la parole et le texte (*RPa* et *RPt*) peuvent indiquer que le document est à indexer sous le concept du prénom d'une personne prononcée dans le commentaire vocal ou présente dans le commentaire textuel.
- Contexte : *Cont*. Les outils d'indexation qui traitent *Hr*, *L* ainsi que la base déjà indexée peuvent indiquer la similarité en terme d'indexation par rapport à ce document. Dans ce cas, le document sera indexé sous le même concept.

Ces trois sources d'informations vont permettre de générer des métadonnées quant à l'indexation de la base E .

3.4.2 Génération des métadonnées

Le mécanisme de génération des métadonnées est présenté à partir d'un exemple qui consiste à indexer un ensemble d'images relatives à la famille Simpson⁵. Notre base de documents comporte cinq personnes connues dans la famille comme l'illustre la figure 3.10. Ainsi, la constitution du

⁵http://fr.wikipedia.org/wiki/Les_Simpson pour les non-initiés.

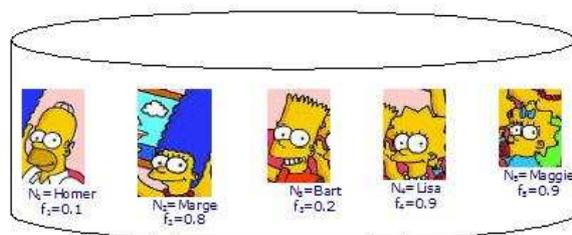


Figure 3.10 : La base des personnes de la famille Simpson contient 5 personnes connues.

carnet d'adresses des personnes connues dans la base est la suivante :

$$B = \{Homer, Marge, Bart, Lisa, Maggie\}. \quad (3.8)$$

Soit I_k , le document à indexer présenté à la figure 3.11, composé d'une image, de trois commentaires vocaux et de deux commentaires textuels. Supposons que le système **FUSINDEX** a



Figure 3.11 : Document à indexer composé d'une image avec la détection de deux zones, de deux commentaires textuels et trois commentaires vocaux.

obtenu les valeurs de $\tau_{k,i}$ suivantes :

$$T_k = \begin{pmatrix} 0.8 \\ 0.1 \\ 0.95 \\ 0.75 \\ 0.2 \end{pmatrix} \quad (3.9)$$

où chaque valeur de $\tau_{k,i}$ représente la probabilité pignistique sur le fait que le concept i de la base B des prénoms peut être associé au document I_k . On a donc $(\tau_{3,k} > \tau_{1,k} > \tau_{4,k} > \tau_{5,k} > \tau_{2,k})$. La

question qui se pose est désormais de savoir comment présenter les métadonnées à l'utilisateur pour qu'il valide l'indexation proposée. S'il n'y a pas de seuil sur la valeur de la probabilité pignistique du vecteur T_k de l'équation (3.9), toutes les métadonnées trouvées seront affichées et triées selon leurs degrés de confiance. Pour l'exemple du document de la figure 3.11, on obtient les métadonnées suivantes dans l'ordre :

1. Indexer sous le *topic Bart* : oui/non ?
2. Indexer sous le *topic Homer* : oui/non ?
3. Indexer sous le *topic Lisa* : oui/non ?
4. Indexer sous le *topic Maggie* : oui/non ?
5. Indexer sous le *topic Marge* : oui/non ?

Cette proposition d'indexation est à valider par l'utilisateur. Pour mieux filtrer ces propositions de métadonnées, il est nécessaire de définir un seuil sur le degré de pertinence des métadonnées à afficher. Il serait même intéressant de créer automatiquement l'annotation lorsque ce degré possède une valeur suffisante. Ce calcul de degré de pertinence est illustré dans les paragraphes suivants.

3.4.3 Évaluation des degrés de pertinence associés aux métadonnées

Le problème consiste à évaluer les degrés de pertinence des différentes métadonnées produites par le système qui seront proposées à l'utilisateur. Nous avons choisi d'évaluer ce degré de pertinence par une fonction de croyance qui permet de représenter l'incertitude associée. Toujours dans le cadre de l'exemple associé à la famille Simpson et si l'on suppose que le carnet d'adresses est identique à celui défini à l'équation (3.8), le système est en mesure de proposer à l'utilisateur de une à cinq métadonnées suivant leurs degrés de pertinence respectifs. Pour traiter ce problème, on définit un cadre de discernement, noté Ψ , d'une cardinalité égale à deux de la forme :

$$\Psi = \{Per, \overline{Per}\}. \quad (3.10)$$

où Per représente l'hypothèse de pertinence indiquant que la métadonnée est pertinente d'un point de vue de l'indexation du document I_k . L'hypothèse \overline{Per} est le complémentaire de Per dans Ψ et représente la non pertinence. Ce cadre de discernement va permettre de quantifier le degré de croyance associé à une métadonnée qui sera proposée à l'utilisateur.

Les trois sources d'information : Image, Commentaires et Contexte, respectivement notées I , Com et $Cont$ vont permettre, pour chaque personne P_i du carnet d'adresses B d'évaluer le degré de croyance sur le fait que cette personne est représentative quand à l'indexation du document.

$A \subseteq \Psi$	Contenu		Contexte
	$m^\Psi[I, P_i]$	$m^\Psi[Com, P_i]$	$m^\Psi[Cont, P_i]$
Per	0.80	0.70	0.50
\overline{Per}	0.00	0.00	0.40
Ψ	0.20	0.30	0.10

Tableau 3.1 : Exemple de fonctions de masse associées aux sources I , Com et $Cont$ pour une personne P_i du carnet d'adresse B .

Ainsi, chacune des sources donne naissance à une allocation de masse définie sur le référentiel Ψ , pour tout $P_i \in B$ par :

$$m^\Psi[I, P_i](\cdot) \quad \text{pour la source } I \quad (3.11)$$

$$m^\Psi[Com, P_i](\cdot) \quad \text{pour la source } Com \quad (3.12)$$

$$m^\Psi[Cont, P_i](\cdot) \quad \text{pour la source } Cont. \quad (3.13)$$

En ce qui concerne les sources relatives au contenu du document, I et Com n'ont pas le pouvoir de confirmer l'absence d'une personne dans l'image ou le fait qu'une personne n'ait pas été prononcé dans le commentaire vocal ou que cette personne n'apparaisse pas dans le commentaire textuel. En effet, si 'L'anniversaire d'*Homer*' apparaît dans les commentaires, rien ne peut permettre d'affirmer que *Lisa* ne se trouve pas dans l'image attachée au document. En fait, I et Com sont des sources non exhaustives qui ne peuvent avoir comme effet que de renforcement dans la pertinence d'une métadonnée mais ne sont pas en mesure de la contester. Ainsi, leurs fonctions de masse respectives seront déduites en allouant de la masse aux deux éléments focaux suivants (cf. section 2.2.5) :

$$N[I] = \{Per, \Psi\}$$

$$N[Com] = \{Per, \Psi\}.$$

Où $N[\cdot]$ est le noyau. Contrairement aux sources qui s'attachent à produire des métadonnées sur le contenu du document, la source relative au contexte $Cont$ peut, quant à elle, venir contester la pertinence d'une métadonnée. Ainsi, sa fonction de masse admet les trois éléments focaux suivants :

$$N[Cont] = \{Per, \overline{Per}, \Psi\}.$$

Un exemple de fonctions de masse associées aux sources I , Com et $Cont$ pour une personne P_i du carnet d'adresses B est proposé au tableau 3.1. Seule la source d'information $Cont$ est capable d'allouer une masse non nulle (0.40 dans notre exemple) lorsque P_i n'est pas pertinente en terme d'indexation. Ces mécanismes d'allocations des masses (pour les fonctions définies aux

$A \subseteq \Psi$	Contenu		Contexte	Après fusion	
	$m^\Psi[I, P_i]$	$m^\Psi[Com, P_i]$	$m^\Psi[Cont, P_i]$	$m^\Psi[., P_i]$	$BetP_{m^\Psi[., P_i]}$
Per	0.80	0.70	0.50	0.85	0.865
\overline{Per}	0.00	0.00	0.40	0.12	0.135
Ψ	0.20	0.30	0.10	0.03	-

Tableau 3.2 : Exemple de fonctions de masse associées aux sources I , Com et $Cont$, fusion $m^\Psi[., P_i]$ et probabilité pignistique $BetP_{m^\Psi[., P_i]}$ avec $\alpha_I = \alpha_{Com} = \alpha_{Cont} = 1$.

$A \subseteq \Psi$	Contenu		Contexte	Après fusion	
	$m_{0.1}^\Psi[I, P_i]$	$m_{0.1}^\Psi[Com, P_i]$	$m^\Psi[Cont, P_i]$	$m^\Psi[., P_i]$	$BetP_{m^\Psi[., P_i]}$
Per	0.08	0.07	0.50	0.54	0.59
\overline{Per}	0.00	0.00	0.40	0.36	0.41
Ψ	0.92	0.93	0.10	0.10	-

Tableau 3.3 : Exemple de fonctions de masse associées aux sources I , Com et $Cont$, fusion $m^\Psi[., P_i]$ et probabilité pignistique $BetP_{m^\Psi[., P_i]}$ avec $\alpha_I = \alpha_{Com} = 0.1$ et $\alpha_{Cont} = 1$.

équations (3.11), (3.12) et (3.13)) seront détaillés dans les chapitres 4 et 5. Le paragraphe suivant indique comment sont fusionnées ces fonctions et comment sont prises en compte leurs fiabilités.

3.4.4 Prise en compte de la fiabilité et Fusion Contenu et Contexte

A partir des fonctions de masses $m^\Psi[I, P_i]$, $m^\Psi[Com, P_i]$ et $m^\Psi[Cont, P_i]$, les fiabilités respectives des différentes sources sont prises en compte. Ceci constitue une des particularités du système **FUSINDEX** qui est un avantage certain avant le mécanisme de fusion. Ainsi, les fonctions de masse définies aux équations (3.11), (3.12) et (3.13) sont affaiblies pour donner naissance aux fonctions de masse $m_{\alpha_I}^\Psi[I, P_i]$, $m_{\alpha_{Com}}^\Psi[Com, P_i]$ et $m_{\alpha_{Cont}}^\Psi[Cont, P_i]$ où α_I , α_{Com} et α_{Cont} représentent des coefficients $\in [0, 1]$ permettant de modéliser la fiabilité des sources I , Com et $Cont$. Les tableaux 3.2 et 3.3 donnent deux exemples de fonctions de masse associées aux sources I , Com et $Cont$ pour des valeurs de coefficients de fiabilité $\alpha_I = \alpha_{Com} = \alpha_{Cont} = 1$ pour le tableau 3.2 et $\alpha_I = \alpha_{Com} = 0.1$, $\alpha_{Cont} = 1$ pour le tableau 3.3. Dans le cas où les informations extraites du contenu sont peu fiables ce qui correspond à $\alpha_I = \alpha_{Com} = 0.1$, on peut s'apercevoir que les masses ont été transférées à l'incertitude totale (tableau 3.3, $m_{0.1}^\Psi[I, P_i](\Psi) = 0.92$ et $m_{0.1}^\Psi[Com, P_i](\Psi) = 0.93$). Ainsi, la contribution des sources I et Com est diminuée dans le processus de fusion.

$A \subseteq \Psi$	$m^\Psi[., Homer]$	$m^\Psi[., Marge]$	$m^\Psi[., Bart]$	$m^\Psi[., Lisa]$	$m^\Psi[., Maggie]$
Per	0.78	0.07	0.94	0.55	0.15
\overline{Per}	0.18	0.87	0.04	0.05	0.75
Ψ	0.04	0.06	0.02	0.40	0.10

Tableau 3.4 : Exemple de fonctions de masse $m^\Psi[., P_i]$ associées aux différents prénoms du carnet d'adresses $B = \{Homer, Marge, Bart, Lisa, Maggie\}$

	$BetP_{m^\Psi[., Homer]}$	$BetP_{m^\Psi[., Marge]}$	$BetP_{m^\Psi[., Bart]}$	$BetP_{m^\Psi[., Lisa]}$	$BetP_{m^\Psi[., Maggie]}$
Per	0.80	0.10	0.95	0.75	0.20
\overline{Per}	0.20	0.90	0.05	0.25	0.80

Tableau 3.5 : Exemple de probabilités pignistiques $BetP_{m^\Psi[., P_i]}$ associées aux fonctions de masse du tableau 3.4 pour les prénoms du carnet d'adresses B

Ces trois fonctions sont ensuite fusionnées avec l'opérateur de combinaison de Dempster défini par :

$$m^\Psi[., P_i] = m_{\alpha_I}^\Psi[I, P_i] \oplus m_{\alpha_{Com}}^\Psi[Com, P_i] \oplus m_{\alpha_{Cont}}^\Psi[Cont, P_i]. \quad (3.14)$$

La fonction de masse issue de la fusion $m^\Psi[., P_i]$ ainsi que sa probabilité pignistique $BetP_{m^\Psi[., P_i]}$ associée sont respectivement données pour des valeurs de coefficients de fiabilité $\alpha_I = \alpha_{Com} = \alpha_{Cont} = 1$ pour le tableau 3.2 et $\alpha_I = \alpha_{Com} = 0.1, \alpha_{Cont} = 1$ pour le tableau 3.3. Dans le cas où les sources sont fiables, le degré de pertinence associé à la personne P_i est très élevé ($BetP_{m^\Psi[., P_i]}(Per) = 0.95$, tableau 3.2) ce qui n'est pas le cas lorsque le contenu extrait du document n'est pas fiable ($BetP_{m^\Psi[., P_i]}(Per) = 0.59$, tableau 3.3).

Finalement les métadonnées (les personnes) seront listées et présentées à l'utilisateur en fonction de leur degré de pertinence qui est donc calculé via la probabilité pignistique de l'hypothèse Per . D'une manière générale pour l'image à indexer I_k , la variable $\tau_{k,i}$ du vecteur T_k est définie par :

$$\tau_{k,i} = BetP_{m^\Psi[., P_i]}(Per). \quad (3.15)$$

avec laquelle on peut associer l'intervalle de confiance ($bel^\Psi[., P_i](Per), pl^\Psi[., P_i](Per)$) pour chacune des personnes P_i de la base B . Deux exemples de fonctions de masse $m^\Psi[., P_i]$ liées aux différents prénoms du carnet d'adresses $B = \{Homer, Marge, Bart, Lisa, Maggie\}$ ainsi que les probabilités pignistiques $BetP_{m^\Psi[., P_i]}$ associées sont donnés aux tableaux 3.4 et 3.5. Les métadonnées présentées à l'utilisateur sont ensuite triées en fonction de leur degré de pertinence en supposant qu'en dessous d'un niveau donné S (fixé à 0.6 dans le cadre de cet exemple), la métadonnée n'est plus pertinente. Pour le tableau 3.5, on aura les métadonnées suivantes à présenter

à l'utilisateur :

1. Indexer sous le *topic Bart* : oui/non ? (Pertinence = 0.95)
2. Indexer sous le *topic Homer* : oui/non ? (Pertinence = 0.80)
3. Indexer sous le *topic Lisa* : oui/non ? (Pertinence = 0.75)

ce qui pourrait correspondre au vecteur T_k de la forme $T_k = (0.8, 0.1, 0.95, 0.75, 0.2)^t$.

Ce mécanisme d'évaluation de la pertinence des métadonnées et de prise en compte de la fiabilité des sources dans le processus de fusion confère au système **FUSINDEX** un avantage indéniable d'un point de vue de l'évaluation. En effet, si, pour différentes raisons, la pertinence attachée à la génération d'une métadonnée n'est pas suffisante, le système évite de donner une réponse en terme d'indexation trop peu fiable pour qu'elle puisse être présentée à l'utilisateur. Ceci nous paraît être important dans un système semi-automatique d'indexation où il peut être demandé à l'utilisateur d'annoter lui-même les documents qui posent problème quant à leur indexation.

3.5 Conclusion

Dans ce chapitre, nous avons présenté de manière générale l'architecture d'un système de FUSion d'INDEX, baptisé **FUSINDEX** qui permet l'indexation de documents multimédia. Dans cette description, nous nous sommes volontairement limités à la production de métadonnées liées aux personnages d'un carnet d'adresses. Dans ce sens, notre problématique consiste à produire un ensemble de prénoms qui permettent d'indexer le document sous un *topic* prédéfini par l'utilisateur.

Le système exploite des données fournies par un ensemble de moteurs d'indexation qui ont été brièvement décrits. Ces moteurs sont capables d'extraire de l'information à partir de l'analyse de l'image et des commentaires vocaux et textuels attachés au document, rendant le système flexible à l'intégration de différents médias. Le système tire profit d'une organisation hiérarchique des informations qui permet d'aggréger des informations extraites du contenu et du contexte multimédia. En ce qui concerne le contexte, nous avons choisi de représenter les informations provenant des images déjà indexées en s'appuyant sur les dates et les lieux comme informations supplémentaires. Le processus pour produire et quantifier la pertinence des métadonnées est basé sur la représentation des incertitudes par des fonctions de croyance dans un cadre de discernement adapté.

Néanmoins, l'application des fonctions de croyance dans un système d'indexation de documents multimédia soulève plusieurs problèmes :

- Il n'existe pas de méthode générique pour modéliser les fonctions de croyance. Si la plupart des travaux traitant des applications de la théorie des fonctions de croyance définissent des modèles adaptés au problème spécifique, très peu tentent de proposer une méthode ou une approche plus générale. Nous verrons que cette modélisation reste adéquate aux informations disponibles en sortie des moteurs d'indexation.
- La gestion du conflit lors de la combinaison de fonctions de croyance reste un problème non totalement traité. Les contradictions entre les sources, qui peuvent provenir d'un problème mal posé ou de sources non fiables, engendrent une masse conflictuelle (conflit) lors de la combinaison. L'analyse du conflit peut permettre de lever certaines ambiguïtés et de prendre une décision plus robuste. Cette information sera exploitée dans le système d'indexation proposé.

Dans les chapitres suivants, nous présenterons en détails les techniques d'exploitation du contenu (cf. chapitre 4) et du contexte (cf. chapitre 5) pour quantifier la pertinence des métadonnées. Celles-ci passent par des étapes de modélisation des fonctions de croyance (calcul de la masse à partir des distances ou des scores, de l'affaiblissement, ...), de fusion d'information et de gestion du conflit associé.

Chapitre 4

FUSINDEX et l'indexation par le contenu

Dans ce chapitre, nous allons décrire la façon dont les éléments d'information relatifs au contenu d'un document peuvent être modélisés sous la forme de fonctions de croyance. Notons que ce genre d'approche a été utilisé avec succès dans d'autres applications de recherche d'information [72], notamment pour l'aide à l'indexation de documents XML. Les paragraphes qui suivent décrivent les méthodologies exploitées pour transformer les scores et les distances en degrés de croyance et ainsi représenter les incertitudes liées à chaque source d'information I (cf. section 4.2) et Com (cf. section 4.3). En premier lieu, on aborde les moteurs d'indexation qui analysent l'image associée au document comme DV , RV et H/F . Une nouvelle stratégie de la gestion entre les zones détectées est également proposée comme source d'information IZ pour améliorer l'indexation. On détaille ensuite les moteurs qui analysent les commentaires RPa et RPt en précisant la technique mise en œuvre pour la détection des nouveaux prénoms dans **FUSINDEX**. Afin d'illustrer l'extraction du contenu de chaque document, un exemple d'application numérique est enfin proposé à la section 4.4 avant de conclure (cf. section 4.5).

4.1 Introduction

Nous proposons ici, de détailler les étapes d'extraction du contenu de chaque document dans la base. Nous verrons comment sont modélisées les informations sous forme de fonctions de croyance pour chacun des moteurs d'indexation. Les notations suivantes seront utilisées :

$$E = \{I_1, I_2, \dots, I_k, \dots, I_K\}. \quad (4.1)$$

pour la base des documents à indexer et

$$B = \{P_1, P_2, \dots, P_i, \dots, P_n\}. \quad (4.2)$$

pour la base des personnes connues également appelée 'carnet d'adresses'. Chaque personne P_i peut être représentée dans la base B par le triplet d'informations suivant :

$$P_i = (V_i, N_i, F_i). \quad (4.3)$$

avec :

- $V_i = \{V_{i,1}, V_{i,2}, \dots, V_{i,h(i)}\}$ l'ensemble des $h(i)$ visages de la personne P_i ;
- N_i : le nom, le prénom ou le surnom de la personne P_i ;
- F_i : la probabilité que la personne P_i soit une femme ;
- i : l'identifiant de la personne P_i pour $i = 1, 2, \dots, n$.

Les visages $V_{i,}$ sont utilisés par le reconnaiseur de visages en terme d'apprentissage. Dans ce but et à partir de la base des personnes connues B , on construit un ensemble, appelé cadre de discernement, noté Ω , de la façon suivante :

$$\Omega = \{P_1, P_2, \dots, P_i, \dots, P_n, Inconnu, Inconnue, *\}. \quad (4.4)$$

où *Inconnu* (respectivement *Inconnue*) représente une personne inconnue (c'est-à-dire une personne hors du carnet d'adresses) de sexe masculin (respectivement féminin) et où le symbole $\{*\}$ est utilisé pour matérialiser le fait que la zone détectée par le moteur d'indexation DV ne contient pas de visage, afin que Ω soit exhaustif. Autrement dit, on va travailler dans le monde fermé.

Dans le cadre de ce travail de thèse, les modèles choisis pour les allocations de masses sont fortement inspirés du modèle des distances de *Th. Denœux* [73] qui évalue l'appartenance d'un document I_k à un concept d'indexation par une fonction Φ de la distance d définie par :

$$\Phi(d(I_k, P_i)) = \tau \exp^{-\delta \cdot d^r}. \quad (4.5)$$

où r , τ et δ sont deux paramètres à déterminer. La fonction 4.5, qui possède des propriétés intéressantes (décroissante, comprise entre 0 et 1, ...), a été utilisée avec succès dans plusieurs applications [74, 75]. Dans notre application, la distance d peut être remplacées par un score. Nous verrons comment ce modèle peut être exploité pour modéliser le contenu d'un document multimédia en vue de son indexation dans le cadre d'une base de 200 images associées à un carnet d'adresses de 15 personnes. En phase d'apprentissage, chacune des personnes possède quatre images de son visage dans la base qui est utilisée par le moteur d'indexation RV .

4.2 Modélisation du contenu à partir de l'image : I

L'exploitation des informations extraites de l'image I est détaillée dans ce paragraphe. Elle est basée sur les résultats fournis par les moteurs d'indexation DV , RV et H/F mais également sur un moteur permettant d'analyser les conflits entre zones détectées IZ .

4.2.1 Détecteur de visage : DV

Soit une zone Z_j détectée dans l'image associée au document I_k . On cherche à représenter l'incertitude inhérente à l'information fournie par le détecteur DV par une fonction de masse qui sera notée $m_{\alpha_{DV}}^{\Omega}[DV, Z_j]$ définie sur Ω . Le moteur DV donne un score $s_{DV}(Z_j) \geq 0$ sur l'hypothèse qu'une zone donnée Z_j englobe effectivement un visage. Plus la valeur de ce score est élevée, plus la détection est bonne.

La figure 4.1 montre le résultat des tests sur la base d'apprentissage comportant 171 zones détectées. Sur ce graphique, chaque zone est présentée par un point dans un espace à deux

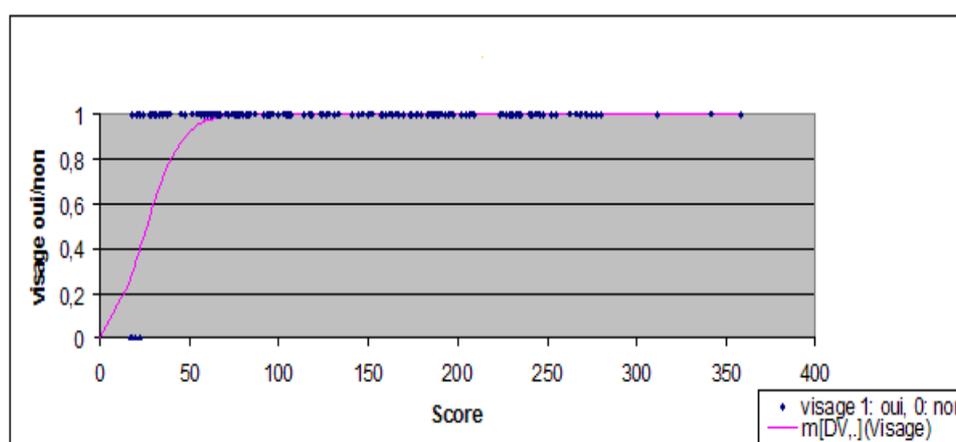


Figure 4.1 : Evaluation du détecteur de visages DV en fonction du score s_{DV} fourni par le moteur, allocation de masse associée au détecteur DV .

dimensions où l'axe des abscisses représente les scores attribués aux zones par le détecteur DV et l'axe des ordonnées représentent les valeurs binaires de bonne détection. On attribue la valeur binaire 1 en cas de bonne détection, 0 sinon. A partir de la figure 4.1, on remarque que :

- le taux d'erreur pour des valeurs de scores compris dans l'intervalle $[17, 23]$ est de $2/3$;
- qu'à partir d'une valeur de score supérieure à 30, le détecteur DV commet peu d'erreurs ;
- qu'au delà d'un score $s_{DV} > 100$, le pourcentage de bonne détection est de 100%.

On signale ici qu'un seuil sur les scores a été déjà fixé (≈ 15) par les concepteurs de l'outil *DV* pour lequel la détection n'est pas fiable.

Ces constats nous amènent à évaluer le degré de croyance associé au détecteur de visage *DV* dans le cadre de discernement Ω . Parmi les parties de Ω , on appelle *Visage* le sous-ensemble défini par :

$$Visage = \{P_1, P_2, \dots, P_i, \dots, P_n, Inconnu, Inconnue\}. \quad (4.6)$$

Ainsi, l'allocation de masse générée par le *DV* pour la zone Z_j peut être formulée par :

$$m^\Omega[DV, Z_j](Visage) = 1 - \exp^{-\eta \cdot s_{DV}(Z_j)^2} \quad (4.7)$$

$$m^\Omega[DV, Z_j](*) = \exp^{-\eta \cdot s_{DV}(Z_j)^2}. \quad (4.8)$$

avec η un paramètre positif à déterminer. A partir des remarques précédentes sur les résultats de la base d'apprentissage, nous avons choisi de fixer la valeur :

$$\eta = -\frac{\ln(\frac{2}{3})}{20^2}.$$

qui permet d'ajuster la courbe en fonction du taux de bonne reconnaissance. Cette dernière fonction est illustrée sur le graphique de la figure 4.1. La prise en compte de la fiabilité du détecteur est assurée par l'affaiblissement de la fonction de masse $m^\Omega[DV, Z_j]$ qui permet de modéliser la croyance dans le fait que la zone Z_j est bien relative à la présence d'un visage. La fonction de masse obtenue, notée $m_{\alpha_{DV}}^\Omega[DV, Z_j]$, peut s'écrire :

$$m_{\alpha_{DV}}^\Omega[DV, Z_j](A) = \alpha_{DV} m^\Omega[DV, Z_j](A) \quad \forall A \subsetneq \Omega \quad (4.9)$$

$$m_{\alpha_{DV}}^\Omega[DV, Z_j](\Omega) = 1 - \alpha_{DV}. \quad (4.10)$$

où $\alpha_{DV} \in [0, 1]$ représente le coefficient de fiabilité associé au moteur d'indexation *DV* (cf. section 2.2.4).

4.2.2 Reconnaisseur de visage : *RV*

Pour l'apprentissage du moteur d'indexation, on dispose d'un ensemble d'images de visages V_i pour chaque personne P_i du carnet d'adresses. A partir de cette base d'apprentissage, le moteur fournit une distance $d_{i,j}$ entre la zone Z_j (détectée et localisée par le détecteur *DV*) et l'ensemble des visages V_i de la personne P_i [66, 70]. Plus la distance $d_{i,j}$ est petite, plus la reconnaissance de la personne P_i dans la zone Z_j est bonne.

Pour évaluer cet outil, nous avons réalisé un test sur la base présentée précédemment. L'histogramme de la figure 4.2 fait apparaître les pourcentages de bonne reconnaissance selon le rang

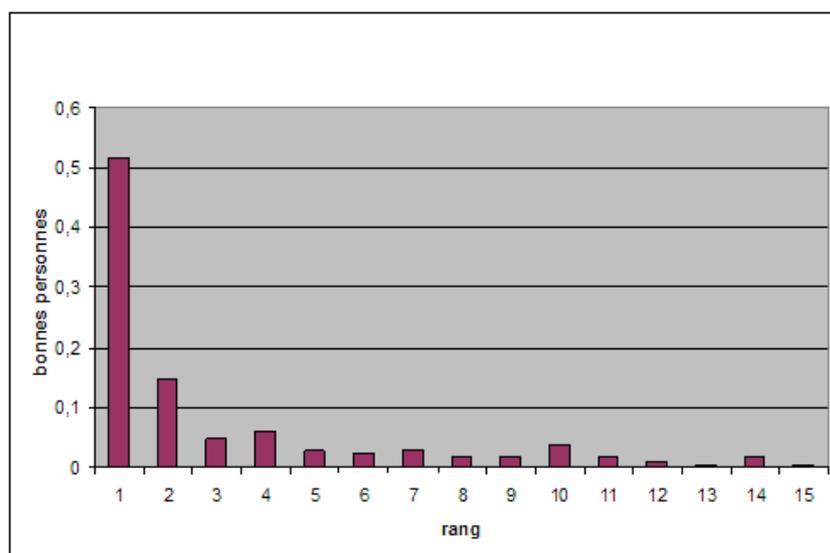


Figure 4.2 : Taux de bonne reconnaissance en fonction du rang des visages bien reconnus par le reconnaiseur de visages *RV*.

calculé par le moteur *RV*. On constate, à partir de cet histogramme, que plus de 50% des réponses sont classées en premier rang. Même si ce taux de bonne reconnaissance varie d'une base d'apprentissage à une autre, le problème parmi ces 50% de visages bien reconnus provient du fait que les distances entre les personnes bien classées et celles qui sont classées en deuxième ou troisième rang sont très peu différentes. Ceci est illustré à la figure 4.3 où les rangs fournis par le moteur sont ordonnés en fonction de la distance calculée.

Pour le calcul du taux de bonne reconnaissance, nous avons choisi de découper la mesure de distance fournie par le *RV* en intervalles réguliers. Dans chaque intervalle, on peut comptabiliser le nombre de réponses et parmi celles-ci le nombre de réponses classées en premier rang. L'évaluation du taux de bonne reconnaissance dans chaque intervalle est présentée dans le tableau 4.1 dont la synthèse est effectuée dans le graphique de la figure 4.4. Ce graphique représente le taux de bonne reconnaissance calculé pour le moteur *RV* dans chaque intervalle de distance où les approximations dans les trois derniers intervalles ($[7.10^5, 8.10^5]$, $]8.10^5, 9.10^5]$ et $]9.10^5, 10.10^5]$) ont été corrigées car elles ne contiennent pas beaucoup de valeurs (cf. les trois dernières lignes du tableau 4.1). Ceci permet de calculer le pourcentage de bonne reconnaissance en fonction de la distance fournie par régression. La fonction à estimer est de la forme :

$$\tau \exp^{-\delta \cdot d(Z_j, P_i)} . \quad (4.11)$$

ce qui permet d'obtenir $\tau = 2.254$ et $\delta = 4.10^{-6}$. Ainsi, le degré de croyance associé au fait que la personne P_i a été reconnue dans la zone Z_j sera calculé à partir d'un score, noté $s(Z_j, P_i)$,

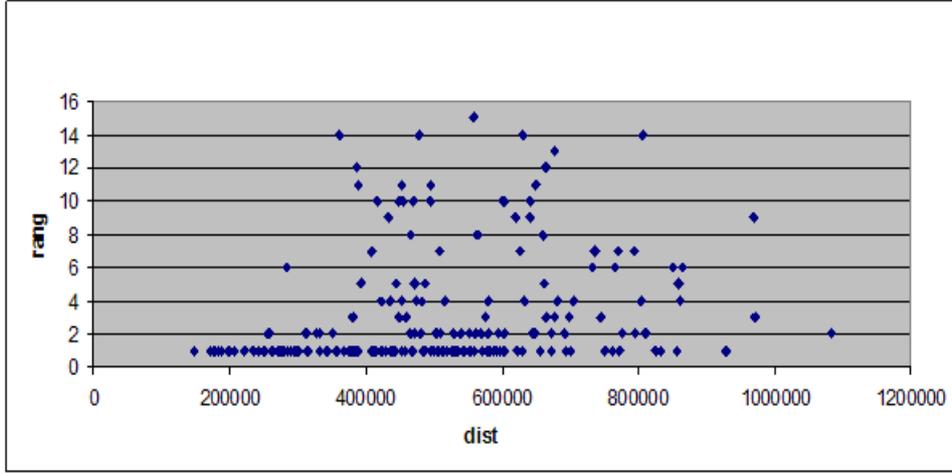


Figure 4.3 : Evolution du rang de classement du reconnaiseur de visage en fonction de la distance fournie par le moteur d'indexation RV .

défini par :

$$s(Z_j, P_i) = \min(2.254 \exp^{-4.10^{-6}d(Z_j, P_i)}, 1). \quad (4.12)$$

qui est compris dans l'intervalle $[0, 1]$. Cette fonction de score, calculée pour toutes les personnes P_i , peut être interprétée comme une distribution de possibilité qui sera notée $\pi(Z_j)$. Cette distribution $\pi(Z_j)$ pour chaque zone Z_j peut être transformée en une fonction de masse dont les éléments focaux sont emboîtés [76]. Soient $\pi_1(Z_j) > \dots > \pi_r(Z_j)$ les valeurs distinctes prises par $\pi(Z_j)$ triées en ordre décroissant avec $\pi_{r+1}(Z_j) = 0$ par convention. Alors la fonction de masse m_π^Ω induite par la distribution de possibilité $\pi(Z_j)$ peut être définie pour tout sous-ensemble $A \subseteq \Omega$ non vide par :

$$m_\pi^\Omega(A) = \begin{cases} \pi_l(Z_j) - \pi_{l+1}(Z_j) & \text{si } A = A_l, l = 1, \dots, r \\ 0 & \text{sinon.} \end{cases} \quad (4.13)$$

avec A_l la π_l coupe du sous-ensemble flou correspondant. On peut remarquer que $\sum_{A \subseteq \Omega} m_\pi^\Omega(A) = \pi_1(Z_j)$ et donc que si la valeur maximale des scores est inférieure à 1, la fonction de masse m_π^Ω obtenue ne sera pas normalisée. A titre d'exemple, soient le cadre de discernement $\Omega = \{P_1, P_2, P_3, P_4\}$ et les scores $\pi(Z_j) = (0.1, 0.5, 0.9, 0)$. Les π_l coupes de niveau 0.9, 0.5 et 0.1 sont respectivement $\{P_3\}$, $\{P_2, P_3\}$ et $\{P_1, P_2, P_3\}$ ce qui conduit au jeu de masses suivant :

$$\begin{aligned} m_\pi^\Omega(P_3) &= 0.9 - 0.5 = 0.4 \\ m_\pi^\Omega(P_2 \cup P_3) &= 0.5 - 0.1 = 0.4 \\ m_\pi^\Omega(P_1 \cup P_2 \cup P_3) &= 0.1 - 0 = 0.1 \\ m_\pi^\Omega(\emptyset) &= 1 - \pi_1(Z_j) = 0.1. \end{aligned}$$

Intervalle de distances	Nbr. Rép.	Nbr. Bon. Rép.	Taux de Bon. Rec.
$[1.10^5, 2.10^5]$	6	6	1
$[2.10^5, 3.10^5]$	24	21	0.875
$[3.10^5, 4.10^5]$	25	15	0.6
$[4.10^5, 5.10^5]$	40	18	0.45
$[5.10^5, 6.10^5]$	23	39	0.59
$[6.10^5, 7.10^5]$	24	7	0.3
$[7.10^5, 8.10^5]$	13	4	0.3→0.2
$[8.10^5, 9.10^5]$	3	1	0.33→0.1
$[9.10^5, 10.10^5]$	1	0	0→0.05

Tableau 4.1 : Intervalle de distance et bonnes réponses associées.

Afin de normaliser la fonction de masse issue de la distribution de possibilité, nous avons opté pour la solution suivante. Il faut, en effet, prendre en compte le fait que la base d'apprentissage du reconnaiseur de visages peut ne pas être exhaustive. Ainsi, si tous les scores obtenus sont petits devant 1, il est possible que le visage présent dans l'image ne soit pas un visage connu. Au niveau de la fonction de masse m_π^Ω , ceci se matérialise par une valeur importante sur l'ensemble vide. Puisque nous avons prévu cette hypothèse dès l'étape de modélisation ($\{Inconnue, Inconnue, *\}$), nous avons donc choisi de modéliser le jeu de masses de la façon suivante :

$$m^\Omega[RV, Z_j](A) = m_\pi(A) \quad A \subseteq \Omega, A \neq \emptyset \quad (4.14)$$

$$m^\Omega[RV, Z_j](Inconnu \cup Inconnue \cup *) = m_\pi(\emptyset). \quad (4.15)$$

Enfin, la fiabilité du reconnaiseur permet d'affaiblir $m^\Omega[RV, Z_j]$ pour obtenir la fonction de masse $m_{\alpha_{RV}}^\Omega[RV, Z_j]$ qui peut s'écrire :

$$m_{\alpha_{RV}}^\Omega[RV, Z_j](A) = \alpha_{RV} m^\Omega[RV, Z_j](A) \quad \forall A \subsetneq \Omega \quad (4.16)$$

$$m_{\alpha_{RV}}^\Omega[RV, Z_j](\Omega) = 1 - \alpha_{RV} + \alpha_{RV} m^\Omega[RV, Z_j](\Omega). \quad (4.17)$$

où $\alpha_{RV} \in [0, 1]$ représente la fiabilité du moteur d'indexation RV (cf. section 2.2.4).

4.2.3 Reconnaisseur de genre homme/femme : H/F

L'outil d'indexation H/F consiste à identifier le sexe de la personne qui se trouve dans la zone Z_j afin de favoriser les personnes du carnet d'adresses B de même sexe. Pour quantifier l'incertitude inhérente au sexe de la personne, nous allons définir son degré de croyance dans un

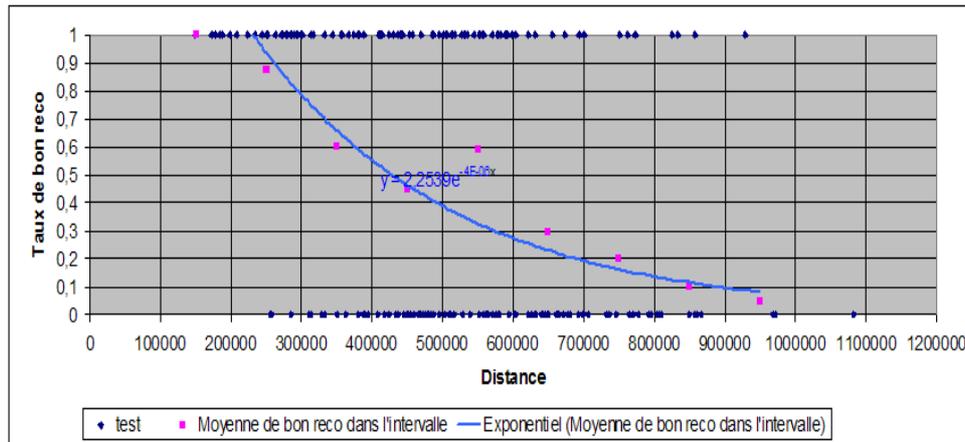


Figure 4.4 : Taux de bonne reconnaissance calculé pour le moteur *RV* dans chaque intervalle de distance et courbe de régression choisie pour approximer ce taux.

cadre de discernement, noté Θ , défini par :

$$\Theta = \{H, F\}. \quad (4.18)$$

où H représente l'hypothèse d'un visage masculin et F un visage féminin. Pour déterminer ce degré de croyance, l'outil de reconnaissance de visages a été un peu détourné afin de réaliser un reconnaiseur de genre homme/femme (H/F). Pour ce faire, on demande au *RV* de reconnaître le visage entre un groupe d'hommes et un groupe de femmes. Initialement, le sexe des personnes contenues dans le carnet d'adresses est inconnu. Afin d'utiliser l'outil H/F dans le processus d'identification de la zone Z_j , il faut donc une étape préliminaire (qui se déroule hors-ligne) consistant à déterminer le genre H/F des personnes du carnet d'adresses. Plus précisément, cette étape détermine la probabilité qu'une personne connue P_i soit une femme F_i ou un homme $H_i = 1 - F_i$. Dans ces paragraphes, on illustre en premier lieu le processus d'identification du genre H/F dans la zone Z_j en supposant la connaissance des probabilités F_i pour chaque personne connue P_i du carnet d'adresses. Une méthode pour déterminer ces probabilités F_i est ensuite proposée.

Identification du genre H/F dans la zone Z_j

L'outil d'indexation H/F donne une information sur le sexe de la personne dans une zone Z_j mais n'est pas capable de donner une information plus précise sur son identité. En effet, si l'outil émet l'hypothèse que la personne détectée est du sexe masculin, on ne peut pas favoriser dans la reconnaissance de l'identité toutes les personnes de sexe masculin dans le carnet d'adresses. Par contre, l'idée défendue avec l'utilisation de cet outil est de défavoriser les personnes de sexe

féminin du carnet d'adresses. Nous allons voir comment ceci est possible avec les fonctions de croyance.

Soient $d_{h,j}$ et $d_{f,j}$ les distances fournies par l'outil H/F représentant respectivement la distance entre la personne détectée dans la zone Z_j et le groupe des hommes (respectivement des femmes) dans la base des visages connus. A partir de ces deux distances, on définit l'écart absolu, noté $d_{hf,j}$ par :

$$d_{hf,j} = |d_{h,j} - d_{f,j}|. \quad (4.19)$$

Nous avons choisi de travailler à partir de cet écart suite aux expérimentations faites sur le moteur. Sur la figure 4.5, nous avons représenté sur les 200 images testées la valeur binaire 1/0 ("1" en cas de bonne détection du genre H/F , "0" sinon) en fonction de cet écart $d_{hf,.}$. On

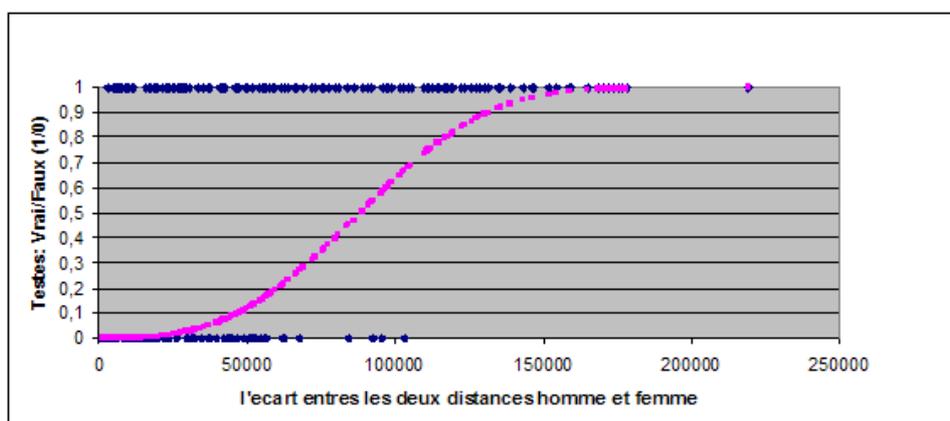


Figure 4.5 : Bonne reconnaissance en fonction de la différence entre les distances respectives 'homme' et 'femme' $|d_{h,j} - d_{f,j}|$.

constate visuellement que, pour des valeurs inférieures à 57000, il y a une grande incertitude sur les résultats donnés par l'outil H/F . Par contre, à partir de cette valeur, l'incertitude commence à diminuer jusqu'à sa disparition à partir de 100000 où le taux de bonne détection est de 100%. L'idée consiste à prendre en compte cette fiabilité variable (en fonction de la distance $d_{hf,.}$) dans la modélisation de la sortie du moteur H/F . Il s'agit d'allouer un degré de croyance à l'hypothèse F (ou H) dont la valeur augmente en fonction de la fiabilité et donc de la distance $d_{hf,.}$ comme il est illustré sur la courbe de la figure 4.5. Ainsi, plus la distance $d_{hf,.}$ est grande, plus l'incertitude sur la sortie fournie par l'outil est faible et plus le système est fiable. Il nous faut donc modéliser via une fonction de croyance définie sur Θ (permettant de représenter cette incertitude) le genre de la personne détectée dans une zone Z_j . Puisque cette fonction est définie sur Θ , elle sera complètement indépendante du carnet d'adresses. Nous avons choisi de modéliser

le jeu de masses associé par :

$$m^\Theta[H/F, Z_j](A) = 1 - \rho_{H/F} \exp^{-\mu_{H/F} d_{h,f,j}^{r_{H/F}}} \quad (4.20)$$

$$m^\Theta[H/F, Z_j](\Theta) = 1 - m^\Theta[H/F, Z_j](A). \quad (4.21)$$

où $A \subseteq \Theta$ désigne le singleton $\{H\}$ ou $\{F\}$ dont la distance $d_{h,j}$ respectivement $d_{f,j}$ est la plus faible. Dans l'implémentation, on prend les valeurs suivantes : $r_{H/F} = 3$, $\rho_{H/F} = 1$ et $\mu_{H/F} = 10^{-7}$. On déduit de cette fonction la probabilité pignistique sur l'hypothèse F pour la zone Z_j :

$$\text{Bet}P_{m^\Theta[H/F, Z_j]}(\{F\}). \quad (4.22)$$

qui permet de quantifier la probabilité que la personne détectée dans la zone soit de sexe féminin.

Passage au cadre de discernement Ω

Le schéma de la figure 4.6 montre les étapes du calcul pour l'outil H/F pour identifier la zone Z_j . La seconde étape dans la reconnaissance du genre consiste, comme nous l'avons expliqué, à

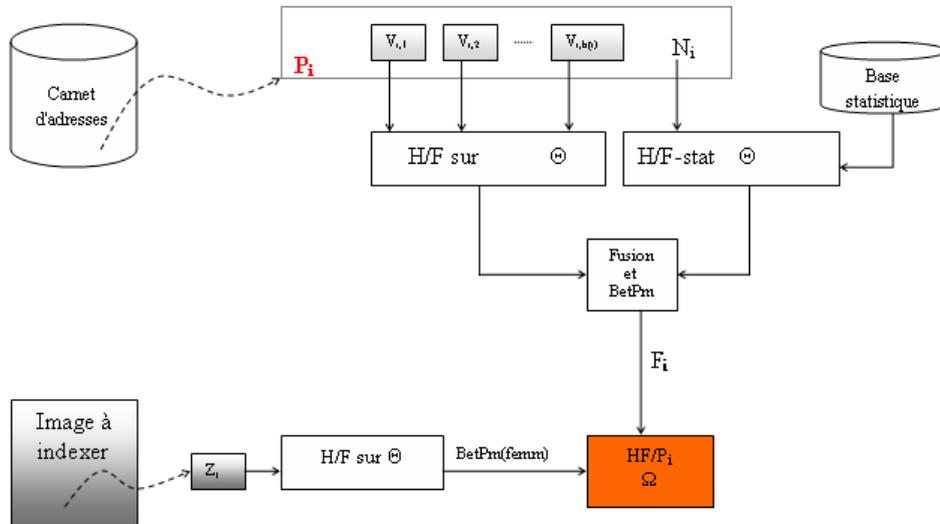


Figure 4.6 : Étapes pour construire le jeu de masses lié au reconnaiseur HF/P_i .

défavoriser les personnes (du carnet d'adresses) de sexe différent du sexe détecté dans la zone Z_j . Le but est donc d'utiliser l'outil H/F dans le processus d'identification des personnes, autrement dit dans Ω .

A ce stade, les informations disponibles sont les probabilités F_i associées aux personnes P_i de Ω (y compris les hypothèses $\{Inconnue\}$ et $\{Inconnu\}$) et la probabilité pignistique

$BetP_{m^\ominus[H/F, Z_j]}$ quantifiant le genre H/F détecté dans la zone Z_j . A partir de ces informations, nous allons construire un jeu de masses, noté $m^\Omega[HF/\omega, Z_j]$ pour chaque singleton ω de Ω excepté $\{*\}$. Chaque jeu de masses, qui permet d'opposer la personne ω à laquelle elle est associée, est défini par :

$$m^\Omega[HF/\omega, Z_j](\bar{\omega}) = \begin{cases} -4 \kappa_\omega & \text{si } \kappa_\omega \leq 0; \\ 0 & \text{sinon;} \end{cases} \quad (4.23)$$

$$m^\Omega[HF/\omega, Z_j](\Omega) = 1 - m^\Omega[HF/\omega, Z_j](\bar{\omega}). \quad (4.24)$$

avec $\kappa_\omega = (F_i - 1/2)(BetP_{m^\ominus[H/F, Z_j]}(\{F\}) - 1/2)$ un coefficient défini pour tout $\omega \in_* \Omega$ y compris pour $\{Inconnue\}$ et $\{Inconnu\}$ avec $F_{Inconnue} = 1$ et $F_{Inconnu} = 0$. Cette allocation de masses respecte les trois axiomes suivants :

- plus l'incertitude sur le genre H/F de la personne candidate P_i est grande (ce qui se traduit par l'équiprobabilité sur F_i) moins l'opposition est forte (source d'information neutre) :

$$F_i \rightarrow 1/2 \quad \Rightarrow \quad m^\Omega[HF/\omega, Z_j](\Omega) \rightarrow 1.$$

- plus l'incertitude sur le genre H/F détecté dans la zone Z_j est grande moins l'opposition est forte (source d'information neutre) :

$$BetP_{m^\ominus[H/F, Z_j]}(\{F\}) \rightarrow 1/2 \quad \Rightarrow \quad m^\Omega[HF/\omega, Z_j](\Omega) \rightarrow 1.$$

- plus le genre H/F détecté dans la zone Z_j est différent du genre H/F de la personne candidate P_i , plus l'opposition est forte :

$$(BetP_{m^\ominus[H/F, Z_j]}(\{F\}) \rightarrow 0, F_i \rightarrow 1) \quad \Rightarrow \quad m^\Omega[HF/\omega, Z_j](\bar{\omega}) \rightarrow 1$$

$$(BetP_{m^\ominus[H/F, Z_j]}(\{F\}) \rightarrow 1, F_i \rightarrow 0) \quad \Rightarrow \quad m^\Omega[HF/\omega, Z_j](\bar{\omega}) \rightarrow 1.$$

On construit enfin le jeu de masses final associé à l'outil H/F pour chaque zone détectée de l'image à indexer par la fusion des toutes les sources HF/ω qui permet d'obtenir :

$$m_{\alpha_{H/F}}^\Omega[H/F, Z_j] = \bigoplus_{\omega \in_* \Omega} m^\Omega[HF/\omega, Z_j]. \quad (4.25)$$

avec $\alpha_{H/F}$ un coefficient compris dans $[0, 1]$ qui représente la fiabilité de l'outil H/F .

Calcul de la probabilité F_i

Nous allons détailler la façon dont peut être calculée la probabilité pour qu'une personne P_i du carnet d'adresses soit du sexe féminin F_i . Dans ce but, on utilise :

- une base statistique générale BS sur le genre homme/femme des porteurs des prénoms usuels¹. Cette source d'information notée " $H/F - stat$ " se base sur les statistiques pour calculer la probabilité F_i de la personne P_i à partir de son prénom N_i .

¹www.linternaute.com/femmes/prenoms/prenoms/1/1940-2003/5/france.shtm.

- les images d'apprentissage $V_{i,}$ concernant les visages. Par conséquent, la probabilité F_i est mise à jour à chaque fois qu'un nouveau visage de la personne P_i a été identifié.

Ces deux étapes dans le calcul de F_i sont illustrées à la figure 4.7. Ainsi, les sources d'information

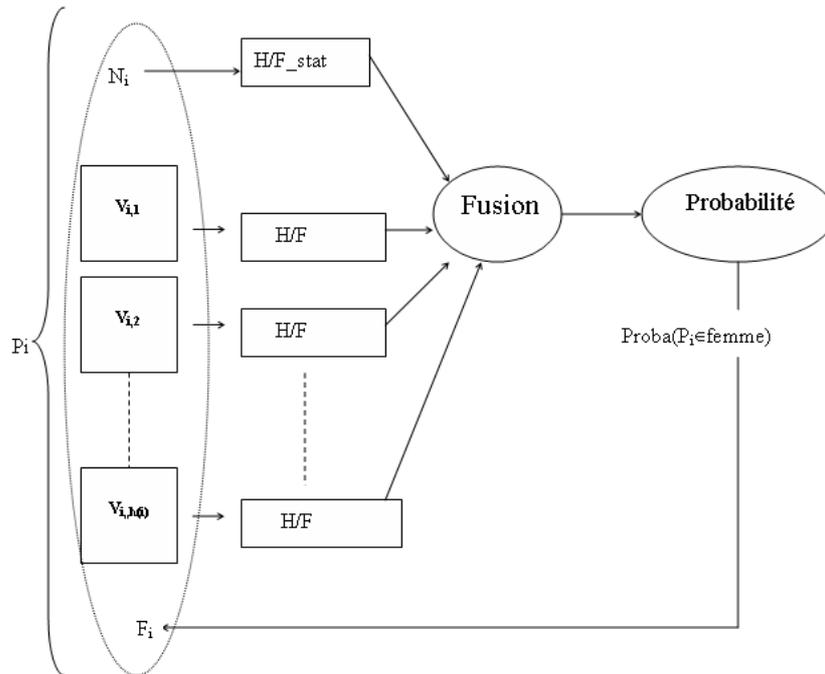


Figure 4.7 : Calcul et mise à jour de la probabilité F_i sur le sexe féminin de la personne P_i .

disponibles pour déterminer le genre Homme/Femme de la personne connue P_i sont :

- la base statistique via la source $H/F - stat$ qui permet d'obtenir une fonction de masse qui sera notée $m^\Theta[H/F - stat, N_i]$ pour chaque prénom N_i ;
- les $h(i)$ images de visages $V_{i,}$ utilisés par le moteur H/F . A chaque visage $V_{i,}$ est associé un jeu de masses $m_{\alpha_{H/F}}^\Theta[H/F, V_{i,}]$;

Pour chaque personne $P_i \in B$, nous avons $(h(i) + 1)$ sources d'informations sur le sexe de cette personne correspondant aux $h(i)$ visages dans la base d'apprentissage et à la base statistique.

La base statistique (notée BS) sur l'ensemble des prénoms de la population française nous fournit pour un prénom donné N_i les nombres de femmes et d'hommes qui portent effectivement ce prénom selon les années. A partir de ces informations, on peut calculer le degré de croyance pour que la personne qui porte ce prénom soit une femme en sachant qu'on a observé o_i fois le prénom N_i , et que sur les o_i observations on a eu o_{hi} hommes et o_{fi} femmes avec $o_i = o_{hi} + o_{fi}$. Pour construire une fonction de croyance définie sur Θ , la masse peut être formulée de la façon

suivante [77, 78] :

$$m^\Theta[H/F - stat, N_i](F) = \frac{o_{fi}}{o_i + 1} \quad (4.26)$$

$$m^\Theta[H/F - stat, N_i](H) = \frac{o_{hi}}{o_i + 1} \quad (4.27)$$

$$m^\Theta[H/F - stat, N_i](\Theta) = \frac{1}{o_i + 1}. \quad (4.28)$$

On remarque que cette fonction de masse possède les caractéristiques d'une probabilité lorsque o_i tend vers l'infini et permet de quantifier l'incertitude si le nombre de prénoms observé n'est pas suffisant. Pour le calcul de $m_{\alpha_{H/F}}^\Theta[H/F, V_{i,q}]$ (avec $q = 1, 2, \dots, h(i)$), on utilise le même principe que pour l'identification du sexe H/F dans une zone Z_j . Ainsi, l'équation (4.20) est détournée pour permettre cette fois d'identifier le sexe associé au visage $V_{i,q}$ dans le référentiel Θ .

Ces deux fonctions, l'une $m_{\alpha_{H/F}}^\Theta[H/F, V_{i,q}]$ associée à P_i via son visage $V_{i,q}$ et l'autre $m^\Theta[H/F - stat, N_i]$ associée à P_i via son prénom N_i vont permettre d'identifier le sexe de la personne P_i . De la fusion des fonctions de croyance précédemment définies résulte le jeu de masses suivant :

$$m^\Theta[H/F, P_i] = m_{\alpha_{H/F}}^\Theta[H/F, V_{i,1}] \oplus \dots \oplus m_{\alpha_{H/F}}^\Theta[H/F, V_{i,h(i)}] \oplus m^\Theta[H/F - stat, N_i]. \quad (4.29)$$

quantifiant le degré de croyance relatif au sexe de P_i ayant observé les informations H/F disponibles. A partir de cette dernière équation, on peut obtenir la probabilité F_i pour chaque personne P_i par la transformation pignistique définie par :

$$F_i = BetP_{m^\Theta[H/F, P_i]}(\{F\}). \quad (4.30)$$

qui permet d'obtenir une information fiable sur le sexe de la personne P_i .

4.2.4 Source d'information I_Z pour la gestion des zones détectées

Dans ce paragraphe, on cherche à définir une stratégie pour faire propager l'information entre les différentes zones de l'image. On considère l'image comme l'union des zones détectées $Z = \bigcup_{j=1}^z Z_j$ et d'une zone, notée Z^C qui représente le complémentaire de Z dans l'image. En d'autres termes, cette zone Z^C correspond à la région de l'image où le moteur DV n'a détecté aucun visage. Chaque zone Z_j pourrait contenir au maximum un seul visage de par la nature de la technologie employée par le moteur. En revanche, la zone Z^C peut contenir plusieurs visages dans le cas où le détecteur de visage aurait commis des non-détections.

L'idée développée ici est basée sur le principe qu'il est impossible d'avoir la même personne P_i dans deux zones différentes de l'image qui ont été détectées par le DV . Pour soutenir cette idée, on va donc disposer pour chaque zone Z_j :

- des informations apportées par les différents moteurs d'indexation (DV , RV et H/F) sur la zone Z_j elle-même ;
- des informations apportées par les autres zones $Z - Z_j$.

Ces deux catégories de sources d'information sont détaillées dans les paragraphes qui suivent.

Fusion des informations extraites de l'image pour la zone Z_j

Dans un premier temps, chaque zone détectée ne possède que sa propre information. Cette information s'appuie uniquement sur les outils d'indexation DV , RV et H/F pour identifier la personne qu'elle englobe. On désigne les trois outils d'indexation DV , RV et H/F par la source d'information I' . Ainsi, $m^\Omega[I', Z_j]$ désigne le jeu de masses issu de la fusion entre les trois fonctions relatives à ces moteurs par l'opérateur de combinaison de *Dempster*, calculé pour chaque zone Z_j par :

$$m^\Omega[I', Z_j] = m_{\alpha_{DV}}^\Omega [DV, Z_j] \oplus m_{\alpha_{RV}}^\Omega [RV, Z_j] \oplus m_{\alpha_{H/F}}^\Omega [H/F, Z_j]. \quad (4.31)$$

Cette étape de fusion permet d'agrèger les propositions générées par les trois moteurs sur chaque zone Z_j .

Fusion des informations extraites des zones voisines

Dans un second temps, on prend en compte les informations issues des zones voisines pour diminuer l'incertitude pour la zone Z_j . L'idée consiste à utiliser le fait que si une personne P_i est dans la zone Z_j , elle ne peut pas se trouver dans une autre zone $Z_{j'}$ ($j' \neq j$). Les informations extraites des zones voisines vont être propagées entre les zones afin de diminuer l'ambiguïté sur certaines zones. Soit $m_{\alpha_{Z_{j'}}}^\Omega [Z_{j'}, Z_j]$ le jeu de masse extrait de la source d'information "zone $Z_{j'}$ " avec $j' \in \{1, 2, \dots, z\}$ et $j \neq j'$. Nous avons choisi de calculer ce jeu de masse de la façon suivante :

$$m_{\alpha_{Z_{j'}}}^\Omega [Z_{j'}, Z_j](\bar{P}_i) = \alpha_{Z_{j'}} \text{bel}^\Omega [I', Z_{j'}](P_i) \quad (4.32)$$

$$m_{\alpha_{Z_{j'}}}^\Omega [Z_{j'}, Z_j](\Omega) = 1 - m_{\alpha_{Z_{j'}}}^\Omega [Z_{j'}, Z_j](\bar{P}_i). \quad (4.33)$$

où $\alpha_{Z_{j'}}$ est un coefficient de fiabilité associé à la source d'information "zone $Z_{j'}$ ". Dans l'équation (4.32), $\text{bel}^\Omega [I', Z_{j'}](P_i)$ représente la croyance totale dans le fait que P_i soit dans la zone $Z_{j'}$ basée sur les informations extraites des sources DV , RV et H/F (cf. équation (4.31)). Les informations issues de la zone $Z_{j'}$ doivent être affaiblies en fonction :

- de la valeur du conflit interne $m_{\odot}^\Omega [I', Z_{j'}](\emptyset)$ (équation (4.31) ; avant normalisation)

- de la différence entre les probabilités pignistiques $BetP_{m^\Omega[I', Z_{j'}]}$ des deux personnes les mieux classées dans la zone $Z_{j'}$.

Ainsi, la fiabilité $\alpha_{Z_{j'}}$ est définie par :

$$\alpha_{Z_{j'}} = (1 - m_{\emptyset}^\Omega[I', Z_{j'}](\emptyset)) \cdot (BetP_{m^\Omega[I', Z_{j'}]}(P_{i'}) - BetP_{m^\Omega[I', Z_{j'}]}(P_{i''})). \quad (4.34)$$

où $P_{i'}$ et $P_{i''}$ (avec $(i', i'') \in \{1, 2, \dots, n\}^2$) sont les personnes les mieux classées dans l'ordre obtenu par le tri des probabilités pignistiques. Enfin, la source IZ rassemble toutes les zones de l'image sauf la zone Z_j elle-même ce qui conduit à :

$$m^\Omega[IZ, Z_j] = \bigoplus_{j'=1,2,\dots,z/j' \neq j} m_{\alpha_{Z_{j'}}}^\Omega[Z_{j'}, Z_j]. \quad (4.35)$$

On déduit enfin la fonction de croyance issue des outils d'indexation capables d'extraire l'information dans une zone à partir de l'image par sa fonction de masse associée :

$$m^\Omega[I, Z_j] = m^\Omega[I', Z_j] \oplus m^\Omega[IZ, Z_j]. \quad (4.36)$$

Cette fonction permet de quantifier la croyance dans le fait qu'une personne est effectivement présente au sein de la zone Z_j .

4.2.5 Croyances associées aux métadonnées issues de l'image I

Les outils d'indexation DV , RV et H/F s'attachent à fournir des index pour chaque zone isolée Z_j de l'image. Cependant, notre objectif d'indexation consiste à générer les métadonnées pour l'image entière ce qui nécessite d'agrégier les informations en provenance de chaque zone Z_j , ($j = 1, \dots, z$) détectée. Il s'agit donc de passer de la croyance sur le fait qu'une personne est présente dans l'image (quantifiée par les fonctions de masse $m^\Omega[I, Z_j]$ pour chaque zone Z_j et définies dans le cadre de discernement Ω) au cadre de discernement $\Psi = \{Per, \overline{Per}\}$ qui permet de quantifier la pertinence de chacune des métadonnées proposées.

A la section 3.4.3, nous avons insisté sur le fait que la source d'information I n'était pas une source exhaustive pour laquelle les éléments focaux de la fonction de masse $m^\Psi[I, P_i]$ associée ne peuvent être que $\{Per, \Psi\}$. Nous rappelons également qu'il est nécessaire d'évaluer la pertinence pour chaque personne P_i du carnet d'adresses. Pour répondre à ce problème, la solution la plus intuitive serait de combiner (avec un opérateur conjonctif ou disjonctif) les différentes fonctions de masse $m^\Omega[I, Z_j]$ pour chaque zone. L'utilisation d'une telle solution n'est pas envisageable pour plusieurs raisons :

- La combinaison conjonctive des jeux de masse n'est pas adéquate car elle fusionne des sources qui ne traitent pas de la même zone cible. Cette fusion serait donc à l'origine de forts conflits qui n'ont pas lieu d'être ;

- La combinaison disjonctive qui semblerait plus adaptée ne permet pas de distinguer le degré de croyance associé à chacune des personnes P_i . En effet, le résultat de cette combinaison va aboutir à allouer de la masse à un sous-ensemble du type $A = \bigcup P_i$ qui ne permettra plus de distinguer avec quelle croyance chacune des personnes P_i a été identifiée. Ceci semble peu raisonnable au regard des affectations différentes de masse qui ont été déduites.

La solution envisagée consiste à utiliser tout de même un opérateur disjonctif de type max basé sur la probabilité pignistique $BetP_{m^\Omega[I, Z_j]}$ évaluée pour chaque zone Z_j . Cette solution permet de conserver une information sur le degré avec lequel la personne a été reconnue.

Ainsi, pour fournir les métadonnées pour l'image entière, on construit pour chaque personne P_i un jeu de masses $m^\Psi[I, P_i]$ qui permet d'évaluer la pertinence de la métadonnée à partir du maximum de la probabilité pignistique $BetP_{m^\Omega[I, Z_j]}$ déduite de chaque zone. Cette fonction peut être définie par :

$$m^\Psi[I, P_i](Per) = \max_j \{BetP_{m^\Omega[I, Z_j]}(P_i)\} \quad (4.37)$$

$$m^\Psi[I, P_i](\Psi) = 1 - m^\Psi[I, P_i](Per). \quad (4.38)$$

Cette solution inspirée de la théorie des possibilités [79] semble assez raisonnable puisqu'il suffit que la personne P_i ait été bien identifiée dans une zone pour dire que cette personne se trouve dans l'image.

$$m_{\alpha_I}^\Psi[I, P_i](A) = \alpha_I m^\Psi[I, P_i](A) \quad \forall A \subsetneq \Psi \quad (4.39)$$

$$m_{\alpha_I}^\Psi[I, P_i](\Psi) = 1 - \alpha_I + \alpha_I m^\Psi[I, P_i](\Psi). \quad (4.40)$$

où $\alpha_I \in [0, 1]$ représente la fiabilité du d'information "image" I (cf. section 2.2.4).

4.3 Modélisation du contenu à partir des commentaires : *Com*

Dans cette section, les résultats fournis par les deux outils d'indexation *RPa* et *RPt* sont exploités. Pour la modélisation du contenu à partir des commentaires *Com*, on utilise directement le cadre de discernement Ψ permettant de quantifier la pertinence des métadonnées. Ce choix est motivé par le fait que le document peut contenir plusieurs commentaires vocaux et textuels dans lesquels on peut retrouver plusieurs prénoms détectés. Par conséquent, l'utilisation du cadre de discernement Ω augmente la complexité en terme de calculs et nécessite la division des commentaires en plusieurs segments pour vérifier la condition d'exclusivité. Nous proposons d'utiliser les degrés de recouvrement entre les mots prononcés ou notés dans les commentaires pour réduire l'incertitude sur les métadonnées proposées à l'utilisateur.

4.3.1 Définition d'une matrice de recouvrement

L'un des problèmes rencontrés avec l'outil d'indexation *RPa* concerne les prénoms qui se prononcent presque de la même manière (ex : Gaëlle et Gaël, Joëlle et Joël, ...). Les prénoms composés (ex : Jean-Paul, Jean-Pierre, Frédérique, ...) posent également souci aux techniques utilisées pour la reconnaissance audio [80]. Les mots usuels comme fils ou ami peuvent quant à eux cibler plusieurs personnes du carnet d'adresses à la fois. Pour pallier ces inconvénients, nous proposons une solution originale basée sur la gestion de la matrice de recouvrement, notion que nous définissons dans les paragraphes suivants pour le moteur *RPa* et pour le moteur *RPt*.

Recouvrements par intervalles de temps

Pour un commentaire vocal donné, on suppose que les deux prénoms N_i et $N_{i'}$ ont été détectés dans deux intervalles de temps distincts $[t_1, t_2]$ pour N_i et $[t'_1, t'_2]$ pour $N_{i'}$. Le degré de recouvrement de N_i sur $N_{i'}$ se calcule de la manière suivante :

$$rec_e(P_i/P_{i'}) = rec_{ee}(N_i/N_{i'}) = \frac{t_\cap}{t'_2 - t'_1}. \quad (4.41)$$

avec t_\cap la longueur de l'intersection entre les deux intervalles de temps $[t_1, t_2]$ et $[t'_1, t'_2]$. Dans cette dernière équation, le temps t_\cap sera nul s'il n'y a pas de recouvrement entre les deux intervalles, soit :

$$t_\cap = 0 \text{ si } [t_1, t_2] \cap [t'_1, t'_2] = \emptyset.$$

De manière similaire, on peut définir le degré de recouvrement de $N_{i'}$ sur N_i qui est calculé par :

$$rec_e(P_{i'}/P_i) = rec_e(N_{i'}/N_i) = \frac{t_\cap}{t_2 - t_1}. \quad (4.42)$$

On peut considérer qu'à partir d'un certain pourcentage de recouvrement entre deux prénoms, les mots interférés sont en conflit total (par exemple dans le prénom composé Jean-Pierre, le prénom Jean est en conflit total avec Jean-Pierre). Pour modéliser ce conflit total, nous allons binariser chaque degré de recouvrement de la façon suivante :

$$rec_e(P_i/P_{i'}) = \begin{cases} 1 & \text{si } rec_e(P_i/P_{i'}) \geq \text{seuil}B \\ 0 & \text{sinon.} \end{cases} \quad (4.43)$$

L'indice e désigne une situation conflictuelle, tel que chaque détection est forcément appartient à une seule situation conflictuelle.

Recouvrements par détection de mots ciblant plusieurs métadonnées

Pour un commentaire vocal ou textuel donné, on suppose qu'un nom W (ex : prénom, frère, ami, ...) détecté cible à la fois les deux personnes P_i et $P_{i'}$. Le degré de recouvrement entre les deux personnes P_i et $P_{i'}$ relatif à la détection du mot W est défini par :

$$rec_e(P_i/P_{i'}) = \begin{cases} 1 & \text{si } (W \text{ cible } P_i \text{ et } P_{i'}) \\ 0 & \text{sinon.} \end{cases} \quad (4.44)$$

Ce degré de recouvrement prend donc des valeurs binaires contrairement au degré de recouvrement défini par intervalles de temps. A partir de ces degrés évalués pour une série d'éléments e , nous allons créer une matrice de recouvrement qui sera utilisée pour la gestion des mots composés ou ciblant plusieurs métadonnées.

Matrice de recouvrements

Pour assurer cette gestion, une matrice de recouvrement est créée à chacune des détections dans les commentaires vocal et textuel. Ainsi, on fait correspondre à une détection, notée e correspondant aux indices précédemment décrits une matrice de recouvrement notée Mr_e . Cette matrice rassemble l'ensemble des degrés de recouvrement calculés (et éventuellement binarisés) pour chaque paire de métadonnées P_i du carnet d'adresses. Elle est donc définie par :

$$Mr_e = \begin{pmatrix} & P_1 & P_2 & \cdots & P_i & \cdots & P_n \\ P_1 & * & rec_e(P_2/P_1) & \cdots & rec_e(P_i/P_1) & \cdots & rec_e(P_n/P_1) \\ P_2 & rec_e(P_1/P_2) & * & \cdots & rec_e(P_i/P_2) & \cdots & rec_e(P_n/P_2) \\ \vdots & & \vdots & & \vdots & & \vdots \\ P_n & rec_e(P_1/P_n) & rec_e(P_n/P_2) & \cdots & rec_e(P_i/P_n) & \cdots & * \end{pmatrix}. \quad (4.45)$$

Dans le cas où la détection (prénom ou mot ciblant plusieurs prénoms) ne génère pas de recouvrement, cette matrice présentera uniquement des valeurs nulles. Il est à noter que l'on disposera d'autant de matrices de recouvrement que de mots détectés dans les commentaires.

4.3.2 Scores

Les scores donnés par les moteurs RPa et RPt à une personne P_i du carnet d'adresses seront notés $RPa(P_i)$ et $RPt(P_i)$. Pour chaque détection dans un commentaire, l'outil d'indexation RPa (si le commentaire est vocal) ou RPt (si le commentaire est textuel) fournit un score non nul qui varie dans l'intervalle $]0, 1]$. La valeur de ce score est directement affectée à la personne P_i correspondant à la détection, le reste des personnes du carnet d'adresses obtenant un score nul.

Pour chaque détection e d'un mot dans les commentaires, on associe donc un vecteur de scores, noté $RP_e(\cdot)$, fourni par le moteur (RPa ou RPt selon le cas) dont chaque élément représente le score attribué à la personne P_i . Ces vecteurs de score RP_e pourront être rassemblés dans une matrice dont la taille dépendra du nombre de détections trouvées :

$$Scores = \begin{pmatrix} RP_1(P_1) & RP_1(P_2) & \cdots & RP_1(P_i) & \cdots & RP_1(P_n) \\ RP_2(P_1) & RP_2(P_2) & \cdots & RP_2(P_i) & \cdots & RP_2(P_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \quad (4.46)$$

4.3.3 Fusion inter-commentaires

Pour un document à indexer et à partir des matrices de recouvrement Mr_e et des vecteurs de score RP_e , il nous faut modéliser l'incertitude inhérente à l'identification d'une personne P_i pour l'indexation ayant observé les commentaires. Nous avons choisi le principe d'allocation de masses suivant :

$$m^\Psi[Com', P_i] = \bigoplus_{e, i'} m^\Psi[Mr_e^i, P_{i'}]. \quad (4.47)$$

où la fonction $m^\Psi[Com', P_i]$ quantifie la pertinence dans Ψ de la personne P_i ayant observé les commentaires Com' . Dans l'équation, on remarque que cette fonction résulte de la fusion de fonctions de masse $m^\Psi[Mr_e^i, P_{i'}]$ déduite de la matrice de recouvrement Mr_e et des scores RP_e . Cette allocation permet de quantifier la pertinence d'une métadonnée avec la gestion des situations conflictuelles (prénoms composés ou mots ciblant plusieurs métadonnées). Elle est définie de la façon suivante :

- si $P_i = P_{i'}$, c'est-à-dire lorsqu'il n'y a pas de conflit :

$$m^\Psi[Mr_e^i, P_{i'}](Per) = RP_e(P_{i'}) \quad (4.48)$$

$$m^\Psi[Mr_e^i, P_{i'}](\Psi) = 1 - RP_e(P_{i'}). \quad (4.49)$$

- si $P_i \neq P_{i'}$, c'est-à-dire en présence d'un conflit et si le prénom $P_{i'}$ a été détecté dans un commentaire sans ambiguïté :

$$m^\Psi[Mr_e^i, P_{i'}](\overline{Per}) = \begin{cases} RP_e(P_{i'}) & \text{si } \sum Mr_e(\cdot, i) > 0 \text{ et } \sum Mr_e(i', \cdot) = 0 \\ 0 & \text{sinon.} \end{cases} \quad (4.50)$$

$$m^\Psi[Mr_e^i, P_{i'}](\Psi) = 1 - m^\Psi[Mr_e^i, P_{i'}](\overline{Per}). \quad (4.51)$$

Cette dernière allocation permet de défavoriser un prénom lorsqu'il a été détecté dans une situation conflictuelle et qu'il ne correspond pas à un prénom réel du carnet d'adresses.

4.3.4 Détection des nouveaux prénoms

Les reconnaisseurs de prénoms audio et texte (*RPa* et *RPt*) peuvent détecter dans les commentaires des nouveaux prénoms qui ne se trouvent pas dans le carnet d'adresses. Pour présenter ces nouveaux prénoms dans les résultats de **FUSINDEX**, nous utilisons une approche qui utilise le degré de croyance alloué à l'hypothèse ($Inconnu \cup Inconnue$) par les outils d'indexation qui traitent l'image. Cette idée est basée sur le principe suivant : si un nouveau prénom a été détecté dans les commentaires, alors il est fort probable que ce prénom corresponde au prénom de la personne inconnue détectée par les outils *DV*, *RV* et *H/F*.

On suppose que les moteurs *RPt* ou *RPa* ont détecté un nouveau prénom, noté N_{n+1} , relatif à une personne P_{n+1} . Notre idée consiste à ajouter cette nouvelle personne dans le carnet d'adresses Ω à la place de l'hypothèse $\{Inconnu, Inconnue\}$ conduisant à un cadre de discernement, noté Ω' , défini par :

$$\Omega' = \{P_1, P_2, \dots, P_n, P_{n+1}, *\}.$$

Le cadre de discernement est ainsi modifié afin que la nouvelle personne détectée P_{n+1} reçoive la quantité de croyance dédiée à l'hypothèse $\{Inconnu, Inconnue\}$ dans le cadre initial Ω . En effet, la seule différence entre une personne connue P_i et une personne initialement inconnue P_{n+1} réside dans le principe d'allocation des masses pour le reconnaisseur de visages. Ceci est dû au fait qu'il n'y a pas de visages de cette personne dans la base d'apprentissage du moteur *RV* et donc qu'il n'est pas possible d'utiliser la même stratégie. Il est pourtant envisageable de calculer la probabilité pour cette personne inconnue d'être une femme F_{n+1} déduite de son prénom. Ainsi, les étapes décrites dans les paragraphes précédents sont suivies scrupuleusement excepté pour deux points particuliers.

- Le premier point concerne l'allocation de masses du moteur d'indexation *RV*, où l'équation (4.15) pour une zone Z_j devient :

$$m_{\alpha_{RV}}^{\Omega'}[RV, Z_j](P_{n+1} \cup *) = m_{\alpha_{RV}}^{\Omega}[RV, Z_j](Inconnue \cup Inconnu \cup *). \quad (4.52)$$

- Le second point concerne la fusion des informations extraites des zones voisines. En effet, le processus exige la certitude sur l'appartenance d'une personne à une zone donnée ce qui n'est pas le cas pour les nouveaux prénoms détectés. De façon à neutraliser cette étape dans la fusion, nous avons choisi de modifier la fonction de masse définie aux équations (4.32) et (4.33). Ainsi, cette fonction devient la masse vide définie par :

$$m_{\alpha_{Z_{j'}}}^{\Omega'}[Z_{j'}, Z_j](\overline{P_{n+1}}) = 0 \quad (4.53)$$

$$m_{\alpha_{Z_{j'}}}^{\Omega'}[Z_{j'}, Z_j](\Omega) = 1. \quad (4.54)$$

qui est l'élément neutre de l'opérateur de combinaison de Dempster.

4.3.5 Croyances associées aux métadonnées issues des commentaires Com

Le jeu de masses $m^\Psi[Com', P_i]$ admet l'hypothèse \overline{Per} comme élément focal. Néanmoins, les commentaires ne peuvent pas informer sur l'absence d'une personne dans le document à indexer car ils correspondent à des sources non exhaustives (exemple : rien ne dit que *Homer* n'est pas dans une photo qui aurait été commenté par Anniversaire de *Bart*). Par conséquent, la quantité de croyance affectée à l'hypothèse \overline{Per} se doit d'être transférée à l'incertitude (Ψ). On en déduit le jeu de masse $m_{\alpha_{Com}}^\Psi[Com, P_i]$ de cette façon :

$$m_{\alpha_{Com}}^\Psi[Com, P_i](Per) = m^\Psi[Com', P_i](Per) \quad (4.55)$$

$$m_{\alpha_{Com}}^\Psi[Com, P_i](\Psi) = m^\Psi[Com', P_i](\Psi) + m^\Psi[Com', P_i](\overline{Per}). \quad (4.56)$$

Nous obtenons ainsi un jeu de masses pour les commentaires associés au document dans la référence Ψ qui permet de quantifier la pertinence de chacune des métadonnées et qui sera ensuite fusionné avec l'image.

4.4 Exemple d'indexation par le contenu d'un document multimédia

Dans ces paragraphes, nous allons illustrer l'extraction des informations concernant le contenu d'un document multimédia en vue de son indexation et de la présentation de métadonnées à l'utilisateur. Soit le carnet d'adresses suivant contenant 5 personnes connues de la famille Simpson :

P_1 le père "Homer".

P_2 la mère "Marge".

P_3 le fils "Bart".

P_4 la fille "Lisa".

P_5 la fille "Maggie".

La figure 4.8 illustre ces personnes connues du carnet d'adresses d'Homer. les images de cette figure serviront à l'apprentissage du reconnaisseur RV . Les prénoms seront utilisés dans notre exemple tandis que nous supposerons connus les sexes des différents personnages. La figure 4.9 représente le document multimédia à indexer avec son image et ses commentaires associés.

Pour des questions de simplicité et de compréhension de l'exemple proposé, si le degré de croyance dépasse 0.99 sur un élément focal, il sera automatiquement arrondi à 1. De la même



Figure 4.8 : Personnages du carnet d'adresses utilisés pour l'indexation



Figure 4.9 : Document à indexer : Image et Commentaires.

façon, s'il est inférieur à 0.01, il sera automatiquement mis à 0 pour permettre une réduction des éléments focaux. On fixe également les coefficients de fiabilité $\alpha_{DV} = \alpha_{RV} = \alpha_{H/F} = 1$.

4.4.1 Analyse du contenu de l'image : I

Détecteur de visage : DV

En sortie du détecteur de visage DV , on a obtenu les scores $s_{DV}(Z_1) = 35$ et $s_{DV}(Z_2) = 40$ pour deux zones détectées. Les fonctions de masse calculées à partir des équations (4.7) et (4.8)

pour les deux zones Z_1 et Z_2 sont respectivement :

$$\begin{aligned} m_{\alpha_{DV}}^{\Omega}[DV, Z_1](P_1 \cup P_2 \cup P_3 \cup P_4 \cup P_5 \cup \text{Inconnu} \cup \text{Inconnue}) &= 0.7; \\ m_{\alpha_{DV}}^{\Omega}[DV, Z_1](*) &= 0.3; \\ m_{\alpha_{DV}}^{\Omega}[DV, Z_2](P_1 \cup P_2 \cup P_3 \cup P_4 \cup P_5 \cup \text{Inconnu} \cup \text{Inconnue}) &= 0.8; \\ m_{\alpha_{DV}}^{\Omega}[DV, Z_2](*) &= 0.2. \end{aligned}$$

Ces deux allocations de masse sont obtenues après affaiblissement des fonctions $m^{\Omega}[DV, Z_1](\cdot)$ et $m^{\Omega}[DV, Z_2](\cdot)$ en considérant un coefficient de fiabilité α_{DV} égal à 1. Elles matérialisent la croyance sur le fait que le moteur d'indexation DV a détecté deux visages dans les zones Z_1 et Z_2 (cf. figure 4.9) avec un certain degré d'incertitude.

Reconnaisseur de visage : RV

Les distances obtenues par le reconnaisseur de visages RV pour les deux zones sont présentées dans le tableau 4.2. A partir des équations (4.13), (4.14) et (4.15), on peut calculer la croyance sur

$d(Z_j, P_i)$	P_1	P_2	P_3	P_4	P_5
Z_1	210000	258900	800000	900000	850000
Z_2	220000	950000	292000	230000	950000

Tableau 4.2 : Distances obtenues par le reconnaisseur de visages RV .

le fait que les personnes P_i du carnet d'adresses se trouvent dans l'image. Les fonctions de masse obtenues à partir des distances fournies par le moteur dans les zones Z_1 et Z_2 sont données dans le tableau 4.3. De la même façon, ces deux allocations de masse sont obtenues après affaiblissement

$m_{\alpha_{RV}}^{\Omega}[RV, Z_j](\cdot)$	P_1	P_2	$P_1 \cup P_2$	$P_1 \cup P_3 \cup P_4$	$P_1 \cup P_4$
Z_1	0	0.2	0.8	0	0
Z_2	0.1	0	0	0.7	0.2

Tableau 4.3 : Fonctions de masse $m_{\alpha_{RV}}^{\Omega}[RV, Z_j](\cdot)$ obtenues par le reconnaisseur de visages RV .

des fonctions $m^{\Omega}[RV, Z_1](\cdot)$ et $m^{\Omega}[RV, Z_2](\cdot)$ en considérant un coefficient de fiabilité α_{RV} égal à 1. On remarque facilement que les éléments focaux sont emboîtés ce qui permet d'obtenir des fonctions consonantes. Dans la zone Z_1 , le moteur RV favorise P_2 au détriment de P_1 et dans la zone Z_2 , RV classe dans l'ordre P_1, P_4 puis P_3 . Nous allons voir comment ce classement évolue en fonction des informations apportées par les autres moteurs.

Reconnaisseur de genre homme/femme : H/F

Le moteur d'indexation H/F quantifie la croyance pour que la personne détectée dans chaque zone soit du sexe féminin ou masculin permettant de favoriser les personnes du carnet d'adresses de même sexe. On suppose, dans notre exemple, que les probabilités F_i sont connues avec précision telles que $F_1 = 0$, $F_2 = 1$, $F_3 = 0$, $F_4 = 1$ et $F_5 = 1$ qui permettent d'obtenir les jeux de masses suivants :

$$\begin{aligned} m^\Omega[HF/P_1, Z_1](\overline{P_1}) &= 0, & m^\Omega[HF/P_1, Z_1](\Omega) &= 1; \\ m^\Omega[HF/P_2, Z_1](\overline{P_2}) &= 1, & m^\Omega[HF/P_2, Z_1](\Omega) &= 0; \\ m^\Omega[HF/P_3, Z_1](\overline{P_3}) &= 0, & m^\Omega[HF/P_3, Z_1](\Omega) &= 1; \\ m^\Omega[HF/P_4, Z_1](\overline{P_4}) &= 1, & m^\Omega[HF/P_4, Z_1](\Omega) &= 0; \\ m^\Omega[HF/P_5, Z_1](\overline{P_5}) &= 1, & m^\Omega[HF/P_5, Z_1](\Omega) &= 0. \end{aligned}$$

pour la zone Z_1 et :

$$\begin{aligned} m^\Omega[HF/P_1, Z_2](\overline{P_1}) &= 0, & m^\Omega[HF/P_1, Z_2](\Omega) &= 1; \\ m^\Omega[HF/P_2, Z_2](\overline{P_2}) &= 1, & m^\Omega[HF/P_2, Z_2](\Omega) &= 0; \\ m^\Omega[HF/P_3, Z_2](\overline{P_3}) &= 0, & m^\Omega[HF/P_3, Z_2](\Omega) &= 1; \\ m^\Omega[HF/P_4, Z_2](\overline{P_4}) &= 1, & m^\Omega[HF/P_4, Z_2](\Omega) &= 0; \\ m^\Omega[HF/P_5, Z_2](\overline{P_5}) &= 1, & m^\Omega[HF/P_5, Z_2](\Omega) &= 0. \end{aligned}$$

pour la zone Z_2 . La fusion de ces différentes fonctions de croyance (cf. équation 4.25) permet d'obtenir des informations plus précises sur l'identification des personnes dans les zones Z_1 et Z_2 :

$$\begin{aligned} m_{\alpha_{H/F}}^\Omega[H/F, Z_1](P_1 \cup \text{Inconnu} \cup *) &= 1, \\ m_{\alpha_{H/F}}^\Omega[H/F, Z_2](P_1 \cup P_3 \cup \text{Inconnu} \cup *) &= 1. \end{aligned}$$

Source d'information IZ

La source d'information IZ permet d'assurer la gestion des zones détectées entre elles. La fusion entre les trois moteurs DV , RV et H/F précédemment décrits permet d'obtenir, à partir de l'équation (4.31), les fonctions de masse suivantes pour les deux zones Z_1 et Z_2 :

$$\begin{aligned} m^\Omega[I', Z_1](P_1) &= 1; \\ m^\Omega[I', Z_2](P_1 \cup P_3) &= 0.7; \\ m^\Omega[I', Z_2](P_1) &= 0.3. \end{aligned}$$

Le tableau 4.4 présente les valeurs de probabilités pignistiques associées à ces fonctions de masse permettant le calcul des coefficients de fiabilité α_{Z_1} et α_{Z_2} .

$BetP_{m^\Omega[I', Z_j]}(P_i)$	P_1	P_2	P_3	P_4	P_5
Z_1	1	0	0	0	0
Z_2	0.65	0	0.35	0	0

Tableau 4.4 : Probabilités pignistiques associées aux fonctions de masse $m^\Omega[I', Z_j](P_i)$.

Avant l'étape de normalisation, le conflit engendré par la fusion est calculé pour la gestion inter-zones avec $m_{\odot}^\Omega[I', Z_1](\emptyset) = 0.5$ pour la zone Z_1 et $m_{\odot}^\Omega[I', Z_2](\emptyset) = 0.2$ pour la zone Z_2 . A partir de ces valeurs, l'équation (4.34) permet le calcul des coefficients de fiabilité $\alpha_{Z_1} = (1 - 0.5) \times (1 - 0) = 0.5$ pour la zone Z_1 et $\alpha_{Z_2} = (1 - 0.2) \times (0.65 - 0.3) = 0.28$ pour la zone Z_2 .

Le calcul de ces coefficients et les crédibilités (cf. tableau 4.5) permettent de calculer les fonctions de croyance relatives à la source d'information IZ . Pour la zone Z_1 , seule P_1 a une

$bel^\Omega[I', Z_j](P_i)$	P_1	P_2	P_3	P_4	P_5
Z_1	1	0	0	0	0
Z_2	0	0	0.3	0	0

Tableau 4.5 : Crédibilités associées aux fonctions de masse $m^\Omega[I', Z_j](P_i)$.

crédibilité non nulle ($bel^\Omega[I', Z_1](P_1) = 1$) ce qui permet de contester sa présence dans la zone Z_2 et conduit à la fonction de masse suivante :

$$\begin{aligned} m_{\alpha_{Z_1}}^\Omega[Z_1, Z_2](\overline{P_1}) &= 0.5 \times 1 = 0.5 \\ m_{\alpha_{Z_1}}^\Omega[Z_1, Z_2](\Omega) &= 1 - 0.5 = 0.5. \end{aligned}$$

Pour la zone Z_2 , P_3 est la seule personne qui possède une crédibilité non nulle dans la zone Z_2 ($bel^\Omega[I', Z_2](P_3) = 0.3$). Ceci permet, dans une moindre contribution, de contester dans la zone Z_1 (cf. équation 4.32) la personne P_3 :

$$\begin{aligned} m_{\alpha_{Z_2}}^\Omega[Z_2, Z_1](\overline{P_3}) &= 0.28 \times 0.3 = 0.084 \\ m_{\alpha_{Z_2}}^\Omega[Z_2, Z_1](\Omega) &= 1 - 0.084 = 0.916. \end{aligned}$$

Fusion des informations I et pertinence des métadonnées

Nous illustrons dans ce paragraphe la fusion d'informations extraites de l'image via $I' = DV \oplus RV \oplus H/F$ et de la gestion inter-zones IZ . A partir des équations (4.35) et (4.36), on

obtient les allocations de masses suivantes :

$$m^\Omega[I, Z_1](P_1) = 1$$

pour la zone Z_1 et

$$m^\Omega[I, Z_2](P_1 \cup P_3) = 0.41$$

$$m^\Omega[I, Z_2](P_1) = 0.18$$

$$m^\Omega[I, Z_2](P_3) = 0.41.$$

pour la zone Z_2 . Le tableau 4.6 présente les valeurs de probabilités pignistiques associées à ces fonctions de masse. On peut ainsi constater que les informations obtenues dans la zone Z_1

$BetP_{m^\Omega[I, Z_j]}(P_i)$	P_1	P_2	P_3	P_4	P_5
Z_1	1	0	0	0	0
Z_2	0.385	0	0.615	0	0

Tableau 4.6 : Probabilités pignistiques associées aux fonctions de masse $m^\Omega[I, Z_j](P_i)$.

n'ont pas subi de modification après intégration de la source IZ (P_1 restant toujours la seule pertinente). Ceci est dû au fait que les informations issues de I' étaient certaines d'une part et que l'influence de la source est très faible (source d'information presque neutre : cf. $m_{\alpha_{Z_2}}^\Omega[Z_2, Z_1]$). Par contre, après la fusion avec la source d'information IZ dans la zone Z_2 , la bonne personne P_3 devient la métadonnée la mieux classée vis-à-vis de la probabilité pignistique.

A partir de ces probabilités définies dans Ω , on calcule la pertinence de chaque métadonnée définie dans Ψ à l'aide des équations (4.37) et (4.38) en considérant un coefficient de fiabilité α_I égal à 1 ce qui permet d'obtenir les résultats suivants :

métadonnée P_1 $BetP_{m_{\alpha_I}^\Psi[I, P_1]}(Per) = 1,$

métadonnée P_2 $BetP_{m_{\alpha_I}^\Psi[I, P_2]}(Per) = 0,$

métadonnée P_3 $BetP_{m_{\alpha_I}^\Psi[I, P_3]}(Per) = 0.615,$

métadonnée P_4 $BetP_{m_{\alpha_I}^\Psi[I, P_4]}(Per) = 0,$

métadonnée P_5 $BetP_{m_{\alpha_I}^\Psi[I, P_5]}(Per) = 0.$

4.4.2 Analyse des commentaires du document : *Com*

A partir des commentaires du document de la figure 4.9, les moteurs d'indexation ont détecté 3 événements conduisant aux matrices de recouvrement suivantes :

Commentaire textuel : "ma fille"

$$Mr_1 = \begin{pmatrix} & P_1 & P_2 & P_3 & P_4 & P_5 \\ P_1 & * & 0 & 0 & 0 & 0 \\ P_2 & 0 & * & 0 & 0 & 0 \\ P_3 & 0 & 0 & * & 0 & 0 \\ P_4 & 0 & 0 & 0 & * & 1 \\ P_5 & 0 & 0 & 0 & 1 & * \end{pmatrix}.$$

permettant de s'apercevoir que le mot détecté peut cibler P_4 et P_5 (*Lisa* et *Maggie*),

Commentaire textuel : "Maggie"

$$Mr_2 = \begin{pmatrix} & P_1 & P_2 & P_3 & P_4 & P_5 \\ P_1 & * & 0 & 0 & 0 & 0 \\ P_2 & 0 & * & 0 & 0 & 0 \\ P_3 & 0 & 0 & * & 0 & 0 \\ P_4 & 0 & 0 & 0 & * & 0 \\ P_5 & 0 & 0 & 0 & 0 & * \end{pmatrix}.$$

pour lequel aucun conflit n'a été détecté et

Commentaire vocal "Maggie ∨ Marge"

$$Mr_3 = \begin{pmatrix} & P_1 & P_2 & P_3 & P_4 & P_5 \\ P_1 & * & 0 & 0 & 0 & 0 \\ P_2 & 0 & * & 0 & 0 & 1 \\ P_3 & 0 & 0 & * & 0 & 0 \\ P_4 & 0 & 0 & 0 & * & 0 \\ P_5 & 0 & 1 & 0 & 0 & * \end{pmatrix}.$$

résultant de la détection de P_2 ou P_5 (*Maggie* ou *Marge*).

Les scores associés aux deux moteurs d'indexation RPa et RPt sont présentés dans la matrice suivante :

$$Scores = \begin{pmatrix} & P_1 & P_2 & P_3 & P_4 & P_5 \\ Mr_1 & 0 & 0 & 0 & 0.4 & 0.4 \\ Mr_2 & 0 & 0 & 0 & 0 & 0.8 \\ Mr_3 & 0 & 0.6 & 0 & 0 & 0.7 \end{pmatrix}.$$

Ces informations permettent de calculer les fonctions de masse $m^\Psi[Mr_e^i, P_{i'}]$ qui sont données dans la liste ci-dessous :

pour P_1 aucun score ;

pour P_2 deux sources :

$$m^\Psi[Mr_3^2, P_2](Per) = 0.6, m^\Psi[Mr_3^2, P_2](\Psi) = 0.4$$

$$m^\Psi[Mr_2^2, P_5](\overline{Per}) = 0.8, m^\Psi[Mr_2^2, P_5](\Psi) = 0.2.$$

P_2 sera contesté par P_5 car P_2 est en conflit avec P_5 et ce dernier a été détecté sans recouvrement dans la situation conflictuelle 2.

pour P_3 aucun score ;

pour P_4 deux sources :

$$m^\Psi[Mr_1^4, P_4](Per) = 0.4, m^\Psi[Mr_1^4, P_4](\Psi) = 0.6$$

$$m^\Psi[Mr_2^4, P_5](\overline{Per}) = 0.8, m^\Psi[Mr_2^4, P_5](\Psi) = 0.2.$$

P_4 sera contesté par P_5 car P_4 est en conflit avec P_5 et ce dernier a été détecté sans recouvrement dans la situation conflictuelle 2.

pour P_5 trois sources :

$$m^\Psi[Mr_1^5, P_5](Per) = 0.4, m^\Psi[Mr_1^5, P_5](\Psi) = 0.6$$

$$m^\Psi[Mr_2^5, P_5](Per) = 0.8, m^\Psi[Mr_2^5, P_5](\Psi) = 0.2$$

$$m^\Psi[Mr_3^5, P_5](Per) = 0.7, m^\Psi[Mr_3^5, P_5](\Psi) = 0.3.$$

P_5 sera renforcée par les trois situations conflictuelles où P_5 a été détecté.

Le résultat de la fusion s'obtient par le calcul relatif à l'équation (4.47). Les valeurs de la fonction de masse $m^\Psi[Com', P_i](\cdot)$ sont données dans le tableau 4.7. Les équations (4.55) et (4.56) sont

$m^\Psi[Com', P_i](\cdot)$	P_1	P_2	P_3	P_4	P_5
Per	0.00	0.23	0.00	0.12	0.964
\overline{Per}	0.00	0.61	0.00	0.70	0.000
Ψ	1.00	0.16	1.00	0.18	0.036

Tableau 4.7 : Fonctions de masse $m^\Psi[Com', P_i](\cdot)$ pour les éléments focaux Per , \overline{Per} et Ψ .

utilisées pour permettre la quantification des métadonnées selon les commentaires Com . Ceci est présenté dans le tableau 4.8. On peut remarquer que la métadonnée P_5 (correspondant à *Maggie*) possède un fort degré de croyance dû à sa détection dans les commentaires (vocal et textuel).

$m^\Psi[Com, P_i](.)$	P_1	P_2	P_3	P_4	P_5
<i>Per</i>	0.00	0.23	0.00	0.12	0.964
Ψ	1.00	0.77	1.00	0.88	0.036

Tableau 4.8 : Fonctions de masse $m^\Psi[Com, P_i](.)$ pour les éléments focaux *Per* et Ψ .

4.4.3 Fusion entre image *I* et commentaires *Com*

Après extraction des informations issues de l'image *I* et des commentaires *Com*, il s'agit d'en réaliser la fusion pour présenter les métadonnées à l'utilisateur. Le tableau 4.9 présente les fonctions de masse $m^\Psi[I \oplus Com, P_i]$ quantifiant les degrés de pertinence pour chacune des métadonnées du carnet d'adresses.

$m^\Psi[I \oplus Com, P_i](.)$	P_1	P_2	P_3	P_4	P_5
<i>Per</i>	1.00	0.23	0.615	0.12	0.964
Ψ	0.00	0.77	0.385	0.88	0.036

Tableau 4.9 : Fonctions de masse $m^\Psi[I \oplus Com, P_i](.)$ pour les éléments focaux *Per* et Ψ .

Après la fusion entre les informations issues de l'image et les informations issues des commentaires, on peut calculer la probabilité pignistique de l'hypothèse "*Per*" pour chaque métadonnée. Ce calcul est présenté dans le tableau 4.10. Dans notre exemple, pour aider l'utilisateur à anno-

$BetP_{m^\Psi[I \oplus Com, P_i]}$	P_1	P_2	P_3	P_4	P_5
<i>Per</i>	1.00	0.6	0.8	0.56	0.99

Tableau 4.10 : Probabilité pignistique $BetP_{m^\Psi[I \oplus Com, P_i]}(Per)$.

ter ce document, les métadonnées permettant l'indexation à partir du contenu seront triées et affichées de la façon suivante :

1. Indexer sous le *topic Homer* : oui/non ? (Pertinence $P_1 = 1.00$),
2. Indexer sous le *topic Maggie* : oui/non ? (Pertinence $P_5 = 0.99$),
3. Indexer sous le *topic Bart* : oui/non ? (Pertinence $P_3 = 0.80$).

les autres métadonnées n'obtenant pas un niveau de pertinence suffisant pour permettre d'être proposées à l'utilisateur (en fonction du seuil de pertinence posé).

4.5 Conclusion

Dans ce chapitre, les principes d'indexation par le contenu d'un document multimédia par les fonctions de croyance ont été présentés. Les techniques employées pour transformer les scores et les distances des différents moteurs en degrés de croyance pour chacune des sources d'informations *I* et *Com* ont été détaillées. Elles permettent de représenter les incertitudes liées à chacune des sorties des moteurs.

Les contributions de ce travail sont nombreuses et concernent :

- la solution proposée pour la gestion des zones détectées par le détecteur *DV* qui permet de diminuer l'incertitude sur la reconnaissance fournie par le moteur *RV* ;
- la détection de nouveaux prénoms présents dans les commentaires vocaux et/ou textuels ;
- la gestion des mots dans les commentaires permettant d'assurer la prise en charge de prénoms composés ou des mots ciblant plusieurs prénoms (ma fille, ma soeur, ...).

Nous avons illustré l'extraction du contenu d'un document au travers d'un exemple et d'une application numérique. D'autres cas d'étude concrets et une évaluation plus complète de **FUSINDEX** pour l'indexation par le contenu seront présentés au chapitre 6.

Nous avons enfin mis en évidence les avantages de modéliser les informations extraites du document par des fonctions de croyance :

- les fonctions de croyance sont un cadre fédérateur qui permet de quantifier imprécision et incertitude dans un même objet mathématique. Dans le cadre des métadonnées présentées à l'utilisateur, cet avantage permet d'associer un niveau de pertinence et donc de satisfaire la maxime qui dit qu'*Il vaut mieux ne rien dire que de se tromper...* ;
- les fonctions de croyance permettent la fusion des informations par l'opérateur de combinaison de Dempster et la détection du conflit existant entre les sources.

Dans le chapitre suivant, nous nous intéressons au contexte associé à chaque document et à son extraction dans un objectif d'indexation. Une fois encore, la contribution de ce travail réside dans la définition de fonctions de croyance permettant de modéliser le contexte pour diminuer l'incertitude associée à l'indexation du document.

Chapitre 5

FUSINDEX et l'indexation par le contexte

Dans le système **FUSINDEX**, les informations provenant des documents de la base indexée sont exploités en terme de contexte. On entend par base indexée la base des documents dont l'indexation a été validée par l'utilisateur mais aussi les documents indexés par le contenu qui n'ont pas forcément été validés par l'utilisateur. Dans ce chapitre, on associe à chacun des documents multimédia un certain nombre d'informations qui permettent de caractériser le contexte dans lequel le document a été produit. Les informations extraites sont ensuite fusionnées avec les informations extraites de contenu du document afin d'améliorer la qualité des décisions sur les métadonnées pertinentes.

Souvent les images prises par les appareils numériques sont regroupées dans des petits intervalles de temps [4]. L'idée défendue dans cette thèse est qu'il est utile de faire propager des métadonnées validées dans certaines images pour les renforcer dans des images acquises à des instants proches. Puisque le système proposé indexe les documents selon les personnes, on peut également supposer raisonnable qu'une même personne ne peut se trouver dans deux lieux différents en même temps. Après avoir étudié le contexte spatio-temporel entre deux documents (cf. section 5.2), ces deux hypothèses sont utilisées pour permettre soit de renforcer soit de contester une métadonnée dans la section 5.3. Nous présentons dans la section 5.4, un exemple qui permet d'illustrer le processus d'extraction du contexte d'un document multimédia. Dans la section 5.5, une conclusion permet de mettre en exergue les avantages d'utiliser le contexte associé à un document en vue de son indexation.

5.1 Introduction

Comme nous l'avons évoqué dans le chapitre 1, plusieurs types d'informations contextuelles peuvent être utilisés comme par exemple des données temporelles, des données spatiales et des informations de nature sociale [2, 36]. Dans le système **FUSINDEX**, nous avons choisi d'exploiter la date et le lieu. Ainsi, le processus d'extraction du contexte nécessite les données suivantes :

- **Image à indexer**
 - la date de capture de l'image,
 - le lieu de capture de l'image ;
- **Image de la base indexée**
 - la date de capture de l'image,
 - le lieu de capture de l'image,
 - les degrés de pertinence des métadonnées proposées par l'indexation du contenu de l'image. Ces valeurs sont disponibles si l'image a été déjà traitée par le système ou si elle a été annotée par l'utilisateur.

Cependant, d'autres informations peuvent être utilisées dans l'extraction du contexte comme par exemple la co-présence de personnes (cette information peut être directement exploitée à partir de la source d'information *IZ* décrite précédemment), la date de connexion à l'interface IHM, etc. Dans le système **FUSINDEX**, le lieu est identifié par ses coordonnées géographiques : altitude et longitude tandis que la date est codée dans un format standard.

Soient I_k un document à indexer et deux documents $I_{k'}$ et $I_{k''}$ de la base indexée des documents multimédia avec $((k', k'') \in \{1, 2, \dots, K\}^2)$ et $k \neq k' \neq k''$. Ces trois documents sont

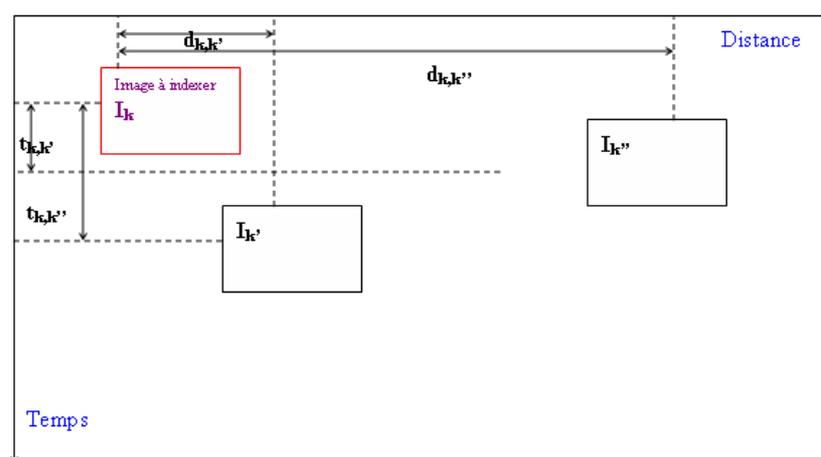


Figure 5.1 : Les documents I_k , $I_{k'}$ et $I_{k''}$ dans un espace spatio-temporel permettant de définir les concepts de proximité en terme de distance et de temps.

illustrés à la figure 5.1 dans un espace à deux dimensions représentant la distance et l'intervalle de temps entre les documents. On utilise les notations suivantes :

$dist_{k,k'}$: pour désigner la distance spatiale entre le document à indexer I_k et le document $I_{k'}$ (la distance entre deux lieux géographiques où les deux images ont été capturées) ;

$dist_{k,k''}$: pour désigner la distance spatiale entre le document à indexer I_k et le document $I_{k''}$;

$t_{k,k'}$: pour désigner l'intervalle de temps entre l'image à indexer I_k et l'image $I_{k'}$ (Le décalage temporel entre les captures d'images) ;

$t_{k,k''}$: pour désigner l'intervalle de temps entre l'image à indexer I_k et l'image $I_{k''}$;

$Score(P_i, I_{k'})$: pour désigner le score de la personne P_i dans le document $I_{k'}$;

$Score(P_i, I_{k''})$: pour désigner le score de la personne P_i dans le document $I_{k''}$.

Les scores $Score(., I_{k'})$ et $Score(., I_{k''})$ des métadonnées P_i se calculent à partir des fonctions de croyance utilisées pour quantifier la pertinence des documents $I_{k'}$ et $I_{k''}$ à partir de leurs contenus respectifs. Ces scores résultent donc de la fusion entre les informations extraites de l'image et de ses commentaires associés (cf. chapitre 4) et sont définis par :

$$Score(P_i, I_{k'}) = bel^\Psi[I \oplus Com, P_i](Per). \quad (5.1)$$

Si le document $I_{k'}$ a été annoté par l'utilisateur en validant la métadonnée P_i , le score $Score(P_i, I_{k'})$ vaut "1". Dans le cas contraire, si l'utilisateur estime que la personne P_i n'indexe pas en bon terme le document $I_{k'}$, le score $Score(P_i, I_{k'})$ vaut "0".

Dans ce chapitre, nous allons exploiter et modéliser par des fonctions de croyance les deux idées suivantes :

1. Si une personne P_i a été reconnue dans une image donnée alors la croyance dans le fait que cette personne se trouve également dans les images proches dans le temps et dans l'espace est augmentée. Ce mécanisme sera appelé renforcement dans la suite ;
2. Si une personne P_i a été reconnue dans une image donnée alors la croyance dans le fait que cette personne se trouve dans les images proches dans le temps et loin dans l'espace est diminuée. Ce mécanisme sera appelé contestation dans la suite.

A partir des notations et de la figure 5.1, la première idée se formalise de la façon suivante. Chaque personne (métadonnée) ayant servi à indexer le document $I_{k'}$ (temps et lieux proches de I_k) devra être renforcée pour l'indexation du document à indexer I_k . En ce qui concerne le processus de contestation, une personne ayant servi à indexer le document $I_{k''}$ (temps proches et lieux éloignés de I_k), pourrait se voir contester dans le document à indexer I_k . Dans la section suivante, ces deux processus (renforcement et contestation) sont décrits en détails.

5.2 Etude du contexte spatio-temporel

Avant de décrire les processus de renforcement et de contestation d'une métadonnée, il faut tout d'abord définir par des fonctions numériques les concepts de proximité temporelle et géographique entre deux documents. Il est à noter que les fonctions qui seront choisies dans les expérimentations se doivent d'être ajustées. Ceci peut être fait via un système d'apprentissage en analysant la distribution usuelle des images dans l'espace : (temps, distance).

5.2.1 Concept de proximité temporelle

La notion de proximité en terme d'intervalle de temps entre deux documents nécessite la définition d'une échelle temporelle qui sera utilisée dans les mécanismes de renforcement et de contestation. Ainsi, pour définir ce concept de proximité temporelle, une fonction notée g_T est utilisée. Elle est définie par :

$$g_T : T \rightarrow [0, 1] \quad (5.2)$$

$$t_{k,k'} \rightarrow g_T(t_{k,k'}). \quad (5.3)$$

où $t_{k,k'}$ représente l'intervalle de temps séparant les dates d'acquisition de l'image à indexer I_k et l'image $I_{k'}$. Cette fonction décroissante permet de quantifier la notion de "date proche/loin" dans l'intervalle $[0, 1]$. Ainsi, deux dates proches conduiront à une valeur de $g_T \simeq 1$. Dans les expériences qui seront menées dans le chapitre 6, nous avons choisi la fonction définie par :

$$g_T(t_{k,k'}) = \exp^{-\left(\frac{t_{k,k'}}{3600}\right)^3}.$$

La figure 5.2 illustre la fonction obtenue. Le choix de cette fonction est simplement justifié par la propriété qu'elle possède de fournir des valeurs entre 0 et 1. D'autres fonctions similaires pourraient être envisagées.

5.2.2 Concept de proximité géographique

En ce qui concerne la notion de proximité géographique, il est nécessaire de définir ce concept à partir de deux fonctions notées respectivement g_{Lp} et g_{Le} définies par :

$$g_{Lp} : Distance \rightarrow [0, 1] \quad (5.4)$$

$$dist_{k,k'} \rightarrow g_{Lp}(dist_{k,k'}). \quad (5.5)$$

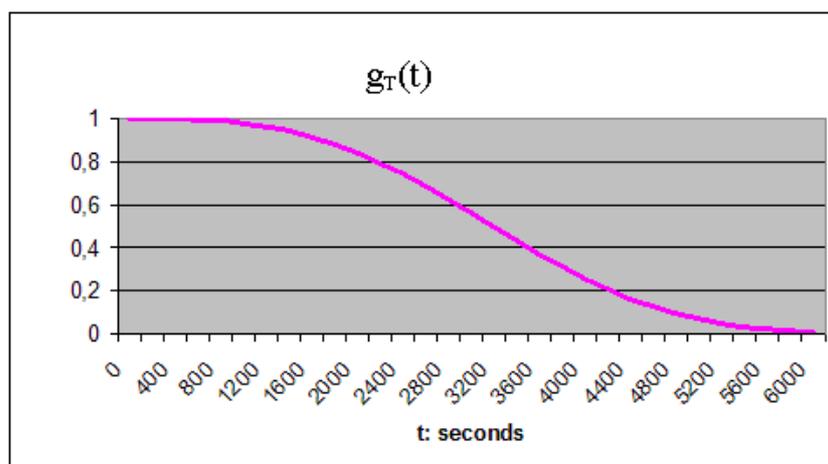


Figure 5.2 : Concept de proximité temporelle et fonction g_T Choisie.

pour la notion de lieu proche et

$$g_{Le} : Distance \rightarrow [0, 1] \quad (5.6)$$

$$dist_{k,k'} \rightarrow g_{Le}(dist_{k,k'}). \quad (5.7)$$

pour la notion d'éloignement. Dans ces équations, la variable $dist_{k,k'}$ désigne la distance spatiale entre le document à indexer I_k et le document $I_{k'}$. Ces fonctions décroissantes pour g_{Lp} et croissante pour g_{Le} permettent de quantifier dans l'intervalle $[0, 1]$ les notions de proximité et d'éloignement géographiques. Dans notre implémentation, nous avons utilisé les fonction suivantes :

$$g_{Lp}(dist_{k,k'}) = \exp^{-\left(\frac{dist_{k,k'}}{100}\right)^2} \quad (5.8)$$

$$g_{Le}(dist_{k,k'}) = 1 - \exp^{-\left(\frac{dist_{k,k'}}{10000}\right)^3}. \quad (5.9)$$

Ces deux fonctions sont respectivement présentées aux figures 5.3 et 5.4. Les paramètres d'ajustement de ces fonctions sont modifiables pour décider à partir de quel intervalle de temps et de quelle distance les images sont considérées proches (ou éloignées).

Prise en compte de la vitesse de déplacement

Si on suppose que les personnes du carnet d'adresses se déplacent à une vitesse moyenne v , les concepts de proximité géographique sont à modifier. Nous avons choisi de modifier la distance entre les lieux associés aux deux documents I_k et $I_{k'}$ de la façon suivante :

$$dist_{k,k'} = \min[0, dist_{k,k'} - v t_{k,k'}]. \quad (5.10)$$

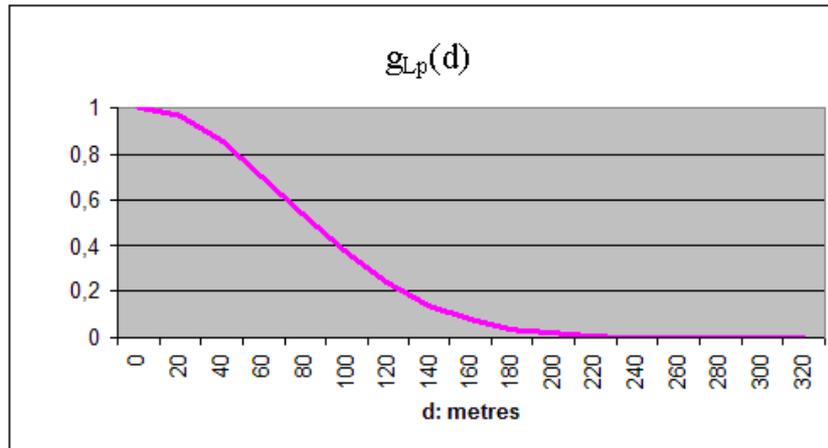


Figure 5.3 : Concept de proximité géographique et fonction g_{Lp} Choisie.

où v $t_{k,k'}$, représente la distance qui pourrait être parcourue par une personne se déplaçant à la vitesse v dans un intervalle de temps $t_{k,k'}$. Ainsi, on corrige la notion de proximité en tenant compte du déplacement de la personne. Ainsi, deux images prises dans le TGV (avec une vitesse de déplacement de 300 km/h) ne seront pas traitées de la même manière que deux images prises au cours d'une excursion touristique (avec une vitesse de déplacement moyenne de 10 km/h).

5.3 Renforcement et contestation

5.3.1 Extraction du contexte pour l'indexation d'un document

La mise en œuvre des deux mécanismes de renforcement et de contestation liés au contexte est assurée en modélisant l'incertitude relative à chaque métadonnée P_i . Cette incertitude est représentée sous la forme d'une fonction de masse, notée $m^\Psi[Cont, P_i]$ définie dans Ψ . Chaque image $I_{k'}$ dans la base indexée apporte ses informations contextuelles pour le renforcement ou la contestation d'une métadonnée P_i en vue de l'indexation du document I_k . Ainsi, le jeu de masses issu du contexte pour quantifier la pertinence de la métadonnée P_i est de la forme :

$$m^\Psi[Cont, P_i] = \bigoplus_{k', k' \neq k} m_{\alpha_{k',i}}^\Psi[I_{k'}, P_i]. \quad (5.11)$$

où $m_{\alpha_{k',i}}^\Psi[I_{k'}, P_i]$ désigne la contribution du document $I_{k'}$ à l'indexation par le contexte du document I_k . Cette dernière fonction résulte de l'affaiblissement de $m_{\alpha_{k',i}}^\Psi[I_{k'}, P_i]$ par un coefficient de fiabilité $\alpha_{k',i}$ qui s'attache à la fois à l'image $I_{k'}$ et à la métadonnée P_i comme nous le verrons dans les sections suivantes.

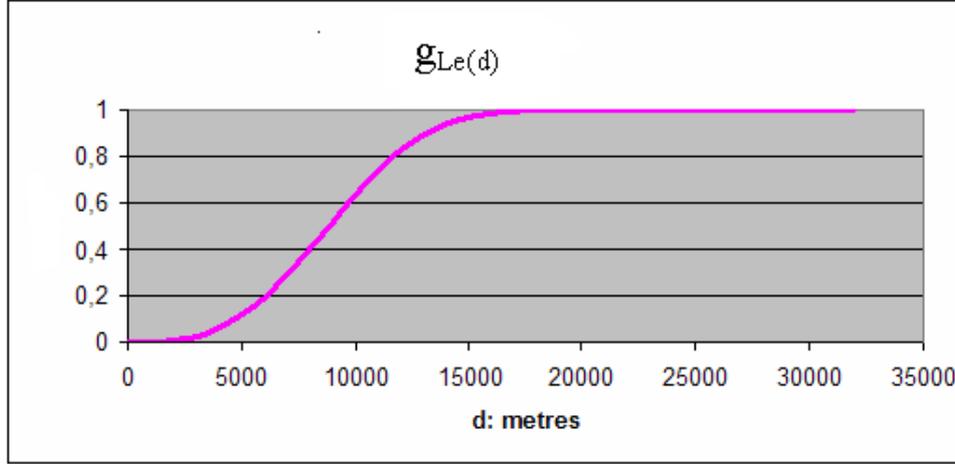


Figure 5.4 : Concept d'éloignement géographique et fonction g_{Le} Choisie.

5.3.2 Incertitude associée au contexte

Nous avons choisi de définir le jeu de masses $m^\Psi[I_{k'}, P_i](\cdot)$ associé au contexte du document $I_{k'}$ de la façon suivante :

$$m^\Psi[I_{k'}, P_i](Per) = Score(P_i, I_{k'}) g_T(t_{k,k'}) g_{Lp}(dist_{k,k'}) \quad (5.12)$$

$$m^\Psi[I_{k'}, P_i](\overline{Per}) = Score(P_i, I_{k'}) g_T(t_{k,k'}) g_{Le}(dist_{k,k'}) \quad (5.13)$$

$$m^\Psi[I_{k'}, P_i](\Psi) = 1 - \sum_{A \subseteq \Psi} m^\Psi[I_{k'}, P_i](A). \quad (5.14)$$

où le score $Score(P_i, I_{k'})$ est calculé à partir du contenu de $I_{k'}$ par l'équation (5.1) et mesure la pertinence de la métadonnée P_i . Cette façon de choisir le jeu de masses $m^\Psi[I_{k'}, P_i]$ satisfait les conditions suivantes :

- Si la personne P_i n'a pas été reconnue dans l'image associée à $I_{k'}$ alors $m^\Psi[I_{k'}, P_i]$ doit être une source d'information neutre :

$$Score(P_i, I_{k'}) \rightarrow 0 \quad \Rightarrow \quad m^\Psi[I_{k'}, P_i](\Psi) \rightarrow 1.$$

- Si l'image associée au document $I_{k'}$ est éloignée temporellement par rapport à I_k alors $m^\Psi[I_{k'}, P_i]$ doit être une source d'information neutre :

$$t_{k,k'} \rightarrow \infty \quad \Rightarrow \quad g_T(t_{k,k'}) \rightarrow 0 \quad \Rightarrow \quad m^\Psi[I_{k'}, P_i](\Psi) \rightarrow 1.$$

- Si l'image de $I_{k'}$ est loin géographiquement de celle de I_k alors $I_{k'}$ ne devrait pas favoriser P_i (renforcer la pertinence relative à P_i dans Ψ) dans le document à indexer I_k . Au contraire, si $I_{k'}$ est proche dans le temps, elle pourrait défavoriser P_i pour I_k :

$$dist_{k,k'} \rightarrow +\infty \quad \Rightarrow \quad g_{Lp}(dist_{k,k'}) \rightarrow 0 \quad \Rightarrow \quad m^\Psi[I_{k'}, P_i](Per) \rightarrow 0.$$

et :

$$(dist_{k,k'} \rightarrow +\infty, t_{k,k'} \rightarrow 0, Score(P_i, I_{k'}) \rightarrow 1) \Rightarrow m^\Psi[I_{k'}, P_i](\overline{Per}) \rightarrow 1.$$

- Si l'image associée au document I_k est proche géographiquement alors $I_{k'}$ ne devra pas défavoriser P_i . Au contraire, si $I_{k'}$ est proche dans le temps, elle pourrait renforcer P_i pour l'image à indexer I_k :

$$dist_{k,k'} \rightarrow 0 \Rightarrow g_{Le}(dist_{k,k'}) \rightarrow 0 \Rightarrow m^\Psi[I_{k'}, P_i](\overline{Per}) \rightarrow 0.$$

et :

$$(dist_{k,k'} \rightarrow 0, t_{k,k'} \rightarrow 0, Score(P_i, I_{k'}) \rightarrow 1) \Rightarrow m^\Psi[I_{k'}, P_i](Per) \rightarrow 1.$$

Ceci permet de représenter l'incertitude liée au contexte de chaque document et de mettre en œuvre les mécanismes de renforcement et contestation.

5.3.3 Calcul des coefficients de fiabilité

Comme nous l'avons fait remarquer à plusieurs reprises, l'image associée à chaque document peut contenir plusieurs personnes dont la présence peut être plus ou moins ambiguë selon les informations disponibles qui ne sont pas toutes validées par l'utilisateur. De plus, chacune des images de la base indexée est considérée mathématiquement comme une source d'information dont la fiabilité varie en fonction de l'écart temporel avec l'image à indexer. Cette hypothèse conduit à un nombre important de sources. Par conséquent, ces considérations peuvent engendrer des contestations et des renforcements 'parasites' pour la génération des métadonnées. C'est pourquoi nous avons choisi de faire propager le plus possible, les informations certaines et de négliger les informations incertaines par le truchement d'un processus d'affaiblissement.

Dans ce cadre, on cherche à propager les informations contextuelles du document $I_{k'}$ vers le document à indexer I_k . Ce paragraphe détaille la façon dont est déterminé le coefficient de fiabilité $\alpha_{k',i}$ qui est relatif à la fois à l'image $I_{k'}$ et à la métadonnée P_i . L'idée consiste à mesurer la crédibilité maximale accordée à une métadonnée P_i par l'analyse de l'image I d'une part et l'analyse des commentaires Com d'autre part.

Analyse de l'image I

L'analyse des informations image I appliquée sur l'image du document $I_{k'}$ donne le jeu de masses $m^\Omega[I, Z_j]$ défini sur le référentiel Ω (cf. chapitre 4). A partir de ce jeu de masses, on calcule la crédibilité pour chaque personne $P_i \subseteq \Omega$ et pour chaque zone de l'image $I_{k'}$. Le maximum des

fonctions de croyance $bel^\Omega[I, Z_j]$ pour chaque zone Z_j défini de la façon suivante :

$$\max_j (bel^\Omega[I, Z_j](P_i)). \quad (5.15)$$

va permettre le calcul du coefficient de fiabilité pour chaque personne P_i . Cette quantité représente la croyance maximale allouée à la métadonnée P_i sur l'ensemble de l'image associée au document $I_{k'}$.

Analyse des commentaires *Com*

Contrairement à l'analyse de l'image, l'incertitude liée aux commentaires est directement représentée dans Ψ . Ainsi, on perd l'information qui permet de discerner si une quantité de croyance a été allouée exactement sur la métadonnée P_i ou non. Pour récupérer cette information, on fait appel aux matrices de recouvrement et à la matrice de scores. La valeur suivante est calculée pour chaque métadonnée P_i :

$$\max_e \left(\prod_{i'=1/i' \neq i}^n (1 - rec_e(P_{i'}/P_i)) \times RP_e(P_i) \right). \quad (5.16)$$

et va permettre le calcul du coefficient de fiabilité pour chaque personne P_i . Elle représente la croyance maximale allouée à la métadonnée P_i pour la totalité des commentaires associés au document $I_{k'}$.

Coefficient de fiabilité

Le calcul du coefficient de fiabilité tire profit de l'analyse de l'image et des commentaires. Soit $maxBel_{k',i}$ la fonction qui représente la crédibilité maximale allouée à la métadonnée P_i par les outils qui traitent l'image et les commentaires :

$$maxBel_{k',i} = \max \left[\max_j (bel^\Omega[I, Z_j](P_i)), \max_e \left(\prod_{i'=1/i' \neq i}^n (1 - rec_e(P_{i'}/P_i)) \times RP_e(P_i) \right) \right]. \quad (5.17)$$

Dans l'équation (5.17), la quantité $maxBel_{k',i}$ représente la croyance maximale allouée à la métadonnée P_i ayant observé le contenu du document $I_{k'}$. A partir de cette mesure de confiance, on propose de calculer le coefficient de fiabilité $\alpha_{k',i}$ sous la forme :

$$\alpha_{k',i} = 1 - \exp^{-\left(\frac{maxBel_{k',i}}{0.35}\right)^3}. \quad (5.18)$$

qui permet d'obtenir un coefficient dans l'intervalle $[0, 1]$. La forme de la courbe représentant la valeur de $\alpha_{k',i}$ en fonction de $maxBel_{k',i}$ est proposée à la figure 5.5. Enfin, le coefficient de

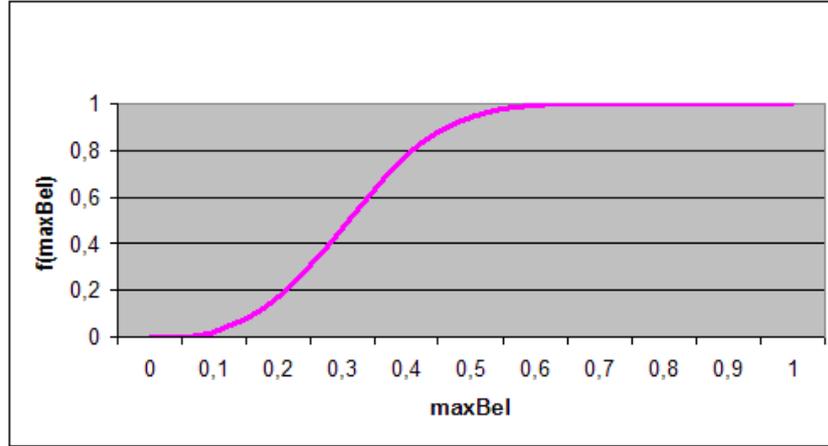


Figure 5.5 : Allure de la courbe représentant le coefficient $\alpha_{k',i}$ en fonction de $\max Bel_{k',i}$.

fiabilité de cette source d'information contextuelle permet d'affaiblir $m^\Psi[I_{k'}, P_i]$ pour obtenir la fonction de masse $m_{\alpha_{k',i}}^\Psi[I_{k'}, P_i]$ qui peut s'écrire :

$$m_{\alpha_{k',i}}^\Psi[I_{k'}, P_i](A) = \alpha_{k',i} m^\Psi[I_{k'}, P_i](A) \quad \forall A \subsetneq \Omega \quad (5.19)$$

$$m_{\alpha_{k',i}}^\Psi[I_{k'}, P_i](\Omega) = 1 - \alpha_{k',i} + \alpha_{k',i} m^\Psi[I_{k'}, P_i](\Omega). \quad (5.20)$$

5.4 Exemple d'indexation par le contexte d'un document multimédia

Pour illustrer le processus d'extraction du contexte d'un document multimédia, l'exemple introduit dans le chapitre 4 (cf. section 4.4) est repris avec le même carnet d'adresses :

$$\Omega = \{Homer, Marge, Bart, Lisa, Maggie, Inconnu, Inconnue, *\}.$$

et trois documents : un document à indexer noté I_1 et deux de la base indexée notés I_2 et I_3 . Ainsi, dans cet exemple on cherche à faire propager les informations contextuelles des documents I_2 et I_3 pour l'indexation du document I_1 .

5.4.1 Extraction du contenu du document à indexer I_1

Avant de pouvoir utiliser le contexte pour l'indexation du document I_1 , il nous faut d'abord extraire son contenu par les techniques décrites dans le chapitre 4. Le document est présenté à la figure 5.6. Comme on peut le remarquer, il n'y a aucun commentaire associé à l'image dans



Figure 5.6 : Document à indexer I_1 : image associée, aucun commentaire vocal ou textuel.

I_1 . Pour cette image, on suppose que le détecteur de visage (DV) n'a détecté qu'une seule zone Z_1 et que l'on a obtenu le jeu de masses suivant :

$$\begin{aligned} m^\Omega[I, Z_1](P_1, P_3) &= 0.4; \\ m^\Omega[I, Z_1](\Omega) &= 0.6. \end{aligned}$$

En ce qui concerne les commentaires, on obtient une source d'information neutre conduisant à $m_{\alpha_{Com}}^\Psi[Com, \cdot](\Psi) = 1$. La fonction de masse $m^\Psi[I \oplus Com, P_i](\cdot)$ associée au document I_1 est présentée au tableau 5.1. Elle permet d'obtenir les degrés de pertinence associés à chacune des

$m^\Psi[I \oplus Com, P_i](\cdot)$	P_1	P_2	P_3	P_4	P_5
Per	0.21	0.01	0.21	0.01	0.01
Ψ	0.79	0.99	0.79	0.99	0.99

Tableau 5.1 : Fonction de masse $m^\Psi[I \oplus Com, P_i](\cdot)$ associée au document I_1 .

métadonnées par la transformation pignistique suivante :

- $BetP_{m^\Psi[I \oplus Com, P_1]}(Per) = 0.6$;
- $BetP_{m^\Psi[I \oplus Com, P_2]}(Per) = 0.51$;
- $BetP_{m^\Psi[I \oplus Com, P_3]}(Per) = 0.6$;
- $BetP_{m^\Psi[I \oplus Com, P_4]}(Per) = 0.51$;
- $BetP_{m^\Psi[I \oplus Com, P_5]}(Per) = 0.51$.

A partir de ces degrés de pertinence et par l'unique analyse du contenu du document I_1 , les métadonnées $P_1 = Homer$ et $P_3 = Bart$ seraient proposées à l'utilisateur en terme d'indexation. Nous allons voir comment ces métadonnées vont évoluer en y intégrant les informations contextuelles liées aux documents I_2 et I_3 .

5.4.2 Extraction du contexte des documents I_2 et I_3

Ce paragraphe présente l'extraction du contexte des documents I_2 et I_3 de la base indexée. Dans l'image du document I_2 , on suppose que le détecteur de visage (DV) a détecté deux zones Z_1 et Z_2 et qu'on a obtenu les jeux de masses suivants :

$$\begin{aligned} m^{\Omega_1}[I, Z_1](P_1, P_2) &= 0.5; \\ m^{\Omega_1}[I, Z_1](\Omega) &= 0.5. \end{aligned}$$

pour la zone Z_1 et

$$\begin{aligned} m^{\Omega}[I, Z_2](P_2) &= 0.3; \\ m^{\Omega}[I, Z_2](\Omega) &= 0.7. \end{aligned}$$

pour la zone Z_2 . La figure 5.7 illustre le document I_2 de la base indexée. Cette image n'a pas

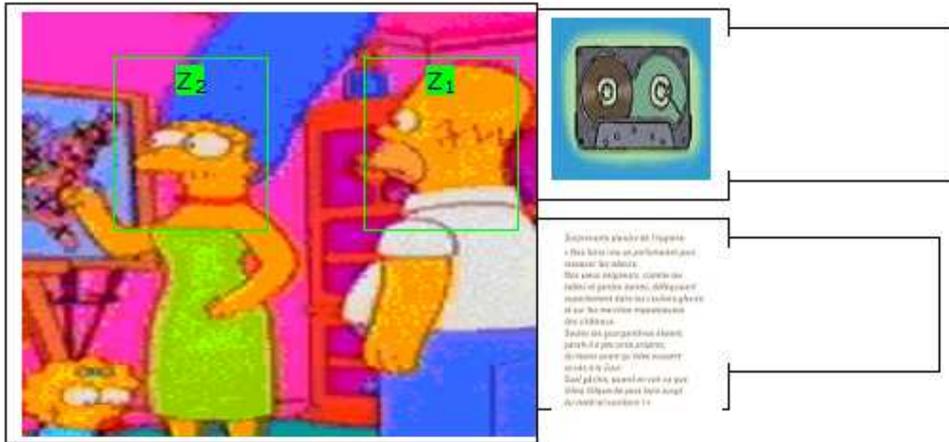


Figure 5.7 : Document I_2 de la base indexée : Image associée, pas de commentaire vocal ou textuel.

été commentée ce qui conduit à la fonction de masse $m_{\alpha_{Com}}^{\Psi}[Com, .](\Psi) = 1$. A partir de ces informations, on peut déduire les scores qui seront manipulés par le processus d'extraction du contexte (cf. équation(5.1)) :

$$\begin{aligned} Score(P_1, I_2) &= bel^{\Psi}[I \oplus Com, P_1](Per) = bel^{\Psi}[I, P_1](Per) = 0.31; \\ Score(P_2, I_2) &= bel^{\Psi}[I \oplus Com, P_2](Per) = bel^{\Psi}[I, P_2](Per) = 0.4; \\ Score(P_3, I_2) &= bel^{\Psi}[I \oplus Com, P_3](Per) = bel^{\Psi}[I, P_3](Per) = 0.31; \\ Score(P_4, I_2) &= bel^{\Psi}[I \oplus Com, P_4](Per) = bel^{\Psi}[I, P_4](Per) = 0.06; \\ Score(P_5, I_2) &= bel^{\Psi}[I \oplus Com, P_5](Per) = bel^{\Psi}[I, P_5](Per) = 0.06. \end{aligned}$$

Via l'équation (5.17), la fonction $maxBel$ est calculée. Les valeurs de cette fonction pour chaque métadonnée sont présentées ci-dessous :

$$maxBel_{2,1} = \max(bel^\Omega[I, Z_1](P_1), bel^\Omega[I, Z_2](P_1), 0) = 0 ;$$

$$maxBel_{2,2} = \max(bel^\Omega[I, Z_1](P_2), bel^\Omega[I, Z_2](P_2), 0) = 0.3 ;$$

$$maxBel_{2,3} = \max(bel^\Omega[I, Z_1](P_3), bel^\Omega[I, Z_2](P_3), 0) = 0 ;$$

$$maxBel_{2,4} = \max(bel^\Omega[I, Z_1](P_4), bel^\Omega[I, Z_2](P_4), 0) = 0 ;$$

$$maxBel_{2,5} = \max(bel^\Omega[I, Z_1](P_5), bel^\Omega[I, Z_2](P_5), 0) = 0.$$

A partir de ces valeurs, il en résulte que seul le coefficient de fiabilité $\alpha_{2,2}$ est non nul et égal à $\alpha_{2,2} = 0.47$.

Pour le document I_3 , qui est présentée à la figure 5.8, on dispose d'une image, d'un commentaire vocal et d'un commentaire textuel. Dans l'image associée, on suppose que le détec-

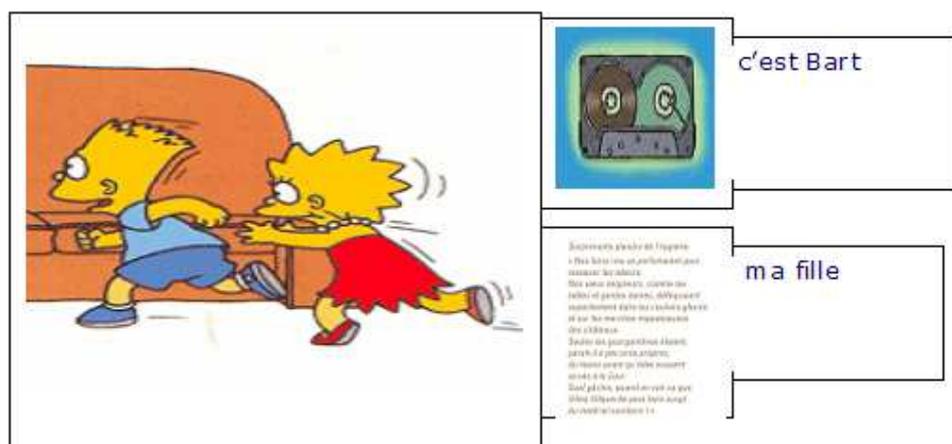


Figure 5.8 : Document I_3 de la base indexée : Image et commentaires Associés.

teur de visage (DV) n'a rien détecté. Le moteur RPa a détecté le mot "Bart" avec le score $RPa(P_3) = 0.6$ sans recouvrement et l'outil RPt a détecté le mot "ma fille" qui engendre deux scores $RPt(P_4) = RPt(P_5) = 0,8$ avec un recouvrement qui conduit aux matrices suivantes :

Matrice de recouvrement pour le mot "Bart" :

$$Mr_1 = \begin{pmatrix} & P_1 & P_2 & P_3 & P_4 & P_5 \\ P_1 & * & 0 & 0 & 0 & 0 \\ P_2 & 0 & * & 0 & 0 & 0 \\ P_3 & 0 & 0 & * & 0 & 0 \\ P_4 & 0 & 0 & 0 & * & 0 \\ P_5 & 0 & 0 & 0 & 0 & * \end{pmatrix}. \quad (5.21)$$

Matrice de recouvrement pour le mot "ma fille" :

$$Mr_2 = \begin{pmatrix} & P_1 & P_2 & P_3 & P_4 & P_5 \\ P_1 & * & 0 & 0 & 0 & 0 \\ P_2 & 0 & * & 0 & 0 & 0 \\ P_3 & 0 & 0 & * & 0 & 0 \\ P_4 & 0 & 0 & 0 & * & 1 \\ P_5 & 0 & 0 & 0 & 1 & * \end{pmatrix}. \quad (5.22)$$

Matrice des scores :

$$Scores = \begin{pmatrix} & P_1 & P_2 & P_3 & P_4 & P_5 \\ RP_1(P_i) & 0 & 0 & 0.6 & 0 & 0 \\ RP_2(P_i) & 0 & 0 & 0 & 0.8 & 0.8 \end{pmatrix}. \quad (5.23)$$

Le détecteur de visage DV n'ayant détecté aucune zone, la fonction de masse résultante correspond à une source d'information neutre d'où $m_{\alpha_I}^\Psi[I, \cdot](\Psi) = 1$. Ceci permet de déduire les scores qui seront manipulés par le processus d'exploitation des informations contextuelles :

$$Score(P_1, I_3) = bel^\Psi[I \oplus Com, P_1](Per) = bel^\Psi[Com, P_1](Per) = 0;$$

$$Score(P_2, I_3) = bel^\Psi[I \oplus Com, P_2](Per) = bel^\Psi[Com, P_2](Per) = 0;$$

$$Score(P_3, I_3) = bel^\Psi[I \oplus Com, P_3](Per) = bel^\Psi[Com, P_3](Per) = 0.7;$$

$$Score(P_4, I_3) = bel^\Psi[I \oplus Com, P_4](Per) = bel^\Psi[Com, P_4](Per) = 0.8;$$

$$Score(P_5, I_3) = bel^\Psi[I \oplus Com, P_5](Per) = bel^\Psi[Com, P_5](Per) = 0.8.$$

De la même façon que pour le document I_2 , on obtient :

$$maxBel_{3,1} = \max(O, \max_{e=1,2}(\prod_{i'=1/i' \neq i}^5 (1 - rec_e(P_{i'}/P_i)) \times RP_e(P_i))) = 0;$$

$$maxBel_{3,2} = \max(O, \max_{e=1,2}(\prod_{i'=1/i' \neq i}^5 (1 - rec_e(P_{i'}/P_i)) \times RP_e(P_i))) = 0;$$

$$maxBel_{3,3} = \max(0, \max_{e=1,2}(\prod_{i'=1/i' \neq i}^5 (1 - rec_e(P_{i'}/P_i)) \times RP_e(P_i))) = 0.7;$$

$$\max Bel_{3,4} = \max(0, \max_{e=1,2}(\prod_{i'=1/i' \neq i}^5 (1 - rec_e(P_{i'}/P_i)) \times RP_e(P_i))) = 0;$$

$$\max Bel_{3,5} = \max(0, \max_{e=1,2}(\prod_{i'=1/i' \neq i}^5 (1 - rec_e(P_{i'}/P_i)) \times RP_e(P_i))) = 0.$$

ce qui conduit au coefficient de fiabilité $\alpha_{3,3} = 0.99$, les autres coefficients étant tous nuls.

5.4.3 Informations contextuelles associées aux documents

De façon à extraire les informations contextuelles des différents documents utilisés dans cet exemple, nous avons choisi de dater et de localiser ces documents. Les différents paramètres liés aux notions de proximité temporelle et géographique sont présentés dans le tableau 5.2. Si on

Paramètre	$t_{1..}$	$dist_{1..}$	$dist_{1..}$	$g_T(t_{1..})$	$g_{Lp}(dist_{1..})$	$g_{Le}(dist_{1..})$
I_1 et I_2	600s	1700m	33m	$\simeq 1$	$\simeq 1$	$\simeq 0$
I_1 et I_3	500s	15000m	13611m	$\simeq 1$	$\simeq 0$	$\simeq 1$

Tableau 5.2 : Paramètres liés aux notions de proximité temporelle et Géographique.

prend $v = 10\text{km/h}$, les distances ont subies des modifications (cf. 3ème colonne du tableau 5.2). On remarque d'après les valeurs des fonctions g_T , g_{Lp} et g_{Le} que les images relatives aux documents I_1 et I_2 sont voisines d'un point de vue temporel (et géographique pour I_1). Les documents I_1 et I_3 quant à eux sont proches d'un point de vue temporel mais éloigné d'un point de vue géographique. Ainsi, selon le processus proposé dans ce chapitre, le document I_2 devrait renforcer les métadonnées pour l'indexation de I_1 tandis que le document I_3 devrait les contester.

Les tableaux 5.3 et 5.4 présentent respectivement les fonction de masse contextuelles $m_{\alpha_{2,i}}^\Psi [I_2, P_i]$ et $m_{\alpha_{3,i}}^\Psi [I_3, P_i]$ associées aux documents I_2 et I_3 pour l'indexation du document I_1 . Ces fonctions

$m_{\alpha_{2,i}}^\Psi [I_2, P_i](.)$	P_1	P_2	P_3	P_4	P_5
<i>Per</i>	0.00	0.19	0.00	0.00	0.00
Ψ	1.00	0.81	1.00	1.00	1.00

Tableau 5.3 : Fonction de masse contextuelle $m_{\alpha_{2,i}}^\Psi [I_2, P_i]$ associée au document I_2 .

permettent d'envisager la fusion qui représente l'incertitude liée au contexte. Celle-ci est présentée au tableau 5.5. A partir de ce tableau, on remarque que la métadonnée P_2 va être renforcée par la connaissance et la modélisation du contexte tandis que P_3 va être contestée.

$m_{\alpha_{3,i}}^{\Psi}[I_3, P_i](\cdot)$	P_1	P_2	P_3	P_4	P_5
\overline{Per}	0.00	0.00	0.60	0.00	0.00
Ψ	1.00	1.00	0.40	1.00	1.00

Tableau 5.4 : Fonction de masse contextuelle $m_{\alpha_{3,i}}^{\Psi}[I_3, P_i]$ associée au document I_3 .

$m_{\alpha_{Cont}}^{\Psi}[Cont, P_i]$	P_1	P_2	P_3	P_4	P_5
Per	0.00	0.19	0.00	0.00	0.00
\overline{Per}	0.00	0.00	0.60	0.00	0.00
Ψ	1.00	0.81	0.40	1.00	1.00

Tableau 5.5 : Fonction de masse $m_{\alpha_{Cont}}^{\Psi}[Cont, P_i]$ associée au document I_1 .

5.4.4 Processus de fusion du contenu et du contexte

La dernière étape du processus dans le système **FUSINDEX** consiste à fusionner contenu et contexte. Le tableau 5.6 présente la fonction de masse $m^{\Psi}[I \oplus Com \oplus Cont, P_i]$ associée au document I_1 pour chacune des métadonnées P_i obtenue après fusion. Les métadonnées seront

$m^{\Psi}[I \oplus Com \oplus Cont, P_i]$	P_1	P_2	P_3	P_4	P_5
Per	0.21	0.22	0.10	0.01	0.01
\overline{Per}	0.00	0.00	0.54	0.00	0.00
Ψ	0.79	0.78	0.36	0.99	0.99

Tableau 5.6 : Fonction de masse finale $m^{\Psi}[I \oplus Com \oplus Cont, P_i]$ associée au document I_1 .

triées selon leurs probabilités pignistiques respectives sur l'hypothèse "Per" ce qui conduit à :

- Marge : : $BetP_{m^{\Psi}[I \oplus Com \oplus Cont, P_2]}(Per) = 0.61$;
- Homer : : $BetP_{m^{\Psi}[I \oplus Com \oplus Cont, P_1]}(Per) = 0.60$;
- Lisa : : $BetP_{m^{\Psi}[I \oplus Com \oplus Cont, P_4]}(Per) = 0.54$;
- Maggie : : $BetP_{m^{\Psi}[I \oplus Com \oplus Cont, P_5]}(Per) = 0.54$;
- Bart : : $BetP_{m^{\Psi}[I \oplus Com \oplus Cont, P_3]}(Per) = 0.28$.

A partir de ces degrés de pertinence et par l'utilisation du contexte associé au contenu des documents, les métadonnées $P_1 = Homer$ et $P_2 = Marge$ vont être proposées à l'utilisateur en terme d'indexation ce qui correspond effectivement aux personnes identifiées. On remarque ainsi que la métadonnée P_2 a été renforcée tandis que la métadonnée P_3 a été contestée.

5.5 Conclusion

La méthodologie d'exploitation du contexte d'un document multimédia a été présentée dans ce chapitre. Elle est basée sur une étude du contexte spatio-temporel entre le document à indexer et chaque document capable d'apporter des informations dans la base indexée. Encore une fois, les fonctions de croyance permettent de représenter les incertitudes liées aux informations contextuelles et de fusionner ces informations dans un même référentiel qui permet de quantifier le degré de pertinence associé à chacune des métadonnées.

En particulier, la modélisation des informations contextuelles a permis de mettre en place deux scénarii qui permettent respectivement de :

- renforcer une métadonnée. Si une métadonnée est pertinente pour un document donné alors la croyance dans le fait que cette métadonnée indexe correctement les documents proches dans le temps et dans l'espace est augmentée. Ceci permet de renforcer la croyance ;
- contester une métadonnée. Si une métadonnée est pertinente pour un document alors la croyance dans le fait que cette métadonnée est pertinente pour les documents proches dans le temps et loin dans l'espace est diminuée.

Ces deux mécanismes permettent de valider certaines métadonnées au profit d'autres qui seraient moins pertinentes en terme d'indexation du fait de leurs incertitudes plus importantes. En plus, le processus d'affaiblissement proposé dans ce chapitre a permis de réduire le bruit dû aux fausses détections dans les autres documents en transformant les informations incertaines en état de silence (augmenter le degré de croyance posé sur l'incertitude).

Un exemple nous a enfin permis de mettre en évidence les avantages de l'extraction du contexte. Nous allons détailler dans le chapitre suivant une évaluation plus complète du système **FUSINDEX** pour l'indexation multimodale de documents multimédia avec une application à l'identification de personnes dans des albums de photos personnelles.

Chapitre 6

Évaluation des performances du système

Dans ce chapitre, les différentes expérimentations qui ont été réalisées pour évaluer le processus de fusion proposé dans le système **ImagINUM** sont présentées. Il décrit et analyse les résultats obtenus. Ces évaluations ont été conduites au sein de France Telecom en utilisant le système **Someone** pour gérer les métadonnées associées aux images du corpus utilisé. Les métadonnées manipulées dans ces expérimentations correspondent donc à des topics relatifs aux personnes d'un carnet d'adresses donné (cf. chapitre 3). Nous présentons tout d'abord deux exemples détaillés qui illustrent l'utilité de la fusion d'information dans le système **FUSINDEX**. Nous détaillons ensuite le corpus utilisé pour une évaluation plus large et représentative. Ces expérimentations et leurs résultats analysés sont ensuite présentés avant de terminer ce chapitre avec une conclusion.

6.1 Exemples

Dans cette section, on propose de décrire deux exemples qui permettent, dans un cadre applicatif réel, de mettre en évidence les avantages du système de fusion d'index **FUSINDEX**. On suppose, dans ces exemples, que l'on dispose d'un carnet d'adresses de 8 personnes :

P_1 : Olivier T ;

P_2 : Olivier C ;

P_3 : Michel ;

P_4 : Pascal B ;

P_5 : Irène ;

P_6 : Muriel ;

P_7 : Sonia ;

P_8 : Pascal F.

A partir de ce carnet d'adresses, on en déduit le cadre de discernement Ω suivant :

$$\Omega = \{OlivierT, OlivierC, Michel, PascalB, \dots \\ \dots Irene, Muriel, Sonia, PascalF, Inconnu, Inconnue, *\}.$$

Pour simplifier, ces prénoms sont soit masculin, soit féminin. On a donc les probabilités que la personne soit une femme $F_1 = F_2 = F_3 = F_4 = F_8 = 0$ et $F_5 = F_6 = F_7 = 1$. Chaque personne de ce carnet d'adresse possède entre 3 et 6 images dans la base d'apprentissage du reconnaiseur de visages RV :

- $h(1) = |V_1| = 3$;
- $h(2) = |V_2| = 6$;
- $h(3) = |V_3| = 3$;
- $h(4) = |V_4| = 5$;
- $h(5) = |V_5| = 4$;
- $h(6) = |V_6| = 4$;
- $h(7) = |V_7| = 3$;
- $h(8) = |V_8| = 4$.

Pour les deux exemples cités, on suppose que les informations contextuelles ne sont pas disponibles. Pour mieux comprendre et suivre les calculs, le lecteur trouvera les traces complètes des calculs en annexe de ce document.

6.1.1 Exemple 1

Le document traité dans cet exemple est proposé à la figure 6.1. On dispose, comme on peut le remarquer d'une image de *Muriel* pour laquelle le détecteur de visage a détecté la zone avec un score $s_{DV}(Z_1) = 35$ mais le document n'a pas de commentaire associé. Les distances obtenues par le reconnaiseur de visages RV pour cet exemple sont présentées dans le tableau 6.1. Avec

$d(Z_j, P_i)$	Olivier T	Olivier C	Michel	Pascal B	Irene	Muriel	Sonia	Pascal F
Z_1	1306974	801034	1445199	1422361	1434665	976012	1247921	1080409

Tableau 6.1 : Distances obtenues par le reconnaiseur de visages RV pour l'exemple 1.

ce tableau, on peut s'apercevoir que la distance minimale a été trouvée pour *OlivierC* et que *Muriel* n'a été classé qu'au second rang par le moteur RV . Le reconnaiseur de genre H/F

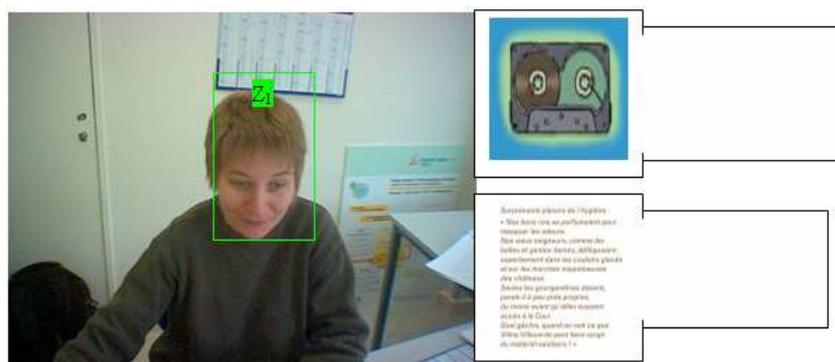


Figure 6.1 : Document de l'exemple 1 à indexer : image sans Commentaire.

fournit les distances suivantes : $d_{f,1} = 224439$ et $d_{h,1} = 360288$ permettant ainsi d'avoir une opinion sur le fait que le visage détecté est celui d'une femme. Nous allons détailler les étapes qui vont permettre à **FUSINDEX** de corriger cette erreur de reconnaissance du moteur *RV*.

Allocation de masses correspondante

Dans les lignes qui suivent, nous présentons les différentes allocations de masse pour chacun des moteurs :

- Jeu de masses issu de *DV* :

$$\begin{aligned} m_{\alpha_{DV}}^{\Omega}[DV, Z_1](\Omega - \{*\}) &= 0.95; \\ m_{\alpha_{DV}}^{\Omega}[DV, Z_1](\Omega) &= 0.05. \end{aligned}$$

- Jeu de masses issu de *RV* :

$$\begin{aligned} m_{\alpha_{RV}}^{\Omega}[RV, Z_1](OlivierC) &= 0.087; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](Muriel \cup PascalF) &= 0.035; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](OlivierT \cup Michel \cup Pascal \cup Irene \cup Sonia) &= 0.01; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](Inconnu \cup Inconnue \cup *) &= 0.26 \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](\Omega) &= 0.608. \end{aligned}$$

- Jeu de masses issu de H/F :

$$\begin{array}{ll}
m^\Omega[HF/OlivierC, Z_1](\Omega - \{OlivierC\}) = 0.92; & m^\Omega[HF/OlivierC, Z_1](\Omega) = 0.08; \\
m^\Omega[HF/OlivierT, Z_1](\Omega - \{OlivierT\}) = 0.92; & m^\Omega[HF/OlivierT, Z_1](\Omega) = 0.08; \\
m^\Omega[HF/Michel, Z_1](\Omega - \{Michel\}) = 0.92; & m^\Omega[HF/Michel, Z_1](\Omega) = 0.08 \\
m^\Omega[HF/PascalB, Z_1](\Omega - \{PascalB\}) = 0.92; & m^\Omega[HF/PascalB, Z_1](\Omega) = 0.08; \\
m^\Omega[HF/PascalF, Z_1](\Omega - \{PascalF\}) = 0.92; & m^\Omega[HF/PascalF, Z_1](\Omega) = 0.08; \\
m^\Omega[HF/Inconnu, Z_1](\Omega - \{Inconnu\}) = 0.92; & m^\Omega[HF/Inconnu, Z_1](\Omega) = 0.08; \\
m^\Omega[HF/Irene, Z_1](\Omega - \{Irene\}) = 0; & m^\Omega[HF/Irene, Z_1](\Omega) = 1; \\
m^\Omega[HF/Sonia, Z_1](\Omega - \{Sonia\}) = 0; & m^\Omega[HF/Sonia, Z_1](\Omega) = 1; \\
m^\Omega[HF/Muriel, Z_1](\Omega - \{Muriel\}) = 0; & m^\Omega[HF/Muriel, Z_1](\Omega) = 1; \\
m^\Omega[HF/Inconnue, Z_1](\Omega - \{Inconnue\}) = 0; & m^\Omega[HF/Inconnue, Z_1](\Omega) = 1.
\end{array}$$

En réalité, le jeu de masses issu de H/F est déduit de la fusion entre toutes les fonctions précédentes. Il conduit, pour cet exemple à une fonction de masse possédant de nombreux éléments focaux (voir annexe) et n'est pas présenté ici. Il faut néanmoins retenir que l'outil H/F défavorise chaque membre masculin du carnet d'adresse y compris l'hypothèse *Inconnu* (cf. section 4.2).

- Jeu de masses issu de $I \oplus Com \oplus Cont$:

On rappelle, dans cet exemple, qu'une seule zone a été détectée, que le document est vide de tout commentaire et que les informations contextuelles ne sont pas prises en considération. Il n'y aura par conséquent pas de fusion due à la gestion des zones détectées (1 seule zone) ni de fusion avec le contexte. Ici, l'agrégation se limite uniquement aux informations sur la zone détectée : $I \oplus Com \oplus Cont = I = I' \oplus IZ = I' = DV \oplus RV \oplus H/F$.

Résultats obtenus par le système FUSINDEX

Le tableau 6.2 présente les crédibilités et les probabilités pignistiques calculées par le système après la fusion d'information sur la zone détectée pour chaque métadonnée : On constate que dans la liste des métadonnées (ordonnées par probabilité pignistique décroissante), *OlivierC* n'occupe désormais plus la première place malgré la reconnaissance de cette personne par le moteur *RV*. Ceci provient de la contestation du reconnaiseur de genre H/F dans la zone détectée. **FUSINDEX** a ainsi permis à la métadonnée *Muriel* d'occuper maintenant la première place dans cette liste de métadonnées ce qui était conforme au résultat attendu.

Prénoms	$bel^\Psi[I \oplus Com \oplus Cont, P_i](Per)$	$BetP_{m^\Psi}[I \oplus Com \oplus Cont, P_i](Per)$
Muriel	0.156	0.578
Olivier C	0.016	0.508
Irène	0.124	0.562
Sonia	0.124	0.562
Michel	0.008	0.504
Pascal B	0.008	0.504
Olivier T	0.008	0.504
Pascal F	0.009	0.504

Tableau 6.2 : Résultats obtenus par le système après la fusion des informations.

6.1.2 Exemple 2

Le document à indexer dans ce second exemple est proposé à la figure 6.2. On suppose

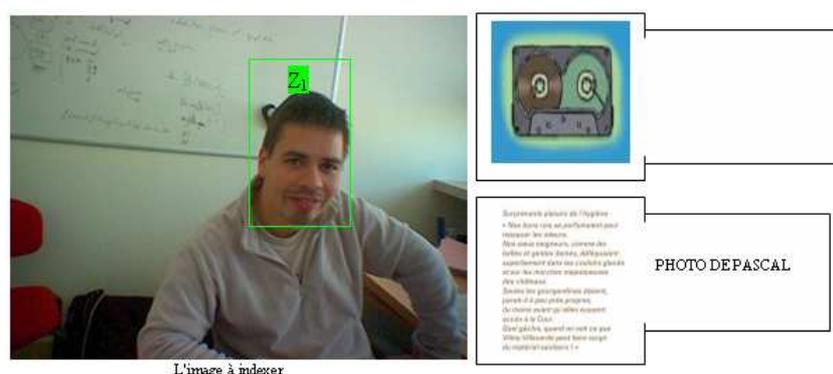


Figure 6.2 : Document de l'exemple 2 à indexer : image avec commentaire textuel.

que le document est constitué d'une image de *Pascal F* associée à un commentaire textuel : "photo de Pascal". Le score obtenu par le moteur *DV* est $s_{DV}(Z_1) = 203$ tandis que les distances obtenues par le reconnaiseur de visages *RV* pour cet exemple sont présentées dans le tableau 6.3. Le reconnaiseur de genre *H/F* fournit les distances suivantes : $d_{f,1} = 315641$ et

$d(Z_j, P_i)$	Olivier T	Olivier C	Michel	Pascal B	Irène	Muriel	Sonia	Pascal F
Z_1	1798729	779543	1280191	1448802	2031266	1214059	1907242	846101

Tableau 6.3 : Distances obtenues par le reconnaiseur de visages *RV* pour l'exemple 2.

$d_{h,1} = 273277$ permettant ainsi d'avoir une opinion sur le fait que le visage détecté est plutôt celui d'un homme. L'analyse du commentaire textuel *RPT* a donné les scores suivants :

- $Rpt(PascalB) = 0.38$, position 10 – 15 ("Pascal");
- $Rpt(PascalF) = 0.38$, position 10 – 15 ("Pascal").

Le système va tacher de représenter les incertitudes liées à ces informations.

Allocation de masses correspondante

- Jeu de masses issu de DV :

$$\begin{aligned} m_{\alpha_{DV}}^{\Omega}[DV, Z_1](\Omega - \{*\}) &= 0.95; \\ m_{\alpha_{DV}}^{\Omega}[DV, Z_1](\Omega) &= 0.05. \end{aligned}$$

- Jeu de masses issu de RV :

$$\begin{aligned} m_{\alpha_{RV}}^{\Omega}[RV, Z_1](OlivierC) &= 0.094; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](PascalF) &= 0.074; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](Michel \cup PascalB \cup Muriel) &= 0.011; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](OlivierT \cup Irene \cup Sonia) &= 0.001; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](Inconnu \cup Inconnue \cup *) &= 0.25; \\ m_{\alpha_{RV}}^{\Omega}[RV, Z_1](\Omega) &= 0.59. \end{aligned}$$

- Jeu de masses issu de H/F :

$$\begin{aligned} m^{\Omega}[HF/OlivierC, Z_1](\Omega - \{OlivierC\}) &= 0; & m^{\Omega}[HF/OlivierC, Z_1](\Omega) &= 1; \\ m^{\Omega}[HF/OlivierT, Z_1](\Omega - \{OlivierT\}) &= 0; & m^{\Omega}[HF/OlivierT, Z_1](\Omega) &= 1; \\ m^{\Omega}[HF/Michel, Z_1](\Omega - \{Michel\}) &= 0; & m^{\Omega}[HF/Michel, Z_1](\Omega) &= 1; \\ m^{\Omega}[HF/PascalB, Z_1](\Omega - \{PascalB\}) &= 0; & m^{\Omega}[HF/PascalB, Z_1](\Omega) &= 1; \\ m^{\Omega}[HF/PascalF, Z_1](\Omega - \{PascalF\}) &= 0; & m^{\Omega}[HF/PascalF, Z_1](\Omega) &= 1; \\ m^{\Omega}[HF/Inconnu, Z_1](\Omega - \{Inconnu\}) &= 0; & m^{\Omega}[HF/Inconnu, Z_1](\Omega) &= 1; \\ m^{\Omega}[HF/Irene, Z_1](\Omega - \{Irene\}) &= 0.073; & m^{\Omega}[HF/Irene, Z_1](\Omega) &= 0.927; \\ m^{\Omega}[HF/Sonia, Z_1](\Omega - \{Sonia\}) &= 0.073; & m^{\Omega}[HF/Sonia, Z_1](\Omega) &= 0.927; \\ m^{\Omega}[HF/Muriel, Z_1](\Omega - \{Muriel\}) &= 0.073; & m^{\Omega}[HF/Muriel, Z_1](\Omega) &= 0.927; \\ m^{\Omega}[HF/Inconnue, Z_1](\Omega - \{Inconnue\}) &= 0.073; & m^{\Omega}[HF/Inconnue, Z_1](\Omega) &= 0.927. \end{aligned}$$

On constate que l'outil H/F n'a pas fortement différencié si l'image contenait un homme ou une femme ce qui aura pour effet de ne pas défavoriser les personnes de sexe féminin. Le moteur d'indexation H/F ne jouera donc pas un rôle décisif dans cet exemple.

- Jeu de masses issu de Com :

$$\begin{aligned}
 m_{\alpha_{Com}}^{\Psi}[Com, OlivierT](\Psi) &= 1; \\
 m_{\alpha_{Com}}^{\Psi}[Com, OlivierC](\Psi) &= 1; \\
 m_{\alpha_{Com}}^{\Psi}[Com, Michel](\Psi) &= 1; \\
 m_{\alpha_{Com}}^{\Psi}[Com, PascalB](Per) &= 0.38, \quad m_{\alpha_{Com}}^{\Psi}[Com, PascalB](\Psi) = 0.62; \\
 m_{\alpha_{Com}}^{\Psi}[Com, Irene](\Psi) &= 1; \\
 m_{\alpha_{Com}}^{\Psi}[Com, Muriel](\Psi) &= 1; \\
 m_{\alpha_{Com}}^{\Psi}[Com, Sonia](\Psi) &= 1; \\
 m_{\alpha_{Com}}^{\Psi}[Com, PascalF](Per) &= 0.38 \quad m_{\alpha_{Com}}^{\Psi}[Com, PascalF](\Psi) = 0.62.
 \end{aligned}$$

- Jeu de masses issu de $I \oplus Com \oplus Cont$:

De la même façon que dans l'exemple 1, on a une seule zone détectée et les informations contextuelles ne sont pas prises en considération. Ceci conduit au processus de fusion suivant : $I \oplus Com \oplus Cont = I \oplus Com = I' \oplus IZ \oplus Com = I' \oplus Com = DV \oplus RV \oplus H/F \oplus Com$.

Résultats obtenus par le système FUSINDEX

Le tableau 6.4 donne la crédibilité et la probabilité pignistique à l'issu du processus de fusion $I \oplus Com$. Pour cette image, l'outil RPt a détecté le prénom $Pascal$ dans le commentaire textuel.

Prénoms	$bel^{\Psi}[I \oplus Com \oplus Cont, P_i](Per)$	$BetP_{m^{\Psi}[I \oplus Com \oplus Cont, P_i]}(Per)$
Pascal F	0.461	0.730
Pascal B	0.418	0.709
Olivier C	0.148	0.574
Michel	0.057	0.528
Olivier T	0.053	0.526
Muriel	0.052	0.526
Irène	0.049	0.524
Sonia	0.049	0.524

Tableau 6.4 : Résultats obtenus par le système après la fusion des informations $I \oplus Com$.

Comme il y a deux personnes dans le carnet d'adresses qui se prénomment $Pascal$ ($PascalB$ et $PascalF$), le RPt propose ces deux personnes avec le même score. L'outil RV a reconnu $PascalB$ à la cinquième place et $PascalF$ en deuxième place. Là encore, l'étape de fusion permet de trouver la bonne personne $PascalF$ (voir tableau 6.4). Cet exemple montre comment la fusion entre I et

Com a permis de lever l'ambiguïté du commentaire textuel. D'autre part, le résultat donné par le système **FUSINDEX** n'a pas été perturbé par la mauvaise reconnaissance du moteur *RV* mais a tiré profit de l'ordre fourni par ce dernier pour trouver la bonne indexation.

6.2 Corpus et méthodologie d'évaluation

6.2.1 Le corpus

Le corpus a été réalisé au sein de France Telecom au début d'une campagne d'évaluation et contient 453 images dont 125 ont été choisies pour construire la base d'apprentissage du reconnaiseur de visage *RV*. Il reste donc 328 images qui permettent de constituer la base à indexer. Dans cette base, 326 images contiennent des commentaires vocaux (un seul commentaire au maximum par image). Ces images sont prises par un cameraphone (SPV M2000) dans des conditions réelles : caméra non stable, résolution faible 640×400 , commentaires bruités, etc. Les participants à la campagne d'évaluation représentent les personnes du carnet d'adresses et ont chacun commenté les images d'une autre personne.

Notre but dans ces expérimentations n'est pas d'évaluer les moteurs d'indexation eux-mêmes, mais d'évaluer leur apport et leur adaptations à notre application (albums de photos personnelles). En effet, les performances des outils *RV*, *H/F*, etc. varient d'une application à l'autre et dépendent énormément des conditions d'usages (albums photos, caméra de surveillance, identification, ...). Nous avons donc veillé au maximum à éviter l'adaptation de notre corpus de test aux outils d'indexation mais au contraire de définir ce corpus de manière à ce qu'il représente le mieux possible les conditions d'exploitation visées. Dans ce but, un faible nombre d'images d'apprentissage pour chaque personne a été volontairement choisi pour le moteur *RV*. De même, nous avons évité d'acquérir uniquement des commentaires vocaux non bruités pour améliorer les taux de reconnaissance du moteur *RPa*.

Le carnet d'adresses choisi contient 25 personnes ($n = 25$) connues dont 11 de sexe féminin et 14 de sexe masculin. Les personnes connues sont identifiées par leurs id ($i = 1, 2 \dots 25$) leurs prénoms N_i et (lorsqu'elles sont disponibles) les images d'apprentissage $V_i = \{V_{i,1}, \dots, V_{i,h(i)}\}$. Le tableau 6.5 illustre les différentes personnes du carnet d'adresses avec leurs prénoms et leurs images d'apprentissage associées. Pour s'approcher de la réalité, le nombre d'images d'apprentissages n'est pas identique pour toutes les personnes. Dans ce tableau, on remarque que $h(23) = 0$ ce qui signifie que cette personne *Magalie* ne possède aucune image d'apprentissage dans la base. Pourtant, puisqu'elle se trouve dans le carnet d'adresses, c'est une personne connue qui pourrait être identifiée uniquement via son prénom dans un commentaire.

id : i	Prénoms : N_i	Images d'apprentissage : V_i	Probabilité : F_i
1	Lisa		1
2	Saïd		0
3	Olivier T		0
4	Olivier C		0
5	Gilles		0
6	Guy		0
7	Irène		1
8	Jérôme		0
9	Johannes		0
10	Michel		0
11	Morgane		1
12	Muriel		1
13	Patrick		0
14	Pascal B		0
15	Pascal F		0
16	Pierre		0
17	Valérie		1
18	Laurent		0
19	Sonia		1
20	Frédérique		0
21	Céline		1
22	Gérard		0
23	Magalie	pas de visages	1
24	Liv		1
25	Vanessa		1

Tableau 6.5 : Les 25 personnes connues du carnet d'adresses.

6.2.2 Conditions d'évaluations

Nous détaillons les conditions dans lesquelles ont été réalisées les expérimentations.

Mesure d'évaluation

Chaque moteur d'indexation trie les métadonnées selon leurs degrés de pertinence respectifs. Le but du moteur est alors de présenter la bonne métadonnée en listant le moins de propositions possibles. La mesure d'évaluation retenue est donc le taux de bonne réponse en fonction du nombre de métadonnées proposée. Une réponse sera considérée comme bonne si et seulement si la liste proposée par le moteur contient au moins une personne présente dans l'image. Il est à remarquer que, dans notre corpus, une seule personne est présente par image.

Ainsi, pour un moteur qui propose un ensemble de Y métadonnées, on dit que la réponse du moteur est bonne si et seulement si parmi ces Y métadonnées, il y a la bonne métadonnée désirée. Pour mieux comprendre cette mesure, considérons l'exemple suivant. Soit le carnet d'adresses suivant $\{\text{Michel}, \text{Patrick}, \text{Saïd}\}$ et soit une image de *Michel*. Soient trois moteurs d'indexation qui donnent les scores suivants :

- Moteur1 : $\text{score}(\text{Saïd}) = \text{score}(\text{Michel}) = \text{score}(\text{Patrick})$;
- Moteur2 : $\text{score}(\text{Saïd}) > \text{score}(\text{Michel}) > \text{score}(\text{Patrick})$;
- Moteur3 : $\text{score}(\text{Michel}) > \text{score}(\text{Patrick}) > \text{score}(\text{Saïd})$.

Les réponses obtenues pour différentes valeurs de Y sont présentées dans la liste suivante :

Cas $Y = 1$:

- Moteur1 propose $\{\}$ \mapsto erreur
- Moteur2 propose $\{\text{Saïd}\}$ \mapsto erreur

Cas $Y = 2$:

- Moteur1 propose $\{\}$ \mapsto erreur
- Moteur2 propose $\{\text{Saïd}, \text{Michel}\}$ \mapsto bonne réponse
- Moteur3 propose $\{\text{Michel}, \text{Patrick}\}$ \mapsto bonne réponse

Cas $Y = 3$:

- Moteur1 propose $\{\text{Patrick}, \text{Saïd}, \text{Michel}\}$ \mapsto bonne réponse
- Moteur2 propose $\{\text{Saïd}, \text{Michel}, \text{Patrick}\}$ \mapsto bonne réponse
- Moteur3 propose $\{\text{Michel}, \text{Patrick}, \text{Saïd}\}$ \mapsto bonne réponse

Le Moteur1 n'a rien proposé dans les cas $Y = 1$ et $Y = 2$ car il a attribué le même score aux trois personnes *Michel*, *Patrick* et *Saïd* leur permettant uniquement d'être classées au troisième rang.

Dans les expérimentations proposées, nous nous intéressons surtout à la détection des personnes du carnet d'adresses. Ainsi, le nombre maximum de métadonnées pouvant être proposées par un des moteurs d'indexation correspond à la taille du carnet d'adresses (25 dans le cas de notre corpus).

Nombre de métadonnées proposées

Suivant le moteur d'indexation, les réponses peuvent être de nature différente. Le détecteur de visage (*DV*) propose soit toutes les personnes connues soit il ne propose rien. Pour cette raison, ses performances ne pourront être affichées. Le reconnaiseur de genre *H/F* peut proposer les 11 femmes du carnet d'adresses ou au contraire les 14 hommes du carnet d'adresses mais peut également proposer les 25 personnes du carnet d'adresses s'il détecte les deux genres avec le même score. Le moteur *RPa* fournit seulement les deux métadonnées qui ont le meilleur score. Le *RV* ordonne toutes les métadonnées selon leurs scores respectifs. Ainsi il peut proposer n'importe quel nombre des métadonnées (entre 1 et 25). En conséquence, le nombre de métadonnées proposées par *I*, *Com* et *Cont* dépend des outils d'indexation et du résultat de la fusion.

Moteur *DV* et évaluation

Dans le corpus choisi, toutes les images ne contiennent qu'un seul visage d'une personne présente dans le carnet d'adresses. L'outil *DV* a correctement détecté 305 visages de la base avec un pourcentage de réussite de 93%. Il a donné 16 fausses alertes sur l'existence d'un deuxième visage avec un pourcentage d'erreur de 5%. Nous n'avons pas choisi d'évaluer séparément le détecteur de visage *DV* car il aurait fallu intégrer l'hypothèse $\{*\}$ (qui correspond à l'événement 'pas de visage') dans les métadonnées. Ce choix peut se justifier par les bonnes performances du détecteur *DV* sur les images testées. Par contre, les informations issues du *DV* sont fusionnées avec *RV*, *H/F* et *IZ* pour obtenir l'analyse du contenu de l'image *I*.

Le reconnaiseur de genre *H/F*

Le reconnaiseur de genre *H/F* s'inspire de l'outil *RV* en définissant deux classes de personnes "homme" et "femme". Pour la classe "homme", on considère toutes les images des hommes de la base d'apprentissage (75 images) de façon identique à la classe "femme" (50 images). Pour différentes raisons, nous avons créé un moteur d'indexation simulé homme/femme qui sera noté *H/F'* pour visualiser le gain maximum qui pourrait être engendré par un tel moteur.

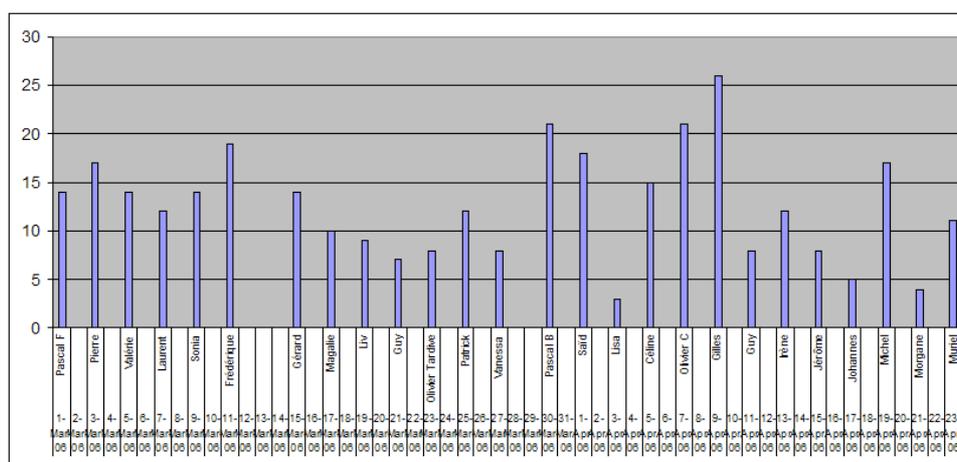
Commentaire vocal

Les commentaires audios du corpus ont été pris dans des conditions réelles avec un SPV M2000. Or, il s'avère que pour environ 60% des commentaires, le *RPa* ne propose aucune méta-donnée car le signal est trop bruité ou trop faible. Ceci s'explique par un problème d'ergonomie de l'outil d'annotation vocale du SPV révélé en observant les participants de l'expérimentation. En effet, les personnes annotent vocalement une photo en la gardant sous les yeux. Ainsi, ils tiennent le terminal mobile à distance pour regarder la photo laissant le micro trop éloigné pour enregistrer une annotation vocale de qualité. Pour garder le réalisme du corpus, nous n'avons pas cherché à acquérir de nouveaux commentaires plus exploitables.

6.2.3 Les dates et les coordonnées géographiques

Pour l'évaluation du processus d'intégration d'informations contextuelles *Cont*, nous avons besoin d'images avec leurs dates de capture et les coordonnées géographiques (longitude et latitude) qui désignent l'endroit où l'image a été capturée. Ces données ont été simulées. Nous avons attribué les dates et les coordonnées géographiques de manière statistique suivant une logique qui permette de tester les scénarii respectifs de renforcement et contestation. On s'est fixé tout d'abord une date centrale. Ensuite, on a tiré aléatoirement des dates générées autour de cette date centrale dans un intervalle selon une loi uniforme. Ce tirage a été réalisé pour chaque ensemble d'image (en pratique, toutes les images incluses dans le même sous-répertoire). Par exemple, pour les images de *Michel* qui se trouvent dans un sous-répertoire, on a fixé une date centrale à 12h00 et la taille de l'intervalle de 20 min. Ainsi, toutes les images de ce répertoire possèdent des dates dans l'intervalle [11h50-12h10]. Cette même opération de tirage statistique sera appliquée pour la distribution des longitudes et des latitudes. La taille d'intervalle étant fixée à ϵ , toutes les images ont alors une latitude *lat* tirée uniformément dans $[lat - \frac{\epsilon}{2}, lat + \frac{\epsilon}{2}]$ et une longitude *long* dans $[long - \frac{\epsilon}{2}, long + \frac{\epsilon}{2}]$.

Dans notre corpus, nous avons attribué de deux manières différentes les dates et les coordonnées géographiques : **DatesCordonnéesA** et **DatesCordonnéesB** respectivement. Les histogrammes présentés aux figures 6.3 et 6.4 illustrent ces distributions. Dans la distribution **DatesCordonnéesA**, il n'y a pas d'interférences entre les personnes dans le temps. Ceci nous permet de tester le processus de renforcement. Dans la distribution **DatesCordonnéesB**, pour chaque intervalle de temps, on a choisi des images de deux personnes chacune dans un lieu différent. Ceci nous permet de tester le mécanisme de renforcement mais aussi la contestation. La personne *Guy* est la seule personne qui possède des images dans deux lieux et deux intervalles de temps différents car son corpus d'apprentissage a été découpé en deux ensembles d'images.



- l’analyse de l’image en fusionnant DV , RV et H/F à l’aide d’un système de vote majoritaire.
- l’analyse de l’image fusionnée avec l’analyse de commentaires : $I \oplus Com$.
- l’analyse de l’image en fusionnant DV , RV , H/F et Com à l’aide d’un système de vote majoritaire.
- la fusion du contexte avec analyse de l’image : $Cont \oplus I$.
- la fusion du contexte avec analyse de l’image et des commentaires : $Cont \oplus I \oplus Com$.

Chaque évaluation sera représentée par une courbe dans un graphique qui permet de montrer comment le processus de fusion améliore le taux de bonnes métadonnées proposées. Pour mener à bien ces expérimentations, nous avons réalisé 3 évaluations qui sont décrites dans les paragraphes suivants.

Pour évaluer notre système Fusindex, un système de vote majoritaire basé sur l’algorithme de Borda (cf. section 2.1.1) sera utilisé dans cette évaluation. Ce système a été comparé avec succès à d’autres systèmes de vote majoritaire [81].

6.3.1 Évaluation 1 : Fusion du contenu

Les performances des différents moteurs et des différentes fusions sont données à la figure 6.5. L’analyse de ces résultats conduit aux constatations suivantes. On remarque que les résultats de la fusion I sont presque identiques à ceux du RV seul. Ceci reste vrai même lorsque H/F possède un taux de bonne réponse de 67%. Si on analyse plus finement ce résultat, on constate que le moteur H/F donne souvent des bonnes réponses avec des écarts de distances entre femmes et hommes suffisamment grandes sur les images où le RV a déjà bien reconnu la personne. Ceci est probablement dû au fait que les bases d’apprentissage utilisées par les moteurs RV et H/F sont quasi-identiques d’un point de vue de leurs compositions. D’après le graphique de la figure 6.5, il est à noter qu’à partir du deuxième rang, le taux de bonne réponse pour l’analyse des commentaires Com est de 40%, pour le RV de 58% et celui de la fusion $I \oplus Com$ est de 76% ce qui nous semble être un bon exemple de l’intérêt de fusionner ces deux types d’information.

Contrairement au Fusindex qui montre une certaine intolérance au bruit de l’outil H/F , la performance du système de vote est négativement influencée par les erreurs apportées par H/F .

6.3.2 Évaluation 2 : Utilisation d’un reconnaiseur H/F simulé

Pour cette évaluation, nous avons utilisé un reconnaiseur de genre homme/femme simulé afin de s’approcher au taux de bonne réponse des outils existants [82, 83], qui varie entre 86% et

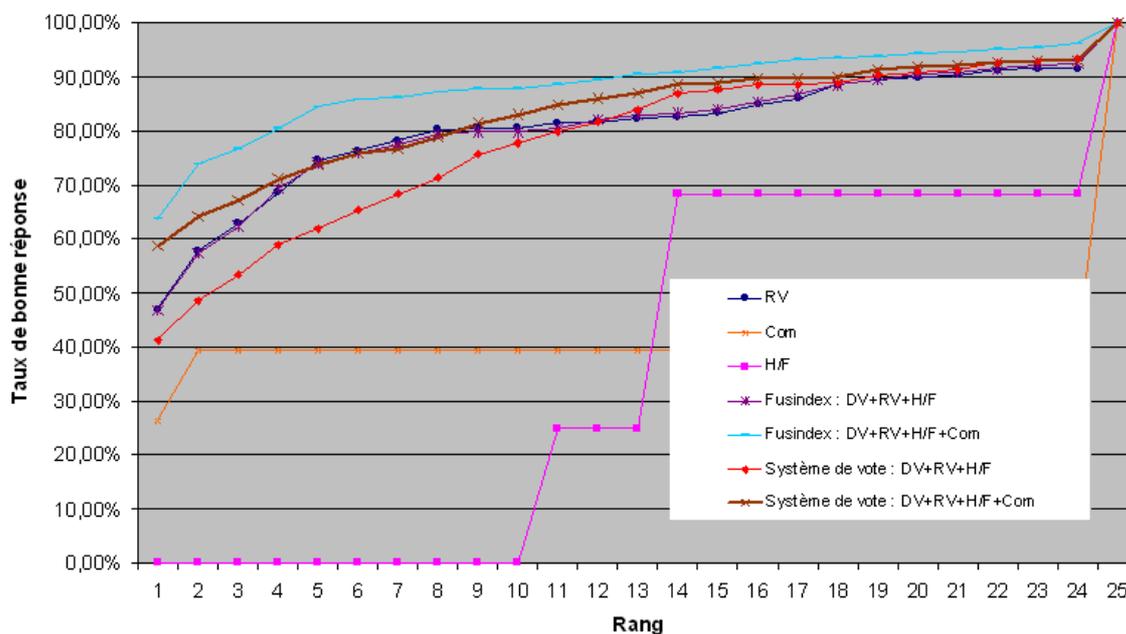


Figure 6.5 : Performances en phase d'évaluation 1 : Fusion du contenu.

94%. Ensuite, valoriser l'utilité que pourrait apporté un tel outil. L'idée consiste à simuler un bon outil de discrimination H/F (indépendant du RV) à l'aide de la loi normale avec un taux de bonne réponse d'environ 92%. Les résultats sont présentés à la figure 6.6. Ce graphique montre que la fusion du RV avec un bon H/F (I avec H/F') améliore les performances du système de 7% à 12% suivant le nombre de métadonnées proposées. De plus, la fusion avec le Com ($I \oplus Com$ avec H/F') continue d'améliorer les performances du système.

On constate ici que la courbe du système de vote majoritaire est presque identique à celui de Fusindex, due à l'imperfection de H/F' .

6.3.3 Evaluation 3 : Fusion contenu et contexte

Après avoir effectué cette panoplie de tests sur le contenu, cette campagne d'évaluation a consisté à utiliser les informations contextuelles extraites des images voisines (dans le temps) afin de les fusionner avec celles des contenus des images à indexer. Ces tests ont été réalisés pour deux affectations aléatoires des dates et des coordonnées géographiques : $DatesCoordonneesA$ où seuls les renforcements sont testés et $DatesCoordonneesB$ où les deux cas de figure (renforcement

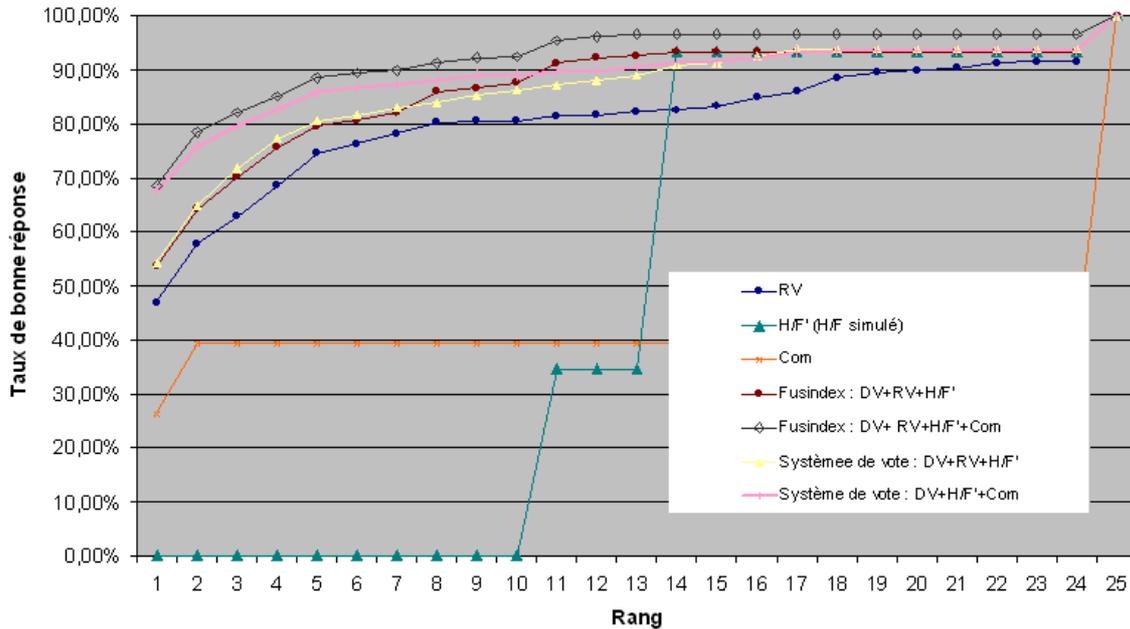


Figure 6.6 : Performances en phase d'évaluation 2 : Fusion du contenu avec H/F' (H/F simulé).

et contestation) sont évalués. Les résultats de ces deux évaluations sont proposés aux figures 6.7 et 6.8.

Ils démontrent que les fusions ($I \oplus Cont$) et ($I \oplus Com \oplus Cont$) avec un reconnaiseur H/F normal ou simulé, améliore de manière remarquable la précision du système. On remarque effectivement sur les courbes de performance que la fusion ($I \oplus Com \oplus Cont$) avec *DatesCordonneesB* donne une meilleure précision que le système avec *DatesCoordoneesA*. En effet, les affectations de date *DatesCoordoneesB* contiennent des informations supplémentaires et le scénario concernant la contestation peut permettre de lever certaines ambiguïtés. Les résultats obtenus ici en considérant le contexte dépendent fortement de la manière dont les dates et coordonnées géographiques ont été générées. Ces résultats nous semblent très satisfaisants et tendent à démontrer l'intérêt de fusionner le contenu et le contexte dans une offre d'indexation multimodale.

6.3.4 Évaluation 4 : Recherche d'images

Cette fois-ci, au lieu d'évaluer la précision de reconnaissance de personnes dans les images, on évalue la précision de retrouver les images correspondant à une personne donnée. Ainsi, la

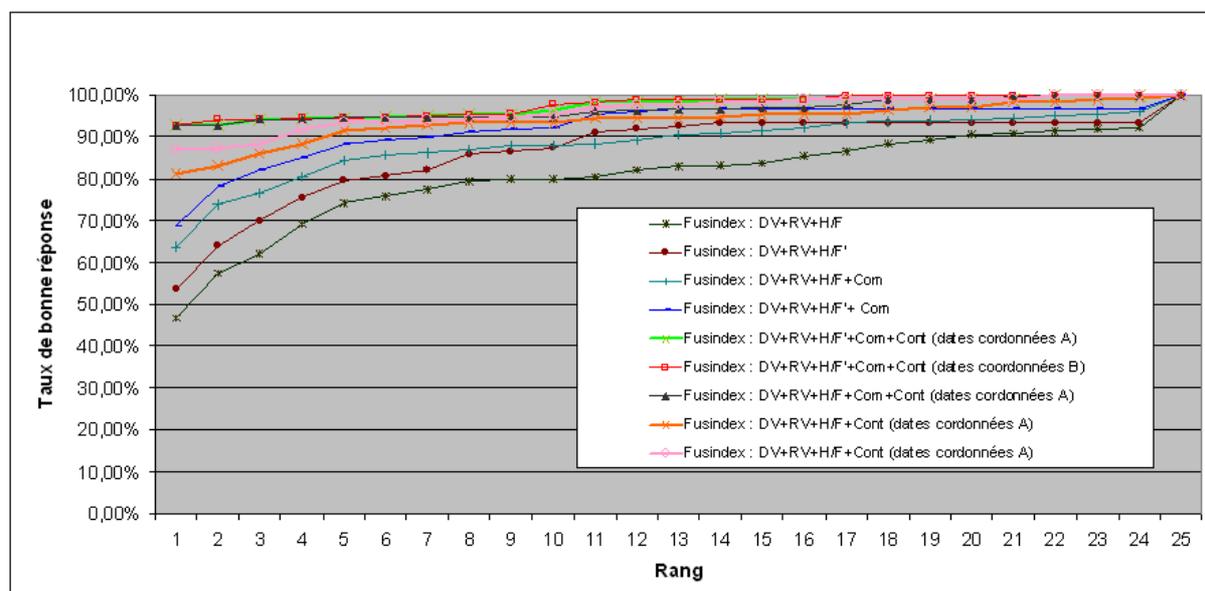


Figure 6.7 : Performances en phase d'évaluation 3 : contenu et contexte.

requête prend la forme d'une identité d'une personne (connue) et la réponse est les images de la base rangées selon les degrés d'appartenance (vis-à-vis du moteur d'indexation ou technique de fusion utilisé) de cette personne requête à chacune de ces images. Ce qui donne 25 requêtes et 328 réponses possibles pour chaque requête. Dans cette évaluation on s'appuie uniquement sur le contenu (I et Com).

Les courbes rappel/précision de les figures 6.8 et 6.9 montrent le gain remarquable apporté par Fusindex par apport aux moteurs d'indexations basé sur une seule modalité (RV , Com : RPa) et par apport au système de vote majoritaire.

6.4 Conclusion

Dans ce chapitre, une description fine de deux exemples issus d'un contexte applicatif réel a été détaillée. Ils ont notamment clairement mis en évidence les avantages du système de fusion d'index **FUSINDEX**. La fusion des index dont l'incertitude a été modélisée par des fonctions de croyance a permis de corriger des erreurs dont l'origine provient de mauvaises détections réalisées par certains outils d'indexation. Ainsi, la contribution d'un moteur d'indexation fournissant une réponse insuffisamment certaine pour être prise en compte est atténuée dans le mécanisme de fusion.

Nous avons ensuite présenté les évaluations faites sur le corpus de documents multimédia

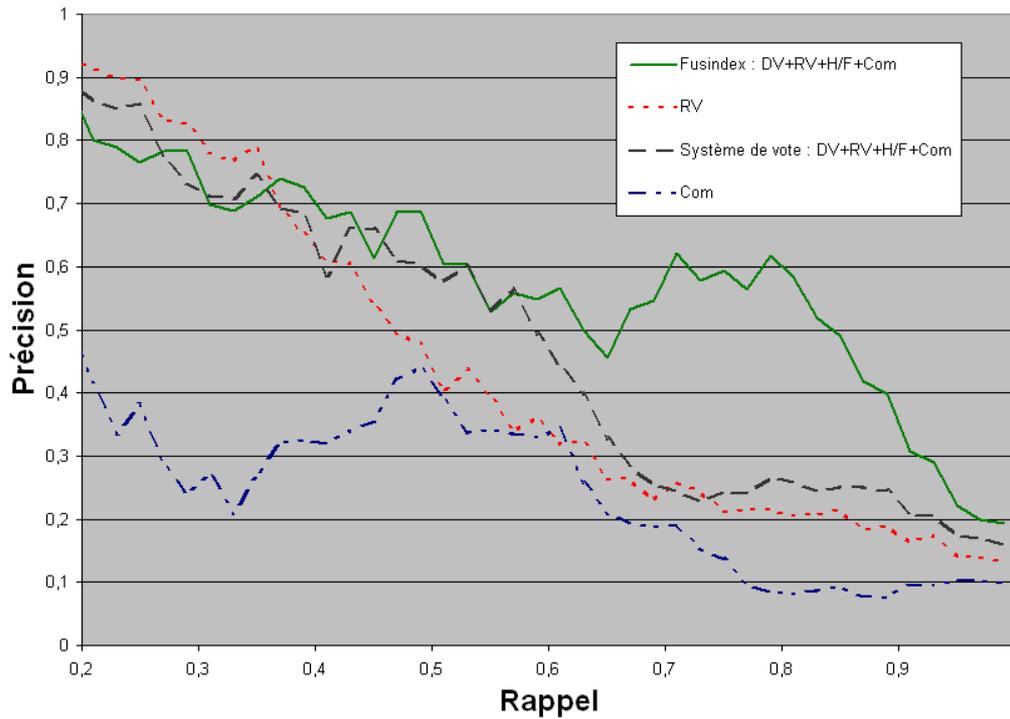


Figure 6.8 : Performance en phase d'évaluation 4 : courbes rappel/précision de *RV*, *Com*, *Fusindex* et système de vote.

recueillis lors d'une campagne d'évaluation qui a duré plusieurs semaines chez France Telecom. Le système **FUSINDEX** a apporté un gain important en ce qui concerne les taux de bonne réponse et a permis d'améliorer la qualité du système d'indexation semi-automatique **Imaginum**. Ce gain est essentiellement dû à la fusion des informations dont l'incertitude et l'imprécision ont été quantifiées par des fonctions de croyance. Cette stratégie a permis de réduire la propagation des erreurs et de lever certaines ambiguïtés causées par l'indexation de certains moteurs.

Ces résultats préliminaires semblent suffisamment convaincants pour qu'il soit mené une plus vaste campagne d'évaluation. Dans le cadre d'un programme de recherche plus conséquent initié à l'instar de l'Agence de l'Innovation Industrielle, les retombées scientifiques de ce travail de thèse devraient être mis à profit. Il est également à noter que, à partir des développements réalisés dans le cadre de ce travail de thèse, "France Telecom" a déposé une demande de brevet français le 17 octobre 2006 sous le numéro 0609084 intitulé *Procédé et système de détermination d'une probabilité de présence d'une personne dans au moins une partie d'une image et programme d'ordinateur correspondant*.

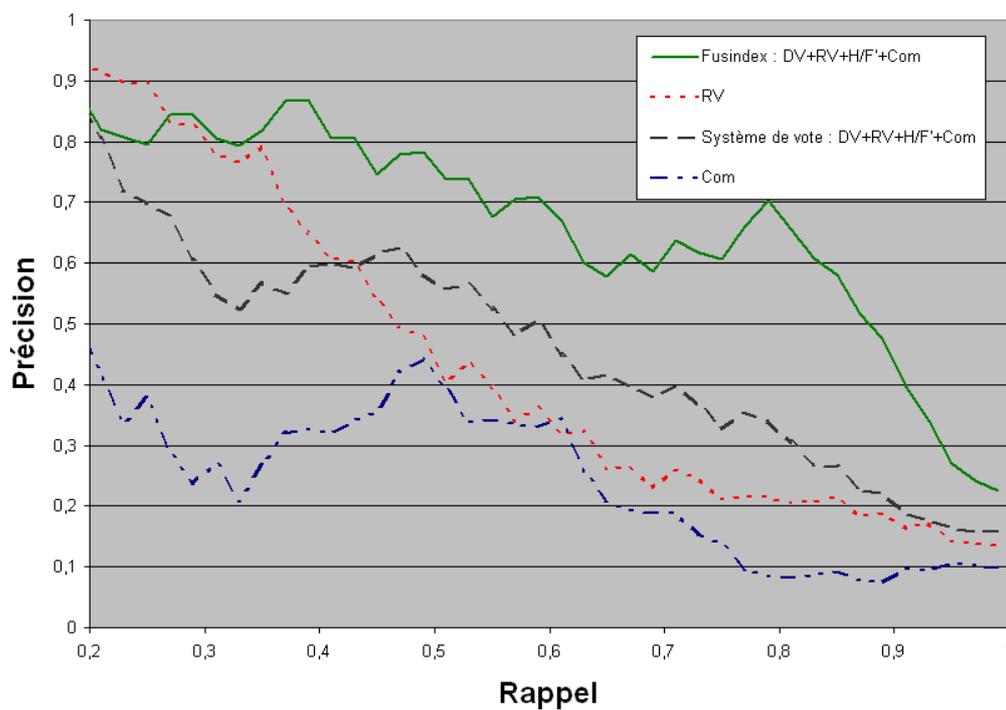


Figure 6.9 : Performance en phase d'évaluation 5 : courbes rappel/précision de *Fusindex* et système de vote avec et sans *H/F* simulé.

Conclusion et Perspectives

L'indexation de documents multimédia est un problème complexe pour lequel il est nécessaire de mettre en œuvre des méthodologies adaptées au regard de l'objectif visé. Dans le cadre de cette thèse, le travail réalisé a permis de montrer comment l'application des fonctions de croyances à l'indexation multimodale de photos personnelles répond aux nombreuses problématiques, en particulier :

- L'imprécision et l'incertitude des résultats d'analyse des contenus et des contextes ;
- La gestion d'hypothèses en monde fermé et ouvert ;
- La diversité des informations disponibles ;
- La gestion du conflit entre les sources d'information ;
- L'accumulation et la propagation d'erreurs ;
- l'exploitation des contenus contextuels.

Plus précisément, nous avons montré dans les chapitres 4 et 5 comment les différents concepts de cette théorie peuvent être utilisés dans le processus de fusion. Différents éléments peuvent être soulignés dans cette conclusion.

Les différents outils disponibles pour manipuler les fonctions de croyance nous ont permis de distinguer les fusions avec et sans conflits. Dans le système **FUSINDEX** développé dans le cadre de cette thèse, les sources de conflit sont principalement :

- Le conflit entre les outils analysant les zones détectées dans l'image (cf. section 4.2) ;
- Le conflit inter-images apporté par l'analyse du contexte (cf. section 5.3.3) ;
- Le conflit entre les références à des personnes détectées dans les commentaires lorsque celles-ci ont un recouvrement (cf. section 4.3).

Deux constats nous ont amené à gérer un processus de fusion sans conflit en particulier lors de l'agrégation des informations issues de l'image et des commentaires. Le premier concerne la non-exhaustivité des commentaires tandis que le second est l'incapacité de détecter l'ensemble des personnes dans une image à partir du détecteur de visage. A chaque étape de fusion, il a aussi fallu distinguer différents cadres de discernement et nous avons défini différents modes

d'allocations de masse pour passer d'un cadre à un autre. Cette allocation de masse n'exploite que les informations disponibles en reportant au plus tard toute décision simplificatrice. Notamment, le report de l'incertitude d'une information sur l'ensemble du cadre de discernement et pas sur son complémentaire, comme ce serait le cas dans le cadre probabiliste, permet de tenir compte du fait que ne pas avoir reconnu une personne ne veut pas dire qu'elle n'est pas dans l'image. L'utilisation de la probabilité pignistique sur un cadre de discernement permettant de quantifier la pertinence des informations fournies par notre système a été choisi comme critère de décision. Ce critère, qui permet de conserver ou au contraire d'éliminer une métadonnée, assure la prise en compte de la plausibilité et de la crédibilité de chaque métadonnée de manière indépendante de celles des autres métadonnées potentielles. Nous avons montré dans le chapitre 6 comment cette fusion améliorerait le taux de rappel de notre système, c'est-à-dire sa capacité à fournir les bonnes métadonnées. La précision de ces métadonnées peut alors être optimisée en fonction du taux de rappel acceptable.

Il est également important de souligner que les stratégies et les algorithmes de fusion développés dans cette thèse permettent facilement d'intégrer d'autres outils d'indexation en s'inspirant des différentes méthodes présentées. Ceci permet à notre système d'être extensible. Il suffit de trouver une méthode pour passer du cadre de discernement initial de ce nouvel outil au cadre quantifiant la pertinence. Le passage à ce cadre de discernement permet à chaque nouvel outil d'indexation de participer avec les autres moteurs déjà intégrés au calcul des pertinences des métadonnées proposées. Au delà du cadre théorique utilisé, cette thèse contribue aussi à mettre en évidence l'intérêt de la fusion multimodale de sources de métadonnées. Elle a aussi montré comment il pouvait être intéressant de fusionner aussi avec d'autres contenus provenant du contexte de l'image. Ceci est plutôt original par rapport aux autres systèmes qui se limitent souvent soit au contexte, soit aux contenus (cf. chapitre 1).

Comme tous travaux de recherche, ceux-ci ont de nombreuses perspectives et pistes d'améliorations. Pour améliorer le système, il serait intéressant d'étudier l'effet d'un système d'apprentissage des coefficients de fiabilité des sources d'informations en utilisant les annotations effectives des utilisateurs à partir des métadonnées proposées. Comme signalé plus haut, il serait aussi intéressant d'intégrer d'autres outils d'indexation d'images ayant des propriétés et qualités différentes, par exemple d'autres reconnaisseurs de visage ou d'autres reconnaisseurs de genre homme/femme. Suite à cette étude, il semble aussi que le calcul de métadonnées globales à l'image imposé par le cadre applicatif visé de cette étude, nous fasse perdre de l'information sur la plausibilité de métadonnées. Il serait donc intéressant de faire évoluer le système pour calculer les métadonnées pour chaque zone. Ceci revient à étudier comment fusionner les commentaires globaux à l'image avec chaque zone de visage détecté. Nous avons aussi fait l'hypothèse que le détecteur de visage n'était pas suffisamment fiable pour lister exhaustivement l'ensemble des

personnes présentes sur une photo. Il serait aussi intéressant d'étudier la remise en cause de cette hypothèse dans le processus de fusion des commentaires et de l'image. Des évaluations plus larges en faisant varier différentes propriétés du corpus utilisé pour évaluer indépendamment les qualités du système développé en fonction de chaque difficulté identifiée pourraient également être menées. Ces tâches d'évaluation sont néanmoins coûteuses en temps et en disponibilité de personnes et de contenus. Pour évaluer l'intérêt de l'exploitation du contexte, il serait intéressant d'exploiter des corpus issus de services de partages de photos afin de profiter au mieux des réseaux sociaux de chaque individu comme sources de métadonnées mais aussi comme sources de contexte.

Les méthodes et les techniques développées dans cette thèse peuvent aussi être appliquées à d'autres contenus multimédia comme des pages web ou surtout des vidéos. Une application prometteuse est la reconnaissance de personnes dans une vidéo en intégrant des outils de reconnaissance du locuteur, de reconnaissance de visages, d'extraction d'entités nommées dans des transcriptions du flux parlé et de reconnaissance de caractères appliqués aux incrustations textuelles dans les vidéos. Un flux vidéo est aussi un media temporel qui permet de fournir naturellement un contexte d'images important à chaque image analysée. Ainsi, nos techniques de fusion contenu/contexte présentées au chapitre 6 doivent pouvoir facilement exploiter des informations sur des images avant et après chaque image analysée.

Bibliographie

- [1] R. Gossweiler and J. Tyler. PLOG : Easily create digital picture stories through cell phone cameras. In *First International Workshop on Ubiquitous Computing (IWUC 2004)*, pages 94–103, Porto, Portugal, April 2004. INSTICC Press.
- [2] M. Davis and al. MMM2 : Mobile media metadata for media sharing. In *Proceedings of 13th Annual ACM International Conference on Multimedia*, pages 267–268, Singapore, 2005. ACM Press.
- [3] M. Naaman, Y-J. Song, A. Pepcke, and H. Garcia-Molina. Automatically generating metadata for digital photographs with geographic coordinates. In *Proceedings of the Thirteenth International World Wide Web Conference (WWW2004)*, pages 244–246, New York, New York, USA, May 2004.
- [4] M. Naaman, R.B. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. In *Proceedings of the Fifth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2005)*, pages 178–187, Denver, Colorado, USA, June 2005.
- [5] K. Toyama, R. Logan, A. Roseway, and P. Anandan. Geographic location tags on digital images. In ACM, editor, *Proceedings of the Eleventh ACM International Conference on Multimedia*, New York, NY, USA, November 2003. ACM Press.
- [6] J.C. Platt, M. Czerwinski, and B.A. Field. PhotoTOC : Automatic clustering for browsing personal photographs. In *Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia*, volume 1, pages 6–10, December 2003.
- [7] L. Wilcox and J. Boreczky. Annotation and segmentation for multimedia indexing and retrieval. In *Proceedings of the 31st Annual Hawaii International Conference on System Sciences*, volume 2, pages 259–267, Hawaii, USA, February 1998.
- [8] W. Adams, G. Iyengary, C. Linz, M. Naphade, C. Neti, H. Nock, and J. Smith. Semantic indexing of multimedia content using visual, audio and text cues. *Eurasip Journal on Applied Signal Processing*, 2 :1–16, Feb 2003.
- [9] J.R. Smith C. Lin, B.L. Tseng. VideoAnnEx : IBM MPEG-7 annotation tool for multimedia indexing and concept learning. In *IEEE International Conference on Multimedia & Expo*, page cdrom, july 2003.

- [10] M. Davis, S. King, N. Good, and R. Sarvas. From context to content leveraging context to infer media metadata. In ACM, editor, *Proceedings of the 12th Annual ACM International Conference*, pages 188–195, New York, New York, USA, October 10-16 2004.
- [11] G. Shafer. A mathematical theory of evidence. *Princeton University Press*, 1976.
- [12] P. Vannoorenberghe. Un état de l’art sur les fonctions de croyance appliquées au traitement de l’information. *Revue I3*, 3(3(2)) :9–45, 2003.
- [13] W. H. Adams, G. Iyengar, C.Y. Lin, M.R. Naphade, C. Neti, H.J. Nock, and J. R. Smith. Content-based multimedia indexing and retrieval. *Eurasip Journal on Applied Signal Processing*, pages 170–185, 2003.
- [14] R.L. de Mántaras and E. Plaza. *Machine Learning*. Springer, 2001.
- [15] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning*. Springer, 2001.
- [16] B.E. Boser, I.M. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pages 144–152, New York, NY, USA, 1992. ACM Press.
- [17] L.D.F. Costa and R. Marcondes Cesar Junior. *Shape Analysis and Classification*. CRC Press, Janvier 2001.
- [18] Y.R. Huang, T.S. Ortega, and M. Mehrotra. Relevance feedback : A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5) :644–655, Septembre 2001.
- [19] Q. Iqbal and J.K. Aggarwal. Feature integration, multi-image queries and relevance feedback inImage retrieval. In *6th International Conference on Visual Information Systems (VISUAL 2003)*, pages 467–474, Miami, Florida, USA, September 2003.
- [20] H-W. Yoo, D-S. Jang, S-H. Jung, J-H. Park, and K-S. Songl. Visual information retrieval system via content-based approach. *Pattern Recognition*, 35(3) :749–769, 2002.
- [21] J. Fournier. *Indexation d’images par le contenu et recherche interactive dans les bases généralistes*. Thèse, Université de Cergy -Pontoise, 2002.
- [22] A.W.M.Smeulders, M.Worrying, S. Santini, A.Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12) :1349–1380, December 2000.
- [23] I. Pitas. *Digital Image Processing Algorithms and Applications*. Wiley-IEEE, 2000.
- [24] M. Ferecatu, M. Boujemaa, and M. Crucianu. Hybrid visual and conceptual image representation within active relevance feedback context. In *7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pages 209–216, Singapore, November 2005. ACM Press.

- [25] S. Philipp-Foliguet, G. Logerot, P. Constant, P.H. Gosselin, and C. Lahanier. Multimedia indexing and fast retrieval based on a vote system. In *In IEEE International Conference on Multimedia & Expo*, pages 1781–1784, Toronto, Ontario, Canada, July 2006.
- [26] R.K. Srihari. Automatic indexing and content-based retrieval of captioned images. *Computer*, 28(9) :49–56, September 1995.
- [27] F. Florea, A. Rogozan, A. Bensrhair, J.-N. Dacher, and S. Darmoni. Modality categorization by textual annotations interpretation in medical imaging. In *Connecting Medical Informatics and Bio-Informatics, Medical Informatics Europe*, pages 1270–1275. R.E. et al., 2005.
- [28] Y. Wu, B.L. Tseng, and J. R. Smith. Ontology-based multi-classification learning for video concept detection. In *IEEE International Conference on Multimedia & Expo (ICME'2004)*, volume 2, pages 1003–1006, Taipei, Taiwan, June 2004. IEEE.
- [29] Y. Wu, E.Y. Chang, K.C.-C. Chang, and J. R. Smith. Optimal multimodal fusion for multimedia data analysis. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 572–579, New York, NY, USA, October 2004. ACM Press.
- [30] J.R. Smith. MARVEL : Multimedia analysis and retrieval system. Technical report, IBM T. J. Watson Research Center, Hawthorne, NY, USA, 2004.
- [31] C. Neti, B. Maison, A. Senior, G. Iyengar, P. DeCuetos, S. Basu, and A. Verma. Joint processing of audio and visual information for multimedia indexing and human-computer interaction. In *Proceedings of Recherche d'Informations Assistée par Ordinateurs*, page cdrom, Paris, France., April 2000.
- [32] J.-P. Chevallet and Joo-Hwee Lim. SnapToTell : Accès ubiquitaire à de l'information multimédia à partir d'un téléphone portable. Application sur une base d'images de Singapour. In *Conférence En Recherche d'Informations et Applications, CORIA '2005*, pages 245–260, Grenoble, France, 9–11 Mars 2005.
- [33] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2004)*, pages 53–62, Tucson, Arizona, USA, June 2004.
- [34] K. Sahr and D. White. Discrete global grid systems. In *Proceedings of the 13th Symposium on the Interface, Computing Science and Statistics*, pages 269–278, 1998.
- [35] A. Graham, H.G. Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In ACM, editor, *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 326–335, Portland, Oregon, USA, July 2002. ACM Press.

- [36] M. Davis and al. MMM2 : Mobile media metadata for media sharing. In *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI 2005)*, pages 1335–1338, Portland, Oregon, April 2-7 2005. ACM Press.
- [37] M. Davis, M. Smith, F. Stentiford, A. Bambidele, J. Canny, N. Good, S. King, and R. Janakiraman. Using context and similarity for face and location identification. In *Proceedings of the IS& T/SPIE 18th Annual Symposium on Electronic Imaging Science and Technology Internet Imaging VII*, page cdrom, San Jose, California, 2006. IS& T/SPIE Press.
- [38] R. Nair and M. Davis. Bluetooth pooling to enrich co-presence information. In *Adjunct Proceedings of the Seventh International Conference on Ubiquitous Computing*, page cdrom, Tokyo, Japan, 2005.
- [39] J. Canny. Collaborative filtering with privacy via factor analysis. In *ACM Conference on Research and Development in Information Retrieval, SIGIR'2002*, pages 238–245, Tampere, Finland, August 2002. ACM.
- [40] M. Windhouwer, A. Schmidt, and M. Kersten. ACOI : A system for indexing multimedia objects. In *International Workshop on Information Integration and Web-Based Applications & Services*, page cdrom, Yogyakarta, Indonesia, November 1999.
- [41] S. Ayache, G. Quénot, and Shin'Ichi Satoh. Context-based conceptual image indexing. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'06*, Toulouse, France, May 15–19 2006.
- [42] P. Jourlin, S.E. Johnson, K.S. Jones, and P.C. Woodland. Spoken document representations for probabilistic retrieval. *Speech communication*, 32(1-2) :21–36, 2000.
- [43] S.K.M. Wong and Y.Y. Yao. A probability distribution model for information retrieval. *Information Processing & Management*, 25(1) :39–53, 1989.
- [44] J.-C. d. Borda. Memoire sur les elections au scrutin. *Histoire de l'Academie Royale des Sciences*, 1781.
- [45] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, AMS 38 :325–339, 1967.
- [46] L.A. Zadeh. Fuzzy sets. *Journal of Information and Control*, 8 :338–353, 1965.
- [47] L. A. Zadeh. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23 :421–427, 1968.
- [48] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.
- [49] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3) :244–264, 1988.

- [50] I.A. Azzam, C.H.C. Leung, and J.F. Horwood. A fuzzy expert system for concept-based image indexing and retrieval. In *Proceedings of the 11th International Multimedia Modelling Conference (MMM'05)*, pages 452–457. IEEE Computer Society, 2005.
- [51] Y. Chen and J.Z. Wang. A region-based fuzzy feature matching approach to Content-based image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9) :1252–1267, 2002.
- [52] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Royal Statistical Society*, 39(1) :1–38, November 1977.
- [53] Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2) :191–234, 1994.
- [54] K. Mellouli Z. Elouedi and P. Smets. Color-based retrieval using spatial-chromatic histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 34 :782–787, 2004.
- [55] M. Arif and N. Vincent. Comparison of three data fusion methods for an offline signature verification problem. Technical report, Laboratoire d'Informatique de Tours, Université François-Rabelais, Tours, 2003.
- [56] L. Rujie and Y. Baozong. Information fusion modal and multi-perception system. In *Fifth Asia-Pacific Conference on Communications*, volume 2, pages 985–989, Beijing, China, 1999.
- [57] Y. Alp Aslandogan, C. Yu, and L. Robert. Experiments in using visual and textual clues for image hunting on theWeb. In *Advances in Visual Information Systems*, pages 108–119, Lyon, France, November 2000.
- [58] Y. Alp Aslandogan and C. Yu. Multiple evidence combination in image retrieval : Diogenes searches for people on the web. In *Proceedings of the 23rd Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, pages 88–95, Athens, Greece, June 2000.
- [59] S. Kharbouche, P. Vannoorenberghe, C. Lecomte, and P. Miche. An automatic relevance feedback in image retrieval using belief functions. In LNCS Springer Verlag, editor, *Advances in Visual Computing, ISVC05*, pages 704–710, NV, USA, Decembre 2005.
- [60] A.Trémeau, C. Fernandez-Maloigne, and P.Bonton. *Image couleur : de l'acquisition au traitement*. Editions Dunod, 2004.
- [61] Y. Gong, C. Chuan, and G. Xiaoyi. Image indexing and retrieval using color histograms. *Multimedia Tools and Applications*, 2(2) :133–156, March 1996.
- [62] J.R. Smith and S.F.Chang. VisualSEEK : A fully automated content-based image query system. In *ACMMM'96 Multimedia*, pages 87–98. ACM, 1996.
- [63] S. Ayache, G. Quénot, J. Gensel, and Shin'Ichi Satoh. Using topic concepts for semantic video shots classification. In *International Conference on Image and Video Retrieval, CIVR'06*, Tempe, AZ, USA, July 13–15 2006.

- [64] L. Agosto, M. Plu, P. Bellec, and L. Vignollet. Someone : A cooperative system for personalized information exchange. In *International Conference of Enterprise Information Systems*, pages 71–78, Angers, France, 2003.
- [65] K. Rodden and K.R. Wood. How do people manage their digital photographs? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 409–416, Ft Lauderdale, Florida, USA, 2003.
- [66] M. Visani. *Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages*. Thèse, Institut National des Sciences Appliquées de Lyon, 2005.
- [67] D. Charlet and al. Neologos : an optimized database for the development of new speech processing algorithms. In *Proceedings of the 9th European Conference on Speech Communication and Technology INTERSPEECH'2005*, pages 1549–1552, Lisbon, Portugal, September 2005.
- [68] G. Smits, M. Plu, and P. Bellec. Personal semantic indexation of images using textual annotations. In *The First International Conference on Semantics and Digital Media Technology, SAMT2006*, pages 71–85, Athens, Greece, 2006.
- [69] C. Garcia and M. Delakis. Convolutional face finder : A neural architecture for fast and robust face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11) :1408–1423, 2004.
- [70] M. Visani, C. Garcia, and C. Laurent. Comparing robustness of two-dimensional PCA and eigenfaces for face recognition. In *Proceedings of the International Conference on Image Analysis and Recognition*, volume 3212/2004, pages 717–724, 2004.
- [71] A. Pigeau and M. Gelgon. Incremental statistical geo-temporal structuring of a personal camera phone image collection. In *Proceedings of the 17th International Conference on Pattern Recognition, ICPR'2004*, volume 3, pages 878–881, 2004.
- [72] M. Lalmas and P. Vannorenberghe. Indexation et recherche de documents XML par les fonctions de croyance. In *Conférence En Recherche d'Informations et Applications CORIA*, pages 143–160, Toulouse, France, Mars 2004.
- [73] T. Dencœur. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transaction on Systems, Man and Cybernetics*, 25(5) :804–813, 1995.
- [74] A. Martin. Fusion d'information haut niveau application à la classification d'images sonar. In *Atelier : Fouille de Données Complexes, EGC'05*, page cdrom, Paris, France, Janvier 2005.
- [75] S. Kharbouche and P. Vannorenberghe. Un filtre crédal pour la segmentation d'images. In *IEEE Troisième Conférence Internationale Francophone d'Automatique, CIFA'2004*, page cdrom, Douz, Tunisie, Novembre 2004.

- [76] D. Dubois and H. Prade. *On several representations of an uncertainty body of evidence*, chapter Fuzzy Information and Decision Processes, pages 167–181. M.M. Gupta and E. Sanchez, 1982.
- [77] Ph. Smets. Belief induced by the partial knowledge of the probabilities. In D. Heckerman and Al., editors, *Uncertainty in Artificial Intelligence, UAI'94*, pages 523–530, San Mateo, 1994. Morgan Kaufmann.
- [78] P. Vannoorenberghe and T. Denœux. Handling uncertain labels in multiclass problems using belief decision trees. In *Proceedings of IPMU'2002*, pages 1919–1926, Annecy, France, 2002.
- [79] D. Dubois and H. Prade. Combinaison of information in the framework of possibility theory. *Data fusion in robotics and machine intelligence*, pages 481–505, 1992.
- [80] D. Charlet, G. Mercier, and D. Jouvét. On combining confidence measures for improved rejection of incorrect data. In *EuroSpeech'2001*, page cdrom, Aalborg, Danemark, 2001.
- [81] M. van Erp, L. Vuurpijl, and L. Schomaker. An overview and comparison of voting methods for pattern recognition. In *IEEE ighth International Workshop on Frontiers in Handwriting Recognition(IWFHRŠ02)*, pages 195–200, 2002.
- [82] Y. D. Cheng, A. J. O'Toole, and H. Abdi. Classifying adults' and children's faces by sex : computational investigations of subcategorical feature encoding. *Cognitive science*, 25(5) :659–731, 2001.
- [83] B. Golomb and T. Sejnowski. Sex recognition from faces using neural networks. *Application of Neural Netwoks*, pages 71–92, 1995.

Annexe

Traces de calculs pour l'Exemple 1 du chapitre 6

SCORES DES OUTILS D'INDEXATION

----- DV::

Score Z1 = 189.83999633789062

----- H/F ::

homme: 360288.9375

femme: 224439.71875

-- Note : Bonne reconnaissance de femme

----- RV :

RV :: Olivier T : 1306974.0
 RV :: Olivier C : 801034.8125
 RV :: Michel : 1445199.375
 RV :: Pascal : 1422361.0
 RV :: Irene : 1434665.125
 RV :: Muriel : 976012.875
 RV :: Sonia : 1247921.75
 RV :: Frederique : 9999999.0
 RV :: Pascal F : 1080409.75

-- Note : La bonne personne (Muriel) s'est arrangée au 2ème rang par RV derrière Olivier C

ALLOCATION DES JEUX DE MASSE

----- le jeu de masses de la source DV pour la Z1 -----

$m[DV,Z1](,Olivier\ T,Olivier\ C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal\ F,Inconnu,Inconnue)= 0.9499999999999998$
 $m[DV,Z1](,*)= 1.0547118733938987E-16$
 $m[DV,Z1](,Olivier\ T,Olivier\ C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal\ F,Inconnu,Inconnue,*)= 0.0500000000000000044$

----- le jeu de masses issu de la source RV pour la Z1 -----

$m[RV,Z1](,Olivier\ C)= 0.08691970692950039$
 $m[RV,Z1](,Muriel,Pascal\ F)= 0.03503259043978889$
 $m[RV,Z1](,Olivier\ T,Michel,Pascal,Irene,Sonia)= 0.008876674855377182$
 $m[RV,Z1](,Inconnu,Inconnue,*)= 0.26522790073186026$
 $m[RV,Z1](,Olivier\ T,Olivier\ C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal\ F,Inconnu,Inconnue,*)= 0.6039431270434732$

----- jeux de masses issu de HF pour la Z1-----

----- le jeu de masses issu de la source HF/Olivier T pour la Z1-----

$m[HF/Olivier\ T,Z1](,Olivier\ C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal\ F,Inconnu,Inconnue,*)= 0.9184956366756245$
 $m[HF/Olivier\ T,Z1](,Olivier\ T,Olivier\ C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal\ F,Inconnu,Inconnue,*)= 0.0815043633243755$

-- Note : Olivier C est bien contesté par H/F : la masse allouée au seul élément focal contenant Oliver C est moins importante – ce sera le cas ci-dessous pour chaque prénom masculin

----- le jeu de masses issu de la source HF/Olivier C pour la Z1-----

m[HF/Olivier C,Z1](,Olivier T ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.9184956366756245
 m[HF/Olivier C,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0815043633243755

----- le jeu de masses issu de la source HF/Michel pour la Z1-----

m[HF/Michel,Z1](,Olivier T ,Olivier C ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.9184956366756245
 m[HF/Michel,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0815043633243755

----- le jeu de masses issu de la source HF/Pascal pour la Z1-----

m[HF/Pascal,Z1](,Olivier T ,Olivier C ,Michel ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.9177693096211257
 m[HF/Pascal,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.08223069037887432

----- le jeu de masses issu de la source HF/Irene pour la Z1-----

m[HF/Irene,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0
 m[HF/Irene,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.0

-- Note : Irène étant un prénom 100% féminin , la masse est allouée sur l'élément focal contenant ce prénom

----- le jeu de masses issu de la source HF/Muriel pour la Z1-----

m[HF/Muriel,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0
 m[HF/Muriel,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.0

----- le jeu de masses issu de la source HF/Sonia pour la Z1-----

m[HF/Sonia,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0
 m[HF/Sonia,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.0

----- le jeu de masses issu de la source HF/Frederique pour la Z1-----

m[HF/Frederique,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0
 m[HF/Frederique,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.0

----- le jeu de masses issu de la source HF/Pascal F pour la Z1-----

m[HF/Pascal F,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Inconnu ,Inconnue ,*)= 0.9184956366756245
 m[HF/Pascal F,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0815043633243755

----- le jeu de masses issu de la source HF/Inconnu pour la Z1-----

m[HF/Inconnu,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnue ,*)= 0.9184956366756245
 m[HF/Inconnu,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0815043633243755

----- le jeu de masses issu de la source HF/Inconnue pour la Z1-----

m[HF/Inconnue,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,*)= 0.0
 m[HF/Inconnue,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.0

----- le jeu de masse issu de la fusion des sources DV, RV et HF pour la zone Z1

m[DV+RV+HF,Z1](,Olivier C)= 0.007698986849352014
 m[DV+RV+HF,Z1](,Muriel)= 0.03496904843516707
 m[DV+RV+HF,Z1](,Muriel ,Pascal F)= 0.0031030414462100402
 m[DV+RV+HF,Z1](,Irene ,Sonia)= 0.007469172737272115
 m[DV+RV+HF,Z1](,Pascal ,Irene ,Sonia)= 6.692261598925171E-4
 m[DV+RV+HF,Z1](,Michel ,Irene ,Sonia)= 6.627904850092823E-4

m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Sonia)= 5.9384987695177645E-5
 m[DV+RV+HF,Z1](,Olivier T,Irene,Sonia)= 6.627904850092822E-4
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Sonia)= 5.938498769517763E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Sonia)= 5.8813906502218234E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Sonia)= 5.269633757477137E-6
 m[DV+RV+HF,Z1](,Inconnue)= 0.2515094896234198
 m[DV+RV+HF,Z1](,Inconnu,Inconnue)= 0.022318147199903264
 m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.37408744445472747
 m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 0.03319532262371867
 m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 0.033195322623718666
 m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 0.0029456466941812196
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 0.03351764816833983
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 0.0029742488314680278
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 0.0029742488314680278
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.6392532277502836E-4
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.033195322623718666
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 0.0029456466941812196
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 0.0029456466941812196
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.6138726064801036E-4
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.0029742488314680278
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.6392532277502836E-4
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.639253227750284E-4
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.3419888499215405E-5
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.03319532262371867
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 0.0029456466941812204
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 0.0029456466941812204
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.6138726064801047E-4
 m[DV+RV+HF,Z1](,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.0029742488314680286
 m[DV+RV+HF,Z1](,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.6392532277502847E-4
 m[DV+RV+HF,Z1](,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.6392532277502847E-4
 m[DV+RV+HF,Z1](,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.3419888499215408E-5
 m[DV+RV+HF,Z1](,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.00294564669418122
 m[DV+RV+HF,Z1](,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.6138726064801047E-4
 m[DV+RV+HF,Z1](,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.6138726064801047E-4
 m[DV+RV+HF,Z1](,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.3194668988658966E-5
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnue)= 2.6392532277502847E-4
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.3419888499215408E-5
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.3419888499215408E-5
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.0782059544290845E-6
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.03319532262371868
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 0.002945646694181221
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 0.002945646694181221
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.613872606480105E-4
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.0029742488314680295
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.639253227750285E-4
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.639253227750285E-4
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.3419888499215415E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.002945646694181221
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.613872606480105E-4
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.613872606480105E-4
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.3194668988658976E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnue)= 2.639253227750285E-4
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.3419888499215415E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.3419888499215415E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.078205954429085E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frederique,Inconnue)= 0.00294564669418122
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.6138726064801047E-4
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.6138726064801047E-4
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.3194668988658966E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Inconnue)= 2.6392532277502847E-4
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.3419888499215408E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.3419888499215408E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.0782059544290845E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Inconnue)= 2.6138726064801047E-4
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.3194668988658966E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.319466898865897E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 2.0582206958354034E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnue)= 2.3419888499215408E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue)= 2.0782059544290845E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnue)= 2.0782059544290845E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frederique,Pascal F,Inconnu,Inconnue)= 1.8441334548493648E-7
 m[DV+RV+HF,Z1](,*)= 9.962619455984485E-17
 m[DV+RV+HF,Z1](,Inconnue,*)= 0.013237341559127373
 m[DV+RV+HF,Z1](,Inconnu,Inconnue,*)= 0.0011746393263106995
 m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frederique,Inconnue,*)= 0.01968881286603831
 m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frederique,Inconnu,Inconnue,*)= 0.0017471222433536164

m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 0.0017471222433536164
 m[DV+RV+HF,Z1](,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 0.0017640867457020986
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.5653941218252796E-4
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.5653941218252796E-4
 m[DV+RV+HF,Z1](,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.3890806461843616E-5
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 0.0017471222433536166
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Michel,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.56539412182528E-4
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.3890806461843617E-5
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.3890806461843616E-5
 m[DV+RV+HF,Z1](,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.2326257104850227E-6
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 0.0017471222433536166
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.56539412182528E-4
 m[DV+RV+HF,Z1](,Olivier C,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.3890806461843617E-5
 m[DV+RV+HF,Z1](,Olivier C,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.3890806461843616E-5
 m[DV+RV+HF,Z1](,Olivier C,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.2326257104850227E-6
 m[DV+RV+HF,Z1](,Olivier C,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Olivier C,Michel,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier C,Michel,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier C,Michel,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.220772052034684E-6
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.3890806461843617E-5
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.2326257104850229E-6
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.2326257104850227E-6
 m[DV+RV+HF,Z1](,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.093792607594256E-7
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 0.0017471222433536166
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Olivier T,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.56539412182528E-4
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.3890806461843617E-5
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.3890806461843616E-5
 m[DV+RV+HF,Z1](,Olivier T,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.2326257104850227E-6
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.220772052034684E-6
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.3890806461843617E-5
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.2326257104850229E-6
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.2326257104850227E-6
 m[DV+RV+HF,Z1](,Olivier T,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.093792607594256E-7
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.5503403653585387E-4
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.3757224244632145E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.220772052034684E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.3890806461843617E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.2326257104850229E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.2326257104850227E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 1.093792607594256E-7
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.3757224244632146E-5
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.220772052034684E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.0832740504396873E-7
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Irene,Muriel,Sonia,Frédérique,Inconnue,*)= 1.2326257104850229E-6
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Inconnu,Inconnue,*)= 1.093792607594256E-7
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnue,*)= 1.093792607594256E-7
 m[DV+RV+HF,Z1](,Olivier T,Olivier C,Michel,Pascal,Irene,Muriel,Sonia,Frédérique,Pascal F,Inconnu,Inconnue,*)= 9.705965551838774E-9

----- Probabilité pignistique pour la Z1 issue de la fusion Image -----

-- Note : Résultats de la fusion pour la zone Z1 dans la référence "Omega", ici la bonne personne (Muriel) a la probabilité pignistique la plus grande

BetPm(Olivier T) = 0.00861048275872515

BetPm(Olivier C) = 0.016057935796106544

BetPm(Michel) = 0.008610482758725146

BetPm(Pascal) = 0.008688111479275892

BetPm(Irene) = 0.12464652377789305

BetPm(Muriel) = 0.1567221209603339

BetPm(Sonia) = 0.12464652377789305

BetPm(Frederique) = 0.120201551802062

BetPm(Pascal F) = 0.009910469669859553

BetPm(Inconnu) = 0.01990956898880973

BetPm(Inconnue) = 0.3898803322471003

BetPm(*) = 0.012115895983216078

----- FUSION Image et commentaire -----

----- allocation de masse sur PSI = { Pertinent , Non-Pertinent }-----

----- le jeu de masses issu de l'Image (dans la reference 'psi') -----

$m[I, Olivier T](Per) = 0.00861048275872515$, $m[I, Olivier T](psi) = 0.9913895172412749$
 $m[I, Olivier C](Per) = 0.016057935796106544$, $m[I, Olivier C](psi) = 0.9839420642038934$
 $m[I, Michel](Per) = 0.008610482758725146$, $m[I, Michel](psi) = 0.9913895172412749$
 $m[I, Pascal](Per) = 0.008688111479275892$, $m[I, Pascal](psi) = 0.9913118885207242$
 $m[I, Irene](Per) = 0.12464652377789305$, $m[I, Irene](psi) = 0.8753534762221069$
 $m[I, Muriel](Per) = 0.1567221209603339$, $m[I, Muriel](psi) = 0.8432778790396661$
 $m[I, Sonia](Per) = 0.12464652377789305$, $m[I, Sonia](psi) = 0.8753534762221069$
 $m[I, Frederique](Per) = 0.120201551802062$, $m[I, Frederique](psi) = 0.879798448197938$
 $m[I, Pascal F](Per) = 0.009910469669859553$, $m[I, Pascal F](psi) = 0.9900895303301405$

le jeu de masses issu de la fusion des commentaires vocaux et textuels (dans la référence 'psi')

$m[Com, Olivier T](Per) = 0.0$, $m[Com, Olivier T](psi) = 1.0$
 $m[Com, Olivier C](Per) = 0.0$, $m[Com, Olivier C](psi) = 1.0$
 $m[Com, Michel](Per) = 0.0$, $m[Com, Michel](psi) = 1.0$
 $m[Com, Pascal](Per) = 0.0$, $m[Com, Pascal](psi) = 1.0$
 $m[Com, Irene](Per) = 0.0$, $m[Com, Irene](psi) = 1.0$
 $m[Com, Muriel](Per) = 0.0$, $m[Com, Muriel](psi) = 1.0$
 $m[Com, Sonia](Per) = 0.0$, $m[Com, Sonia](psi) = 1.0$
 $m[Com, Frederique](Per) = 0.0$, $m[Com, Frederique](psi) = 1.0$
 $m[Com, Pascal F](Per) = 0.0$, $m[Com, Pascal F](psi) = 1.0$

-- Note : Il n'y a pas de commentaire

le jeu de masses issu de la combinaison entre l'Image et les commentaires (dans la référence 'psi')

$m[+Com, Olivier T](,Per) = 0.00861048275872515$
 $m[+Com, Olivier T](,Per ,non-Per) = 0.9913895172412749$

$m[+Com, Olivier C](,Per) = 0.016057935796106544$
 $m[+Com, Olivier C](,Per ,non-Per) = 0.9839420642038934$

$m[+Com, Michel](,Per) = 0.008610482758725146$
 $m[+Com, Michel](,Per ,non-Per) = 0.9913895172412749$

m[+Com,Pascal](,Per)= 0.008688111479275892
 m[+Com,Pascal](,Per ,non-Per)= 0.9913118885207242

m[+Com,Irene](,Per)= 0.12464652377789305
 m[+Com,Irene](,Per ,non-Per)= 0.8753534762221069

m[+Com,Muriel](,Per)= 0.1567221209603339
 m[+Com,Muriel](,Per ,non-Per)= 0.8432778790396661

m[+Com,Sonia](,Per)= 0.12464652377789305
 m[+Com,Sonia](,Per ,non-Per)= 0.8753534762221069

m[+Com,Frederique](,Per)= 0.120201551802062
 m[+Com,Frederique](,Per ,non-Per)= 0.879798448197938

m[+Com,Pascal F](,Per)= 0.009910469669859553
 m[+Com,Pascal F](,Per ,non-Per)= 0.9900895303301405

la Pertinence des meta-données après l'Image et les commentaires: [Bel, PI] et BetPm :

Metadata : Olivier C	[Bel, PI]: [0.00861048275872515 , 1.0],	BetPm:0.5043052413793626
Metadata : Michel	[Bel, PI]: [0.016057935796106544 , 1.0],	BetPm:0.5080289678980533
Metadata : Pascal	[Bel, PI]: [0.008610482758725146 , 1.0],	BetPm:0.5043052413793626
Metadata : Irene	[Bel, PI]: [0.008688111479275892 , 1.0],	BetPm:0.5043440557396379
Metadata : Muriel	[Bel, PI]: [0.12464652377789305 , 1.0],	BetPm:0.5623232618889465
Metadata : Sonia	[Bel, PI]: [0.1567221209603339 , 1.0],	BetPm:0.578361060480167
Metadata : Frederique	[Bel, PI]: [0.12464652377789305 , 1.0],	BetPm:0.5623232618889465
Metadata : Pascal F	[Bel, PI]: [0.120201551802062 , 1.0],	BetPm:0.560100775901031
Metadata : Pascal F	[Bel, PI]: [0.009910469669859553 , 1.0],	BetPm:0.5049552348349298

-- Note: La métadonnée la plus Pertinente est Muriel au niveau de sa crédibilité et sa probabilité pianistique .

***** **Fin** *****

Traces de calculs pour Exemple 2 du chapitre 6

SCORES DES OUTILS D'INDEXATION

**** Résultats Reconnaissance de Prénom textuel RPT****

[
score:0.38370574 : Pascal B
score:0.38370574 : Pascal F

*-- Note : RPT a détecté le mot " Pascal" avec un score de détection de 0.383. Donc la personne " Pascal B" aura le score 0.383 et, "Pascal F" aura aussi le même score 0.383.
On a donc l'ambiguïté entre Pascal B et Pascal F.*

----- DV:
Score Z1 = 203.4600067138672

----- H/F :
: homme: 273277.94106323953
femme: 315641.00104047067

----- RV :
RV :: Olivier T : 1798729.8799241066
RV :: Olivier C : 779543.7940266493
RV :: Michel : 1280191.7038551534
RV :: Pascal : 1448802.2024423357 *-- Note : la personne identifiée Pascal est en fait Pascal B*
RV :: Irene : 2031266.441839766
RV :: Muriel : 1214059.9830930089
RV :: Sonia : 1907242.903040803
RV :: Frederique : 9999999.0
RV :: Pascal F : 846101.1804348761

-- Note La personne la mieux reconnue (distance la plus faible) est Olivier C, dans la liste des personne en ordre décroissant des distance la bonne personne (Pascal F) est au 2 ème rang par RV et Pascal B est au 5 ème rang .

--- ALLOCATION DES JEUX DE MASSE

----- le jeu de masses de la source DV pour la Z1 -----

m[DV,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.95
m[DV,Z1](,*)= 0.0
m[DV,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.050000000000000044

----- le jeu de masses issu de la source RV pour la Z1 -----

m[RV,Z1](,Olivier C)= 0.09472224360364681
m[RV,Z1](,Pascal F)= 0.07258217090979173
m[RV,Z1](,Michel ,Pascal ,Muriel)= 0.011153115186428612
m[RV,Z1](,Olivier T ,Irene ,Sonia)= 0.0010196625445862868
m[RV,Z1](,Inconnu ,Inconnue ,*)= 0.25325686191203567
m[RV,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.5672659458435109

----- jeux de masses issus de la source HF

----- le jeu de masses issu de la source HF/Olivier T pour la Z1-----

m[HF/Olivier T,Z1](,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0

m[HF/Olivier T,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.0

...
idem pour les sources relatives aux prénoms 100% masculin
...

----- le jeu de masses issu de la source HF/Irene pour la Z1-----

m[HF/Irene,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.07320786220788722
m[HF/Irene,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.9267921377921128

...
idem pour les sources relatives aux prénoms 100% féminin
...

----- le jeu de masses issu de la source HF/Inconnu pour la Z1-----

m[HF/Inconnu,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnue ,*)= 0.0
m[HF/Inconnu,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.0

----- le jeu de masses issu de la source HF/Inconnue pour la Z1-----

m[HF/Inconnue,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,*)= 0.07320786220788722
m[HF/Inconnue,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.9267921377921128

----- le jeu de masse issu de la fusion des sources DV, RV et HF pour la zone Z1

m[DV+RV+HF,Z1](,Olivier C)= 0.09472224360364681
m[DV+RV+HF,Z1](,Pascal F)= 0.07258217090979173
m[DV+RV+HF,Z1](,Michel ,Pascal)= 8.164957197567602E-4
m[DV+RV+HF,Z1](,Michel ,Pascal ,Muriel)= 0.010336619466671856
m[DV+RV+HF,Z1](,Olivier T)= 5.464770355292776E-6
m[DV+RV+HF,Z1](,Olivier T ,Sonia)= 6.918254470732376E-5
m[DV+RV+HF,Z1](,Olivier T ,Irene)= 6.918254470732376E-5
m[DV+RV+HF,Z1](,Olivier T ,Irene ,Sonia)= 8.758326848163465E-4
m[DV+RV+HF,Z1](,Inconnu)= 0.017613373777555317
m[DV+RV+HF,Z1](,Inconnu ,Inconnue)= 0.22298064503887857
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Pascal F ,Inconnu)= 1.1331789757278176E-6
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Pascal F ,Inconnu ,Inconnue)= 1.4345745576254688E-5
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Frederique ,Pascal F ,Inconnu)= 1.434576105517924E-5
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 1.816135337872491E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Pascal F ,Inconnu)= 1.4345745576254687E-5
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Pascal F ,Inconnu ,Inconnue)= 1.816133378281657E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Frederique ,Pascal F ,Inconnu)= 1.8161353378724909E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.0022991792159248515
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Pascal F ,Inconnu)= 1.4345745576254688E-5
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Pascal F ,Inconnu ,Inconnue)= 1.8161333782816573E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Frederique ,Pascal F ,Inconnu)= 1.8161353378724914E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.002299179215924852
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Pascal F ,Inconnu)= 1.8161333782816573E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Pascal F ,Inconnu ,Inconnue)= 0.002299176735134779
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu)= 0.002299179215924852
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.029106999664096377
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Pascal F ,Inconnu)= 1.4345745576254692E-5
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Pascal F ,Inconnu ,Inconnue)= 1.8161333782816576E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Frederique ,Pascal F ,Inconnu)= 1.8161353378724917E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.0022991792159248524
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Pascal F ,Inconnu)= 1.8161333782816573E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Pascal F ,Inconnu ,Inconnue)= 0.002299176735134779
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Frederique ,Pascal F ,Inconnu)= 0.002299179215924852
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.029106999664096377
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Pascal F ,Inconnu)= 1.8161333782816573E-4
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Pascal F ,Inconnu ,Inconnue)= 0.002299176735134779
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Frederique ,Pascal F ,Inconnu)= 0.002299179215924852
m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.029106999664096377

m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Pascal F ,Inconnu)= 0.002299176735134779
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Pascal F ,Inconnu ,Inconnue)= 0.029106968257951386
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu)= 0.02910699966409638
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue)= 0.36848690331645634
 m[DV+RV+HF,Z1](,Inconnu ,*)= 9.270196725029123E-4
 m[DV+RV+HF,Z1](,Inconnu ,Inconnue ,*)= 0.011735823423098883
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Pascal F ,Inconnu ,*)= 5.964099872251679E-8
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Pascal F ,Inconnu ,Inconnue ,*)= 7.550392408555108E-7
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Frederique ,Pascal F ,Inconnu ,*)= 7.550400555357504E-7
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 9.558607041434175E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Pascal F ,Inconnu ,*)= 7.550392408555106E-7
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Pascal F ,Inconnu ,Inconnue ,*)= 9.558596727798204E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Frederique ,Pascal F ,Inconnu ,*)= 9.558607041434175E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.2100943241709758E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Pascal F ,Inconnu ,*)= 7.550392408555106E-7
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Pascal F ,Inconnu ,Inconnue ,*)= 9.558596727798204E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Frederique ,Pascal F ,Inconnu ,*)= 9.558607041434175E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.2100943241709758E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Pascal F ,Inconnu ,*)= 9.558596727798206E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Pascal F ,Inconnu ,Inconnue ,*)= 1.21009301849199E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,*)= 1.2100943241709758E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0015319473507419158
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Pascal F ,Inconnu ,*)= 7.550392408555106E-7
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Pascal F ,Inconnu ,Inconnue ,*)= 9.558596727798204E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Frederique ,Pascal F ,Inconnu ,*)= 9.558607041434175E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 1.2100943241709758E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Pascal F ,Inconnu ,*)= 9.558596727798206E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Pascal F ,Inconnu ,Inconnue ,*)= 1.21009301849199E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Frederique ,Pascal F ,Inconnu ,*)= 1.2100943241709758E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0015319473507419158
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Pascal F ,Inconnu ,*)= 9.558596727798206E-6
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Pascal F ,Inconnu ,Inconnue ,*)= 1.21009301849199E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Frederique ,Pascal F ,Inconnu ,*)= 1.2100943241709759E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.001531947350741916
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Pascal F ,Inconnu ,*)= 1.2100930184919902E-4
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0015319456977869164
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,*)= 0.0015319473507419163
 m[DV+RV+HF,Z1](,Olivier T ,Olivier C ,Michel ,Pascal ,Irene ,Muriel ,Sonia ,Frederique ,Pascal F ,Inconnu ,Inconnue ,*)= 0.0193940475429714

----- Probabilité pignistique pour la Z1 issue de la fusion Image -----

BetPm(Olivier T) = 0.05365561334538429

BetPm(Olivier C) = 0.14801126540569648

BetPm(Michel) = 0.05714280948415189

BetPm(Pascal) = 0.05714280948415189

BetPm(Irene) = 0.049343973529470765

BetPm(Muriel) = 0.05246297785106894

BetPm(Sonia) = 0.049343973529470765

BetPm(Frederique) = 0.04901744226506414

BetPm(Pascal F) = 0.1258711927118412

BetPm(Inconnu) = 0.18676816907632868

BetPm(Inconnue) = 0.1644197016893173

BetPm(*) = 0.006820071628054129

-- Note : La fusion entre DV, RV, et H/F garde toujours la bonne personne Pascal F au 2 ème rang et Pascal F au 5 ème rang.

----- **FUSION Image et commentaire** -----

---- **allocation de masse sur PSI = {Pertinent, Non-Pertinent}** -----

----- **le jeu de masses issu de l'Image (dans la référence 'psi')** -----

m[I,Olivier T](Per) = 0.05365561334538429, m[I,Olivier T](psi) = 0.9463443866546157
 m[I,Olivier C](Per) = 0.14801126540569648, m[I,Olivier C](psi) = 0.8519887345943036
 m[I,Michel](Per) = 0.05714280948415189, m[I,Michel](psi) = 0.9428571905158482
 m[I,Pascal](Per) = 0.05714280948415189, m[I,Pascal](psi) = 0.9428571905158482
 m[I,Irene](Per) = 0.049343973529470765, m[I,Irene](psi) = 0.9506560264705293
 m[I,Muriel](Per) = 0.05246297785106894, m[I,Muriel](psi) = 0.9475370221489311
 m[I,Sonia](Per) = 0.049343973529470765, m[I,Sonia](psi) = 0.9506560264705293
 m[I,Frederique](Per) = 0.04901744226506414, m[I,Frederique](psi) = 0.9509825577349359
 m[I,Pascal F](Per) = 0.1258711927118412, m[I,Pascal F](psi) = 0.8741288072881588

le jeu de masses issu de la fusion des commentaires vocaux et textuels (dans la référence 'psi')

m[Com,Olivier T](Per) = 0.0, m[Com,Olivier T](psi) = 1.0
 m[Com,Olivier C](Per) = 0.0, m[Com,Olivier C](psi) = 1.0
 m[Com,Michel](Per) = 0.0, m[Com,Michel](psi) = 1.0
 m[Com,Pascal](Per) = 0.383705735206604, m[Com,Pascal](psi) = 0.616294264793396
 m[Com,Irene](Per) = 0.0, m[Com,Irene](psi) = 1.0
 m[Com,Muriel](Per) = 0.0, m[Com,Muriel](psi) = 1.0
 m[Com,Sonia](Per) = 0.0, m[Com,Sonia](psi) = 1.0
 m[Com,Frederique](Per) = 0.0, m[Com,Frederique](psi) = 1.0
 m[Com,Pascal F](Per) = 0.383705735206604, m[Com,Pascal F](psi) = 0.616294264793396

le jeu de masses issu de la combinaison entre l'Image et les commentaires (dans la référence 'psi')

m[I+Com,Olivier T](,Per)= 0.05365561334538429
 m[I+Com,Olivier T](,Per ,non-Per)= 0.9463443866546157

m[I+Com,Olivier C](,Per)= 0.14801126540569648
 m[I+Com,Olivier C](,Per ,non-Per)= 0.8519887345943036

m[I+Com,Michel](,Per)= 0.05714280948415189
 m[I+Com,Michel](,Per ,non-Per)= 0.9428571905158482

m[I+Com,Pascal](,Per)= 0.4189225209658685
 m[I+Com,Pascal](,Per ,non-Per)= 0.5810774790341315

m[I+Com,Irene](,Per)= 0.049343973529470765
 m[I+Com,Irene](,Per ,non-Per)= 0.9506560264705293

m[I+Com,Muriel](,Per)= 0.05246297785106894
 m[I+Com,Muriel](,Per ,non-Per)= 0.9475370221489311

m[I+Com,Sonia](,Per)= 0.049343973529470765
 m[I+Com,Sonia](,Per ,non-Per)= 0.9506560264705293

m[I+Com,Frederique](,Per)= 0.045592590817677366
 m[I+Com,Frederique](,Per ,non-Per)= 0.8845373529507728
 m[I+Com,Frederique](,non-Per)= 0.06987005623154989

m[I+Com,Pascal F](,Per)= 0.461279429377616

$m[I+Com,Pascal F](,Per,non-Per)=0.538720570622384$

la Pertinence des meta-données après l'Image et les commentaires: [Bel, PI] et BetPm :

Metadata : Olivier T	[Bel, PI]: [0.05365561334538429 , 1.0],	BetPm:0.5268278066726921
Metadata : Olivier C	[Bel, PI]: [0.14801126540569648 , 1.0],	BetPm:0.5740056327028482
Metadata : Michel	[Bel, PI]: [0.05714280948415189 , 1.0],	BetPm:0.528571404742076
Metadata : Pascal B	[Bel, PI]: [0.4189225209658685 , 1.0],	BetPm:0.7094612604829342
Metadata : Irene	[Bel, PI]: [0.049343973529470765 , 1.0],	BetPm:0.5246719867647354
Metadata : Muriel	[Bel, PI]: [0.05246297785106894 , 1.0],	BetPm:0.5262314889255345
Metadata : Sonia	[Bel, PI]: [0.049343973529470765 , 1.0],	BetPm:0.5246719867647354
Metadata : Frederique	[Bel, PI]: [0.045592590817677366 , 0.9301299437684503],	BetPm:0.4878612672930638
Metadata : Pascal F	[Bel, PI]: [0.461279429377616 , 1.0],	BetPm:0.730639714688808

*-- Note : Dans l'ordre de probabilité pignistique décroissante, la personne " Pascal F " est au 1ere rang,
La fusion du commentaire et de l'Image a permis d'enlever l'ambiguïté sur le commentaire, malgré la mauvaise reconnaissance de RV.*

***** **Fin** *****

