



Modélisation de la pertinence en recherche d'information : modèle conceptuel, formalisation et application

Nathalie Denos

► To cite this version:

Nathalie Denos. Modélisation de la pertinence en recherche d'information : modèle conceptuel, formalisation et application. Interface homme-machine [cs.HC]. Université Joseph-Fourier - Grenoble I, 1997. Français. NNT : . tel-00004938

HAL Id: tel-00004938

<https://theses.hal.science/tel-00004938>

Submitted on 20 Feb 2004

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.

THÈSE

présentée par

Nathalie Denos

pour obtenir le grade de DOCTEUR
de l'UNIVERSITÉ JOSEPH FOURIER - GRENOBLE I
(arrêtés ministériels du 5 juillet 1984 et du 30 Mars 1992)
Spécialité : **Informatique**

**Modélisation de la pertinence en recherche d'information :
modèle conceptuel, formalisation et application**

Date de soutenance : 28 octobre 1997

Composition du jury :

Rapporteurs : Claude Chrisment
Jean-Marie Pierrel
Examineurs : François Rechenmann
Edmond Lassalle
Yves Chiaramella
Catherine Berrut

Thèse préparée au sein du laboratoire
Communication Langagière et Interaction Personne-Système – IMAG
Université Joseph Fourier - Grenoble I

Résumé

Les systèmes de recherche d'information ont pour fonction de permettre à l'utilisateur d'accéder à des documents qui contribuent à résoudre le problème d'information qui motive sa recherche. Ainsi le système peut être vu comme un instrument de prédiction de la pertinence des documents du corpus pour l'utilisateur. Les indices traditionnellement utilisés par le système pour estimer cette pertinence sont de nature thématique, et sont fournis par l'utilisateur sous la forme d'un ensemble de mots-clés : la requête. Ainsi le système implémente une fonction de correspondance entre documents et requête qui modélise la dimension thématique de la pertinence.

Les performances des systèmes de recherche d'information sont traditionnellement évaluées conformément à ce principe de modélisation, en s'appuyant sur une collection test. Pour chaque requête de la collection de tests, un sous-ensemble du corpus est désigné comme l'ensemble des documents pertinents pour cette requête, et cela dans un sens normalisé relatif à la relation thématique que les documents entretiennent avec la requête. Ainsi, un système est d'autant meilleur que la correspondance thématique qu'il établit entre une requête et le corpus est proche de l'ensemble des documents associés à cette requête dans la collection test.

Cependant l'éventail des utilisations et des utilisateurs des systèmes va s'élargissant, de même que la nature des documents présents dans les corpus, qui ne sont plus seulement des documents textuels. Nous tirons deux conséquences de cette évolution. D'une part, l'hypothèse que le facteur thématique de pertinence est prépondérant (et donc seul sujet à modélisation dans les systèmes), ne tient plus. Les autres facteurs, nombreux, de la pertinence interviennent d'une manière telle qu'ils compromettent les performances des systèmes dans le contexte d'une utilisation réelle. Ces autres facteurs dépendent fortement de l'individu et de sa situation de recherche d'information, ce qui remet en cause la conception de la pertinence système comme une fonction de correspondance qui ne prend en compte que les facteurs de la pertinence qui ne dépendent pas de l'utilisateur. D'autre part, la nature de l'utilisation interactive du système contribue à définir la situation de recherche de l'utilisateur, et en cela participe aux performances du système de recherche d'information. Un certain nombre de caractéristiques de l'interaction sont directement liées à la modélisation de la pertinence système et à des préoccupations spécifiques à la problématique de la recherche d'information, comme le nombre de documents retrouvés dont l'utilisateur doit évaluer la pertinence, la précision du résultat, l'exhaustivité du résultat, etc.

Notre thèse s'appuie sur les travaux réalisés sur les facteurs de la pertinence pour un individu, pour définir un modèle de conception de la pertinence système qui prend en compte les facteurs qui relèvent de l'utilisation interactive du système et de la nécessité d'adaptation de la fonction de correspondance à la situation de recherche particulière dans laquelle l'utilisateur se trouve.

Ainsi, nous définissons trois nouvelles fonctions du système de recherche d'information, en termes d'utilisation du système : permettre la détection de la pertinence des documents retrouvés, permettre la compréhension des raisons de leur pertinence système, et permettre de procéder à une reformulation du problème d'information dans le cadre d'un processus itératif de recherche.

La notion de schéma de pertinence se substitue à celle de requête, en tant qu'interface entre la pertinence système et l'utilisateur. Ce schéma de pertinence intègre deux types de paramètres permettant l'adaptation du système à la situation de recherche : d'une part les paramètres sémantiques, qui recouvrent non seulement la dimension thématique de la pertinence mais aussi d'autres critères de pertinence liés aux caractéristiques indexées des documents, et d'autre part les paramètres pragmatiques qui prennent en compte les facteurs de la pertinence liés aux conditions dans lesquelles l'utilisateur réalise les tâches qui lui incombent dans l'interaction.

Nous appliquons ce modèle de conception de la pertinence système dans le cadre d'une application de recherche d'images, dont le corpus est indexé de façon à couvrir plusieurs dimensions de la pertinence outre la dimension thématique. Notre prototype nous permet de montrer comment le système s'adapte en fonction des situations qui se présentent au cours d'une session de recherche.

Sommaire

Partie I	Contexte thématique de notre recherche	11
Chapitre 1	Pertinence et SRI	15
1.1	Pertinence et recherche d'information	16
1.2	Fonction d'un système de recherche d'information	20
1.3	Pertinence système et pertinence utilisateur	21
1.4	Bouclage de pertinence	23
1.5	Performances qualitatives d'un SRI	24
1.6	Objectifs	25
Chapitre 2	Problématique	27
2.1	Interaction utilisateur système	27
2.2	Réduire la distance	30
2.2.1	Schéma de pertinence et classes de pertinence	32
2.2.2	Adaptation du système	34
2.2.3	Démarche	35
2.3	Plan de ce mémoire	35
Partie II	Pertinence utilisateur et pertinence système	37
Chapitre 1	Facteurs de pertinence et conception des SRI	41
1.1	Motivation des deux études	42
1.1.1	Approche classique de modélisation de la pertinence dans les systèmes de recherche d'information	42
1.1.2	Paradoxe de l'évaluation et évolution des besoins	43
1.1.3	Facteurs de pertinence	44
1.1.4	Utilisation des critères de pertinence par les individus	44
1.1.5	Deux axes pour améliorer les performances des systèmes	46
1.2	Étendre la fonction de correspondance	47
1.2.1	Critères de pertinence	47

1.2.2	Éventail des critères et limites de leur prise en compte dans la fonction de correspondance	48
1.2.3	Principe de l'intégration des critères dans la fonction de correspondance	53
1.3	Concevoir l'interaction	53
1.3.1	Principes d'interaction	53
Chapitre 2	Adaptation du système	57
2.1	Jugements de pertinence et bouclages de pertinence	58
2.2	Typologies et stratégies de recherche	58
2.2.1	Déduire de la requête la meilleure stratégie	58
2.2.2	Associer stratégies et typologie	59
2.2.3	Modéliser l'utilisateur	59
2.2.4	Conclusion	60
2.3	Paramétrisation logique de la correspondance	60
2.3.1	Approches axiomatiques de la pertinence	60
2.3.2	Approche abductive de la pertinence	61
Chapitre 3	Résumé et objectifs	65
Partie III	Modèle conceptuel	67
Chapitre 1	Critères élémentaires abstraits	71
1.1	Critère élémentaire, critère élémentaire abstrait	71
1.1.1	Critère élémentaire	72
1.1.2	Critère élémentaire abstrait	72
1.1.3	Modèle de document	73
1.1.4	Évaluation d'un critère élémentaire	74
1.1.5	Ordre sur les critères élémentaires abstraits	75
1.1.6	Choix des critères élémentaires abstraits pour une application donnée	75
1.2	Critère de dimension, critère de dimension abstrait	77
1.2.1	Présentation	77
1.2.2	Définitions	77
Chapitre 2	Schéma de pertinence	79
2.1	Combinaison de critères pour l'interrogation	79
2.1.1	Discussion	80
2.1.2	Notre choix	81
2.2	Statuts de critères	82

2.3	Schéma de pertinence	82
2.3.1	Structure hiérarchique du schéma de pertinence	82
2.3.2	Rôle des éléments du schéma de pertinence	86
2.4	Évaluation de la pertinence système	86
2.4.1	Relation de pertinence	86
2.4.2	Ensemble des documents retrouvés	88
2.5	Classes de pertinence et visualisation	88
2.6	Résumé	89
Chapitre 3 Bilan et application des principes		91
3.1	Mise en forme des résultats : détecter	91
3.2	Paramétrisation et visualisation : comprendre	92
3.3	Bouclage de pertinence généralisé : reformuler	92
3.4	Vers une formalisation	93
Partie IV Formalisation		95
Chapitre 1 Formalisation de la pertinence système		99
1.1	Portée de la formalisation	99
1.1.1	Un ensemble prédéfini de critères abstraits et de dimensions	99
1.1.2	Des statuts à valeur binaire	100
1.2	Structure générale et notations	100
1.3	Langage des schémas de pertinence	101
1.4	Fonction d'évaluation d'un critère élémentaire	102
1.5	Évaluation binaire	103
1.6	Évaluation symbolique	105
1.6.1	Sémantique symbolique du schéma de pertinence	105
1.6.2	Langage des relations de pertinence	106
1.6.3	Conclusion	107
Chapitre 2 Classes de pertinence et schémas homologues		109
2.1	Classes de pertinence	109
2.1.1	Notations	109
2.1.2	Ensemble des classes de pertinence	110
2.2	Classe de pertinence de dimension et schéma homologue	110
2.3	Sous-classe de pertinence et schéma homologue	113
2.4	Degré de pertinence	115
2.5	Classes de pertinence et degrés	117

Chapitre 3	Visualisation selon les classes de pertinence	121
3.1	Classes visuelles de pertinence	121
3.1.1	Définition des classes visuelles de pertinence	121
3.1.2	Propriétés des classes visuelles	124
3.1.3	Conclusion	131
3.2	Classes visuelles et interaction	132
3.2.1	Structure des classes visuelles	132
3.2.2	Modes de visualisation et manipulation	133
Chapitre 4	Utilisation interactive du système	137
4.1	Objectif de l'interaction	138
4.1.1	État de satisfaction optimale et évaluation d'une session	141
4.1.2	Situation de non-satisfaction et situation d'entrave	142
4.1.3	Niveau de notre formalisation du déroulement d'une session de recherche	143
4.2	Situations stéréotypiques et actions	144
4.2.1	Définitions et notations	144
4.2.2	Situations problématiques et actions	146
4.2.3	Bilan	149
4.3	Possibilités d'automatisation de la reformulation	149
4.3.1	Opérations globales	150
4.3.2	Opérations paramétrées	152
4.4	Résumé	153
Chapitre 5	Bilan	155
5.1	Niveau de granularité	155
5.2	Portée de la modification	155
5.3	Degré et conditions d'intervention de l'utilisateur	156
Partie V	Application	157
Chapitre 1	Exemple d'application	161
1.1	Le prototype VIZ	161
1.1.1	Modèle opérationnel	161
1.1.2	Description technique et interface	169
1.2	Propriétés des dimensions	172
1.2.1	Particularités des dimensions	172
1.2.2	Statuts de dimensions et types de recherche	178
1.2.3	Contrôle pour la formulation	179
1.3	Exemples de situations et réactions	181

1.3.1	Pénurie	182
1.3.2	Non-pertinence	191
1.3.3	Non-discrimination	195
1.3.4	Non-focalisation	202
1.4	Exemple de recherche	204
1.4.1	Problème d'information	204
1.4.2	Étape 1	204
1.4.3	Étape 2	205
1.4.4	Étape 3	209
1.4.5	Étape 4	210
Chapitre 2	Perspectives de contrôle de l'interaction	213
2.1	Mode manuel avec formulation contrôlée	213
2.1.1	Correction du schéma de pertinence	215
2.1.2	Types de recherche	215
2.1.3	Heuristiques d'utilisation	216
2.2	Mode automatique avec reformulation automatique	217
2.2.1	Trois types de reformulation	217
2.2.2	Principe de la réinjection	218
2.2.3	Trois modes de reformulation par réinjection	219
2.2.4	Reformulation secondaire	219
2.2.5	Reformulation de transition	220
2.2.6	Boucle de validation de la reformulation	220
Chapitre 3	Champ d'application	223
3.1	Classification des documents retrouvés	223
3.2	Clarification du sens des termes employés dans la requête	226
3.3	Ensemble de critères abstraits	226
3.4	Dimensions d'interrogation	227
3.5	Bilan et perspectives d'application	228
3.5.1	Bilan	228
3.5.2	Perspectives d'application sur le corpus du World Wide Web	229
3.5.3	Perspectives d'application à des corpus de documents non textuels	229
Chapitre 4	Bilan	231
4.1	Bilan	231
4.2	Intégration des techniques existantes	232
4.3	Extensions souhaitables du modèle	233

Partie VI	Conclusion	235
Chapitre 1	Bilan	237
1.1	Contexte	237
1.2	Apports	238
1.3	Problèmes soulevés	239
Chapitre 2	Perspectives	241
2.1	Systèmes de recherche d'information interactifs et adaptatifs	241
2.2	Lien avec les modèles d'indexation avancés	242
Partie VII	Annexe	253
Annexe A	Agrandissements de résultats	i
A.1	Agrandissements – 1	ii
A.2	Agrandissement – 2	xvi

Partie I

Contexte thématique de notre recherche

Systematic and measurable the concept of user-centered relevance may be but this does not imply that the discipline can now “get on with other matters” as Schamber et al. ([SEN90], p.756) may claim. A view of relevance bound by the situationality of the user and determined by individual context is exceedingly problematic for the discipline. How will information science move ahead if the key concept which drives research efforts and recommendations for information system design is so fundamentally individualized and so dynamic?

[Bru94], p. 142.

Les systèmes de recherche d'information permettent à un individu de résoudre un problème d'information au moyen des documents présents dans un corpus. Le concept central que ces systèmes doivent modéliser est appelé «pertinence». Le concept de pertinence a été beaucoup étudié, dans le cadre de la communication entre les individus en général, et dans le cadre de la recherche d'information en particulier.

De nombreux travaux théoriques et expérimentaux ont pour but de définir et de cerner les caractéristiques de la pertinence en recherche d'information, dont la définition la plus consensuelle est aujourd'hui encore opérationnelle : la pertinence en recherche d'information est définie par les jugements de pertinence que portent des individus sur des documents. La complexité de ce concept fait que la finesse et la couverture des études le concernant ne se répercutent que très partiellement dans la conception des systèmes de recherche d'information.

Avec l'importance grandissante des sources d'information électroniques, et la nécessité d'y fournir un accès au plus grand nombre, la recherche s'oriente vers de nouveaux paradigmes de recherche d'information dont l'objectif est de réduire les situations d'échec lors de l'utilisation des systèmes par des utilisateurs non spécialistes. Ainsi la pratique classique de modélisation de la pertinence doit être remise en question pour atteindre une plus grande robustesse en tenant compte de facteurs de la pertinence traditionnellement ignorés.

Cette introduction s'attache à présenter le concept de pertinence dans sa complexité et à montrer les aspects qui sont peu ou pas intégrés dans les systèmes de recherche d'information. La suite de ce mémoire explore les possibilités d'intégrer un plus grand nombre d'aspects de la pertinence dans le but d'améliorer les performances des systèmes.

Chapitre 1

Pertinence et SRI

Une information peut être à la fois disponible et inaccessible. Imaginons un étudiant en droit qui doit étudier un cas, et qui dispose de documents thésaurisant toute la jurisprudence du droit français. Si ces documents sont classés de manière chronologique exclusivement, l'étudiant n'a a priori aucun moyen d'accéder directement aux documents qui recèlent des informations utiles pour résoudre son cas.

Ne disposant que d'un accès chronologique aux documents, il va chercher à identifier, dans les caractéristiques du cas qu'il doit étudier, des éléments qui lui permettent de cerner la période pendant laquelle des actes de jurisprudence qui lui sont *pertinents* ont pu être émis. Par exemple si son étude concerne la nuisance portée par un individu à son voisin par l'intermédiaire de son antenne de télévision personnelle, l'étudiant peut éliminer la jurisprudence antérieure à l'apparition des premiers postes de télévision en France.

Ainsi, pour juger de la pertinence des documents dont il dispose pour résoudre son problème d'information, l'étudiant utilise un *critère* relatif à la date de décret des actes de jurisprudence. Ce critère lui est suggéré par la façon dont les documents sont indexés.

Ce n'est pas pour autant que tous les documents qui vérifient les critères exprimés en termes de date de décret sont pertinents. D'autres critères non exprimables dans ce contexte interviennent aussi.

Par exemple des critères thématiques vont très certainement intervenir : le document doit traiter de nuisance, de téléviseurs ou d'appareils du même type. Par ailleurs, en fonction des connaissances préalables de l'individu, certains documents ne seront pas utiles, car déjà connus. Aussi, si l'individu est pressé, il éliminera certains documents qui demanderaient trop de travail pour être utilisés. Il se peut aussi que l'individu se contente d'un petit nombre de documents pertinents, même s'il en existe de nombreux autres, car ce à quoi l'information doit servir ne requiert peut-être pas une connaissance exhaustive du sujet.

Ces remarques donnent une idée de la grande diversité des facteurs qui inter-

viennent lorsqu'un individu évalue la pertinence d'un document. Ils sont liés à la *situation de recherche d'information* particulière dans laquelle l'individu se trouve, qui inclut l'individu lui-même, le genre de problème d'information qu'il se pose, l'utilisation qu'il projette de faire des documents qu'il trouvera, les conditions concrètes de sa recherche, etc.

Les systèmes de recherche d'information permettent l'expression de requêtes indiquant le contenu sémantique des documents cherchés par l'utilisateur. Étant donné cette requête, le système applique à chaque document une fonction de correspondance qui détermine, du point de vue du système, son adéquation avec la requête.

Il existe une distance plus ou moins grande entre les résultats d'un système et les jugements de pertinence de l'utilisateur. Traditionnellement, l'amélioration des performances des systèmes passe par le raffinement de la fonction de correspondance. Cette thèse explore d'autres points d'attaque possibles pour améliorer les performances, qui font intervenir, conjointement avec la fonction de correspondance modélisant la pertinence pour le système, la nature itérative de l'interaction entre l'utilisateur et le système.

Nous introduisons d'abord la notion de pertinence d'un point de vue cognitif, afin d'en montrer la complexité, et de justifier les concepts relatifs à la pertinence pour un individu que nous introduisons. Puis nous décrivons l'approche classique de modélisation de la pertinence dans les systèmes de recherche d'information. La comparaison entre ces deux points de vue nous permet ensuite d'expliquer le décalage existant entre la pertinence pour un individu (pertinence utilisateur) et la pertinence telle qu'implémentée dans les systèmes de recherche d'information (pertinence système). Enfin, nous abordons le problème de l'évaluation des performances d'un système de recherche d'information, et mettons nos objectifs en perspective.

1.1 Pertinence et recherche d'information

Nous décrivons ici le concept de pertinence indépendamment de l'utilisation d'un système de recherche d'information, en donnant un point de vue cognitif sur les données et les processus qui permettent de décrire la production de jugements de pertinence par les individus.

Étant donné un individu, un document est dit pertinent s'il contribue à le satisfaire dans une situation de recherche d'information.

La figure I.1.1 constitue l'ébauche d'un modèle cognitif de la pertinence ; elle identifie l'*individu*, les *documents* et les *jugements de pertinence* émis par l'individu au vu des documents. Un individu, dans une *situation de recherche d'information* donnée, procède à l'*identification du problème d'information* qu'il se propose de résoudre.

Par exemple, dans une banque, un employé est chargé par son patron de trouver

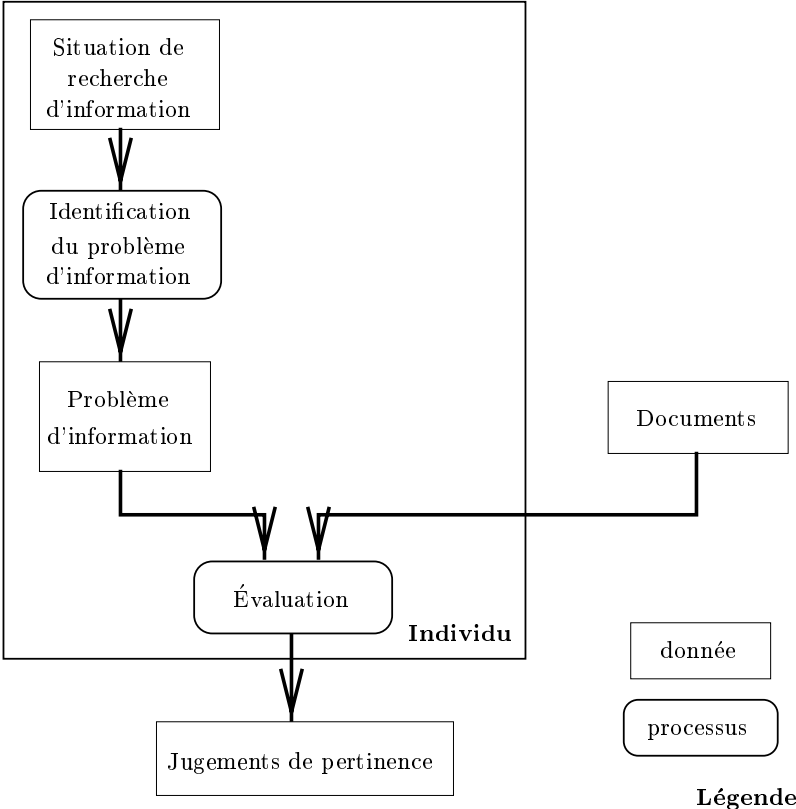


FIG. I.1.1 – *Individu et jugements de pertinence*

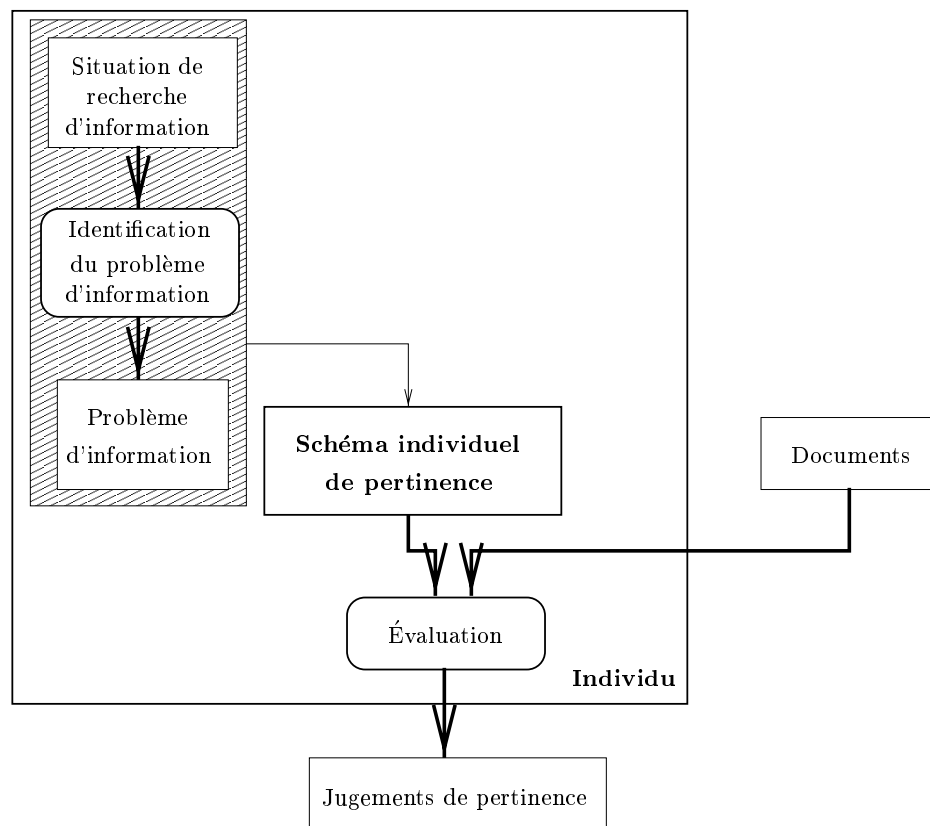
sur le marché un produit permettant de gérer les transactions bancaires de manière plus efficace que celui actuellement utilisé dans la banque, et cela dans un délai de 2 jours. Ces faits résument une partie de sa situation de recherche d'information.

Selon ses connaissances actuelles (qui font partie de la situation de recherche d'information), l'employé peut identifier comme premier problème d'information *que signifie plus efficace?*, d'où les sous-problèmes : *quels sont les critères d'évaluation de l'efficacité d'un système de gestion de transactions bancaires en général?*, *en quoi le système de gestion actuel peut-il être amélioré?*. Si l'employé a déjà résolu ces problèmes grâce à son expérience antérieure, il peut identifier le problème d'information *trouver une liste exhaustive des systèmes de gestion de transactions bancaires vendus aujourd'hui avec leurs caractéristiques techniques*.

Le processus qui consiste, pour cet individu, à évaluer la pertinence d'un document est appelé *évaluation*. L'évaluation consiste à sélectionner certains documents en fonction de leur capacité à résoudre le problème d'information identifié, ce qui produit, pour chaque document, un jugement de pertinence. L'opération de sélection donne une image simplifiée de la pertinence, qui recouvre deux autres aspects. D'une part, les jugements de pertinence peuvent être nuancés, deux documents pertinents ne l'étant pas toujours autant l'un que l'autre, ni pour les mêmes raisons. D'autre part, la pertinence d'un document est relative à l'ensemble des documents consultés, un document pertinent pouvant devenir non pertinent si un autre document plus pertinent que celui-ci est consulté.

Pour les théoriciens de la pertinence, les jugements de pertinence sont les seules traces tangibles de l'évaluation de la pertinence que réalise un individu dans une certaine situation de recherche d'information. Ainsi le concept de pertinence se définit avant tout de manière opérationnelle comme l'ensemble des jugements de pertinence produits par un individu.

Ainsi, à partir d'études expérimentales qui étudient les régularités qui ressortent du comportement des individus en situation de recherche d'information, le concept de *schéma individuel d'évaluation de la pertinence* a été défini. Il s'agit de la structure de la construction mentale que les individus élaborent pour procéder à l'évaluation de la pertinence d'un document, dans une situation donnée. En pratique, ce concept recouvre l'ensemble des raisons qui sous-tendent les jugements de pertinence, et il peut jouer le rôle de réceptacle pour l'ensemble des données et processus cognitifs des individus. Ainsi, comme l'illustre la figure I.1.2, nous laissons dans l'ombre les processus cognitifs qui président à l'évaluation de la pertinence en considérant le schéma individuel de pertinence comme l'unique donnée de l'évaluation.

FIG. I.1.2 – *Schéma individuel d'estimation de la pertinence (suite)*

1.2 Fonction d'un système de recherche d'information

Un système classique de recherche d'information est un système informatique qui fournit un accès à un corpus par l'intermédiaire de critères thématiques que l'utilisateur formule au moyen d'une *requête*. Le résultat fourni par le système est une sélection des documents du corpus qui correspondent à la requête : c'est l'ensemble des *documents retrouvés*.

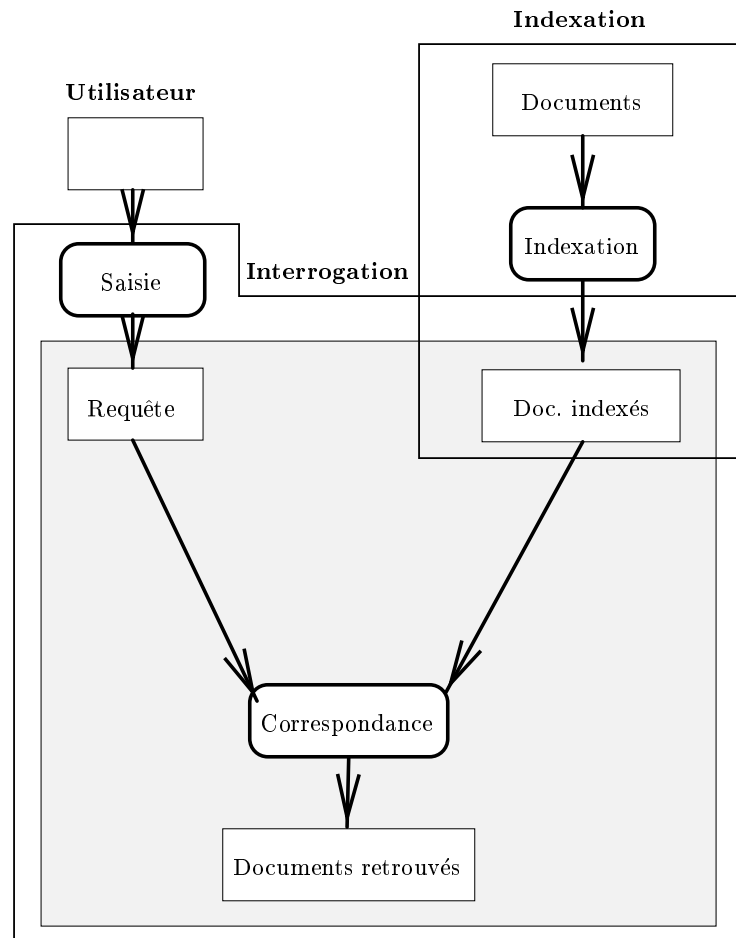


FIG. I.1.3 – Schéma fonctionnel classique des systèmes de recherche d'information

Le schéma fonctionnel classique pour les systèmes de recherche d'information (figure I.1.3) comprend deux fonctions principales : l'indexation et l'interrogation.

La fonction d'indexation Les documents subissent une analyse, manuelle ou automatique, pour construire un ensemble de *documents indexés* sur lequel la recherche est effectuée.

La fonction d'interrogation Cette fonction comprend deux sous-fonctions la saisie de la requête et la fonction de correspondance.

La *saisie* de la requête se définit en termes de langage de requête et d'interface. Le système offre à l'utilisateur un langage de requête plus ou moins souple, qui va du langage des expressions booléennes sur l'ensemble des termes d'indexation, par exemple, à la langue naturelle. Dans le second cas, la fonction de saisie intègre une fonction de transformation de la requête pour la rendre conforme à ce que la fonction de correspondance peut traiter.

La fonction de *correspondance* est la fonction de recherche proprement dite. Le système met en correspondance les documents indexés avec la requête de façon à sélectionner un sous-ensemble des documents du corpus, appelé ensemble des *documents retrouvés*. La fonction de correspondance est souvent assortie d'une fonction de pondération des documents retrouvés selon le degré de la correspondance établie avec la requête.

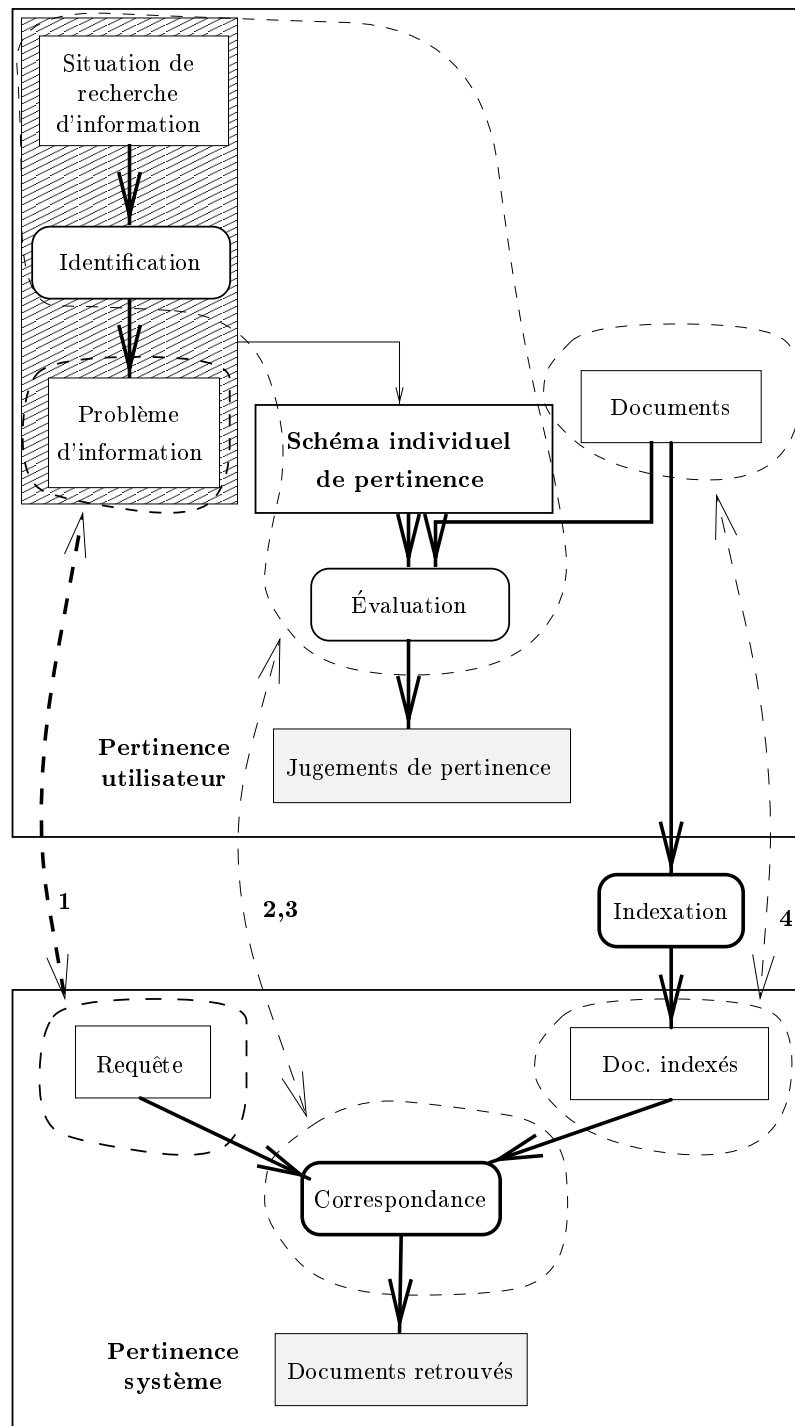
Nous appelons *pertinence système* l'ensemble des principes qui sous-tendent la fonction de correspondance dans un système de recherche d'information, par opposition à la *pertinence utilisateur*, qui correspond à l'ensemble des jugements de pertinence que produit l'individu qui utilise le système.

1.3 Pertinence système et pertinence utilisateur

Si l'on rapproche le schéma fonctionnel d'un système de recherche d'information (figure I.1.3) du schéma associé à la pertinence pour un individu (figure I.1.1), on peut repérer la dualité des fonctions de la pertinence système avec certains éléments relevant de la pertinence utilisateur. Les documents qui constituent l'objet de la recherche pour l'utilisateur, sont remplacés par les documents indexés. Le processus complexe d'évaluation de la pertinence utilisateur, qui met en jeu la situation de recherche d'information, joue un rôle symétrique à celui de la correspondance. Enfin, le problème d'information est exprimé sous la forme d'une requête dans le système.

La distance qui sépare ces entités duales relève de plusieurs facteurs, signalés par des flèches numérotées dans la figure I.1.4 :

- 1° La requête est généralement une représentation formelle du problème d'information conforme au langage de requête du système; elle ne reflète que partiellement le problème d'information original.
- 2° La correspondance est effectuée selon l'hypothèse d'un monde fermé correspondant au corpus, alors que la requête ne peut pas être formulée dans cette

FIG. I.1.4 – *Distance entre pertinence système et pertinence utilisateur*

perspective étant donné que l'utilisateur ne connaît pas le corpus.

- 3° Le calcul de la correspondance, aussi raffiné soit-il, ne constitue qu'une modélisation hypothétique des principes qui sous-tendent la pertinence utilisateur ; elle ne rend pas compte de la complexité de la pertinence utilisateur.
- 4° Les documents indexés sont produits conformément à un langage d'indexation (que nous considérons comme une donnée dans notre étude) qui donne un point de vue nécessairement incomplet sur les informations que les documents véhiculent.

Par conséquent il est rare qu'un utilisateur puisse résoudre son problème d'information à partir d'une seule requête, et l'utilisation d'un système de recherche d'information est plus généralement conçue comme un processus itératif visant à améliorer progressivement l'adéquation entre pertinence système et pertinence utilisateur. Pour ce faire, une troisième fonction est très fréquemment ajoutée au schéma fonctionnel classique : le bouclage de pertinence (*relevance feedback*).

1.4 Bouclage de pertinence

Une fois un premier ensemble de documents retrouvés, l'utilisateur peut émettre des *jugements de pertinence* sur ces documents, jugements qui sont pris en compte pour définir une nouvelle requête : nous appelons la définition de la nouvelle requête *reformulation*. Cette fonction n'est pas toujours automatisée ; une stratégie classique d'utilisation des systèmes de recherche d'information consiste à procéder manuellement à la reformulation de la requête en tenant compte des documents pertinents et non pertinents obtenus à un instant donné.

La technique de reformulation automatique dénommée *relevance feedback* ou bouclage de pertinence a été d'abord introduite par Rocchio dans le système SMART [RJ71]. Elle consiste à fonder l'interaction sur le processus itératif suivant :

- 1° parmi l'ensemble des documents retrouvés, l'utilisateur sélectionne le sous-ensemble des documents qu'il juge pertinents ;
- 2° le système reformule la requête en conséquence et évalue la nouvelle requête obtenue ;
- 3° le système affiche les nouveaux documents retrouvés et passe la main à l'utilisateur (aller en 1).

La requête est modifiée en fonction des termes d'indexation présents dans les documents jugés pertinents.

De nombreuses techniques ont été élaborées sur ce principe de reformulation de la requête, ou d'un autre élément de la modélisation qui peut évoluer avec les informations de pertinence fournies par les utilisateurs. Le principe commun à toutes

ces techniques est que le système s'adapte progressivement à la situation de l'utilisateur. Plus précisément il y a adaptation mutuelle de l'utilisateur et du système, car l'utilisateur s'adapte lui aussi au système, notamment en découvrant les termes qui sont utilisés pour décrire le contenu des documents.

1.5 Performances qualitatives d'un SRI

Le but de l'évaluation des systèmes de recherche d'information est de mesurer la différence entre la pertinence système et la pertinence utilisateur.

La seule technique existante pour évaluer les performances d'un système de recherche d'information est la comparaison de mesures effectuées sur une *collection test*, selon le *paradigme de Cranfield* [SJ81]. Une collection test est un corpus assorti d'un ensemble de requêtes résolues à l'avance manuellement, qui servent de référence de qualité. Pour un système donné, les requêtes résolues sont évaluées et les résultats obtenus sont comparés à la référence ou aux résultats fournis par d'autres systèmes, au moyen de mesures classiques comme par exemple le *rappel* (nombre de documents pertinents retrouvés par rapport au nombre de documents pertinents) ou la *précision* (nombre de documents pertinents retrouvés par rapport au nombre de documents retrouvés). On dit qu'un système obtient de meilleures performances qu'un autre si la moyenne des mesures obtenues sur l'ensemble des requêtes de la collection test est meilleure.

Le paradigme de Cranfield assimile la pertinence utilisateur aux jugements de pertinence définis par une collection de tests, qui constitue une approximation grossière (car consensuelle) de la pertinence utilisateur. De plus, en ne mettant pas en scène l'utilisateur dans l'évaluation des performances des systèmes, il n'évalue pas la capacité d'un système à s'adapter à la situation de recherche d'information particulière d'un utilisateur réel.

Outre sa dépendance vis-à-vis d'un corpus particulier, ce type d'évaluation permet de compenser les mauvaises performances sur certaines requêtes par de bonnes performances sur certaines autres. Cette caractéristique ne convient pas forcément aux exigences du marché, comme le souligne Croft dans [Cro95].

En effet, les industriels, se faisant l'écho des désirs des utilisateurs potentiels de systèmes de recherche d'information, considèrent le bouclage de pertinence comme une caractéristique souhaitable des systèmes dans la stricte mesure où les résultats obtenus sont fiables quelle que soit la requête. Cette condition est assortie d'une seconde condition qui réclame l'intelligibilité du processus de bouclage de pertinence : un utilisateur ne doit pas être confronté à des documents dont il ne comprend pas la relation avec sa requête.

1.6 Objectifs

S'adapter à l'utilisateur de manière robuste et intelligible constitue donc un objectif majeur des systèmes de recherche d'information. Dans notre étude, nous envisageons cet objectif selon un ensemble de principes de conception des systèmes de recherche d'information qui permet de combler la distance qui existe entre pertinence utilisateur et pertinence système.

Nous proposons de considérer la définition de la pertinence système comme relevant non seulement d'une fonction de correspondance, mais surtout des autres éléments qui entrent en jeu dans l'utilisation effective des systèmes : l'utilisateur et son interaction avec le système.

Nos principes de conception montrent comment les connaissances actuelles sur la pertinence en général peuvent être mises à profit dans le cadre de l'utilisation interactive d'un système de recherche d'information. Ceci inclut particulièrement la visualisation des documents retrouvés et la justification de leur pertinence système. Ils s'appuient sur les résultats obtenus par les études expérimentales sur la pertinence telle qu'elle est conçue par les individus en situation de recherche d'information.

Chapitre 2

Problématique

Dans ce chapitre nous présentons la problématique de la modélisation de la pertinence pour les systèmes de recherche d'information, dans la perspective de l'utilisation interactive des systèmes.

Nous montrons d'abord comment l'architecture classique des systèmes de recherche d'information conduit à travailler à l'amélioration de cette modélisation par l'intermédiaire de la fonction de correspondance et de la fonction de prise en compte du bouclage de pertinence, que nous appelons «fonction de reformulation».

Puis nous proposons une autre architecture qui met en évidence les liens entretenus par l'utilisateur et le système pendant l'interaction. À partir de cette architecture, nous montrons comment l'amélioration de la modélisation de la pertinence peut être envisagée sous un autre angle : celui de la relation entre le schéma individuel de pertinence et la pertinence système.

La pertinence système comprend une fonction de correspondance et une fonction de mise en forme qui détermine la façon dont cette réponse sera effectivement visualisée. La fonction de reformulation de l'architecture classique relève désormais des tâches incombant à l'utilisateur (tâches d'évaluation et de reformulation), le système ayant pour but de favoriser la réalisation de ces tâches au travers de la fonction de mise en forme de la réponse du système. Les caractéristiques souhaitées de la mise en forme induisent une conception de la fonction de correspondance que nous développons par analogie avec le schéma individuel d'évaluation de la pertinence décrit précédemment.

2.1 Interaction utilisateur système

La nature interactive des systèmes de recherche d'information a évolué, depuis l'intervention d'intermédiaires de recherche (documentalistes utilisant le système pour le compte d'un utilisateur final), jusqu'à l'utilisation directe du système par

l'utilisateur final de l'information recherchée. C'est dans cette seconde perspective d'utilisation que nous nous situons.

La figure I.2.1 montre le schéma fonctionnel classique d'un système de recherche d'information où intervient un bouclage de pertinence automatique, réalisé par la fonction de reformulation. Les parties grisées désignent les données qui jouent un rôle dans l'interaction : requête, documents retrouvés et jugements de pertinence. La fonction de reformulation réalise la prise en compte automatique des modifications du schéma individuel de pertinence produites par la vue de la réponse, tels que les traduisent les jugements de pertinence recueillis par le système.

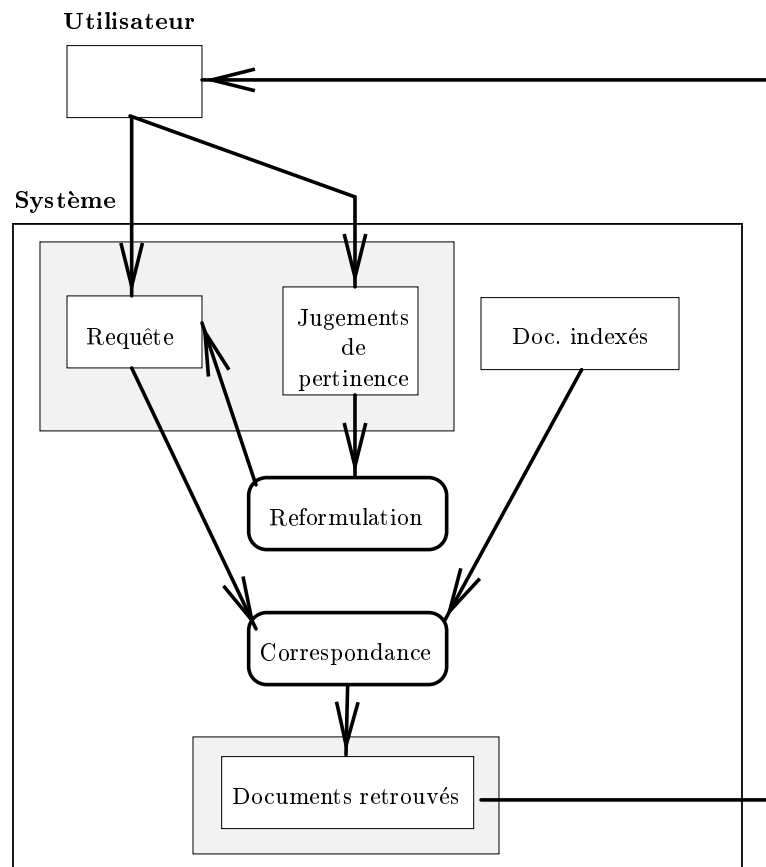


FIG. I.2.1 – *Schéma fonctionnel avec reformulation automatique*

La figure I.2.2 montre les tâches que doit réaliser un utilisateur dans ce cadre d'utilisation d'un système de recherche d'information. On retrouve dans la partie supérieure de la figure les données et processus mentaux d'un individu en situation de recherche d'information mis en présence de documents. La partie inférieure re-

présente l'ensemble des données attendues (requête, jugements de pertinence) ou fournies (documents retrouvés) par un système de recherche d'information¹. Nous précisons la nature de la donnée relative aux documents retrouvés, car dans la pratique, l'utilisateur perçoit les documents retrouvés selon une vue sur ces documents, qui est souvent assortie d'informations complémentaires telles que le degré de pertinence estimé par le système. Ainsi nous parlerons plutôt de *vue sur la réponse* du système.

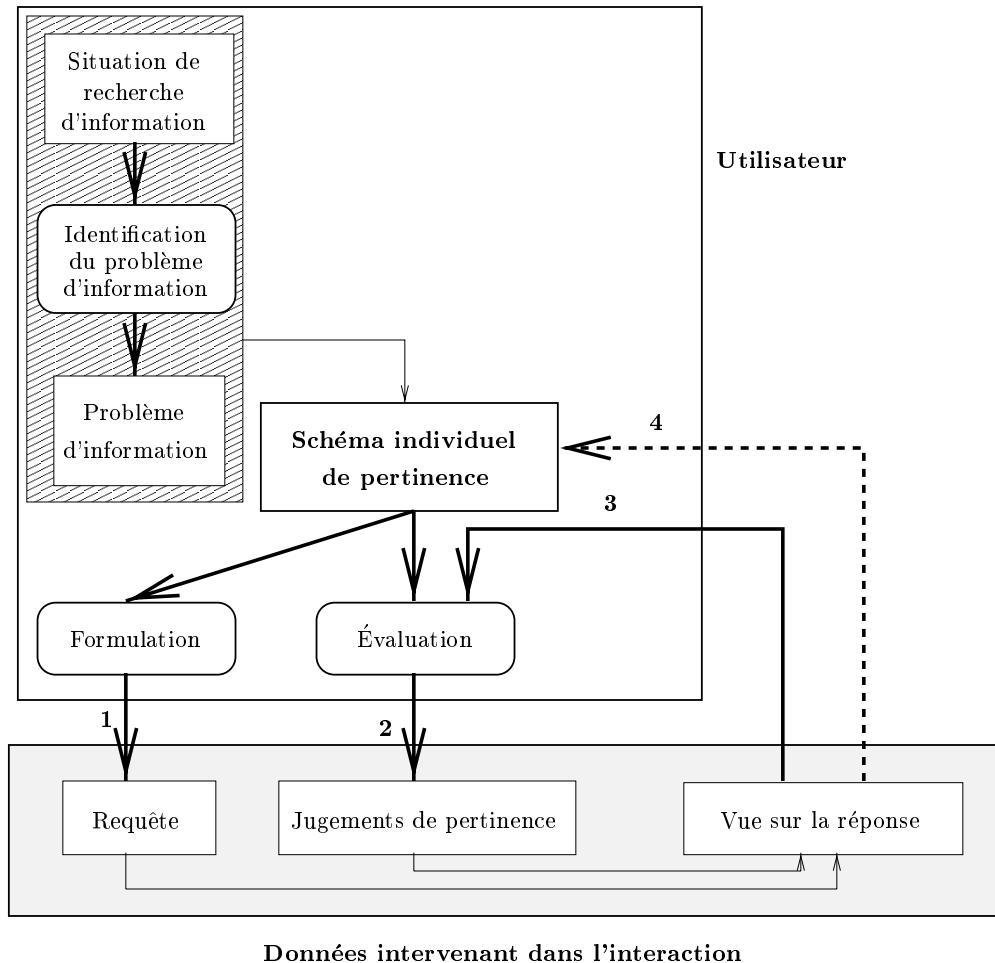


FIG. I.2.2 – Interaction entre l'utilisateur et les données du système en entrée et en sortie

1. Seuls les éléments du système de recherche d'information qui appartiennent à l'interface apparaissent sur ce schéma : ainsi des documents n'apparaît que la vue que fournit le système sur les documents retrouvés.

L'individu, en tant qu'utilisateur d'un système de recherche d'information, a pour tâche la *formulation* d'une requête qui doit traduire son problème d'information pour le système (flèche 1). De plus, il évalue la pertinence (flèche 2) des documents retrouvés par le système sur la base de la *vue sur la réponse* qui lui est fournie (flèche 3).

La vue sur les documents retrouvés intervient directement (flèche 3) dans l'évaluation de la pertinence utilisateur en tant que donnée sur laquelle sont effectués les jugements de pertinence. Elle intervient aussi indirectement (flèche 4) par l'intermédiaire de l'effet que produit la vue du résultat sur le schéma individuel de pertinence. En effet, au vu des documents retrouvés, la situation de recherche d'information de l'utilisateur est modifiée, ce qui se traduit dans notre schéma par la modification du schéma individuel de pertinence.

Dans la figure I.2.3, nous reprenons le schéma fonctionnel classique en mettant en évidence le rôle joué par la fonction de mise en forme de la réponse du système, qui produit une vue sur la réponse sur laquelle intervient effectivement l'interaction.

Sur cette figure, on peut repérer la dualité de la fonction de *reformulation* du système avec les modifications engendrées par la vue des résultats. Du point de vue de l'utilisateur, les processus mentaux qui régissent les modifications du schéma individuel de pertinence en fonction des documents observés sont complexes. Du point de vue du système, la reformulation produit une nouvelle requête (flèche 1') calculée à partir de données très rudimentaires que sont des jugements de pertinence binaires (pertinent vs. non pertinent) sur les documents retrouvés (flèche 4'). Cette modélisation est par conséquent très grossière. De plus la fonction de reformulation est conçue comme une boîte noire pour l'utilisateur ce qui rend impossible le contrôle du sens que le système associe aux jugements de pertinence fournis. Notamment, il est difficile pour l'utilisateur de savoir ce que le système attend en matière d'évaluation.

Nous proposons d'améliorer les performances des systèmes par le biais de la fonction de mise en forme d'une part, et en ménageant un lien entre la vue sur la réponse et la requête, afin de permettre à l'utilisateur de participer à la reformulation de manière active et éclairée.

2.2 Réduire la distance

Réduire la distance entre pertinence système et pertinence utilisateur amène généralement à raffiner la fonction de correspondance. Cependant, ce raffinement est limité par l'indexation du corpus considéré, et par le fait que la tâche de formulation de la requête qui incombe à l'utilisateur doit conserver une complexité limitée. De même, le raffinement de la fonction de reformulation est limité par l'expressivité des jugements de pertinence et par les caractéristiques des langages mis en œuvre (indexation, interrogation).

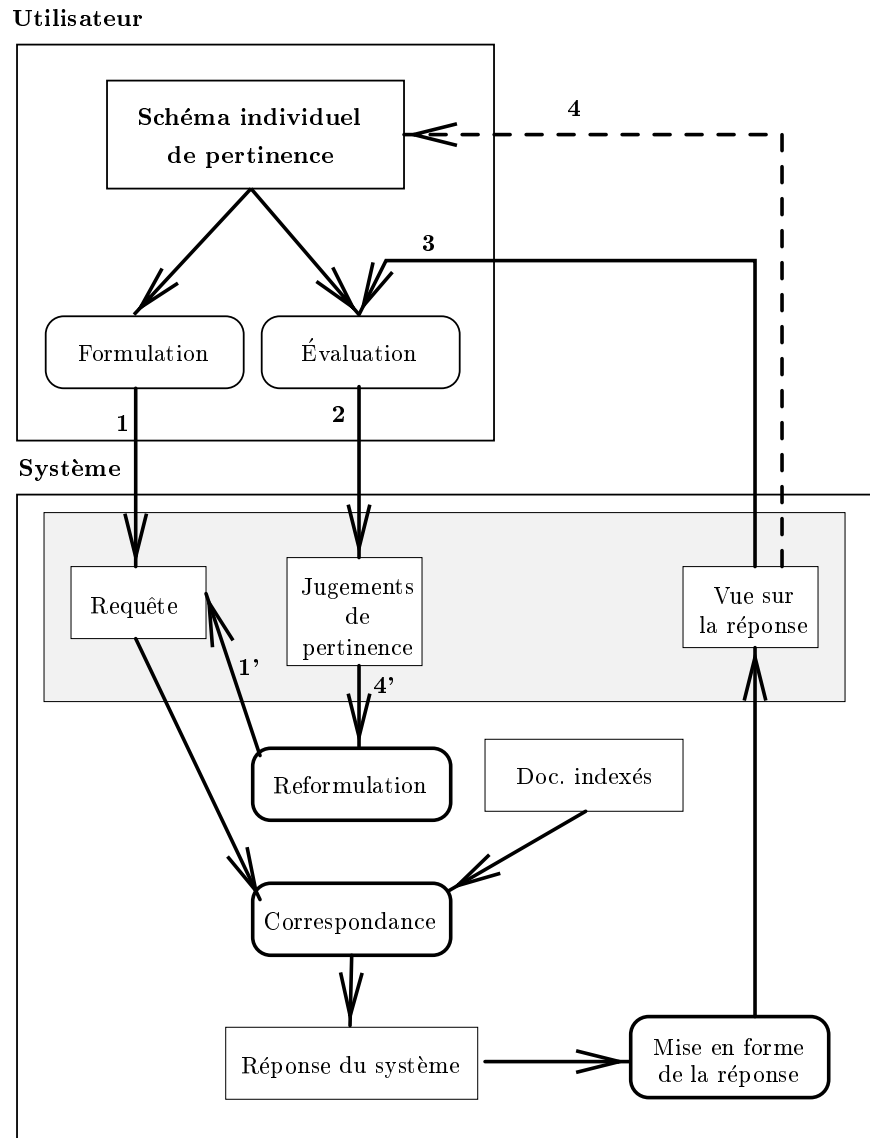


FIG. I.2.3 – Mise en évidence des tâches de l'utilisateur

Nous exposons ici les deux axes selon lesquels nous travaillons pour réduire la distance entre pertinence utilisateur et pertinence système.

2.2.1 Schéma de pertinence et classes de pertinence

La voie que nous explorons consiste à recueillir plus d'informations de l'utilisateur, tant lors de la formulation que lors de la reformulation, afin d'améliorer l'adaptation du système à la pertinence utilisateur et à ses modifications. Nous proposons d'augmenter le pouvoir d'expression des jugements de pertinence pour l'utilisateur, dans la mesure où le système peut mettre à profit ce surcroît d'information pour s'adapter à la situation courante de recherche d'information.

La figure I.2.4 décrit le schéma d'interaction que nous proposons. D'un point de vue général, les deux caractéristiques qui diffèrent sont les suivantes :

- les données qui interviennent dans l'interaction sont fortement liées entre elles du point de vue sémantique, grâce au schéma de pertinence (qui joue le rôle de requête) et aux classes de pertinence ;
- le contrôle de la reformulation revient à l'utilisateur, et ne se distingue pas fondamentalement de l'évaluation, grâce au lien établi dans l'interface entre la vue sur la réponse à évaluer et le schéma de pertinence à reformuler.

Si nous traitons essentiellement de reformulation manuelle dans ce mémoire, l'objectif final d'application de notre modèle consiste à définir une fonction supplémentaire du système qui, à partir d'une évaluation détaillée de la réponse obtenue par un dialogue, permet de guider l'utilisateur vers la reformulation la plus appropriée, afin de réduire la complexité de cette tâche.

L'élément central de notre modélisation est l'introduction d'un *schéma de pertinence*, qui a pour but de recueillir des informations définissant le schéma individuel de pertinence. Le schéma de pertinence, qui joue le rôle de la requête, permet de paramétrer la fonction de correspondance d'une part et la fonction de mise en forme d'autre part.

Les *classes de pertinence* organisent l'ensemble des documents retrouvés en fonction du schéma de pertinence. Elles fournissent à l'utilisateur une vue du corpus structurée en fonction du schéma de pertinence qu'il a formulé. Ainsi l'interface constitue le lieu de la confrontation entre le sens que l'utilisateur veut exprimer (son schéma individuel de pertinence) et le sens que le système est capable de produire à partir de l'expression du schéma de pertinence.

Le regroupement des documents en classes permet à l'utilisateur d'effectuer une opération mentale d'abstraction qui relève de sa compétence [How94] : identifier les similarités entre documents d'une même classe et les différences entre les documents appartenant à des classes distinctes. Cette opération lui permet d'appréhender le

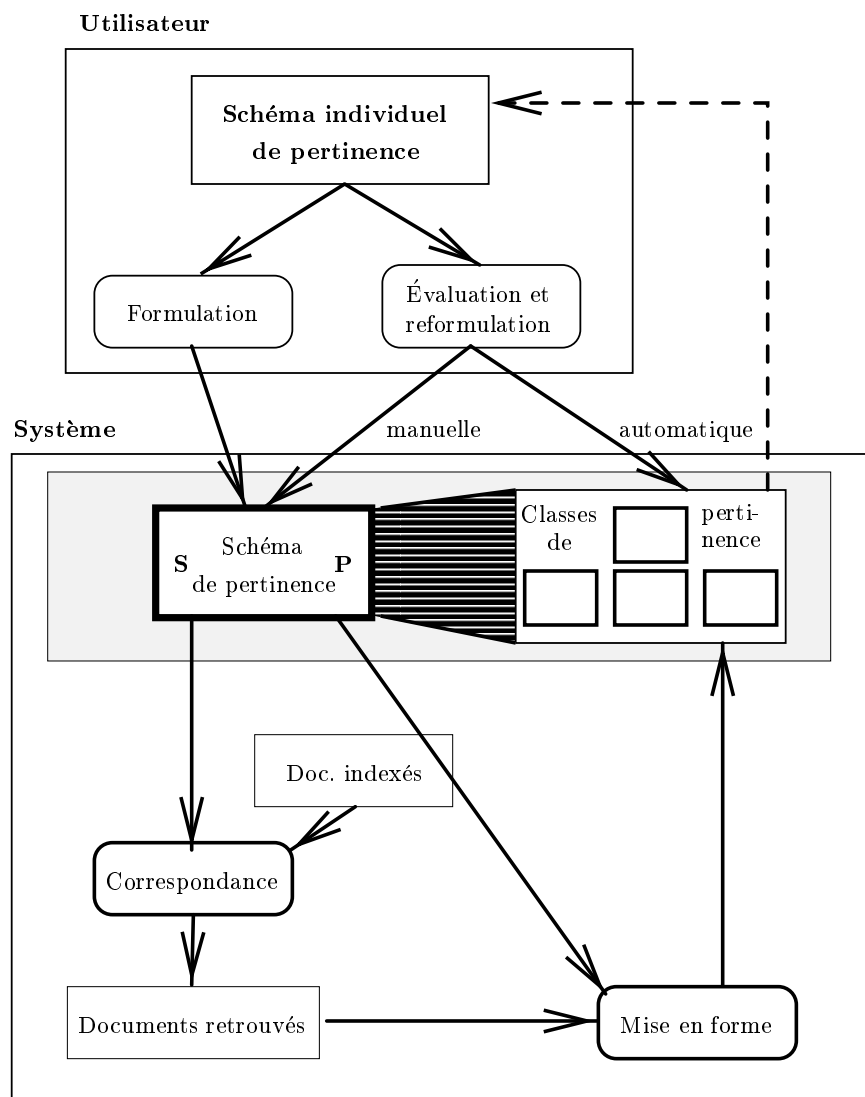


FIG. I.2.4 – Schéma fonctionnel adopté

sens que le système associe à l'expression qu'il a fournie en entrée, à travers les documents eux-mêmes.

Ainsi, le bouclage de pertinence peut se situer au niveau global de l'expression du schéma de pertinence, qui est directement connecté aux documents par l'intermédiaire des classes, plutôt qu'au niveau des documents pris un par un comme c'est le cas dans le bouclage de pertinence classique.

En comparant cette interface avec celle de la figure I.2.3, nous constatons que les divers éléments de l'interface sont connectés entre eux de manière à unifier d'une part la sémantique des données que l'utilisateur doit fournir entre elles (la requête et les jugements de pertinence), et d'autre part la sémantique de ces données fournies en entrée et les données obtenues en sortie (la vue sur les documents retrouvés).

Dans la suite nous détaillons les conséquences de cette définition de l'interface sur la conception du système et sur ses capacités d'adaptation.

2.2.2 Adaptation du système

La formulation du schéma de pertinence incombe à l'utilisateur au même titre que la requête. Il est défini en fonction de la requête et intervient selon deux axes dans le système : la *paramétrisation sémantique de la correspondance* (S), qui donne lieu à un ensemble de fonctions de correspondances possibles, et la *paramétrisation pragmatique de la visualisation* des résultats (P), qui donne lieu à un ensemble de mises en forme possibles pour les résultats de la requête.

Dans le schéma classique (figure I.2.3), la correspondance, la reformulation et la mise en forme des résultats constituent autant de boîtes noires pour l'utilisateur. Dans notre schéma, l'utilisateur dispose d'une vue sur la correspondance par l'intermédiaire du schéma de pertinence. Le schéma de pertinence montre les paramètres sémantiques de la correspondance (S) et permet leur manipulation. Le schéma de pertinence joue ainsi le rôle de définition de l'interprétation à donner de la requête.

Le schéma de pertinence permet aussi la manipulation de paramètres relatifs à la visualisation de la réponse (P) qui permettent à l'utilisateur de saisir plus clairement la relation entre les documents retrouvés et son schéma de pertinence. En cela, ce dernier joue un rôle pratique important dans la mise en place de bonnes conditions d'interaction, qui favorisent un bon déroulement du processus itératif de recherche d'information.

La fonction de reformulation est ainsi remise entre les mains de l'utilisateur. Du fait du lien entre le schéma de pertinence et la réponse du système, évaluer la pertinence d'un document trouve un équivalent sémantique en termes de reformulation du schéma de pertinence.

2.2.3 Démarche

Notre objectif est de définir un modèle de pertinence pour le système, qui fait intervenir conjointement la conception de la fonction de correspondance et la conception de l'interaction, dans le but de favoriser la réalisation des tâches qui incombent à l'utilisateur.

Pour établir les principes de conception de la fonction de correspondance, nous nous appuyons sur les critères de pertinence que les individus sont capables d'énoncer dans le cadre des études expérimentales sur la pertinence. L'objectif est de paramétrer la fonction de correspondance selon des termes ayant un sens pour l'utilisateur du système, et dans un contexte interactif destiné à favoriser l'acquisition de ce sens.

Ainsi notre fonction de correspondance est paramétrée par un ensemble de critères de pertinence, qui sont accessibles au niveau de l'interface du système par l'intermédiaire du schéma de pertinence (ces paramètres sont symbolisés par **S** dans notre schéma d'interaction, figure I.2.4).

Les principes de conception de l'interaction font donc entrer en ligne de compte les conditions qui favorisent une utilisation adéquate du système, par une meilleure maîtrise de la sémantique de la pertinence système. L'objectif est alors de concevoir l'interaction de façon à permettre à l'utilisateur d'élaborer une bonne image mentale de la pertinence système ; nous nous appuyons pour cela sur les connaissances établies quant à l'influence des conditions dans lesquelles s'effectuent les jugements de pertinence sur les jugements eux-mêmes.

Dans ce but, le schéma de pertinence inclut un second type de paramètres (symbolisés par **P** dans notre schéma d'interaction, figure I.2.4), qui définissent les *statuts de critères*. Ces paramètres spécifient la fonction de mise en forme des réponses du système. Leur introduction a pour but de régler les conditions pratiques de l'interaction dans le contexte d'une situation de recherche donnée.

2.3 Plan de ce mémoire

Dans la partie II, nous étudions les caractéristiques connues de la pertinence pour un individu, à partir des études réalisées indépendamment de l'utilisation d'un système de recherche d'information. Notre étude a pour but de resituer ces connaissances dans la perspective de leur prise en compte dans la conception des systèmes de recherche d'information. Dans le chapitre 1 nous identifions les caractéristiques de la pertinence qui peuvent être prises en compte dans la fonction de correspondance. Nous analysons ensuite les caractéristiques relatives aux conditions dans lesquelles sont produits les jugements de pertinence et nous en tirons des conséquences (sous forme de principes) quant aux conditions d'interaction qu'il est souhaitable d'établir, afin d'améliorer les performances des systèmes dans le contexte de la recherche d'in-

formation interactive. Nous présentons dans le chapitre 2 un ensemble de travaux existants permettant d'*adapter le système* à une situation de recherche d'information particulière, et nous replaçons chacune de ces approches dans le cadre de notre schéma d'interaction (figure I.2.4) afin de les situer vis-à-vis de la nôtre.

Dans la partie III, nous décrivons le modèle de pertinence système que nous proposons, d'un point de vue conceptuel. Ce modèle s'appuie sur les principes établis dans la partie précédente relativement à la conception de la fonction de correspondance (chapitre 1) et à la conception de l'interaction d'autre part (chapitre 2). À l'issue de ce chapitre, les liens qu'entretiennent les éléments qui participent à l'interface (le schéma de pertinence et les classes de pertinence), et les fonctions paramétrées du système (la fonction de correspondance et la fonction de mise en forme des résultats) sont spécifiés.

La partie IV propose une formalisation de ce modèle de pertinence système. Le chapitre 1 définit la sémantique du schéma de pertinence. Le chapitre 2 donne la définition des classes de pertinence. Le chapitre 3 complète la formalisation des classes de pertinence dans la perspective de leur visualisation à travers l'interface. Le chapitre 4 donne le principe d'utilisation de notre modèle en montrant comment il permet de décrire et de gérer la dynamique du processus de recherche d'information.

Dans la partie V, nous présentons une évaluation de cette approche. Dans le chapitre 1, nous appliquons les concepts définis à un corpus d'images, et illustrons les notions de situation problématique et de réaction dans ce contexte. Le chapitre 2 met en perspective les possibilités de contrôle de l'interaction qu'offre notre modèle. Le chapitre 3 cerne le champ d'application de notre modèle, et un bilan de cette application est dressé dans le chapitre 4.

Partie II

Pertinence utilisateur et pertinence système

Les études sur la pertinence pour la recherche d'information regroupent deux catégories de travaux : d'une part les études s'appliquant à analyser et modéliser la pertinence telle qu'elle est conçue par un individu en situation de recherche d'information, d'autre part les études s'intéressant à la modélisation de la pertinence pour les systèmes de recherche d'information.

Dans le premier chapitre, nous utilisons les résultats des études expérimentales menées sur les facteurs de pertinence pour discuter d'une part des possibilités et des limites à l'extension de la fonction de correspondance, et d'autre part des principes selon lesquels il est souhaitable que l'interaction des systèmes de recherche d'information soit conçue.

Dans le deuxième chapitre, nous passons en revue les modèles et techniques existantes pour l'adaptation du système à la situation de recherche d'information. À cette occasion nous nous positionnons vis-à-vis de ces approches.

Dans le troisième chapitre, nous résumons les conclusions tirées de ces études et exposons brièvement les choix qui sous-tendent notre approche.

Chapitre 1

Facteurs de pertinence et conception des SRI

Although two users might judge the meaning of the same document differently (whether based upon the full text of the document or representations of the document), it is assumed that both judgements somehow derive from characteristics of that document.

[Bar94], p. 152.

Cette citation de Barry souligne l'hypothèse fondatrice de la conception des systèmes de recherche d'information : les jugements de pertinence d'un utilisateur reposent non seulement sur le contenu objectif des documents, mais aussi sur les caractéristiques des documents qu'il juge en fonction de son propre contexte. Un système de recherche d'information, en indexant certaines caractéristiques des documents, peut définir une fonction de correspondance qui établit, pour chaque document, une relation entre ses caractéristiques indexées et une requête, afin d'approcher la relation de pertinence que l'utilisateur établirait lui-même.

Mais nous allons voir dans ce chapitre que de nombreux facteurs interviennent dans l'évaluation de la pertinence par l'utilisateur, dont seulement une petite partie peut être intégrée directement dans la fonction de correspondance.

Ceci nous conduit à réaliser deux études fondées sur les résultats de travaux expérimentaux relatifs à la notion de pertinence.

La première étude nous permet de cerner les possibilités d'étendre la *fonction de correspondance* afin de permettre une formulation du problème d'information prenant en compte un plus grand nombre d'éléments du schéma individuel de pertinence. Cette étude nous conduit à spécifier la structure générale du schéma de pertinence du système que nous proposons.

Dans la seconde étude, nous reprenons un certain nombre de résultats existants sur la notion de pertinence pour proposer un ensemble de principes de conception des

systèmes de recherche d'information interactifs. Ces principes concernent principalement le processus de *mise en forme* des résultats et les qualités que doit présenter l'interface d'un système de recherche d'information.

1.1 Motivation des deux études

Dans l'optique d'améliorer les performances des systèmes de recherche d'information, nous nous intéressons à ce que peut apporter un rapprochement entre l'évaluation de la pertinence par un individu et la modélisation de la pertinence dans un système de recherche d'information.

Nous montrons d'abord la distance qui existe entre pertinence système et pertinence utilisateur selon la conception classique des systèmes. Puis, partant des connaissances existant sur la pertinence pour un individu, nous identifions deux axes selon lesquels la conception d'un système de recherche d'information peut être discutée, par analogie avec les facteurs de la pertinence.

1.1.1 Approche classique de modélisation de la pertinence dans les systèmes de recherche d'information

Les chercheurs s'accordent, du point de vue théorique, sur une définition du concept de pertinence fondée sur une relation entre un problème d'information et l'information qui permet de résoudre ce problème. Ils s'accordent aussi sur le fait qu'une définition opérationnelle de ce concept passe par la décision que prend un individu d'accepter ou de rejeter un objet d'information qui lui est présenté.

La plupart des modèles théoriques de la pertinence proposent une définition dichotomique de ce concept, qui n'est pas sans rapport avec les problèmes de conception des systèmes de recherche d'information.

Traditionnellement on distingue deux grandes notions de pertinence. En 1959, Vickery [Vic59b, Vic59a] les identifie comme la «pertinence pour un sujet» (*relevance to a subject*), et la «pertinence pour l'utilisateur» (*user relevance*).

La pertinence pour un sujet fait référence à la *relation thématique* entre des objets d'information (*topicality*). Elle constitue la vue informationnelle de la pertinence. Elle est souvent assimilée à une vue de la pertinence *orientée vers le système*, car elle dépend de facteurs liés aux systèmes : elle est associée dans le système à une correspondance directe entre les termes de la requête et les termes qui indexent le document.

La pertinence pour l'utilisateur est fondée sur la décision que l'utilisateur prend de poursuivre ou non sa recherche d'information. Elle constitue la vue situationnelle de la pertinence. Cette notion s'appuie donc sur des facteurs liés aux utilisateurs, et est souvent assimilée à une vue de la pertinence *orientée vers l'utilisateur*.

Cette dichotomie entre la pertinence pour un sujet et la pertinence pour l'utilisateur apparaît sous diverses formes dans de nombreux travaux théoriques sur la pertinence. Cooper [Coo71, Coo73] oppose «pertinence logique» (*logical relevance*) et «utilité» (*utility*), ce dernier terme étant repris par Saracevic [Sar75] ; Maron [Mar77] distingue «à-propos objectif» et «à-propos subjectif» (*objective and subjective aboutness*) ; Wilson [Wil73] parle de «pertinence situationnelle» (*situational relevance*) qui enrichit la notion de relation thématique de paramètres liés à l'individu.

Cette analyse dichotomique du concept de pertinence induit un principe de conception des systèmes de recherche d'information qui l'apparente à une machine [DN86] dans laquelle seule la pertinence pour un sujet est prise en compte dans le système, alors que la pertinence pour l'utilisateur est ignorée.

1.1.2 Paradoxe de l'évaluation et évolution des besoins

La méthodologie d'évaluation des systèmes de recherche d'information reflète ce décalage entre les connaissances sur la pertinence et la modélisation de la pertinence dans les systèmes. En effet, elle repose sur un paradoxe, formulé entre autres par Barry [Bar94], selon lequel les systèmes de recherche d'information modélisent la pertinence comme un processus de correspondance entre termes, tandis qu'ils sont évalués en fonction de la pertinence utilisateur.

C'est ainsi que Barry [Bar94] explique les performances qualitatives limitées des systèmes de recherche d'information. Belkin en 1980 [Bel80] rapporte que les performances des systèmes de recherche d'information se situent en moyenne autour de 60 % de rappel et 40% de précision, ce qui est loin de la configuration idéale correspondant à 100% de rappel et 100% de précision ! Depuis, les progrès obtenus sur ces performances ne sont pas décisifs, et ce n'est pas surprenant car il semble difficile d'améliorer de tels résultats si l'on persiste à évaluer les systèmes sur la base de la réalisation d'une fonction pour laquelle ils n'ont pas été conçus.

Les besoins qui se font sentir dans la demande des entreprises et du grand public permettent d'identifier des critères de performance des systèmes qui ne font pas seulement référence aux mesures de rappel et précision, mais aussi à la robustesse et à l'intelligibilité dans l'utilisation des systèmes. C'est selon ces critères que les systèmes non seulement doivent être évalués, mais aussi doivent être conçus, en intégrant les paramètres liés à la pertinence utilisateur.

Ainsi, la question que nous posons en corollaire à ce problème est la suivante : quels rapports doivent entretenir dans les systèmes, la pertinence telle qu'elle est estimée par la fonction de correspondance (*pertinence système*), et la pertinence telle que l'individu l'évalue (*pertinence utilisateur*) ?

1.1.3 Facteurs de pertinence

Les travaux d'identification des facteurs de pertinence, menés indépendamment du contexte de l'utilisation d'un système, ont produit de nombreuses données que nous reprenons dans les sections 1.2 et 1.3.

Parmi les facteurs de pertinence, on trouve les critères de pertinence, qui sont les facteurs de pertinence que les individus peuvent formuler explicitement a priori. Des catégories de critères de pertinence sont identifiées dans diverses études [Bar94, Par93, Sch91], s'appliquant à un corpus et un type de situation de recherche particulier. La démarche expérimentale de ces travaux consiste à recueillir auprès d'utilisateurs les raisons qui président à leurs jugements de pertinence. Ils permettent d'établir des typologies de critères de pertinence employés par les utilisateurs.

Dans [Bar94], Barry constate que les catégories de critères ainsi identifiées se recouvrent considérablement pour des contextes d'application similaires, et il en déduit que pour un certain champ d'application, ce genre d'étude doit permettre d'établir une typologie exhaustive des critères de pertinence effectivement employés.

1.1.4 Utilisation des critères de pertinence par les individus

Nous rapportons ici les résultats concernant la façon dont les critères sont utilisés par les individus pour évaluer la pertinence d'un document.

Dans son étude expérimentale sur la pertinence telle qu'elle se reflète dans les constructions mentales individuelles [How94], Howard met en évidence l'existence d'un schéma individuel d'évaluation de la pertinence, qu'élabore chaque individu en situation de recherche d'information.

Les constructions mentales [Kel55] sont des canaux psychologiques qu'un individu crée *pendant qu'il interprète des événements récurrents, en anticipant le traitement futur d'événements semblables*¹.

Du point de vue de la méthode expérimentale, les constructions mentales sont collectées selon le processus suivant.

- 1° Chaque individu possède un ensemble de documents qu'il a utilisé pour résoudre un problème d'information. D'abord, l'individu attribue des scores aux documents sur une échelle de pertinence comprenant 13 points. Puis un processus itératif commence.
- 2° L'individu considère trois documents à la fois. On lui demande d'identifier pour chaque triplet une caractéristique partagée par les trois documents, et de grouper les deux documents les plus semblables du point de vue de cette caractéristique.

1. *as he interprets recurring events in anticipation of future processing of similar events*, [How94], p. 173

3° Ainsi on obtient une paire appelée *la paire de similarité*, et un élément isolé appelé *le singleton différent*. L'individu articule alors la base de la similarité pour la paire de similarité et la base de la différence pour le singleton différent.

Les constructions mentales collectées sont assimilés à des critères de pertinence, qui sont groupés en fonction des catégories de critères auxquels elles appartiennent.

Enfin on demande à l'individu d'attribuer des scores aux constructions mentales pour chaque document.

La conclusion que tire Howard de cette étude est que des documents également pertinents ne le sont pas parce qu'il partagent les mêmes scores sur les critères de pertinence. En d'autres termes, pour un individu donné, deux documents peuvent être pertinents au même degré pour des raisons différentes.

Howard tire d'autres conclusions intéressantes de ces données expérimentales, parmi lesquelles les suivantes :

- les individus ont créé des constructions mentales similaires, c'est-à-dire des critères de pertinence similaires (aucun des critères collectés n'a été créé par un seul individu) ;
- les individus appliquent leurs critères de manière assez cohérente ;
- des documents également pertinents peuvent ne pas partager de constructions mentales, et dériver leur pertinence du rôle conjugué de différents critères.

Cette étude non seulement conforte le fait que l'individualité des utilisateurs est cruciale dans les jugements de pertinence, et que les fondements des jugements varient, mais elle montre aussi que ces fondements (les critères de pertinence, ou autrement dit les *raisons individuelles sous-jacentes aux jugements de pertinence*), peuvent être articulés par les utilisateurs.

De plus, les individus emploient ces raisons de manière cohérente. Enfin, ils partagent ces raisons entre eux, ils se comprennent quand ils les évoquent.

Confirmant notre point de vue issu des travaux de Howard, Barry [Bar94] précise que les utilisateurs sont apparemment capables de reconnaître et de discuter des aspects non thématiques des informations et des sources qui influencent leur évaluation de la pertinence².

2. *Users are apparently able to recognize and discuss non topical aspects of information and sources that are influencing their evaluation.*, [Bar94], p. 151

1.1.5 Deux axes pour améliorer les performances des systèmes

La question du rapport entre pertinence utilisateur et pertinence système dans l'utilisation d'un système de recherche d'information met l'accent sur des zones d'ombre de la conception classique des systèmes. Éclaircir ce point doit permettre de tirer un plus grand profit des connaissances dont la communauté dispose sur la pertinence.

Nous proposons d'utiliser la capacité des individus à formuler et discuter les critères de pertinence qui interviennent dans leur évaluation de la pertinence. D'une part, nous proposons de définir la fonction de correspondance en termes de critères de pertinence. D'autre part, nous proposons de fournir, au travers de la fonction de mise en forme de la réponse du système, des moyens interactifs qui permettent à l'utilisateur de contrôler le sens que prennent les critères formulés pour le système, et de visualiser diverses combinaisons de ces critères.

Dans la suite de ce chapitre, nous montrons dans quelle mesure les critères de pertinence identifiés dans les études sur la pertinence peuvent être intégrés à la fonction de correspondance, et en quoi certains facteurs de pertinence également identifiés dans ces études, permettent de déterminer des points critiques de l'interaction avec un système de recherche d'information, et de suggérer ainsi les principes d'interaction qui doivent être appliqués dans la conception des systèmes.

Étendre la fonction de correspondance

Les éléments de la pertinence qui peuvent être intégrés au niveau de la fonction de correspondance sont ceux que l'on peut décrire indépendamment de la situation de recherche d'information, et qui sont en relation directe avec les caractéristiques des documents. Nous les appelons *critères de pertinence*. Parmi ceux-ci, certains font appel à des caractéristiques des documents qui sont toujours indexées, comme les thèmes abordés dans un document textuel, tandis que d'autres font référence à des caractéristiques des documents qui sont rarement ou jamais indexées, comme par exemple la solidité scientifique d'un document textuel, ou la qualité esthétique d'un document iconographique.

Nous décrivons rapidement l'éventail de ces critères au moyen d'exemples. Pour chacun de ces critères nous évoquons la possibilité de leur prise en compte effective dans la fonction de correspondance, qui est conditionnée par l'indexation disponible.

Dans les systèmes classiques, les critères de pertinence utilisés relèvent tous de la relation thématique entre un document et les termes de la requête. Nous montrons dans la partie suivante comment la multiplication des critères que peut calculer la fonction de correspondance entraîne un niveau de modélisation supplémentaire : les critères qui s'appliquent aux termes de la requête doivent être rendus explicites.

Concevoir l'interaction

Parmi les éléments qui font varier la pertinence en fonction de la situation de recherche d'information, un grand nombre ne relève pas directement des caractéristiques des documents (comme c'est le cas des critères de pertinence), mais des caractéristiques de la situation de recherche d'information. Par exemple «l'individu est pressé», ou «un grand nombre de documents est présenté à l'individu» font référence à la situation de recherche d'information.

Ces facteurs de pertinence liés à la situation, qui ont été établis indépendamment du contexte de l'utilisation d'un système, peuvent être replacés dans le contexte de la conception des systèmes de recherche d'information interactifs. La qualité de réalisation des tâches incombant à l'utilisateur dépend au moins en partie de facteurs de pertinence, qui permettent de définir un ensemble de principes de conception des systèmes de recherche d'information interactifs.

1.2 Étendre la fonction de correspondance

Un système de recherche d'information implémente, au travers de la fonction de correspondance, un ou plusieurs critères qui définissent les conditions pour qu'un document soit estimé pertinent par le système, en fonction de la requête et de la forme indexée de ce document. Le choix des critères est limité par l'étendue des caractéristiques des documents effectivement indexées. Étant donné le cas le plus fréquent où un document est indexé uniquement par un ensemble de mots-clés décrivant le contenu thématique du document, la fonction de correspondance traduit un unique critère : l'apparition d'un thème de la requête dans le document indexé.

Les travaux sur la pertinence pour un individu montrent que l'éventail des critères qui sont appliqués par les individus pour évaluer la pertinence d'un document 1) sont variés, et 2) ne sont pas appliqués de la même façon en fonction de la situation de recherche d'information.

L'objectif de cette étude est tout d'abord d'évaluer les capacités d'un système à modéliser d'autres critères que celui d'une correspondance thématique, et ensuite de proposer une approche pour paramétrer la fonction de correspondance en fonction des critères. Ceci permet d'envisager des variations de la définition de la fonction de correspondance en fonction de la situation de recherche d'information.

1.2.1 Critères de pertinence

Lorsqu'un individu énonce un jugement de pertinence, il est capable de donner les raisons de son jugement, sous la forme de critères de pertinence, comme par

exemple

Exemple 1 – *ce document est pertinent parce qu’il traite du sujet qui m’intéresse, c’est-à-dire les éléphants d’Afrique*

Exemple 2 – *ce document est pertinent parce qu’il traite des éléphants, qui est un thème proche de celui qui m’intéresse (les éléphants d’Afrique)*

qui correspondent à des critères thématiques, ou encore

Exemple 3 – *ce document est pertinent parce qu’il est disponible actuellement dans ma bibliothèque de prêt*

Exemple 4 – *ce document est pertinent parce que l’auteur est très compétent dans le domaine*

Exemple 5 – *ce document est pertinent parce qu’il est bien étayé par des exemples*

Exemple 6 – *cette image est pertinente car elle a de belles couleurs*

Exemple 7 – *cette image est pertinente car elle évoque le calme et la tranquillité.*

Pour que le système prenne en compte non seulement les critères de pertinence thématiques mais aussi d’autres critères, plusieurs conditions doivent être réunies : il faut d’une part que l’utilisateur puisse formuler ces critères, et d’autre part que le système puisse évaluer la vérification de ces critères dans les documents à partir des documents indexés.

Dans la suite nous donnons l’éventail des critères de pertinence que les individus formulent pour expliquer leurs jugements de pertinence, en discutant les possibilités de les modéliser dans le système. Puis nous proposons une façon d’étendre la requête afin d’intégrer les divers critères de pertinence.

1.2.2 Éventail des critères et limites de leur prise en compte dans la fonction de correspondance

De nombreuses études se sont appliquées à identifier l’ensemble des facteurs de pertinence, parmi lesquelles [CK67], [RS67], [Coo71], [Coo73], [Tay86], [Par93]. Dans un récent état de l’art [Sch94], Schamber réunit les résultats de ces études sous la forme d’une liste de quatre-vingts catégories de facteurs. Ces catégories de facteurs sont instanciées de manière détaillée dans plusieurs études expérimentales spécifiques (cf. [Bar94], [Par93], [Sch91]), s’appliquant à un corpus et un type de situation de recherche particulier. La démarche expérimentale de ces travaux consiste à recueillir auprès d’utilisateurs les raisons qui président à leurs jugements de pertinence. Ils permettent d’établir des typologies de critères de pertinence employés par les utilisateurs.

Nous prenons les catégories de facteurs réunies par Schamber comme référence pour cette étude, et nous nous appuyons également sur l'étude de [GB95] qui s'intéresse plus précisément aux nuances des critères thématiques dans le cadre de documents textuels.

a. Le critère thématique et ses variantes

Le critère thématique est souvent modélisé sans nuances qualitatives dans les systèmes. Il est parfois strict (le terme de la requête doit être présent dans le document indexé), parfois souple (le terme de la requête doit être présent sous une forme ou une autre dans le document indexé), et parfois contextuel : quand la requête et les documents indexés présentent une structure logique (le terme de la requête doit apparaître dans le document indexé dans le même contexte que dans la requête). Mais malgré cela le critère thématique est souvent unique au sein d'un système.

Il est fréquent que le critère thématique soit nuancé de manière quantitative dans les systèmes, c'est-à-dire que la fonction de correspondance calcule non plus la vérification binaire d'un critère, mais le degré de vérification du critère. Ainsi chaque document se voit associer un degré de pertinence. Les documents peuvent alors être classés par ordre décroissant de pertinence.

Pour ce qui est des critères de pertinence thématiques identifiés, Green et Bean [Gre95, GB95] ont étudié les raffinements qui interviennent au sein d'une relation thématique de pertinence entre deux éléments d'information. Une relation thématique est la relation linguistique entretenue par un thème de la requête et un thème trouvé dans le document. Un critère thématique de pertinence s'exprime en fonction de ces relations de la façon suivante : un document est pertinent s'il existe un thème du document t qui entretient une relation R avec le thème t_0 donné.

Le degré de raffinement que ces auteurs atteignent impose un niveau technique du point de vue linguistique qui semble dépasser considérablement les capacités des utilisateurs à formuler leurs critères de pertinence. Cependant, même si les utilisateurs ne peuvent exprimer ces critères à ce niveau de détail, Green et Bean montrent qu'ils utilisent ces finesses malgré tout.

Les nuances sur les critères thématiques [Gre95, GB95] relèvent de trois catégories de relations thématiques :

- relations d'égalité (*topic matching relationships*) tableau II.1.1 ;
- relations hiérarchiques, tableau II.1.2 ;
- relations structurelles, tableau II.1.3.

Chacune des catégories comprend un grand nombre de relations traduisant les nuances qui peuvent être associées au terme t d'une requête. Nous donnons ici la

taxonomie associée par Green et Bean [GB95] aux relations d'égalité, aux relations hiérarchiques et un extrait de celle associée aux relations structurelles, correspondant à la sous-catégorie *Composants structurels*.

Relations d'égalité	
Référence	Je veux des informations à propos de X
Définition	Que signifie X, qu'est-ce que X
Attributs	
Adjectival	
Caractéristiques	Quelles sont les caractéristiques de X, à quoi ressemble X
Amplitude	Quelle est la taille de X, combien y a-t-il de X
Adverbial	
Conditions temporelles	Quand X intervient-il
Manière	Comment X intervient-il

TAB. II.1.1 – *Relations d'égalité*

Relations hiérarchiques	
Taxonomie	
Classe/sous-classe	Relation hiérarchique stricte avec X
Type/élément	X appartient à une classe
Partonymie	
Activités impliquées	X est une sous-activité d'une activité plus vaste

TAB. II.1.2 – *Relations hiérarchiques*

Les nuances qualitatives évoquées dans ces taxonomies sont souvent trop complexes pour qu'un individu non linguiste puisse les formuler. Cependant, quand un individu en situation de recherche d'information introduit un terme X dans une requête, l'intention qu'il associe implicitement à ce terme en matière de critère thématique peut se traduire par une ou plusieurs de ces relations.

Les relations d'égalité et les relations hiérarchiques sont relativement simples à appréhender par un utilisateur car elles peuvent se traduire simplement par une phrase. Cependant pour les implémenter dans les systèmes, un réseau sémantique normalisé est utilisé pour organiser les thèmes selon leurs relations de spécificité et de généralité, sous la forme d'un thésaurus. Les relations ainsi normalisées ne correspondent pas nécessairement aux relations qu'un individu donné conçoit effectivement entre les thèmes en question, ce qui constitue un premier problème d'adéquation entre la sémantique du système et celle que l'utilisateur a en tête.

Relations structurelles (extrait)

Composants structurels de X	
Objet	Ce à quoi l'action X s'applique
Résultat	Conséquence de l'action X
Objectif/But	Résultat espéré, qui déclenche l'action X
Source	Point d'origine de X
Autorité	Source de décision de X, pouvoir, etc.
Receveur	Personnalité qui acquiert X
Percepteur	Personnalité qui perçoit X
Standard	
Idéal	Point de comparaison idéal pour X
Repère	Point de référence pour X, souvent spatial

TAB. II.1.3 – *Relations structurelles (extrait)*

De plus, les systèmes existants exploitent généralement ces relations de manière interne, en réalisant systématiquement l'union des diverses relations thématiques qu'ils peuvent établir. Ainsi l'effet obtenu du point de vue de l'utilisateur, prend la forme d'une augmentation du rappel dans le résultat.

Si l'on considère que l'utilisateur peut introduire certaines nuances quant à la nature précise de la relation thématique qu'un terme X de sa requête doit entretenir avec les documents pertinents, il est préférable de paramétrer la fonction de correspondance avec les diverses relations possibles, plutôt que de faire systématiquement l'union des nuances possibles. Ainsi l'exemple 1 et l'exemple 2 peuvent être distingués l'un de l'autre, sous les formes *Thème = éléphants d'Afrique* et *Thème proche_de éléphants d'Afrique* ou bien *Thème =_ou_plus_général_que éléphants d'Afrique*.

En raison de l'inadéquation possible entre les relations implémentées et les relations que l'utilisateur a en tête, il est également important de permettre à l'utilisateur de comprendre la sémantique associée à ces relations par le système. En particulier, la relation *Thème proche_de* est en général implémentée comme un voisinage de thèmes au sein d'un thésaurus. Ce voisinage est même souvent quantifié en fonction d'une distance entre les thèmes dans le thésaurus. Le sens intuitif associé à cette relation par un utilisateur est susceptible d'être très différent de celui implémenté.

Cette inadéquation ne suffit pas à nous convaincre de renoncer en pratique aux techniques utilisant des thésaurus, qui s'avèrent souvent utiles, comme le montrent les évaluations qualitatives concernant l'expansion de requête (voir par exemple [QF93, HBW92, SB95]). Cependant, dans le contexte de l'interaction entre l'utilisateur et le système, il est important de mettre en place un dispositif qui clarifie la sémantique associée à ce type de relation dans le système.

b. Les critères non thématiques

Les exemples 3 à 7 font référence à des critères de pertinence qui ne relèvent pas des caractéristiques thématiques des documents. Il est fréquent que les caractéristiques des documents auxquelles ils font référence soient disponibles dans le système d'information qui gère le corpus (nom de l'auteur, présence d'exemples, disponibilité), ce qui permet de les intégrer dans la fonction de correspondance.

Attributs externes Quand un système dispose de caractéristiques des documents relevant des *attributs externes* du document (nom de l'auteur, date de publication, éditeur, type de support, disponibilité, etc.), la modélisation de critères simples du type *Document disponible*, *Auteur = X*, *Publication postérieure_à Y* est immédiate.

Mais l'exemple 4 montre que les critères formulés par les individus font en général référence à des connaissances supplémentaires qui ne sont pas disponibles dans le système. Dans notre exemple, le fait que l'auteur X est compétent dans le domaine des éléphants d'Afrique relève des connaissances (et aussi de la subjectivité) de l'individu.

Ainsi il est impossible que la fonction de correspondance modélise un critère de pertinence traduisant le fait qu'un document est écrit par un auteur compétent. En revanche elle peut permettre à un utilisateur, par l'intermédiaire du critère *Auteur = X*, de faire intervenir ces connaissances en réclamant de préférence des documents écrits par X.

Attributs structurels des documents Dans le contexte d'un corpus dont la structure logique est indexée, comme il est possible de le faire automatiquement pour les corpus textuels dont le format repose sur un langage de marquage du texte comme SGML, le genre de critère apparaissant dans l'exemple 5 peut être modélisé dans la fonction de correspondance [Par96], sous la forme *X au sein d'une structure logique de type Exemple*.

Attributs subjectifs L'exemple 6 montre que certains critères font référence à des attributs des documents qui sont fondamentalement subjectifs. Il est impossible de définir un critère *Belle couleur*. En revanche, l'utilisateur peut essayer d'exploiter au mieux les autres critères offerts par le système pour approcher ce critère, comme par exemple, s'il aime le bleu, *Couleur = bleu*.

L'exemple 7 montre que certains critères subjectifs sur des images sont plus difficiles à caractériser sous la forme de critères plus simples. Pourtant, dans le cas d'une indexation manuelle, l'indexeur peut faire intervenir sa propre subjectivité comme référence.

1.2.3 Principe de l'intégration des critères dans la fonction de correspondance

La forme générale d'un critère de pertinence nécessite non seulement un terme (X), comme c'est le cas dans les requêtes traditionnelles, mais aussi la spécification du critère particulier (ou de la nuance particulière) à appliquer à ce terme : *Auteur =*, *Publication postérieure_à*, *Thème =*, *Thème proche_de*, etc.

Ainsi la fonction de correspondance ne traite pas tous les termes de la requête indifféremment par un test d'égalité ou d'appartenance avec les termes indexant le documents : chaque terme est traité en fonction de la nature du critère de pertinence que l'utilisateur y associe. Les critères qui constituent la requête doivent permettre de traduire une plus grande partie du schéma individuel de pertinence qu'un simple ensemble de termes, même si l'on a pu constater au travers d'exemples que le degré de cette adéquation présente des limites.

1.3 Concevoir l'interaction

Le second axe à explorer pour améliorer les performances des systèmes de recherche d'information est l'axe de l'interaction.

Selon le schéma fonctionnel que nous avons décrit dans la figure I.2.4, les tâches incombant à l'utilisateur sont les suivantes :

Tâche 0 formuler une requête

Tâche 1 évaluer sa satisfaction (détecter la pertinence en parcourant les documents retrouvés ; décider de la poursuite ou de l'arrêt de l'itération)

Tâche 2 formuler des jugements de pertinence ou reformuler manuellement la requête (selon les systèmes)

1.3.1 Principes d'interaction

Pour déterminer ce qui constitue de bonnes conditions de recherche d'information interactive, nous nous appuyons sur les facteurs de pertinence liés à l'interaction de l'individu avec les documents qui lui sont présentés lors des expérimentations [Sch94].

Principe 1 (Détecter) *faciliter la détection de la pertinence*

La *commodité de détection de la pertinence* a été identifiée comme un facteur de pertinence : en fonction de cette commodité, les jugements de pertinence produits par un individu varient. Les paramètres dont elle dépend sont le nombre de documents présentés à la fois, l'ordre dans lequel des documents sont présentés, la

forme sous laquelle les documents sont présentés (le document entier, un résumé du document, le titre seulement, etc.), les informations de pertinence associées aux documents présentés (une estimation de la pertinence des documents censée prédire la pertinence utilisateur, comme par exemple un poids ou un ordre sur les documents), etc.

Tous ces paramètres interviennent traditionnellement dans la fonction de présentation des réponses du système. Il nous paraît utile de faire intervenir ces paramètres dans la conception des systèmes afin de faciliter la détection de la pertinence.

Principe 2 (Comprendre) *clarifier la sémantique de l'interaction*

Dans le contexte de l'utilisation d'un système de recherche d'information, la sémantique des tâches à réaliser doit être explicite. Ainsi divers utilisateurs peuvent élaborer leurs tâches en fonction de l'image mentale qu'ils ont du système. Une bonne image mentale implique une bonne réalisation des tâches, qu'il s'agisse de la formulation ou de la production de jugements de pertinence.

Ainsi la conception des systèmes de recherche d'information doit tenir compte de ce paramètre, au niveau des informations apparaissant dans l'interface. En particulier, les informations associées aux documents retrouvés doivent être choisies de façon à éclairer la sémantique de la pertinence système.

Il est cependant difficile, dans le cadre de l'utilisation autonome d'un système, de rendre explicite la sémantique des tâches à réaliser comme on peut le faire lors d'une expérience menée par un expérimentateur. En revanche, on peut concevoir l'interaction de façon à favoriser l'amélioration de l'utilisation itérative tout au long du processus de recherche d'information.

Principe 3 (Reformuler) *faciliter la reformulation*

La reformulation est rendue nécessaire lorsqu'il existe un décalage entre le sens que l'utilisateur veut exprimer au travers de la requête et celui que le système associe à cette requête. Ce décalage peut se situer à plusieurs niveaux : au niveau du sens des termes employés et de leur combinaison dans la requête, mais aussi (1) au niveau de la dimension du problème d'information qui n'est pas exprimée ou (2) qui n'est pas exprimable dans la requête.

Pour le cas 1, il est difficile pour un individu de cerner objectivement son problème d'information a priori, dans la mesure où sa situation de recherche complète implicitement et inconsciemment le sens qu'il attribue à sa requête. Pour s'apercevoir de l'ambiguïté de cette formulation, il lui faut être confronté à un exemple de document qui satisfait sa requête pour le système mais ne contribue pas à résoudre son problème d'information, ce qui implique une itération.

Pour le cas 2, il se peut que l'utilisateur n'exploite pas au mieux les capacités d'expression du système. Il est bon de favoriser par l'interaction la prise de conscience

des capacités du système. La reformulation sert alors à faire prendre progressivement conscience de ces limites. L'interface joue un rôle central dans la réalisation de cette fonctionnalité.

Enfin un dernier type de raison repose dans le fait que le contenu du corpus n'est pas connu de l'utilisateur. Ainsi la reformulation peut servir, pour l'utilisateur, à établir une confirmation de l'absence de (d'autres) documents pertinents dans le corpus. Pour permettre à l'utilisateur ce genre de confirmation, il faut que l'interface puisse donner une vue globale du corpus.

La *cohérence des informations* présentées à l'individu a été identifié comme un facteur de pertinence. Du point de vue de l'utilisation d'un système, nous traduisons cela par le fait que les données présentées dans l'interface sont cruciales pour un bon déroulement du processus itératif de recherche. Notamment, il est nécessaire de mettre en évidence dans l'interface les liens que le système établit entre les données en entrée et les données en sortie [VB96]).

Chapitre 2

Adaptation du système

Researchers have long been interested in the sort of feedback which the end-user requires in order to make appropriate judgments concerning the relevance of documents that have been retrieved from information systems. This judgment is subject to situational factors which are unique to the user and the context in which the IR interaction takes place, i.e. factors affecting the situational dynamism of user-centered relevance estimation. The way a user judges relevance may change during the process of engaging the information system. A mechanism which allows the user to articulate changes to the cognitive schema for relevance estimation could be used to inform the retrieval engine of an information system and improve the effectiveness of the search process.

[Bru94], p. 148.

Ce chapitre présente trois types de méthodes d'adaptation du système au contexte de la recherche. Nous présentons d'abord les techniques s'apparentant au principe de *bouclage de pertinence*. Puis nous décrivons les approches fondées sur les *stratégies de recherche* et les modèles d'utilisateur. Enfin nous présentons les travaux qui visent à définir une *paramétrisation logique de la correspondance*.

Nous analysons chacune de ces approches dans les termes de notre schéma d'interaction afin de distinguer les divers niveaux d'adaptation et de paramétrisation existants, et de situer notre approche.

Toutes les techniques présentées ci-après s'inscrivent dans le contexte d'un corpus indexé par un ensemble de termes d'indexation atomiques correspondant aux thèmes abordés dans les documents. À ces termes sont associés des poids issus de mesures statistiques sur les documents et l'ensemble du corpus. La requête consiste en un ensemble de termes d'indexation pondérés.

2.1 Jugements de pertinence et bouclages de pertinence

Les techniques de *bouclage de pertinence classique*, modifient automatiquement la requête de l'utilisateur lorsque son problème d'information est mal modélisé du point de vue du système.

Des variantes de cette technique consistent à adapter la requête selon un processus semi-automatique, appelé *aide à la reformulation* de la requête. En l'absence de tout processus de bouclage intégré au système, l'utilisateur lui-même peut procéder à une reformulation manuelle de sa requête. Il est courant d'utiliser un thésaurus des termes d'indexation pour procéder à la reformulation.

Plutôt que de modifier la requête, les jugements de pertinence peuvent aussi servir à adapter le contexte d'évaluation de la correspondance, c'est-à-dire l'espace des termes d'indexation (voir le *logical imaging* de Crestani [CvR96] et Ruthven [Rut96]), ce qui revient indirectement à modifier l'indexation des documents.

Ces techniques reposent en général sur une méthode de révision des probabilités associées aux termes d'indexation, vu comme un espace d'événements (voir le modèle probabiliste [Mar85, Fuh92, vR92]).

2.2 Typologies et stratégies de recherche

Quand la pertinence système est mise en cause dans la non-satisfaction de l'utilisateur, l'adaptation du système consiste en une modification de la fonction de correspondance. C'est le cas des systèmes fondés sur un ensemble de *stratégies de recherche* constituant autant de fonctions de correspondance.

2.2.1 Dédurre de la requête la meilleure stratégie

La première approche conçue dans ce sens [CT84] consiste à intégrer dans un même système plusieurs fonctions de correspondance. Par ailleurs, on identifie des types de requêtes sur la base de caractéristiques syntaxiques des requêtes. Enfin, au moyen d'expérimentations on détermine, pour chaque type de requête, quelle fonction de correspondance est la plus performante. Ainsi chaque fois qu'une requête est formulée, le système détermine son type, et applique la stratégie la mieux adaptée qui lui a été associée à l'avance.

Cette approche s'est révélée infructueuse [MW86], sans doute parce que la requête ne recèle pas les informations qui permettent de distinguer de manière assez fiable des types de requêtes susceptibles de préférer l'une ou l'autre des stratégies prédéfinies. Nous voyons une deuxième raison à cet échec : les stratégies prédéfinies

sont choisies en fonction des formalismes de représentation qui permettent de définir telle ou telle fonction de correspondance, et pas par une approche conceptuelle des variations du concept de pertinence en liaison avec les facteurs de pertinence utilisateur.

2.2.2 Associer stratégies et typologie

En réponse à cet échec, un autre type d'approche est proposé qui définit non plus des types de requêtes selon des critères syntaxiques, mais des types d'utilisateurs, ou des types d'utilisation du système.

À chaque type est associée une stratégie de recherche, dont on peut, cette fois, argumenter de l'adéquation en fonction de la typologie choisie [CDBK86, Sma94].

Cependant, les stratégies appliquées se aux hypothèses des modèles classiques de pertinence système où les objets à mettre en correspondance sont les thèmes ou les attributs externes à apparier. Ainsi les diverses stratégies possibles relèvent de variations sur les interprétations des opérateurs booléens ou des liens sémantiques d'un thésaurus. Ces variations, de nature formelle, sont souvent éloignées de celles que suggèrent les facteurs de pertinence, et le choix des associations est difficile à argumenter. Smaïl [Sma94] propose de remédier à ce problème au moyen d'un processus d'apprentissage des associations.

Ces approches sont orientées vers le système dans la mesure où l'on définit à l'avance un certain nombre de fonctions de correspondance, puis on cherche à déterminer les circonstances dans lesquelles on peut appliquer ces stratégies. Cette dernière étape constitue la pierre d'achoppement de ces techniques pour deux raisons :

- D'une part lors de la conception du système, la distance entre la dimension formelle des fonctions de correspondance et la typologie très qualitative à laquelle il faut les associer, pose problème.
- D'autre part, lors de l'utilisation du système, l'utilisateur doit associer sa situation de recherche, voire ses caractéristiques intrinsèques, à un élément d'une typologie prédéfinie, ce qui est difficile en soi. De plus la distance évoquée ci-dessus entre la typologie et la stratégie qui lui est associée compromet la clarté du dialogue entre l'utilisateur et le système.

2.2.3 Modéliser l'utilisateur

Un autre type d'approche, orientée vers l'utilisateur, consiste à élaborer un modèle d'utilisateur, c'est-à-dire un ensemble de connaissances capables de raisonner

sur l'utilisateur et son comportement en matière de recherche d'information. Ce modèle a pour but de fournir au système un certain nombre d'informations relatives à l'utilisateur afin de compléter la requête.

Daniels [Dan86] a synthétisé les travaux relatifs aux modèles d'utilisateur, et conclut de manière pessimiste sur la faisabilité d'une approche fondée sur une modélisation *a priori* des utilisateurs de systèmes de recherche d'information. En effet si ce type d'approche se montre efficace dans le cadre d'autres applications, la problématique de recherche d'information ne permet pas d'imposer à l'utilisateur des contraintes fortes, en particulier sur la nature de la tâche à accomplir.

2.2.4 Conclusion

Le choix de la stratégie à appliquer relève de la compétence de l'utilisateur lui-même, plutôt que d'une association prédéfinie par le système. Cette tâche peut difficilement être modélisée pour être prédite du fait de l'étendue des connaissances qu'elle implique.

Ces solutions constituent des palliatifs à une prise en compte plus globale de la nécessité de l'adaptation du système à la situation de recherche d'information.

2.3 Paramétrisation logique de la correspondance

Les fonctions de correspondance des divers systèmes existants reposent pour la plupart sur le choix d'un formalisme qui paraît adapté pour modéliser la pertinence. Leur abondance et leur diversité amène à s'interroger sur ce qui les différencie. Il est fréquent que les performances des systèmes qui les implémentent soient statistiquement équivalentes selon le paradigme de Cranfield. Mais quand cela n'est pas le cas, on peut s'interroger sur les raisons qui contribuent à rendre un système meilleur qu'un autre.

Est-ce parce que la fonction de correspondance modélise mieux la pertinence ? Ou est-ce l'association de la fonction de correspondance et du modèle d'indexation qui est plus harmonieuse ? Ou bien encore est-ce que l'ensemble du système est simplement très bien adapté au traitement des requêtes résolues de la collection test ?

2.3.1 Approches axiomatiques de la pertinence

Le besoin d'évaluer séparément la fonction de correspondance d'un système est issu de ces interrogations. Un certain nombre de travaux étudie les propriétés axiomatiques des fonctions de correspondance existantes dans le but de les comparer entre elles *a priori*, et de manière qualitative.

Propriétés logiques structurelles

Wong et Yao [WY89, Yao95] les premiers ont posé le problème suivant. Considérant qu'une fonction de correspondance f doit ordonner un ensemble de documents selon les *préférences d'un utilisateur* se trouvant dans une certaine situation de recherche d'information, quelles sont les propriétés que f doit ou peut vérifier vis-à-vis de l'ordre représentant les préférences de l'utilisateur sur les documents? La fonction de correspondance est vue dans ce contexte comme une fonction ordonnatrice (*ranking function*).

Ces travaux permettent de savoir si une fonction de correspondance quantitative linéaire donnée est compatible avec un certain nombre d'axiomes sur les structures de préférence d'un document sur un autre, telles qu'un utilisateur est susceptible de les concevoir effectivement. Ainsi un pont est jeté entre un concept relevant de la problématique de recherche d'information du point de vue de l'utilisateur, *c'est-à-dire*, la structure des préférences sur les documents, et la fonction ordonnatrice qui modélise ces préférences dans un système.

Propriétés logiques sémantiques

Les travaux de Huibers et Bruza [BH94, Hui96] ont permis d'identifier un ensemble de propriétés logiques fréquemment vérifiées par les fonctions de correspondance. Ainsi l'on dispose d'un ensemble d'*axiomatisations possibles de la relation d'«à-propos»*, sur lequel la comparaison de fonctions de correspondance est possible. Par exemple, Huibers et Bruza [BH95] ont montré que la relation d'«à-propos» est non monotone pour la composition.

Dans tous ces travaux, les règles qui sont identifiées correspondent à des propriétés fondamentales de la pertinence système : ce sont des propriétés structurelles et sémantiques indépendantes du domaine. On peut s'interroger sur l'intérêt d'introduire ces propriétés comme paramètre de la pertinence système au sein d'un même système de recherche d'information.

De plus, il paraît difficile de paramétrer la pertinence système en fonction des propriétés logiques générales telles qu'elles sont abordées dans ces travaux. En effet, les axiomes qui définissent les propriétés d'un système logique ne peuvent pas être aisément manipulés de manière modulaire, car les propriétés qu'un axiome confère à un système dépendent intimement du système auquel il est ajouté.

2.3.2 Approche abductive de la pertinence

Müller et Thiel [MT94] montrent que le mécanisme d'abduction est bien adapté pour représenter plusieurs sémantiques pour une même expression syntaxique. Ce mécanisme est utilisé dans la conception d'un modèle de recherche d'information où

l'utilisateur intervient dans le mécanisme de preuve, donnant ainsi des informations sémantiques qui complètent l'expression syntaxique de sa requête [TGMS95, MK95, TM96].

Mécanisme d'abduction

Une logique abductive est définie de la façon suivante : étant donné une théorie $\mathcal{T}$ et une formule ω qui doit être expliquée en termes de $\mathcal{T}$, l'abduction fournit un ensemble d'hypothèses ϕ tel que $\mathcal{T} \cup \phi \vdash \omega$.

Une théorie $\mathcal{T}$ est un ensemble construit sur un langage logique du premier ordre dont les constantes sont les termes d'indexation. Un ensemble de prédicats s'appliquent à ces constantes ou à des variables sur ces constantes. Un ensemble de règles sont disponibles, de la forme $p_1 \wedge \dots \wedge p_k \rightarrow a_1 \wedge \dots \wedge a_n$, où les p_i et les a_i sont des prédicats, la partie de gauche de la règle pouvant être vide.

Modèle abductif de correspondance

Si l'on prend $\omega = \text{Requête}$, où *Requête* est la description d'un concept que l'utilisateur recherche sous la forme d'une combinaison d'éléments de $\mathcal{T}$, alors le processus de correspondance selon le mécanisme abductif peut s'écrire comme suit : $\mathcal{T} \cup \text{Hypothèses} \vdash \text{Requête}$, qui se lit : «ajouter des hypothèses à la théorie pour qu'elle satisfasse la requête»¹.

La requête est donnée comme une expression intensionnelle du besoin d'information, sous la forme d'une formule quantifiée existentiellement. Reprenons un exemple donné dans [MK95] concernant un corpus de reproductions de peintures modernes. Pour un besoin comme «Que savons-nous du XX^e siècle ; dans quels pays trouvons-nous de l'art abstrait», la requête est formulée comme suit : $R = \exists C : \text{pays}(C)$, à-propos(art abstrait), profession(peintre), à-partir-de-l'année(1900).

Le mécanisme d'abduction cherche à combiner les attributs de cette requête de toutes les manières autorisées par la théorie, et il fournit un ensemble d'hypothèses comme par exemple : $H_1 = \exists A, B : \text{artiste}(A), \text{supérieur}(B, 1900)$, qui sont établies grâce à certaines règles de $\mathcal{T}$, en limitant l'abduction à l'ajout de prédicats simples.

Une hypothèse est donc une expression intensionnelle qui décrit une explication possible de la requête, c'est-à-dire une interprétation de la requête en termes de pertinence. Les documents qui sont des extensions de cette hypothèse seront considérés comme pertinents pour la requête, dans l'interprétation donnée par l'hypothèse.

1. Ce principe n'est pas sans rappeler les modélisations de Nie [Nie90] et Chevallet [Che92] qui utilisent le formalisme des logiques modales pour produire une sémantique analogue, mais moins propice à la génération d'un dialogue que celle de Müller et Thiel.

Interaction et dialogue

Chaque hypothèse constitue donc une interprétation différente de la pertinence pour la requête. Cette interprétation est traduite sous la forme d'un ensemble de règles utilisées et de prédicats obtenus par abduction.

L'utilisateur choisit l'interprétation qui comprend les règles qu'il estime valides. Alors, les règles utilisées pour cette interprétation sont activées, et l'extension de l'hypothèse est calculée en poursuivant l'abduction jusqu'au bout pour arriver aux documents.

Conclusion Ce dernier modèle permet à l'utilisateur de jouer sur le paramètre des règles décrivant les connaissances du domaine : en choisissant une hypothèse (une interprétation de sa requête), il détermine les connaissances qu'il considère comme valides pour l'inférence.

Par le biais d'une théorie qui comprend des règles, certaines données de niveau supérieur au niveau des termes, en l'occurrence les connaissances sur les termes d'indexation et les prédicats, sont ramenées au niveau des données manipulables, par l'intermédiaire de l'association de la requête à son interprétation. L'interprétation s'exprime sur le même vocabulaire de base que celui de la requête, les informations supplémentaires s'exprimant sous la forme de connexions entre ces données simples selon les flèches d'un graphe. S'il est probable que le détail de la sémantique des interprétations échappe à l'utilisateur, il est aussi probable qu'étant donné la simplicité du langage des interprétations, l'utilisateur élabore une image mentale de ce qu'elles représentent. Il serait intéressant d'étudier la cohérence de l'interprétation que l'utilisateur fait de ces flèches.

Chapitre 3

Résumé et objectifs

Les deux études relatives aux facteurs de pertinence (chapitre 1) nous ont permis de montrer les possibilités et les limites de la paramétrisation de la correspondance et de motiver un ensemble de *principes de conception* qu'il est souhaitable d'appliquer dans la conception des systèmes de recherche d'information. Nous faisons l'hypothèse qu'une amélioration des performances des systèmes de recherche d'information passe par l'adaptation du système à la situation de recherche courante, et que les compétences de l'utilisateur sont indispensables pour mener à bien cette adaptation. Par conséquent l'interaction doit faciliter les tâches de l'utilisateur pour qu'il apporte la meilleure contribution possible à l'adaptation.

Les méthodes existantes d'adaptation des systèmes nous permettent d'identifier les problèmes rencontrés (chapitre 2). Le système, pour s'adapter, doit obtenir des informations détenues par l'utilisateur, qui complètent le niveau de formulation du problème d'information qu'exprime la requête. Nous savons qu'il est impossible pour le système de produire lui-même de telles informations à partir de la seule requête. Par ailleurs, quand un ensemble de fonctions de correspondance est défini à un niveau formel comme un ensemble de stratégies orientées vers le système, il est difficile d'une part de déterminer à l'avance les situations de recherche dans lesquelles une stratégie sera meilleure qu'une autre, et d'autre part d'obtenir de l'utilisateur qu'il choisisse lui-même une stratégie à bon escient, en raison de la formulation orientée système des stratégies.

Nous choisissons de définir des éléments de stratégies appelés *critères élémentaires de pertinence* dans des termes aussi proches que possible de ceux dans lesquels se formulent les critères de pertinence utilisateur. Ce principe de modélisation a pour but d'évaluer la capacité d'un utilisateur à exprimer des informations relatives aux critères de pertinence en complément des termes de la requête.

Les critères élémentaires de pertinence constituent les briques de base du *schéma de pertinence* qui organisent l'ensemble des données intervenant dans l'interaction. Pour permettre cette formulation étendue du problème d'information, nous conce-

vons la partie interactive du système de façon à favoriser et faciliter cette formulation, par l'intermédiaire des *classes de pertinence*, qui sont définies en relation directe avec le schéma de pertinence.

Dans la partie qui suit, nous définissons un modèle conceptuel de la pertinence système fondé sur le concept de schéma de pertinence en accord avec les principes de conception énoncés. Ce modèle établit la relation entre les éléments de l'interaction d'une part (schéma de pertinence et classes de pertinence) et les fonctions internes du système (correspondance et mise en forme du résultat).

Nous procédons dans la partie IV à la formalisation de ce modèle qui nous permet de spécifier en détail les propriétés de la visualisation des documents retrouvés selon les classes de pertinence.

Partie III

Modèle conceptuel

Such research might further our understanding of what clues should be presented to users of computerized information retrieval systems, and perhaps suggest areas in which the retrieval mechanism itself could be extended beyond subject matching. [...]

One question that should be asked is whether users could have predicted their responses before examining these materials. In other words, are users able to specify the criteria they desire on an a priori basis, and to what extent? Certainly before there can be any discussion of changing retrieval mechanisms to include user-defined criteria, we should determine whether users can actually identify those criteria as part of the search request.

[Bar94], p. 158.

Cette partie présente le modèle conceptuel que nous proposons pour la pertinence en recherche d'information. Elle présente de manière informelle les concepts qui seront définis formellement dans la partie IV, en argumentant nos choix.

Ce modèle inclut :

- la définition d'un ensemble de fonctions de correspondance définies par une paramétrisation de la pertinence système selon des critères élémentaires de pertinence, dont le pendant dans l'interface est le schéma de pertinence ;
- la définition de la dimension pragmatique de l'interrogation par l'intermédiaire de la mise en forme des résultats, dont le pendant dans l'interface est l'ensemble des classes de pertinence.

Dans le chapitre 1, nous présentons la notion de critère abstrait de pertinence à partir de laquelle sont construits les critères élémentaires de pertinence, et montrons quelle relation ils entretiennent avec le modèle d'indexation.

Puis, dans le chapitre 2, nous discutons du choix d'une sémantique de combinaison des critères élémentaires, pour l'interrogation. Nous définissons ensuite la notion de statut de critère qui, associés aux critères de pertinence, donnent la sémantique de leur combinaison dans le schéma de pertinence. Les statuts déterminent la dimension pragmatique de l'utilisation du système, au travers la visualisation des documents retrouvés selon des classes de pertinence.

Enfin nous montrons dans le chapitre 3 comment l'interaction ainsi définie applique les principes énoncés dans la partie précédente.

Chapitre 1

Critères élémentaires abstraits

Dans ce chapitre nous décrivons les éléments constitutifs du schéma de pertinence en argumentant nos choix. Nous établissons également les notations qui sont utilisées par la suite.

Dans un premier temps, nous définissons les critères élémentaires abstraits de pertinence et leur relation avec l'indexation. Les critères élémentaires de pertinence se définissent en fonction de ces critères abstraits, et constituent les briques de base de l'interrogation. Après avoir défini l'évaluation de la vérification d'un critère élémentaire dans un document, nous évoquons une propriété des critères abstraits qui est susceptible de se présenter fréquemment dans les applications : une relation d'ordre entre critères abstraits. Enfin, nous montrons comment l'ensemble des critères abstraits de pertinence d'un modèle de pertinence système joue un rôle d'interface entre l'indexation et l'interrogation. Dans un second temps, nous définissons les critères de dimension comme une combinaison d'un ensemble de critères élémentaires.

1.1 Critère élémentaire, critère élémentaire abstrait

Pour concevoir un système de recherche d'information pour un corpus donné, il faut connaître les besoins des utilisateurs potentiels en matière d'interrogation sur ce corpus : quel est l'éventail des critères que les utilisateurs emploient pour formuler un problème d'information qu'ils souhaitent résoudre à partir de ce corpus ?

1.1.1 Critère élémentaire

Voici quelques exemples de critères qu'un utilisateur est susceptible d'employer pour formuler un problème d'information :

$$k_1 = \textit{aborde-le-thème-fleur}$$

$$k_2 = \textit{aborde-le-thème-jardinage}$$

$$k_3 = \textit{aborde-un-thème-spécifique-à-jardinage}$$

$$k_4 = \textit{comprend-une-zone-de-couleur-bleu}$$

$$k_5 = \textit{est-paru-en-1992}$$

$$k_6 = \textit{est-paru-après-1992}$$

$$k_7 = \textit{est-disponible-en-rayon}$$

Nous appelons ces critères simples des *critères élémentaires* et nous leur associons le symbole

$$k$$

1.1.2 Critère élémentaire abstrait

Les critères élémentaires k_1 et k_2 sont apparentés car il s'agit pour l'un que le document traite du thème *fleur*, et pour l'autre qu'il traite du thème *jardinage*. L'ensemble des critères k qui partagent cette caractéristique peuvent être notés $k = \xi[t]$, où ξ désigne un même *critère élémentaire abstrait*, noté $\xi_0 = \textit{aborde-le-thème}$, et où t est un terme désignant un thème.

Ainsi les critères élémentaires k_1 et k_2 s'écrivent aussi

$$k_1 = \textit{aborde-le-thème}[\textit{fleur}]$$

$$k_2 = \textit{aborde-le-thème}[\textit{jardinage}]$$

De la même façon, les critères élémentaires k_3 à k_7 donnent lieu à la définition des critères élémentaires abstraits ξ_3 à ξ_7 comme suit.

$$\xi_3 = \textit{aborde-un-thème-spécifique-à}$$

$$\xi_4 = \textit{comprend-une-zone-de-couleur}$$

$$\xi_5 = \textit{est-paru-en}$$

$$\xi_6 = \textit{est-paru-après}$$

$$\xi_7 = \textit{est-disponible-en-rayon}$$

Le critère k_7 s'écrit alors $k_7 = \textit{est-disponible-en-rayon}[\textit{oui}]$.

Nous associons aux critères élémentaires abstraits le symbole

$$\xi$$

D'une manière générale, un critère élémentaire k peut se noter comme l'association d'un critère élémentaire abstrait ξ et d'un terme t :

$$k = \xi[t]$$

1.1.3 Modèle de document

La définition d'un critère élémentaire abstrait requiert une convention quant au modèle de document que nous considérons.

Nous considérons qu'un document indexé est un ensemble de couples **Attribut**[valeur], où un même attribut peut intervenir plusieurs fois avec des valeurs distinctes. Un exemple de document indexé est donné dans la table III.1.1.

Thème-abordé[fleur]
Thème-abordé[jardinage]
Thème-abordé[engrais]
Couleur-présente[bleu]
Taux-de-tournures-passives[.2]
Date-de-parution[1992]
Titre[Mon jardin]
Disponibilité-en-rayon[oui]
Disponibilité-chez-l'éditeur[non]

TAB. III.1.1 – *Exemple de document indexé*

L'idée de décrire un document en associant une valeur (ou plusieurs valeurs) à un attribut, s'apparente à la méthode de classification de documents par facettes [PD91], qui constitue une alternative plus flexible et plus précise aux schémas de classification énumératifs traditionnellement utilisés dans les bibliothèques. Le processus de classification par facette permet une classification souple, précise, et aisément extensible dans le cas d'ajout de nouveaux documents. En revanche il présente l'inconvénient de requérir l'intervention manuelle d'un indexeur.

Le principe de décrire un document selon des facettes répond bien à notre besoin de décrire plusieurs aspects d'un même document pour permettre d'appliquer des types de critères variés pour retrouver un document, en fonction de la situation de recherche.

En outre, ce choix est motivé par la simplicité et la souplesse du formalisme, dans lequel il est possible d'instancier des langages d'indexation plus classiques comme l'indexation vectorielle¹, même s'il s'agit d'une forme dégénérée où une seule facette apparaît : celle du contenu sémantique.

Grâce à ce choix de langage d'indexation, nous conservons une certaine généralité dans nos propos, en évitant de spécifier les propriétés du langage d'indexation au-delà de ce qui est nécessaire pour développer notre modèle de pertinence système. C'est au niveau des critères abstraits que ces propriétés sont prises en compte.

1.1.4 Évaluation d'un critère élémentaire

L'évaluation d'un critère élémentaire k dans un document D est une fonction qui associe une valeur v de réalisation du critère dans le document D , ce que l'on note

$$k(D) = v$$

Par exemple, l'évaluation booléenne du critère *aborde-le-thème*[fleur] à vrai sera notée

$$\text{aborde-le-thème}[\text{fleur}](D) = \text{vrai}$$

Étant donné un document D , nous pouvons définir l'évaluation booléenne du critère abstrait *aborde-le-thème*, comme suit

$$\text{aborde-le-thème}[t](D) = (\text{Thème-abordé}[t] \in D)$$

Il n'existe pas nécessairement une bijection entre les attributs d'indexation et les critères élémentaires abstraits.

Par exemple, l'évaluation de *aborde-un-thème-spécifique-à* s'exprime aussi en fonction de l'attribut *Thème-abordé*.

$$\text{aborde-un-thème-spécifique-à}[t](D) = (\exists \text{thème-abordé}[t'] \in D, \text{Spéc}(t', t))$$

où la relation *Spéc* (X, Y) est définie comme X est un fils de Y ou Y lui-même dans l'arbre des relations hiérarchiques entre les thèmes.

À l'opposé, l'évaluation du critère abstrait *traite-de-manière-exhaustive-de* peut faire appel à plusieurs attributs d'indexation.

$$\begin{aligned} \text{traite-de-manière-exhaustive-de}[t](D) = \\ (\text{Thème-abordé}[t] \in D) \wedge (\text{Genre}[\text{ouvrage-de-référence}]) \end{aligned}$$

1. Dans cet exemple, la valeur est un couple (terme, poids).

Si l'indexation est conforme au modèle vectoriel, le document indexé est noté comme suit

$$D = \{\text{poids-du-terme-}t_1[w_1], \dots, \text{poids-du-terme-}t_N[w_N]\}$$

où w_i est le poids associé au terme d'indexation t_i associé à la dimension i , et l'on pourra procéder à une estimation numérique du critère, comme suit

$$\text{aborde-le-thème}[t_i](D) = \cos([0, \dots, 0, 1, 0, \dots, 0], [w_1, \dots, w_N])$$

où le 1 est positionné à la i ème place du vecteur.

La sémantique de la valeur numérique obtenue dépend directement de la sémantique de l'indexation. Il peut s'agir d'une sémantique statistique, si les poids des termes sont calculés en fonction des mesures statistiques *tf* et *idf* (*term frequency* et *inverse document frequency*, [SM83]), ou bien d'une sémantique probabiliste si les poids résultent d'une indexation probabiliste.

1.1.5 Ordre sur les critères élémentaires abstraits

Certains critères abstraits sont naturellement organisés, en raison de leurs propriétés, selon une relation d'inclusion. Cette relation définit un ordre partiel sur les critères abstraits, car certains critères ne sont pas comparables.

Considérons les critères abstraits définis plus haut. $\xi_3 = \text{aborde-un-thème-spécifique-à}$ est inclus dans $\xi_0 = \text{aborde-le-thème}$, car quelque soit t , tout document vérifiant $\xi_3[t]$ vérifie aussi $\xi_0[t]$. De même, $\xi_5 = \text{est-paru-en}$ est inclus dans $\xi_6 = \text{est-paru-après}$.

Cette relation d'ordre n'implique pas une relation d'ordre en matière de degré de pertinence, comme c'est souvent le cas dans les modèles qui utilisent un thésaurus pour inférer un document à partir d'une requête.

Étant donné un critère abstrait ξ qui est inclus dans un autre ξ' , noté

$$\xi \prec \xi'$$

nous dirons que ξ est *plus strict* que ξ' , et que ξ' est *plus lâche* que ξ . Ainsi nous avons $\xi_3 \prec \xi_0$ et $\xi_5 \prec \xi_6$.

1.1.6 Choix des critères élémentaires abstraits pour une application donnée

Lors de la conception de l'interrogation, le choix des critères élémentaires abstraits est fortement couplé au choix du modèle d'indexation.

Pour le système de recherche d'images par le contenu QBIC^(TM) d'IBM par exemple, les caractéristiques indexées des images déterminent les trois critères d'interrogation offerts (pourcentage de couleur, répartition des couleurs, texture). Dans ce cas

l'interrogation du corpus s'effectue nécessairement par similarité avec une image-requête, car le problème d'information d'un individu peut rarement se définir selon des termes techniques tels que la répartition des couleurs des images recherchées. C'est au processus d'interrogation qu'il incombe d'effectuer cette ré-expression interne du problème en termes conformes au modèle d'indexation, au travers d'un critère élémentaire abstrait.

De façon générale, le choix des critères élémentaires abstraits implique un travail sémantique sur la dimension technique de l'indexation afin de produire les éléments de base de la sémantique d'interrogation destinée à l'utilisateur du système.

Le principe d'interaction *Comprendre* (principe 2) s'instancie dans cet aspect de la modélisation : la qualité du choix des critères élémentaires conditionne la qualité de l'interaction obtenue. Deux tendances antagonistes interviennent dans ce choix.

D'une part ce principe est d'autant mieux vérifié que les critères abstraits sont proches des critères naturellement formulés par la catégorie d'utilisateurs visée pour décrire leur problème d'information.

D'autre part, ce principe est d'autant moins bien vérifié que la sémantique des critères abstraits pour l'interrogation est éloignée de la réalité des calculs qui les sous-tendent. L'exemple suivant de critère thématique illustre ce point dans la mesure où les calculs trahissent le sens intuitif qui lui est associé. Considérons un système dont l'interface réclame un thème en guise de requête, et promet de retrouver les documents textuels abordant ce thème. Les calculs trahissent cette promesse si la notion de thème abordé par un document est implémentée comme une simple recherche dans les documents sur la chaîne de caractères entrée. Ce défaut introduit un biais dans l'utilisation du système, car il favorise l'élaboration d'une représentation mentale erronée de ce que réalise le système, et par conséquent du contenu du corpus, ce qui compromet le bon déroulement du processus itératif de recherche d'information.

Dans tous les systèmes, tôt ou tard l'utilisateur est confronté explicitement ou implicitement à la réalité du langage d'indexation. Nous considérons que ce phénomène gagne à être traité explicitement afin de pénaliser le moins possible la qualité de l'interaction. Cependant la dimension technique du calcul doit être cachée à l'utilisateur, car elle ne relève pas de ses compétences.

Pour appliquer le principe 2, un compromis consiste à définir des critères élémentaires abstraits *dont l'utilisateur peut évaluer lui-même la réalisation* dans les documents tels qu'ils lui sont présentés par le système.

1.2 Critère de dimension, critère de dimension abstrait

1.2.1 Présentation

Les critères élémentaires abstraits fournis dans l'interrogation peuvent être nombreux, mais ils peuvent souvent se grouper logiquement selon des *dimensions d'interrogation* qui correspondent à divers points de vue sur un document. Par exemple, étant donné un corpus de photographies, un utilisateur peut souhaiter formuler son problème d'information en termes des objets que l'on peut voir sur la photographie (Contenu thématique), en termes de Genre (portrait, paysage, scène de rue, etc.), en termes de Morphologie (lumière, prise de vue, etc.), en termes de Caractéristiques physiques (couleur, répartition des couleurs, etc.), ou encore selon les Attributs externes (auteur de la photographie, date de parution, etc.), ou en termes de Disponibilité du document réel (disponibilité en rayon, épuisement chez l'éditeur, etc.).

Une dimension d'interrogation constitue donc une rubrique de l'interrogation. Dans le prototype VIZ que nous présentons dans la partie V, l'interrogation est organisée selon six dimensions d'interrogation notées :

Contenu
Précisions
Connotation
Genre
Morphologie
Auteur

Si nous reprenons nos premiers exemples de critères élémentaires abstraits, les $\xi_1 = \textit{aborde-le-thème}$ et $\xi_3 = \textit{aborde-un-thème-spécifique-à}$ relèvent de la dimension Contenu thématique de l'interrogation. Le critère abstrait $\xi_4 = \textit{comprend-une-zone-de-couleur}$ relève de la dimension Caractéristiques physiques du document. Les critères abstraits $\xi_5 = \textit{est-paru-en}$ et $\xi_6 = \textit{est-paru-après}$ relèvent de la dimension des Attributs externes du document. Le critère abstrait $\xi_7 = \textit{est-disponible-en-rayon}$ relève de la dimension Disponibilité du document.

Nous associons aux dimensions d'interrogation le symbole

Dim

1.2.2 Définitions

Nous définissons ainsi la notion de *critère de dimension* comme le couple constitué du nom de la dimension d'interrogation concernée, et le tuple des critères élé-

mentaires relatifs à une même dimension d'interrogation².

$$K = (\text{Dim}, [k_1, \dots, k_n]) = (\text{Dim}, [\xi_1[t_1], \dots, \xi_n[t_n]])$$

Par exemple, si les trois critères élémentaires k_1 , k_3 et k_5 sont formulés, ils donnent lieu aux deux critères de dimension K_1 (dimension $\text{Dim}_1 = \text{Contenu thématique}$) et K_2 (dimension $\text{Dim}_2 = \text{Attributs-externes}$).

$$K_1 = (\text{Contenu thématique}, [\text{aborde-le-thème}[\text{fleur}], \\ \text{aborde-un-thème-spécifique-à}[\text{jardinage}]])$$

$$K_2 = (\text{Attributs externes}, [\text{est-paru-en}[\text{1992}]])$$

Nous associons aux critères de dimension le symbole

$$K$$

De la même façon qu'un critère élémentaire k est le résultat de l'application d'un critère abstrait ξ à un terme t ,

$$k = \xi[t]$$

un critère de dimension K est le résultat de l'application d'un critère de dimension abstrait Ξ , à l'ensemble des termes $t_1, \dots, t_n$.

Ainsi un critère de dimension est aussi noté

$$K = (\text{Dim}, \Xi[t_1, \dots, t_n]) = (\text{Dim}, [\xi_1[t_1], \dots, \xi_n[t_n]])$$

Nous reprenons ce point en détail lors de la discussion concernant la combinaison des critères élémentaires.

2. Nous raffinons cette définition dans le chapitre suivant.

Chapitre 2

Schéma de pertinence

Dans ce chapitre, nous présentons la notion de schéma de pertinence qui se situe au cœur de l'interrogation dans notre modèle. Un schéma de pertinence est une combinaison de critères élémentaires de pertinence.

Nous discutons d'abord de la sémantique à adopter pour la combinaison de critères élémentaires. Nous définissons ensuite la notion de statut de critère qui, appliqué aux critères élémentaires ou aux critères de dimension, permet de les paramétrer. Nous définissons ensuite le schéma de pertinence comme une combinaison de critères de dimension formés de critères élémentaires. La sémantique de cette combinaison est réglée par les statuts des critères. Nous montrons alors comment le schéma de pertinence est utilisé comme un contexte d'évaluation de la pertinence système d'un document. Enfin nous décrivons le principe selon lequel le schéma de pertinence permet, grâce aux statuts de critères, de définir des classes de pertinence qui seront utilisées pour la visualisation des documents retrouvés.

2.1 Combinaison de critères pour l'interrogation

Un critère élémentaire décrit une contrainte sur les caractéristiques des documents susceptibles d'être pertinents. Un utilisateur peut exprimer plusieurs critères élémentaires pour cerner son problème d'information ; l'ensemble de ces critères constitue le schéma de pertinence. Mais quelle est la sémantique de l'utilisation conjointe de ces critères ?

Dans la suite, nous argumentons et définissons la sémantique que nous attachons à cette combinaison.

2.1.1 Discussion

Quand un utilisateur formule deux critères élémentaires, il peut désirer les voir appliqués aux documents en conjonction, mais aussi en disjonction. Ainsi, une sémantique booléenne pour la combinaison de critères effectifs paraît naturelle. Cependant, deux arguments nous incitent à éviter ce type de formulation.

D'abord nous savons que la sémantique des connectives booléennes n'a pas de correspondant direct intuitif dans la langue naturelle, et conduit souvent les non spécialistes à des utilisations erronées [WB95, AK93, GDCG90, Bor84].

De plus, la notion de disjonction nous semble inadéquate pour la formulation de la sémantique d'un problème d'information. En effet, un utilisateur formule une disjonction de critères dans l'une ou l'autre des intentions suivantes :

- 1° il souhaite des documents vérifiant l'un ou l'autre des critères indifféremment ;
- 2° il ignore lequel de ces deux critères est le mieux adapté pour décrire son problème d'information.

Le OU indifférent

Considérons le premier cas. Si l'utilisateur trouvait intéressant d'obtenir des documents vérifiant les deux critères à la fois, il utiliserait une conjonction. Par conséquent, nous assimilons ce cas d'utilisation de la disjonction à un besoin de définir plusieurs problèmes d'information distincts en parallèle.

Nous préconisons de les traiter séparément, pour la raison suivante : notre processus de formulation étant itératif, il se peut que les modifications à apporter à la formulation de chacun des problèmes d'information pour les améliorer ne soient pas identiques, ni même compatibles entre elles. Par exemple, considérons la combinaison de critères effectifs *aborde-le-thème*[t_1] $\vee$ *aborde-le-thème* [t_2]. Si le corpus comprend de nombreux documents traitant de chacun de ces thèmes, parmi lesquels beaucoup ne sont pas pertinents, il peut arriver dans un corpus suffisamment hétérogène que les raisons sous-jacentes à la pertinence d'un document traitant de t_1 soient très différentes de celles sous-jacentes à la pertinence d'un document traitant de t_2 . Par conséquent l'utilisateur devra mener de front deux tâches dont les objectifs divergent, pour raffiner la formulation de son problème d'information en réaction aux réponses du système, ce qui est difficile en soi. De plus, d'un point de vue technique, la formulation devra reposer sur une structure logique plus complexe, où un parenthésage devra intervenir.

Le OU d'incertitude

Dans le second cas, la disjonction des critères traduit une incertitude de l'utilisateur quant à la sémantique de ses critères vis-à-vis du système. Par exemple, désirant des informations sur la logique du premier ordre et ignorant les capacités du système à établir la synonymie entre l'expression «logique du premier ordre» et l'expression «logique des prédicats», un utilisateur peut être enclin à introduire une disjonction entre ces deux termes. Mais ce réflexe relève d'un modèle mental du système conforme au principe d'appariement exact de chaînes de caractères. L'introduction de critères de pertinence est destiné à éviter ce type de formulation qui dépend des caractéristiques techniques internes de la fonction de correspondance.

Pour cet exemple, si l'on dispose d'un thésaurus comprenant les liens de synonymie $Syn(X, Y)$, l'utilisateur pourra utiliser le critère de pertinence $aborde-le-thème[t]$, défini comme

$$aborde-le-thème[t](D) = \exists \text{Thème-abordé}[t_D] \in D, (t = t_D \vee Syn(t, t_D))$$

ou au contraire le critère de pertinence $contient-la-chaîne[t]$ défini comme

$$contient-la-chaîne[t](D) = (t \text{ sous-chaîne de } C, \text{ avec } \text{Texte}[C] \in D)$$

Si l'on n'en dispose pas, il nous paraît meilleur de compléter le langage d'interrogation avec une interface permettant à l'utilisateur de spécifier les synonymes qu'il connaît d'un terme. De manière interne, ces synonymes pourront être ajoutés à la requête interne de façon à exprimer cette sémantique au mieux en fonction du modèle d'indexation et de la fonction de correspondance.

2.1.2 Notre choix

Étant donné le flou et les variations qui entourent le sens que les utilisateurs associent aux diverses connectives booléennes, nous préconisons, conformément au principe 2, d'établir la sémantique la plus simple possible en matière de combinaison des sous-expressions formulées par l'utilisateur pour décrire son problème d'information. En cela nous rejoignons la position exposée par Danilowicz [Dan94], qui suggère de restreindre la sémantique de la combinaison des éléments de requête à la conjonction.

Finalement, un schéma de pertinence est, du point de vue de la correspondance, un ensemble de critères élémentaires combinés en conjonction. Des variations sur la sémantique de cette combinaison sont permises mais traitées séparément, par l'intermédiaire de *statuts* associés aux critères. Dans le cadre de notre modèle, où l'interaction joue un rôle essentiel, la disjonction peut jouer un rôle pragmatique plutôt que sémantique, pour gérer les situations de recherche problématiques, comme nous le voyons dans la suite.

2.2 Statuts de critères

Considérant qu'un schéma individuel d'estimation de la pertinence intègre une notion de priorités entre critères (voir Howard [How94]), nous proposons de permettre à l'utilisateur d'exprimer ces priorités en associant aux critères des *statuts*. Pour limiter la complexité de la formulation de ces statuts, nous la réduisons à deux valeurs : *obligatoire* et *optionnel*.

Ainsi, un critère associé à un statut obligatoire doit être vérifié dans les réponses données par le système, alors qu'un critère associé à un statut optionnel peut ne pas y être vérifié.

De manière générale, l'association d'un critère élémentaire k à un statut s est notée comme un couple

$$(k, s)$$

Du fait que les statuts ne peuvent prendre que deux valeurs, nous adoptons une notation plus compacte des statuts en décorant les critères par un point pour les critères optionnels, et une barre pour les critères obligatoires.

Ainsi un critère élémentaire k dont le statut est obligatoire est noté

$$(k, \text{obligatoire}) = \bar{k}$$

et un critère élémentaire k dont le statut est optionnel est noté

$$(k, \text{optionnel}) = \dot{k}$$

Pour les critères de dimension, les mêmes notations s'appliquent : $\bar{K}$ dénote l'association du critère de dimension K à un statut obligatoire, et $\dot{K}$ l'association du critère de dimension K à un statut optionnel.

Comme nous le montrons dans la partie suivante, la sémantique des statuts se définit en termes de visualisation des documents retrouvés par l'intermédiaire de classes de pertinence, elles-mêmes définies selon une interprétation booléenne des statuts.

2.3 Schéma de pertinence

2.3.1 Structure hiérarchique du schéma de pertinence

Un schéma de pertinence S est constitué de critères de dimension, eux-mêmes constitués de critères élémentaires de pertinence regroupés dans les critères de dimension.

À chaque critère de dimension est associé un statut, et la combinaison de ces critères de dimension avec statut définit un schéma de pertinence

$$S = [(K_1, s_1), \dots, (K_n, s_n)]$$

où n est le nombre de critères de dimension du schéma de pertinence.

Les critères élémentaires $k_{i,j}$ avec leur statut $s_{i,j}$ sont regroupés selon les critères de dimension qui s'échelonnent de 1 à n . Ainsi, le i ème critère de dimension est la combinaison de ses $i.n_i$ critères élémentaires avec statut, noté¹.

$$K_i = (\text{Dim}, [(k_{i.1}, s_{i.1}), \dots, (k_{i.n_i}, s_{i.n_i})])$$

Ces deux niveaux de critères qui structurent le schéma de pertinence peuvent aisément être portés à trois ou quatre si une application particulière le nécessite², mais il nous semble que deux niveaux sont suffisants dans la plupart des cas, ce pourquoi nous développons notre formalisation sur ces deux niveaux : critères de dimension et critères élémentaires.

Enfin, nous prendrons par défaut la convention qui consiste, pour un critère de dimension donné K_i , à placer en tête les $\dot{n}_i$ critères élémentaires optionnels. Ainsi, le critère de dimension K_i sera noté

$$K_i = [(k_{i.1}, \text{optionnel}), \dots, (k_{i.\dot{n}_i}, \text{optionnel}), \\ (k_{i.\dot{n}_i+1}, \text{obligatoire}), \dots, (k_{i.n_i}, \text{obligatoire})]$$

ce qui donne en notations compactes

$$K_i = [\dot{k}_{i.1}, \dots, \dot{k}_{i.\dot{n}_i}, \bar{k}_{i.\dot{n}_i+1}, \dots, \bar{k}_{i.n_i}]$$

De même, pour un schéma S , on place en tête les $\dot{n}$ critères de dimension optionnels.

$$S = [\dot{K}_1, \dots, \dot{K}_{\dot{n}}, \bar{K}_{\dot{n}+1}, \dots, \bar{K}_n]$$

La figure III.2.1 donne une représentation graphique d'un exemple de schéma de pertinence sous la forme d'un arbre étiqueté. Les deux niveaux des critères élémentaires et des critères de dimension y sont identifiés. Les statuts des critères de dimension apparaissent comme des étiquettes des nœuds racines des critères de dimension.

Pour éviter la multiplication des niveaux, nous ramenons graphiquement au même niveau le nom de la dimension d'un critère de dimension Dim_i et le statut de ce critère de dimension s_i . Ainsi on peut noter ces étiquettes de manière plus compacte en surmontant le nom de la dimension d'une barre pour signifier un statut de critère obligatoire, et d'un point pour signifier un statut de critère optionnel.

L'exemple de schéma de la figure III.2.2 concrétise ces notations compactes.

1. La définition que nous avons donnée dans la partie précédente est donc complétée ici.

2. En particulier si les axes d'indexation sont eux-mêmes structurés; cependant la complexité que cette structure introduit dans l'interface en entrée de notre modèle nous semble peu souhaitable.

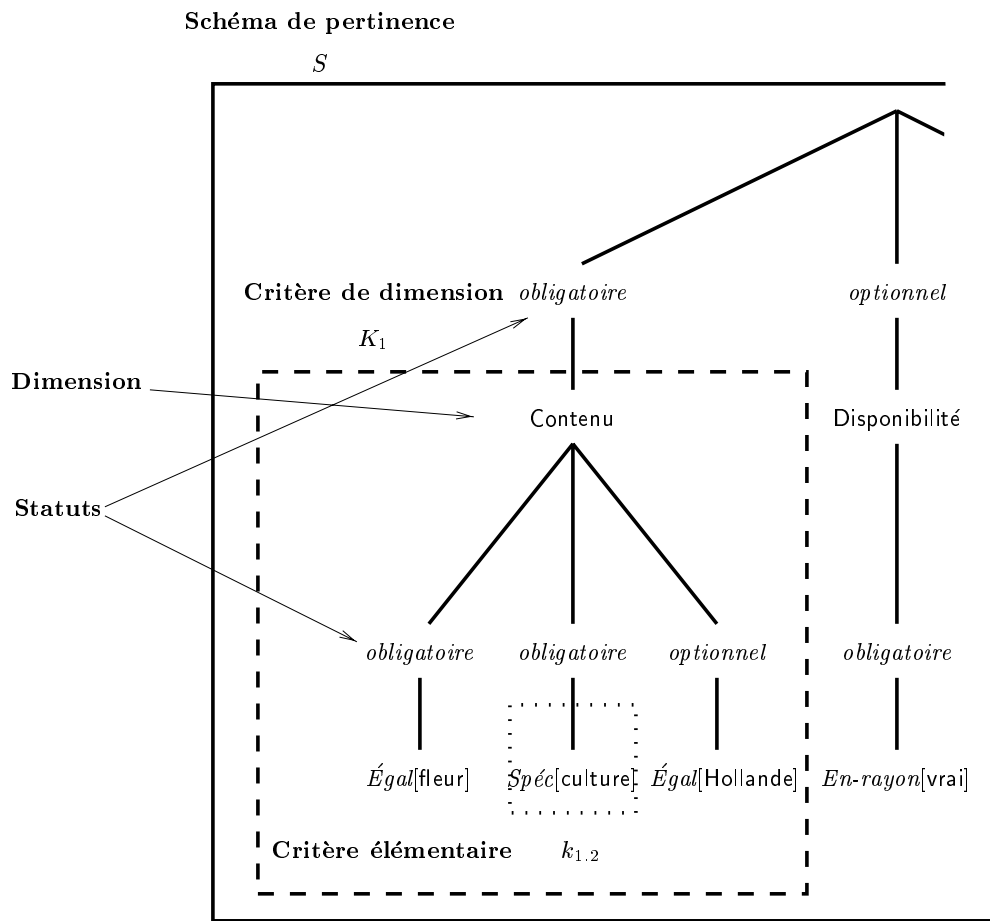


FIG. III.2.1 – Notations pour le schéma de pertinence et ses composants

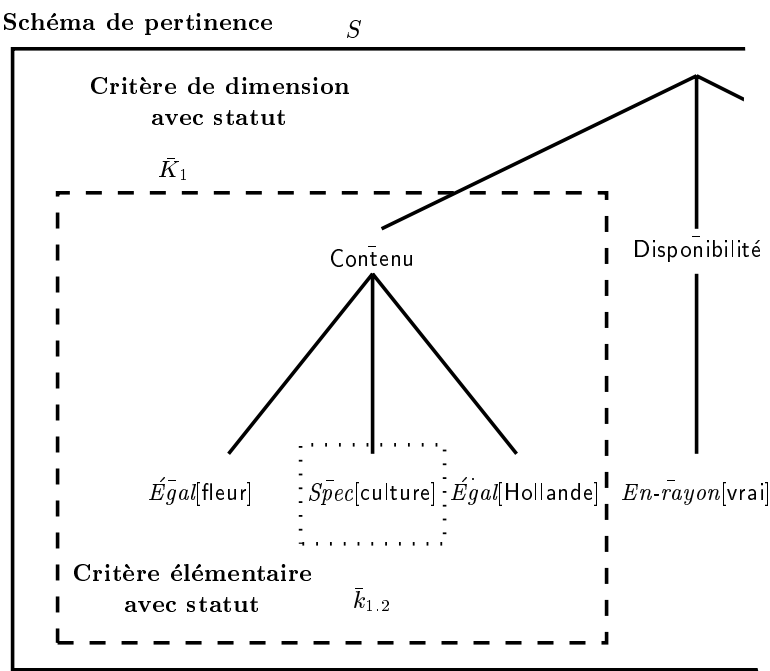


FIG. III.2.2 – Exemple de schéma de pertinence en notations compactes

2.3.2 Rôle des éléments du schéma de pertinence

L'abstraction est une opération algébrique qui consiste à ôter du schéma les termes de la requête. Nous avons déjà évoqué cette notion d'abstraction, qui permet d'obtenir un critère élémentaire abstrait à partir d'un critère élémentaire, ou un critère de dimension abstrait à partir d'un critère de dimension.

Ainsi nous pouvons définir de manière analogue un schéma de pertinence abstrait Σ , comme un schéma de pertinence dont on a ôté les termes. Nous obtenons ainsi d'une part, la liste R des listes de termes $T_i = [t_{i.1}, \dots, t_{i.n_i}]$ associées à chaque dimension Dim_i .

$$R = [[t_{1.1}, \dots, t_{1.n_1}], \dots, [t_{n.1}, \dots, t_{n.n_n}]]$$

et d'autre part le schéma de pertinence abstrait Σ , qui s'applique comme un contexte à R (figure III.2.3). Le schéma de pertinence S se définit alors comme l'application du schéma abstrait Σ à R .

$$S = \Sigma[R]$$

Le contexte qui est appliqué à la requête typique, fournit d'une part l'interprétation sémantique des termes grâce aux critères abstraits de pertinence, et d'autre part l'interprétation pragmatique des termes grâce aux statuts de critères. Cette dernière interprétation pragmatique est destinée à structurer la visualisation des documents retrouvés.

Le schéma de pertinence abstrait n'a de sens qu'une fois appliqué à la liste de termes qu'il interprète. Cette abstraction que nous n'évoquons ici que pour montrer en quoi le schéma de pertinence généralise la requête classique constituée d'un ensemble de termes.

2.4 Évaluation de la pertinence système

2.4.1 Relation de pertinence

Le résultat de l'évaluation de la pertinence système d'un document pour un schéma de pertinence peut être une simple valeur atomique, comme un booléen (vrai pour «pertinent», faux pour «non pertinent»), ou un réel compris entre 0 et 1 (pour traduire un «degré de pertinence»).

Cependant, on peut aussi traduire de manière qualitative cette évaluation, en construisant une valeur symbolique reflétant la nature des critères vérifiés dans le document. C'est ce que nous appelons la *relation de pertinence* entre le schéma de pertinence S et le document D , que nous notons $RP(S, D)$. Nous définissons cette relation dans la partie IV.

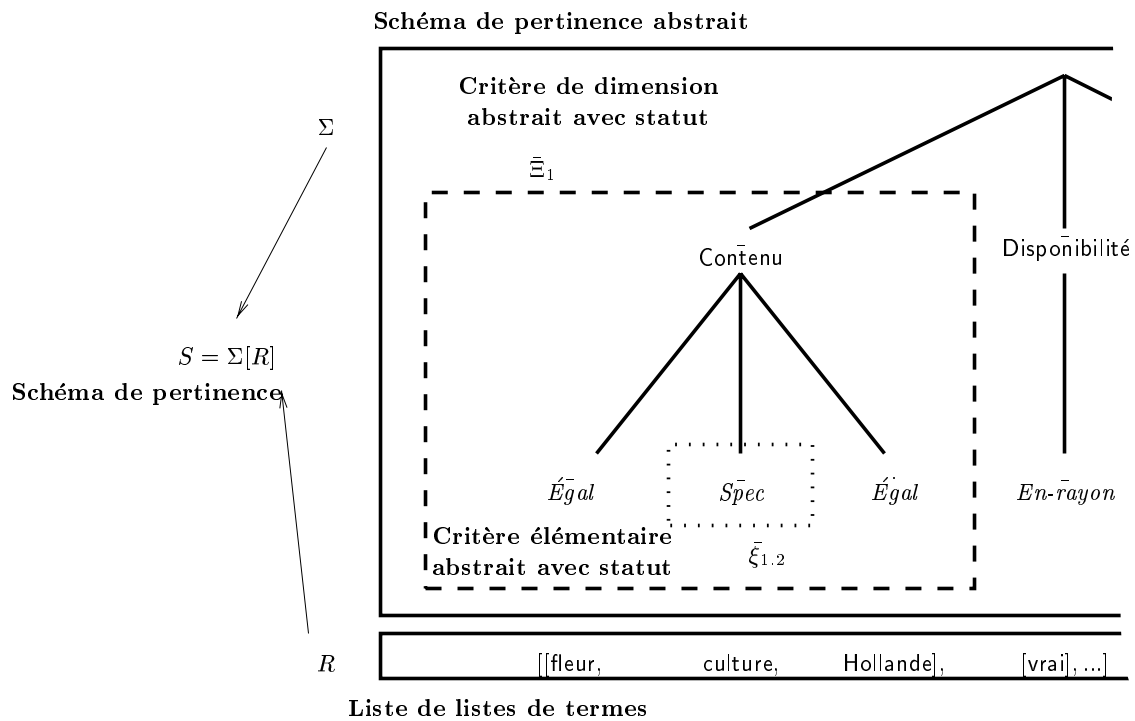


FIG. III.2.3 – Notations compactes pour le schéma de pertinence abstrait et ses parties

2.4.2 Ensemble des documents retrouvés

L'ensemble des documents retrouvés pour le système est l'ensemble des documents du corpus qui vérifient tous les critères de dimension obligatoires du schéma de pertinence. Un document vérifie un critère de dimension s'il vérifie tous ses critères élémentaires obligatoires. Il s'agit donc d'une combinaison conjonctive des critères, où les critères associés à des statuts optionnels sont ignorés.

Étant donné un schéma de pertinence

$$S = [\dot{K}_1, \dots, \dot{K}_{\dot{n}}, \bar{K}_{\dot{n}+1}, \dots, \bar{K}_n]$$

avec pour tout $i \in [1..n]$

$$K_i = (\text{Dim}_i, [\dot{k}_{i,1}, \dots, \dot{k}_{i,\dot{n}_i}, \bar{k}_{i,\dot{n}_i+1}, \dots, \bar{k}_{i,n_i}])$$

et le corpus $\mathcal{E}_D$, l'ensemble des documents retrouvés est défini comme suit :

$$S(\mathcal{E}_D) = \{D, \quad \forall \bar{K}_i, \forall \bar{k}_{i,j} \in \bar{K}_i, k_{i,j}(D) \text{ est vérifié} \}$$

L'expression $k_{i,j}(D)$ est vérifié traduit une évaluation binaire, et non plus symbolique, de la pertinence système, qui sera définie précisément dans la partie suivante.

2.5 Classes de pertinence et visualisation

Soit un schéma de pertinence

$$S = [\dot{K}_1, \dots, \dot{K}_{\dot{n}}, \bar{K}_{\dot{n}+1}, \dots, \bar{K}_n]$$

comprenant n critères de dimension parmi lesquels les $\dot{n}$ premiers sont optionnels et les $n - \dot{n}$ autres sont obligatoires.

L'ensemble des *classes de pertinence* réalise une classification des documents retrouvés selon toutes les combinaisons de vérification des critères de dimension optionnels. Chaque classe de pertinence regroupe les documents qui partagent un certain sous-ensemble des critères de dimension du schéma de pertinence.

L'ensemble des combinaisons est l'ensemble des injections $\iota : [1..\dot{n}] \rightarrow \{0, 1\}$, notées $\iota_{b_1 \dots b_{\dot{n}}}$, avec $\forall b_i, b_i \in \{0, 1\}$.

La classe de pertinence relative à la combinaison $\iota_{b_1 \dots b_{\dot{n}}}$ est notée $C_{\iota_{b_1 \dots b_{\dot{n}}}}$, et est définie comme l'ensemble des documents qui vérifient tous les critères de dimensions obligatoires de S ainsi que tous les critères de dimension K_i pour lesquels $b_i = 1$:

$$C_{\iota_{b_1 \dots b_{\dot{n}}}} = \{D, \quad \forall K_i \in S, K_i(D) \text{ est vérifié et } \forall i, b_i = 1, K_i(D) \text{ est vérifié} \}$$

On notera que les classes de pertinence ne sont pas disjointes. En effet, l'ensemble des documents vérifiant p critères de dimension inclut l'ensemble des documents vérifiant ces p critères plus un.

De manière analogue, nous définissons dans la partie suivante les *sous-classes de pertinence* qui s'expriment en fonction des statuts associés aux critères élémentaires.

Les classes de pertinence sont utilisées pour organiser la visualisation des documents retrouvés, comme nous le montrons dans la suite.

2.6 Résumé

Finalement, le schéma de pertinence de notre modèle joue le rôle de la requête dans la terminologie classique des systèmes de recherche d'information. Il permet d'exprimer explicitement l'interprétation à donner des termes en matière de pertinence, tant du point de vue sémantique (critères abstraits et termes) que du point de vue pragmatique de la mise en forme de la réponse (statuts des critères).

Le fait de disposer de ces éléments de manière explicite, permet de travailler sur le contrôle du déroulement de l'interaction. Il constitue le vocabulaire de base pour définir des associations entre des situations typiques rencontrées lors d'une recherche, et des modifications à apporter au schéma de pertinence pour poursuivre la recherche.

Chapitre 3

Bilan et application des principes

Nous montrons ici comment l'interaction que nous proposons permet de mettre en œuvre les trois principes d'interaction *Détecter*, *Comprendre* et *Reformuler* que nous avons énoncés page 53.

3.1 Mise en forme des résultats : détecter

Le principe *Détecter* est le suivant :

faciliter la détection de la pertinence

L'ensemble des critères élémentaires d'un schéma de pertinence définit un ensemble de documents potentiellement retrouvés : l'ensemble des documents dans lesquels au moins l'un des critères est vérifié. Les statuts définissent quel sous-ensemble de cet ensemble sera effectivement visualisé et organisé.

Les classes visuelles de pertinence réduisent le nombre d'informations que l'utilisateur doit analyser pour évaluer sa satisfaction : par exemple, au lieu de parcourir 100 documents, un utilisateur pourra parcourir 8 classes, et éliminer 6 d'entre elles simplement au vu de deux ou trois documents leur appartenant. Ainsi l'utilisateur peut ne parcourir intégralement que 2 classes de documents qui lui paraissent susceptibles de contenir des documents pertinents.

Elles fournissent par ailleurs un ordre de consultation des documents qui est en relation symbolique explicite avec le schéma de pertinence, ce qui fournit un guide pour, le cas échéant, éviter l'examen de certaines classes de documents.

À plus longue échéance et dans le contexte d'une application particulière pour laquelle une étude peut être menée avec des utilisateurs réels, des connaissances sur les classes de pertinence les plus utiles à visualiser dans une situation donnée, peuvent être établies. Les classes les plus utiles à visualiser dans une situation donnée ne sont pas nécessairement des classes de documents pertinents, qui mènent à la

satisfaction ; il peut s'agir de classes dont l'examen permet de résoudre une situation problématique, comme nous l'illustrons dans la partie V. De telles connaissances peuvent être utilisées pour alléger la tâche de localisation des classes intéressantes.

3.2 Paramétrisation et visualisation : comprendre

Le principe *Comprendre* est le suivant :

clarifier la sémantique de l'interaction

Les critères élémentaires paramètrent la fonction de correspondance d'un point de vue sémantique. Les statuts paramètrent la fonction de mise en forme, d'un point de vue pragmatique.

Nous avons suggéré plus haut que les critères élémentaires soient choisis de façon qu'un individu puisse en évaluer lui-même la vérification sur un document qui lui est présenté. Nous avons vu que le choix des critères est issu d'un compromis : la sémantique des critères requiert en général une clarification pour son bon usage par l'utilisateur. À défaut, un quiproquo peut s'introduire dans l'interaction, ce qui provoque des situations problématiques comme par exemple le fait qu'un individu juge qu'un document retrouvé n'a aucun lien avec la requête qu'il a exprimée.

En organisant les documents retrouvés selon les critères de pertinence, l'individu peut juger de l'adéquation entre l'idée qu'il se fait de la sémantique d'un critère qu'il a utilisé, et la sémantique que lui accorde le système, simplement en observant les caractéristiques communes aux documents d'une classe où ce critère est vérifié.

Ainsi au fur et à mesure du déroulement de l'interaction, l'utilisateur clarifie l'image mentale qu'il a du système, et augmente ainsi ses chances de l'utiliser mieux lors de l'étape suivante.

La paramétrisation par les statuts correspond ainsi à un outil interactif permettant de faire varier rappel et précision. Rappel et précision ne sont plus vus comme des critères qualitatifs de performance, mais comme des paramètres que l'utilisateur peut faire varier pour adapter les caractéristiques de surface (nombre de documents retrouvés, nombre moyen de documents par classe, etc.) de la vue que le système donne du corpus, pour mieux comprendre et détecter la pertinence des documents retrouvés.

3.3 Bouclage de pertinence généralisé : reformuler

Le principe *Reformuler* est le suivant :

faciliter la reformulation

Chaque classe de pertinence est associée, par l'intermédiaire de son lien symbolique avec le schéma de pertinence, aux critères que le système a utilisés pour estimer comme pertinents les documents de cette classe.

Quand un utilisateur juge un document retrouvé non pertinent, il peut analyser les raisons qui font que le système l'a retrouvé tout de même. Ainsi, il peut

- identifier une utilisation erronée d'un terme ou d'un critère, ou bien
- prendre la mesure des limites du système (limites des possibilités d'interrogation), ou encore
- prendre la mesure des limites du corpus (peut-être le corpus ne contient-il pas de document pertinent).

Cette analyse lui permet de réagir à certaines situations problématiques, car elle lui fournit une trame pour reformuler son problème d'information de façon à faire évoluer les paramètres du système dans le sens de son schéma individuel de pertinence, tout en étant conscient des limites du système et du corpus.

Nous reprenons en détail ce dernier point dans le cadre de notre application.

3.4 Vers une formalisation

Dans la partie qui suit, nous donnons une formalisation simple du schéma de pertinence qui nous permet de définir l'ensemble des documents retrouvés, et les classes de pertinence qui organisent cet ensemble. Cela nous permet de définir la fonction de mise en forme du résultat du système, par l'intermédiaire des classes visuelles de pertinence.

Cette formalisation du résultat tel qu'il est présenté à l'utilisateur dans l'interaction nous permet d'appréhender formellement la notion de *vue système d'une situation de recherche d'information*. Nous pouvons ainsi procéder à une analyse formelle du déroulement d'une session de recherche, et ainsi définir les fondements d'une évaluation des systèmes de recherche d'information selon la perspective des trois principes que nous avons définis ci-dessus.

Partie IV

Formalisation

Dans cette partie nous formalisons les notions présentées dans le modèle conceptuel, et nous développons les divers aspects de la conception de l'interaction que notre modèle permet d'appréhender.

Dans le chapitre 1, nous donnons une formalisation du schéma de pertinence. Dans le chapitre 2 nous définissons les classes de pertinence et la notion de schéma homologue à une classe, qui renouvellent la définition de degré de pertinence. Dans le chapitre 3, nous apportons un raffinement à la définition des classes de pertinence de façon à les rendre utilisables concrètement pour la visualisation des documents retrouvés. Dans le chapitre 4, nous montrons comment l'utilisation interactive d'un système conçu selon notre modèle peut être décrite en termes formels. Dans le chapitre 5, nous établissons un bilan de cette formalisation en nous resituant dans la perspective des travaux existants concernant l'adaptation des systèmes de recherche d'information à la situation.

Chapitre 1

Formalisation de la pertinence système

Dans ce chapitre nous donnons une formalisation de notre modèle de pertinence système.

Dans un premier temps nous annonçons la portée de notre formalisation, et fixons les notations employées dans la suite en rappelant la structure du schéma de pertinence. Puis nous définissons le langage des schémas de pertinence, dont nous donnons ensuite une sémantique binaire d'une part, et une sémantique symbolique d'autre part. La première permet de définir l'ensemble des documents retrouvés tandis que la seconde définit la notion de relation de pertinence qui sera utilisée par la suite pour décrire les classes de pertinence.

1.1 Portée de la formalisation

Dans le cadre de cette formalisation, nous considérons que l'ensemble des critères élémentaires abstraits est *donné*, et nous le notons $\mathcal{E}_\varepsilon$. Les dimensions qui groupent ces critères sont elles aussi données.

L'ensemble des valeurs de statuts se réduit à $\{ \textit{obligatoire} , \textit{optionnel} \}$ et nous le notons $\mathcal{E}_s$.

1.1.1 Un ensemble prédéfini de critères abstraits et de dimensions

Pour une application donnée, l'ensemble des critères abstraits et les dimensions associées se définissent à partir des caractéristiques indexées du corpus, des connaissances disponibles sur le corpus (un thésaurus organisant les thèmes de l'indexation,

par exemple), et en fonction des besoins d'interrogation que l'on veut prendre en compte dans le cadre de l'application.

Cela suppose que le corpus peut être effectivement étudié d'un point de vue qualitatif, et que les conditions d'utilisation du système peuvent être étudiées elles aussi, pour identifier les besoins d'interrogation qu'elles suscitent.

Cette hypothèse a donc pour conséquence d'exclure du champ d'application de ce modèle les corpus hétérogènes qui évoluent de manière incontrôlable, et dont les diverses utilisations sont difficiles à cerner, comme c'est le cas du corpus que constitue le World Wide Web dans son entier.

De plus, dans le cas d'indexations réduites à une unique caractéristique, et en l'absence de connaissances permettant de valoriser cette caractéristique pour définir plusieurs dimensions d'interrogation, notre modèle s'instancie de façon dégénérée (une seule dimension), ce qui rend son intérêt limité.

Nous donnons dans la partie V des exemples d'application où notre modèle s'applique dans sa généralité.

1.1.2 Des statuts à valeur binaire

Le choix de n'associer aux statuts que des valeurs binaires (*obligatoire* ou *optionnel*) est motivé par le principe de compréhension que nous avons formulé précédemment. L'objectif est de ne faire figurer dans l'interface que des éléments dont le sens est suffisamment clair pour qu'un utilisateur puisse y associer une image mentale correcte, conformément à ce principe.

De plus, des statuts à valeur binaire se traduisent directement en termes de classes de pertinence et fournissent un outil pratique de manipulation de la réponse du système. Nous revenons cependant sur ce point quand nous introduisons la notion de degré de pertinence (2.4).

1.2 Structure générale et notations

Nous notons N le nombre de dimensions d'interrogation.

Étant donné un schéma de pertinence S , nous notons n le nombre de critères de dimension de ce schéma ($n \leq N$).

La figure IV.1.1 donne la représentation graphique compacte de l'exemple de

schéma de pertinence suivant :

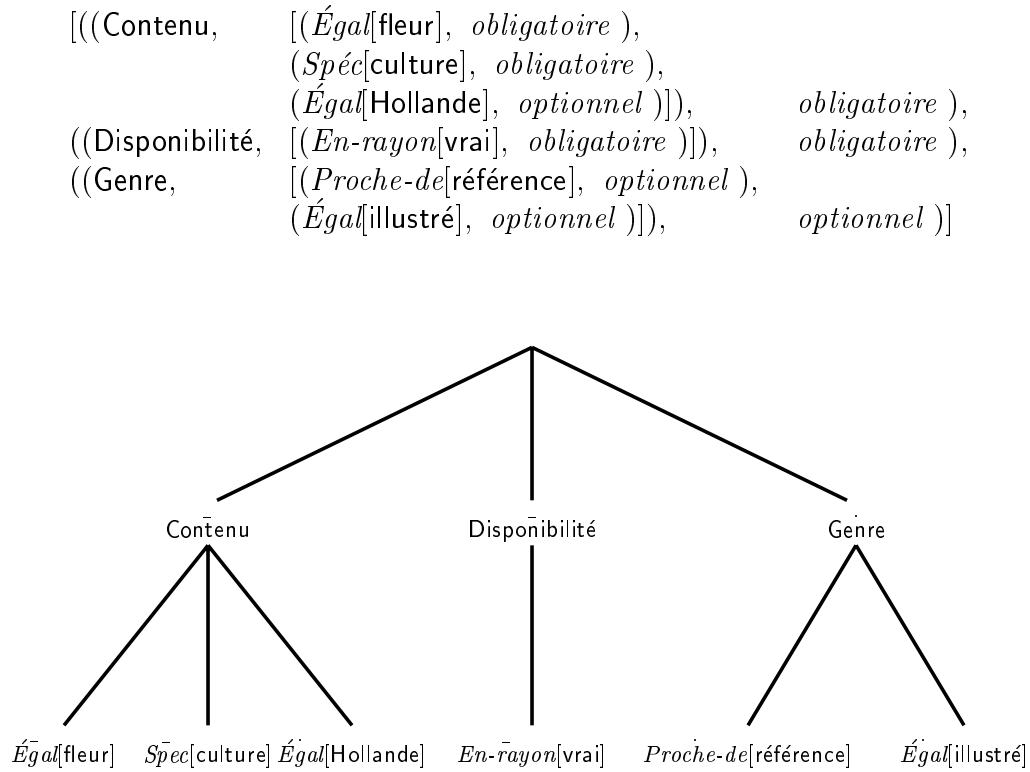


FIG. IV.1.1 – Notation compacte pour l'exemple de schéma de pertinence

1.3 Langage des schémas de pertinence

Un schéma de pertinence est un arbre à deux niveaux, dont les nœuds des feuilles sont étiquetés par des couples (critère élémentaire, statut), et dont les nœuds intermédiaires sont étiquetés par des couples (dimension, statut). Le nœud racine n'a pas d'étiquette.

Définition 1.1 (Langage des critères élémentaires) Soit $\mathcal{E}_\xi$ l'ensemble des critères élémentaires abstraits et $\mathcal{E}_t$ l'ensemble des termes.

L'ensemble des critères élémentaires $\mathcal{E}_k$ est l'ensemble des couples

$$\xi[t]$$

avec $\xi \in \mathcal{E}_\xi$ et $t \in \mathcal{E}_t$.

Ils sont dénotés par le symbole k .

Définition 1.2 (Langage des critères de dimension) Soit $\mathcal{E}_k$ l'ensemble des critères élémentaires, $\mathcal{E}_{Dim}$ l'ensemble des dimensions d'interrogation, et $\mathcal{E}_s$ l'ensemble des valeurs de statut.

L'ensemble des critères de dimension $\mathcal{E}_K$ est l'ensemble des couples

$$(Dim, [(k_1, s_1), \dots, (k_n, s_n)])$$

avec $Dim \in \mathcal{E}_{Dim}$, et pour tout $j \in [1..n]$, $k_j \in \mathcal{E}_k$ et $s_j \in \mathcal{E}_s$.

Ils sont dénotés par le symbole K .

Définition 1.3 (Langage des schémas de pertinence) Soit $\mathcal{E}_K$ le langage des critères de dimension et $\mathcal{E}_s$ l'ensemble des valeurs de statut.

L'ensemble des schémas de pertinence $\mathcal{E}_S$ est l'ensemble des tuples

$$[(K_1, s_1), \dots, (K_n, s_n)]$$

avec pour tout $i \in [1..n]$, $K_i \in \mathcal{E}_K$ et $s_i \in \mathcal{E}_s$.

Ils sont dénotés par le symbole S .

Le schéma de pertinence joue le rôle de la requête, dans la terminologie classique des systèmes de recherche d'information.

1.4 Fonction d'évaluation d'un critère élémentaire

La sémantique du schéma de pertinence est une sémantique logique, c'est-à-dire que la sémantique d'une expression est définie en fonction de la sémantique de ses sous-expressions. Les expressions atomiques sont les critères élémentaires de pertinence.

L'évaluation d'un critère élémentaire se définit par instanciation partielle de la fonction à deux variables qui définit l'évaluation d'un critère élémentaire abstrait.

Définition 1.4 (Fonction d'évaluation d'un critère élémentaire abstrait) Une fonction d'évaluation d'un critère élémentaire abstrait ξ est une fonction $E(\xi) : \mathcal{E}_R \times \mathcal{E}_D \rightarrow \mathcal{E}_v$ qui à tout terme $t \in \mathcal{E}_R$ et à tout document $D \in \mathcal{E}_D$ associe une valeur v dans un ensemble de valeurs $\mathcal{E}_v$ organisées selon un treillis $\langle \mathcal{E}_v, \prec \rangle$.

En fonction du critère élémentaire abstrait concerné, le treillis des valeurs peut varier. Par exemple, dans le cas d'un critère abstrait qui s'évalue en fonction de caractéristiques d'indexation de nature statistique (par exemple issues de statistiques sur l'occurrence de termes dans les documents), les pondérations disponibles dans l'indexation peuvent être mises à profit. Dans ce cas, le treillis des valeurs sera $\langle \mathbb{R}, \leq \rangle$, ou bien $\langle [0, 1], \leq \rangle$. On note $\perp$ le minimum de $\mathcal{E}_v$, quelque soit $\mathcal{E}_v$.

On considère dans le cadre de cette formalisation, que tous les fonctions $E(\xi)$ qui interviennent dans le système de recherche d'information considéré, sont données; toutes ces fonctions ont une valeur résultat en commun qui est $\perp$.

1.5 Évaluation binaire

Nous définissons une évaluation binaire des schémas de pertinence dans les documents selon une sémantique booléenne des connectives et des statuts au moyen de la fonction E_{bin} .

Définition 1.5 (Évaluation binaire d'un critère élémentaire) *L'évaluation binaire E_{bin} d'un critère élémentaire abstrait ξ est une fonction*

$$E_{bin}(\xi) : \mathcal{E}_R \times \mathcal{E}_D \rightarrow \mathcal{E}_{vbin}$$

où $\mathcal{E}_R$ est l'ensemble des termes, $\mathcal{E}_D$ est l'ensemble des documents et $\mathcal{E}_{vbin}$ est le treillis des valeurs booléennes $\langle \{ \text{vrai}, \text{faux} \}, \prec \rangle$ avec $\{ \text{faux} \prec \text{vrai} \}$, aussi noté $\langle \{ \top, \perp \}, \prec \rangle$. Ainsi pour $\mathcal{E}_{vbin}$, on a $\perp = \text{faux}$.

Pour tout $D \in \mathcal{E}_D$

$$E_{bin}(\xi)(t, D) = \begin{cases} \perp & \text{ssi } \xi[t](D) = \perp \\ \top & \text{sinon} \end{cases}$$

Dans la suite nous substituons fréquemment à la notation $E_{bin}(\xi)(t, D)$ la notation $E_{bin}(k)(D)$ avec $k = \xi[t]$.

Définition 1.6 (Sémantique conjonctive pour les dimensions) *L'évaluation binaire d'un critère de dimension est définie sur la base d'une sémantique booléenne conjonctive pour la combinaison des critères élémentaires.*

$$E_{bin}(K) : \mathcal{E}_D \rightarrow \mathcal{E}_{vbin}$$

Pour tout $D \in \mathcal{E}_D$

$$E_{bin}((Dim, [k_1, \dots, k_n]))(D) = \wedge(E_{bin}(k_1)(D), \dots, E_{bin}(k_n)(D))$$

Définition 1.7 (Sémantique conjonctive pour le schéma) *De même l'évaluation binaire d'un schéma de pertinence est définie sur la base d'une sémantique booléenne conjonctive pour la combinaison des critères de dimension.*

$$E_{bin}(S) : \mathcal{E}_D \rightarrow \mathcal{E}_{vbin}$$

Pour tout $D \in \mathcal{E}_D$

$$E_{bin}([K_1, \dots, K_n])(D) = \wedge(E_{bin}(K_1)(D), \dots, E_{bin}(K_n)(D))$$

Nous associons donc aux combinaisons réalisées par les tuples une sémantique conjonctive au sens de la logique booléenne, dans le cadre d'une évaluation binaire ne tenant pas compte des statuts.

Définition 1.8 (Sémantique des statuts des critères élémentaires) *Nous définissons une évaluation binaire booléenne pour les statuts.*

Pour tout $D \in \mathcal{E}_D$

$$E_{bin}([(k_1, s_1), \dots, (k_n, s_n)])(D) = \bigwedge_{\{i \mid \text{statut } s_i = \text{obligatoire}\}} (E_{bin}(k_i)(D))$$

Ainsi les critères élémentaires affectés du statut optionnel n'interviennent pas dans l'évaluation binaire du schéma de pertinence. Par conséquent, si aucun critère n'a de statut obligatoire, on a pour tout $D \in \mathcal{E}_D$

$$E_{bin}([(k_1, s_1), \dots, (k_n, s_n)])(D) = \top$$

Définition 1.9 (Sémantique des statuts des critères de dimension) *Les statuts associés aux critères de dimension ont une sémantique similaire.*

Pour tout $D \in \mathcal{E}_D$

$$E_{bin}([(K_1, s_1), \dots, (K_n, s_n)])(D) = \bigwedge_{\{i \mid \text{statut } s_i = \text{obligatoire}\}} (E_{bin}(K_i)(D))$$

De même, si aucun critère de dimension n'a de statut obligatoire, on a pour tout $D \in \mathcal{E}_D$

$$E_{bin}([(K_1, s_1), \dots, (K_n, s_n)])(D) = \top$$

Définition 1.10 (Ensemble des documents retrouvés) *L'ensemble des documents pour lesquels la pertinence système est évaluée à vrai ($\top$) selon l'évaluation binaire E_{bin} , dans le schéma de pertinence S , est appelé l'ensemble des documents retrouvés et noté $S(\mathcal{E}_D)$.*

$$S(\mathcal{E}_D) = \{D \in \mathcal{E}_D, E_{bin}(S)(D) = \top\}$$

Plus généralement, on notera $K(\mathcal{E}_D)$ l'ensemble des documents vérifiant le critère de dimension K

$$K(\mathcal{E}_D) = \{D, E_{bin}(K)(D) = \top\}$$

et $k(\mathcal{E}_D)$ l'ensemble des documents vérifiant le critère élémentaire k

$$k(D) = \{D, E_{bin}(k)(D) = \top\}$$

1.6 Évaluation symbolique

Le rôle de l'évaluation symbolique est de rendre compte des raisons sous-jacentes à l'évaluation de la pertinence d'un document par le système, dans le but d'aller au-delà d'une évaluation binaire et d'une sémantique booléenne pour la pertinence système.

1.6.1 Sémantique symbolique du schéma de pertinence

Définition 1.11 (Évaluation symbolique d'un critère élémentaire abstrait)

La fonction d'évaluation symbolique $E_{sym}(\xi)$ d'un critère élémentaire abstrait ξ est une fonction

$$E_{sym}(\xi) : \mathcal{E}_R \times \mathcal{E}_D \rightarrow \mathcal{E}_{vsym}$$

où $\mathcal{E}_R$ est l'ensemble des termes de requête, $\mathcal{E}_D$ est l'ensemble des documents du corpus et $\mathcal{E}_{vsym}$ est un ensemble de valeurs symboliques élémentaires. Une valeur symbolique élémentaire est un couple (critère élémentaire, valeur).

Pour tout $D \in \mathcal{E}_D$

$$E_{sym}(\xi)(t, D) = (k, v)$$

avec $k = \xi[t]$ et $v = k(D)$.

Le premier élément du couple désigne le critère élémentaire concerné par l'évaluation tandis que le second représente la réalisation de ce critère dans le document concerné. La sémantique symbolique des critères élémentaires de pertinence est définie par cette fonction d'évaluation.

Définition 1.12 (Évaluation symbolique d'un critère de dimension) *La combinaison des critères élémentaires de pertinence est évaluée de manière symbolique en conservant les termes de la combinaison dans un tuple. Pour distinguer les tuples correspondant aux composants d'un schéma de pertinence de ceux correspondant à une valeur symbolique, nous notons ces derniers en double.*

Pour tout $D \in \mathcal{E}_D$

$$E_{sym}((Dim, [(k_1, s_1) \dots, (k_n, s_n)]))(D) = (Dim, [[E_{sym}(k_1)(D), \dots, E_{sym}(k_n)(D)]])$$

Définition 1.13 (Évaluation symbolique d'un schéma de pertinence) *Pour tout $D \in \mathcal{E}_D$*

$$E_{sym}([(K_1, s_1), \dots, (K_n, s_n)])(D) = [[E_{sym}(K_1)(D), \dots, E_{sym}(K_n)(D)]]$$

À nouveau, la sémantique symbolique du schéma de pertinence est définie par cette fonction d'évaluation.

1.6.2 Langage des relations de pertinence

Définition 1.14 (Relation de pertinence) *La relation de pertinence RP d'un document pour un schéma de pertinence S est le résultat de l'évaluation symbolique du document dans ce schéma.*

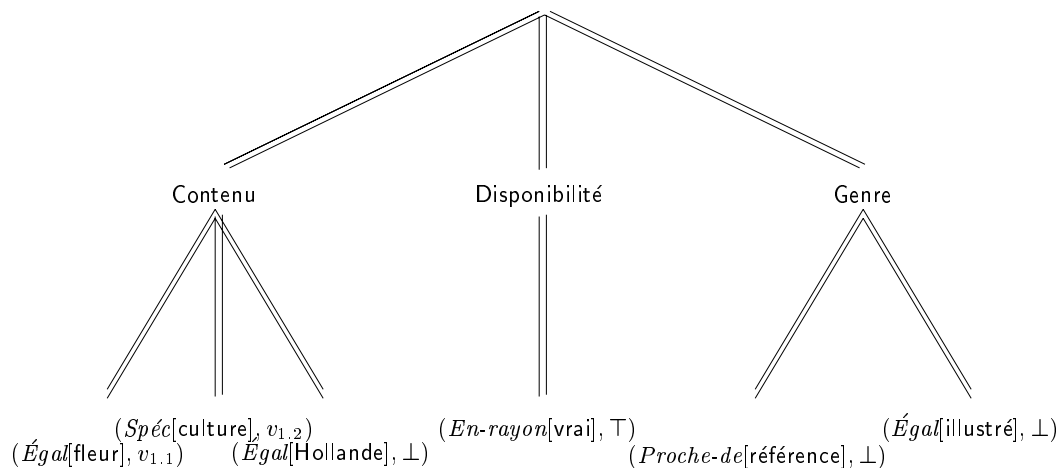


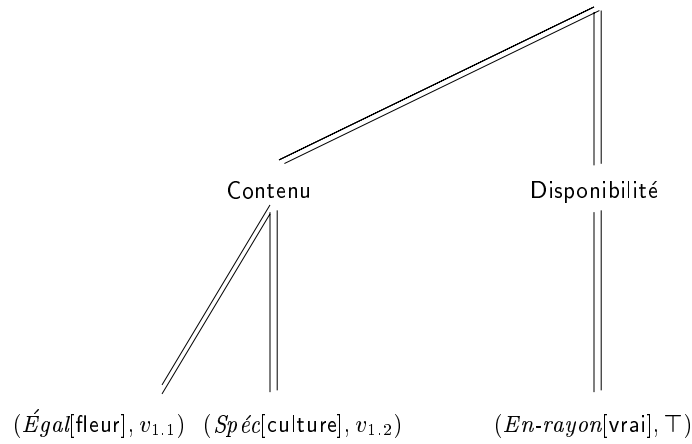
FIG. IV.1.2 – Exemple de relation de pertinence

La figure IV.1.2 donne l'allure d'une relation de pertinence au moyen d'un exemple en précisant quelles valeurs correspondent à la valeur minimum notée $\perp$.

Cet exemple est relatif au schéma de pertinence de la figure IV.1.1. La figure IV.1.3 donne une notation compacte pour cette relation de pertinence : les critères dont l'évaluation donne $\perp$ n'apparaissent plus.

La relation de pertinence intègre une valeur de réalisation de chaque critère dans le document. Dans le cas où l'indexation permet une évaluation quantifiée des critères élémentaires de pertinence, cela permet d'utiliser cette quantité pour ordonner les documents entre eux. Nous verrons dans la suite que les documents sont d'abord regroupés en classes de pertinence, selon des critères purement qualitatifs relatifs à *quels* critères sont vérifiés (voir la notion de schéma homologue), indépendamment de la quantification qui peut être associée à cette vérification.

Conserver la notion de relation de pertinence ménage la possibilité de combiner différemment les valeurs quantifiées des critères pour ordonner les documents de diverses manières, en fonction du contexte de l'interaction. Plus précisément, il s'agit de définir diverses fonctions de classement par ordre (*ranking function*) des documents au sein d'une classe, en fonction de la situation problématique que l'on cherche à résoudre dans l'étape courante. Nous ne traitons pas cet aspect dans le

FIG. IV.1.3 – *Notation simplifiée de relation de pertinence*

cadre de cette thèse, mais nous revenons sur cette idée dans le bilan de la partie Application et dans la conclusion.

1.6.3 Conclusion

Considérant que la tâche principale de l'utilisateur est d'évaluer la qualité du résultat fourni par le système afin d'améliorer la formulation du problème d'information si nécessaire, l'objectif pour le système est de faciliter cette évaluation en organisant les documents retrouvés de manière appropriée.

L'évaluation symbolique a pour but de permettre cette organisation par l'intermédiaire des *classes de pertinence* définies plus loin, et de faire de ces classes les objets de l'interaction.

Chapitre 2

Classes de pertinence et schémas homologues

Dans ce chapitre nous définissons les classes de pertinence comme un ensemble de sous-ensembles des documents retrouvés, en fonction des statuts de critères. Parallèlement, nous définissons la notion de schéma homologue à une classe, qui constitue une caractérisation symbolique des classes. La notion de degré de pertinence peut être généralisée dans ce contexte.

Nous définissons d'abord les classes de dimensions (ou classes), relatives aux statuts des critères de dimension, ainsi que les schémas homologues à ces classes de dimension. Puis nous définissons les sous-classes, relatives aux statuts des critères élémentaires, ainsi que les schémas homologues à ces sous-classes. Enfin, nous donnons notre définition du degré de pertinence d'un document dans les termes du degré de pertinence des classes auxquelles il appartient, ce qui nous permet de montrer un ensemble de propriétés des classes vis-à-vis des degrés de pertinence.

2.1 Classes de pertinence

2.1.1 Notations

Une classe de pertinence est un sous-ensemble des documents retrouvés qui partagent certaines caractéristiques, identifiées par un ensemble de critères de pertinence présents dans le schéma de pertinence.

Les critères dont le statut est obligatoire définissent des caractéristiques de pertinence communes à tous les documents retrouvés. Ceux dont le statut est optionnel définissent des caractéristiques qui servent à organiser les documents retrouvés selon des classes et des sous-classes représentatives des particularités de leur relation de pertinence avec le schéma de pertinence.

Soit un schéma de pertinence S . Nous prenons les conventions suivantes.

$$S = [(\text{Dim}_1, \dot{K}_1), \dots, (\text{Dim}_{\dot{n}}, \dot{K}_{\dot{n}}), (\text{Dim}_{\dot{n}+1}, \bar{K}_{\dot{n}+1}), \dots, (\text{Dim}_n, \bar{K}_n)]$$

comprenant n critères de dimension parmi lesquels les $\dot{n}$ premiers sont optionnels et les $n - \dot{n}$ autres sont obligatoires.

Chaque critère de dimension K_i comprend n_i critères élémentaires, parmi lesquels les $\dot{n}_i$ premiers sont optionnels et les $n_i - \dot{n}_i$ autres sont obligatoires.

$$K_i = [\dot{k}_{i,1}, \dots, \dot{k}_{i,\dot{n}_i}, \bar{k}_{i,\dot{n}_i+1}, \dots, \bar{k}_{i,n_i}]$$

2.1.2 Ensemble des classes de pertinence

L'ensemble des classes de pertinence correspond à l'ensemble de toutes les combinaisons possibles de vérification des critères optionnels pour un niveau donné de critères, c'est-à-dire

- soit le niveau des critères de dimension $\dot{K}_i$, qui définit les *classes de pertinence de dimension* C_B ,
- soit, pour une classe de dimension donnée C_B , le niveau des critères élémentaires optionnels $\dot{k}_{i,j}$ qui définit les *sous-classes de pertinence* $c_{(B,B')}$.

Étant donné $\dot{n}$ le nombre de critères optionnels pour un niveau, l'ensemble $\mathcal{E}_B$ des combinaisons des critères optionnels est l'ensemble des injections $B : [1..\dot{n}] \rightarrow \{0, 1\}$, notées $b_1 \dots b_{\dot{n}}$, avec pour tout $i \in [1..\dot{n}]$, $b_i \in \{0, 1\}$.

Définition 2.1 (Ensemble de classes de pertinence)

$$\mathcal{E}_C = \{C_B, B \in \mathcal{E}_B\}$$

2.2 Classe de pertinence de dimension et schéma homologue

En l'absence de tout critère de dimension optionnel, l'ensemble des classes de pertinence de dimension se réduit à une classe $C = S(\mathcal{E}_D)$ contenant tous les documents retrouvés.

Considérons le cas où il existe au moins un critère de dimension optionnel, et supposons que $\dot{\mathbf{K}} = \{K_i\}_{i \in [1..\dot{n}]}$ est l'ensemble des critères de dimension optionnels.

Définition 2.2 (Classe de pertinence) *La classe de pertinence relative à la combinaison $b_1 \dots b_{\dot{n}}$, notée $C_{b_1 \dots b_{\dot{n}}}$, est définie comme le sous-ensemble des documents*

retrouvés qui vérifient les critères de dimension optionnels $\dot{K}_i$ pour lesquels on a $b_i = 1$ ¹.

$$C_{b_1 \dots b_n} = S(\mathcal{E}_D) \quad \cap \quad \{D \in \mathcal{E}_D, \forall i \in [1..n], (b_i = 1 \implies K_i(D))\}$$

Les relations de pertinence des documents appartenant à une même classe de pertinence $C_{b_1 \dots b_n}$ partagent des caractéristiques. Plus précisément, elles partagent toutes leurs caractéristiques exceptées celles relatives aux valeurs de réalisation des critères. Pour décrire ces caractéristiques communes et ainsi fournir une description des caractéristiques de pertinence d'une classe entière, un sous-langage du langage des schémas de pertinence convient.

En effet, l'ensemble des documents regroupés dans une classe de pertinence $C_{b_1 \dots b_n}$ coïncide avec l'ensemble des documents retrouvés d'un schéma de pertinence particulier, noté $S_{b_1 \dots b_n}$, appelé *schéma de pertinence homologue* à $C_{b_1 \dots b_n}$. Ce schéma homologue est défini à partir du schéma de pertinence en ne conservant que les branches de l'arbre qui constituent un critère de dimension obligatoire ($s_i = \text{obligatoire}$), ou un critère de dimension optionnel pour lequel $b_i = 1$ ($s_i = \text{optionnel} \wedge b_i = 1$).

Les statuts des critères de dimension n'ont plus cours dans le langage des schémas homologues, puisqu'ils doivent définir une seule classe de dimension. Tout se passe comme s'il s'agissait de schémas de pertinence dont tous les critères de dimension sont obligatoires. En revanche les critères élémentaires conservent le statut qui leur est associé dans le schéma de pertinence.

Définition 2.3 (Schéma homologue à une classe de dimension)

$$S_{b_1 \dots b_n} = [K_i]_{i \in I}$$

avec $I = \{i, s_i = \text{obligatoire}\} \cup \{i, s_i = \text{optionnel} \wedge b_i = 1\}$

Le schéma de pertinence proposé dans la figure IV.2.1 permet de définir en tout deux classes de dimension autour de l'unique critère de dimension optionnel $\dot{K}_3$. Ces classes sont notées C_1 et C_0 , et ont pour schéma homologue S_1 et S_0 respectivement.

Un document D_0 entretenant la relation de pertinence présentée dans la figure IV.1.3 avec le schéma de pertinence de la figure IV.1.1 appartient à la classe C_0 et n'appartient pas à la classe C_1 .

Propriété 2.1 (Correction des schémas homologues) *L'ensemble des documents retrouvés par un schéma homologue $S_{b_1 \dots b_n}$ d'une classe $C_{b_1 \dots b_n}$ coïncide avec l'ensemble des documents de cette classe.*

1. Nous rappelons que la notation $K_i(D)$ fait référence à l'évaluation binaire de la pertinence système ; c'est une notation simplifiée pour $\{D, E_{bin}(K_i)(D) = \top\}$

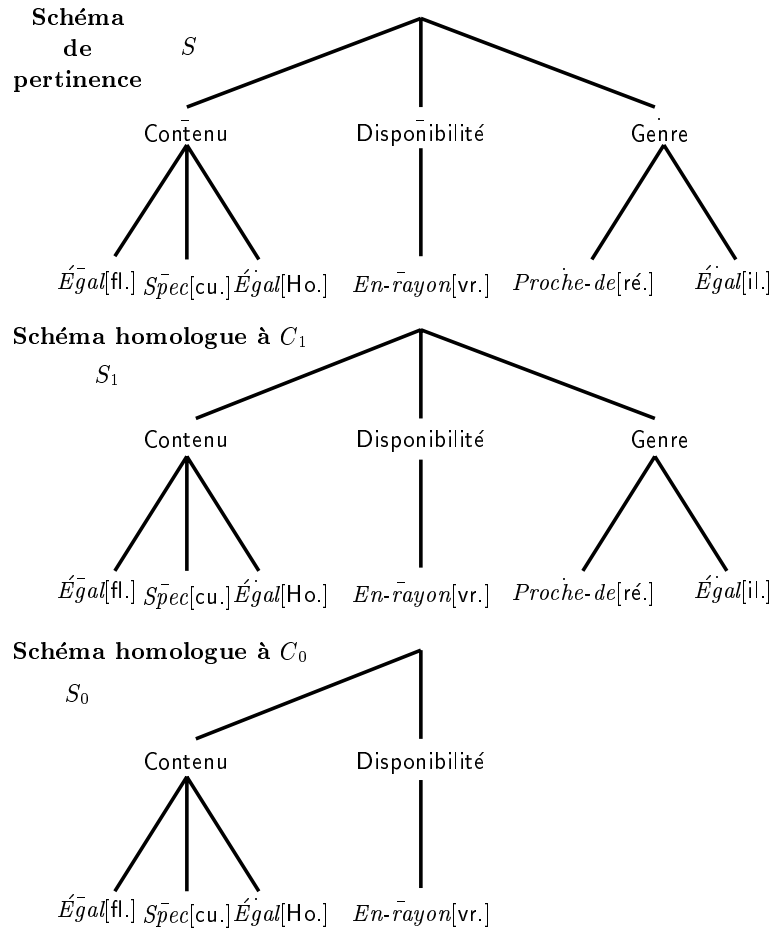


FIG. IV.2.1 – Schémas homologues des deux classes de dimension issues de S

Preuve 2.1 Par équivalences successives :

Soit $D \in S_{b_1 \dots b_{\dot{n}}}(\mathcal{E}_D)$.

$$E_{bin}(S_{b_1 \dots b_{\dot{n}}}(D)) = \top$$

équiv.

$$\forall i \in (\{i, s_i = \text{obligatoire}\} \cup \{i, s_i = \text{optionnel}\} \wedge b_i = 1), E_{bin}(K_i(D)) = \top$$

équiv.

$$(\forall i(s_i = \text{obligatoire}) \Rightarrow E_{bin}(K_i(D)) = \top)$$

$\wedge$

$$(\forall i(s_i = \text{optionnel} \wedge b_i = 1) \Rightarrow E_{bin}(K_i(D)) = \top)$$

équiv.

$$(D \in S_{b_1 \dots b_{\dot{n}}}(\mathcal{E}_D)) \wedge (\forall i \in [1..\dot{n}], (b_i = 1) \Rightarrow K_i(D))$$

Ce qui équivaut, par définition de $C_{b_1 \dots b_{\dot{n}}}$ (définition 2.2), à $D \in C_{b_1 \dots b_{\dot{n}}}$.

■

2.3 Sous-classe de pertinence et schéma homologue

Pour le niveau des critères élémentaires, on considère que le référentiel est non plus l'ensemble des documents retrouvés, mais une classe de pertinence particulière. Ainsi, étant donné une classe de pertinence $C_{b_1 \dots b_{\dot{n}}}$, l'ensemble des sous-classes de cette classe est construit sur l'ensemble des combinaisons de la vérification des critères élémentaires optionnels dans le contexte de cette classe.

L'ensemble des critères élémentaires optionnels pour la classe $C_{b_1 \dots b_{\dot{n}}}$ est l'union des critères élémentaires optionnels relatifs à chaque critère de dimension obligatoire ($\dot{k}_{i,j} \in \bar{K}_i$), ajoutée de l'union des critères élémentaires optionnels relatifs aux critères de dimension $\dot{K}_i$ qui sont nécessairement vérifiés dans cette classe ($\dot{k}_{i,j} \in \dot{K}_i$ et $b_i = 1$).

Ainsi le nombre de critères élémentaires optionnels pris en compte est $\dot{n}'$ défini comme suit.

$$\dot{n}' = \text{Card}(\{\dot{k}_{i,j}\}_{(i,j) \in J})$$

avec $J = \{(i,j), \exists \bar{K}_i \in S, \dot{k}_{i,j} \in \bar{K}_i\} \cup \{(i,j), \exists \dot{K}_i \in S, (\dot{k}_{i,j} \in \dot{K}_i \wedge b_i = 1)\}$

Considérons le cas où $\dot{n}' \geq 1$, et prenons les mêmes notations que précédemment.

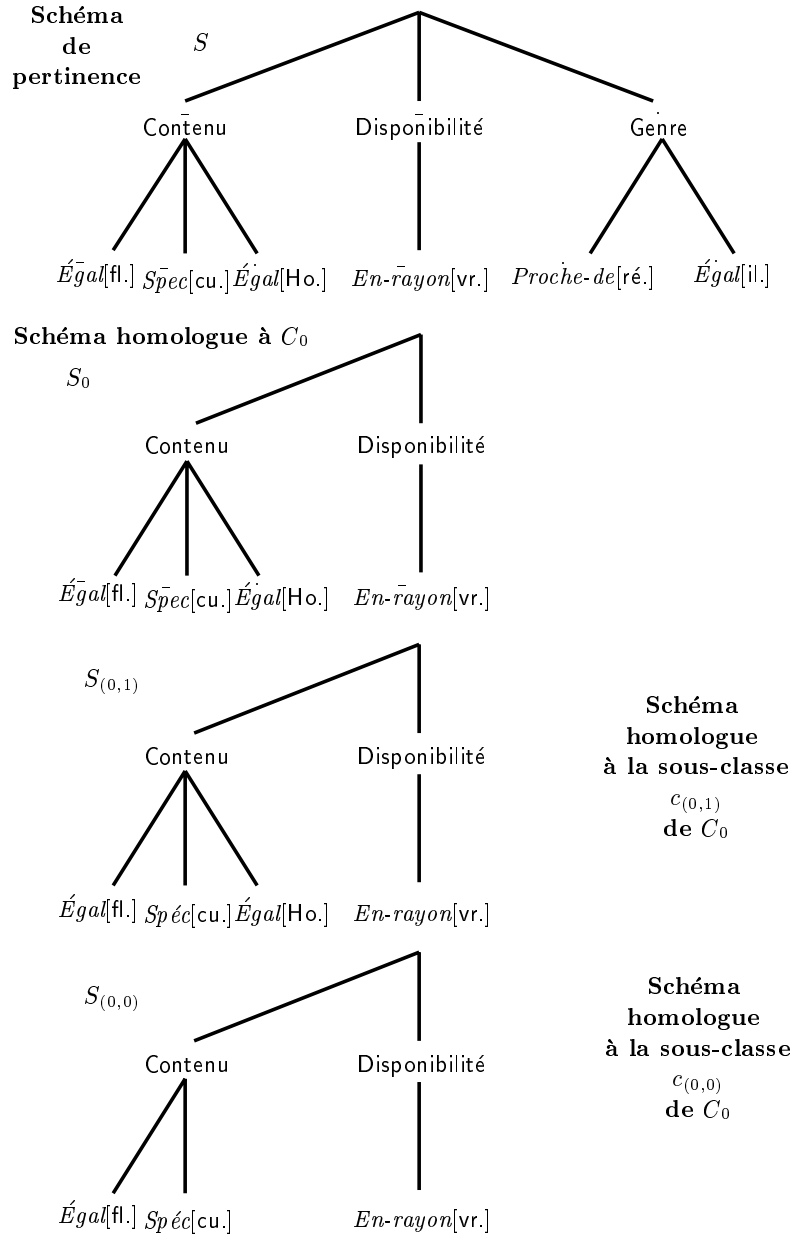


FIG. IV.2.2 – Schémas homologues à des sous-classes

Définition 2.4 (Sous-classe d'une classe de pertinence) *La sous-classe de C_B relative à la combinaison $b'_1 \dots b'_{n'}$ des n' critères élémentaires optionnels est notée $c_{(B, b'_1 \dots b'_{n'})}$, et est définie comme le sous-ensemble des documents de C_B qui vérifient les critères élémentaires optionnels $k_{i,j}$ pour lesquels on a $b'_{i,j} = 1$.*

$$c_{(B, b'_1 \dots b'_{n'})} = C_B \cap \{D \in \mathcal{E}_D, \forall i \in [1..n'], (b'_{i,j} = 1 \implies k_{i,j}(D))\}$$

Définition 2.5 (Schéma de pertinence homologue à une sous-classe) *Le schéma homologue à une sous-classe de pertinence $c_{(B, b'_1 \dots b'_{n'})}$ se définit de façon analogue au cas d'une classe de dimension. Soit C_B la classe dont la sous-classe $c_{(B, b'_1 \dots b'_{n'})}$ est issue et S_B le schéma de pertinence homologue à cette classe. Cette fois, tant les critères de dimension que les critères élémentaires n'ont plus de statuts associés, puisque le schéma homologue doit définir une seule sous-classe.*

$$S_{(B, b'_1 \dots b'_{n'})} = [(Dim_i, [k_{i,j}]_{j \in J_i})]_{i \in I}$$

avec $I = \{i, \exists \bar{K}_i \in S\} \cup \{i, (\exists \dot{K}_i \in SP \wedge b_i = 1)\}$ et
 $J_i = \{j, \exists \bar{k}_{i,j} \in K_i\} \cup \{j, (\exists \dot{k}_{i,j} \in K_i \wedge b'_{i,j} = 1)\}$

Considérons à nouveau un document D_0 vérifiant la relation de pertinence de la figure IV.1.3 avec le schéma de pertinence de la figure IV.1.1. La figure IV.2.2 montre les schémas homologues des sous-classes de la classe C_0 à laquelle appartient le document D_0 (figure IV.2.1). Cette classe présente deux sous-classes, notées $c_{(0,1)}$ et $c_{(0,0)}$ dont les schémas homologues sont $S_{(0,1)}$ et $S_{(0,0)}$.

2.4 Degré de pertinence

Le degré de pertinence d'une classe ordonne les classes entre elles d'une part, et les sous-classes d'une même classe entre elles d'autre part. Dans tous les cas, le degré d'une classe est inversement proportionnel au nombre de critères optionnels qu'elle vérifie à ce niveau. Plus le degré est petit, plus le nombre de critères optionnels vérifiés est grand.

Définition 2.6 (Degré de pertinence d'une classe de dimension) *Soit $C_{b_1 \dots b_n}$ une classe de dimension avec n critères de dimension optionnels. Son degré de pertinence est défini comme suit.*

$$d^o(C_B) = n + 1 - Card(\{b_i, i \in [1..n] \text{ et } b_i = 1\})$$

Définition 2.7 (Degré de pertinence d'une sous-classe) *Étant donné une sous-classe $c_{(B, b'_1 \dots b'_{\dot{n}'})}$ de la classe de dimension C_B , où $\dot{n}'$ est le nombre de critères élémentaires optionnels pris en compte dans cette sous-classe, le degré de pertinence de $c_{(B, b'_1 \dots b'_{\dot{n}'})}$ est défini localement à C_B , comme suit.*

$$d^o(C_{(B, b'_1 \dots b'_{\dot{n}'})}) = \dot{n}' + 1 - \text{Card}(\{b'_i, i \in [1..\dot{n}'] \text{ et } b'_i = 1\})$$

Ce degré traduit la notion de degré de pertinence, selon l'hypothèse que plus de critères du schéma de pertinence sont vérifiés, plus les documents sont pertinents. Cette hypothèse est conforme à l'hypothèse de la sémantique du schéma de pertinence pour l'utilisateur, qui doit représenter les critères de pertinence souhaitables pour un besoin d'information particulier : un schéma de pertinence ne doit pas contenir des critères de pertinence s'appliquant à des besoins distincts, ce qui limite le schéma de pertinence à une sémantique conjonctive.

Les degrés de pertinence sont compris entre 1 et $\dot{n} + 1$. Le degré 1 caractérise la classe vérifiant tous les critères optionnels et le degré $\dot{n} + 1$ la classe qui n'en vérifie aucun. Cette dernière classe contient tout le corpus.

L'ordre entre classes ou entre sous-classes d'une même classe ainsi défini, est un ordre partiel. Plusieurs classes peuvent avoir le même degré. Cela tient au fait que les statuts ne peuvent prendre que deux valeurs.

Si l'on étend l'ensemble des valeurs de statut de façon à permettre d'ordonner totalement les critères de dimension ou les critères élémentaires, on peut définir un ordre total sur les classes et les sous-classes d'une classe donnée, de la façon suivante.

Définition 2.8 (Degré de pertinence étendu d'une classe de dimension) *Soit $C_{b_1 \dots b_{\dot{n}}}$ une classe de dimension avec $\dot{n}$ critères de dimension optionnels, auxquels est associé un ordre de priorité $o : [1..\dot{n}] \rightarrow [1..\dot{n}]$, avec $o(i) = 1$ pour le critère de dimension optionnel K_i préféré et ainsi de suite par ordre décroissant de préférence. Le degré de pertinence étendu de cette classe est défini comme un couple dont le premier élément est donné par la définition précédente, et le second tient compte des priorités entre les critères.*

$$d^o_{\text{étendu}}(C_B) = (d^o(C_B), \sum(\{o(b_i), i \in [1..\dot{n}] \text{ et } b_i = 1\}))$$

Le second élément du couple n'est pas normalisé, mais cela ne nous empêche pas de définir un ordre sur les classes de même degré, conformément à l'ordre de ces entiers.

Cette définition alternative des degrés n'est pas utilisée par la suite, mais elle nous permet de voir que l'extension de l'ensemble des valeurs des statuts afin d'exprimer des priorités entre les critères, ne pose pas de problème majeur, et permet de préciser l'ordre des classes.

2.5 Classes de pertinence et degrés

Propriété 2.2 (Correction des classes de pertinence) *Les classes de pertinence sont incluses dans l'ensemble des documents retrouvés.*

$$\forall B \in \mathcal{E}_B, C_B \subseteq S(\mathcal{E}_D)$$

Preuve 2.2 *Ce résultat est immédiat d'après les définitions 2.2 et 2.4.*

■

Propriété 2.3 (Inclusion des sous-classes) *Les sous-classes d'une classe de pertinence sont incluses dans la classe dont elles sont issues.*

Une sous-classe c d'une classe C est incluse dans C .

Preuve 2.3 *C'est immédiat au vu de la définition 2.4.* ■

La relation d'inclusion entre les classes, ou entre les sous-classes d'une même classe, définit une relation d'ordre partiel.

Propriété 2.4 (Compatibilité des degrés avec l'inclusion) *Les degrés des classes de pertinence munis de la relation $\leq$ sur les entiers définissent une relation d'ordre partiel compatible avec celle induite par la relation d'inclusion.*

$$(C_1 \subseteq C_2) \Rightarrow (d^o(C_1) \leq d^o(C_2))$$

Preuve 2.4 Par l'absurde :

Hypothèse : $\forall D \in C_1, D \in C_2$

On note $\dot{\mathbf{K}}_1$ l'ensemble des critères optionnels obligatoirement vérifiés dans C_1 et $\dot{\mathbf{K}}_2$ l'ensemble des critères optionnels obligatoirement vérifiés dans C_2 .

Supposons que $d^o(C_1) > d^o(C_2)$.

Alors $\exists K_0 \in \dot{\mathbf{K}}_2, K_0 \notin \dot{\mathbf{K}}_1$.

Soit D_0 tel que $\forall K \in \dot{\mathbf{K}}_1, K(D_0)$ et $\neg(K_0(D_0))$.

On a $D_0 \in C_1$ et $D_0 \notin C_2$.

Contradiction ■

Propriété 2.5 *La réciproque est fausse en général.*

Preuve 2.5 Par l'exhibition d'un contre-exemple :

Soit un schéma de pertinence S où trois critères sont optionnels, K_1, K_2 et K_3 .

$C_1 = C_{110}$ est de degré 2 et $C_2 = C_{001}$ est de degré 3; on a $d^o(C_1) \leq d^o(C_2)$.

Soit D_0 tel que $K_1(D_0) \wedge K_2(D_0) \wedge \neg(K_3(D_0))$.

$D_0 \in C_1$, mais comme on n'a pas $K_3(D_0)$, $D_0 \notin C_2$.

Par conséquent on n'a pas $C_1 \subseteq C_2$. ■

Les classes de pertinence ne sont pas disjointes en général. On a cependant les propriétés suivantes, pour deux classes de pertinence C_1 et C_2 relatives au même niveau hiérarchique (ensemble des classes de dimension, ou ensemble des sous-classes d'une classe de dimension donnée)².

Propriété 2.6 (Inclusion pour l'intersection de deux classes 1) *L'intersection de deux classes C_1 et C_2 de même degré de pertinence d est incluse dans l'union des classes de degré strictement inférieur.*

$$d = d^o(C_1) = d^o(C_2) \implies (C_1 \cap C_2) \subseteq \cup_{d^o(C) < d} C$$

Preuve 2.6 *Étant donné deux classes C_1 et C_2 de même degré de pertinence d , il existe au moins un critère de pertinence K_1 qui soit obligatoirement vérifié par les documents de C_1 et ne soit pas obligatoirement vérifié par les documents de C_2 .*

De même, il existe au moins un critère de pertinence K_2 et un seul qui soit obligatoirement vérifié par les documents de C_2 et ne soit pas obligatoirement vérifié par les documents de C_1 .

Ainsi les seuls documents appartenant à la fois à C_1 et C_2 vérifient nécessairement à la fois K_1 et K_2 , en plus des $k_{opt} - d$ critères nécessairement vérifiés. Cela signifie que tout document appartenant à l'intersection de C_1 et C_2 vérifie tous au moins $k_{opt} - d + 1$ critères optionnel, ce qui implique que l'union des classes de degré strictement inférieur les contiennent. ■

Propriété 2.7 (Égalité pour l'intersection de deux classes de même degré) *Plus précisément, il existe exactement une classe de degré strictement inférieur à d qui est égale à l'intersection de C_1 et C_2 .*

2. Ces propriétés sont utilisées dans la section suivante pour montrer la disjonction des classes visuelles.

$$d = d^o(C_1) = d^o(C_2) \implies \exists! C, (d^o(C) < d) \wedge (C_1 \cap C_2) = C$$

Preuve 2.7 *Étant donné deux classes distinctes C_1 et C_2 de même degré de pertinence d , soit $E_1 = \{K_i\}_{i \in [1..l]}$ l'ensemble des critères optionnels obligatoirement vérifiés dans C_1 et $E_2 = \{K_i\}_{i \in [1..l]}$ l'ensemble des critères optionnels obligatoirement vérifiés dans C_2 .*

1° Montrons d'abord $\exists C, (C_1 \cap C_2) \subseteq C$ et construisons ce C

Si $D \in C_1 \cap C_2$, alors D vérifie à la fois les critères de E_1 et ceux de E_2 . Comme C_1 et C_2 sont distinctes, $E_1 \neq E_2$, mais $E_1 \cap E_2$ n'est pas nécessairement vide. Tous les documents de $C_1 \cap C_2$ vérifient nécessairement au moins $\text{Card}(E_1 \cup E_2)$ critères élémentaires optionnels. Par conséquent ils appartiennent à une classe de degré $d - \text{Card}(E_1 \setminus (E_1 \cap E_2))$, plus précisément à la classe $C_{b_1 \dots b_n}$ avec $b_i = 1$ ssi $K \in E_1 \cup E_2$.

2° Montrons enfin que $C_{b_1 \dots b_n} \subseteq (C_1 \cap C_2)$

Soit $D \in C_{b_1 \dots b_n}$. D vérifie tous les critères de E_1 donc $D \in C_1$, et D vérifie tous les critères de E_2 donc $D \in C_2$, donc $D \in C_1 \cap C_2$. ■

Définition 2.9 (Degré d'un document) *Étant donné un document retrouvé D , parmi les classes de dimension contenant D , il existe une unique classe C_0 de degré minimal (propriété 2.6). Nous définissons le degré $d^o(D)$ d'un document D comme un couple dont le premier élément est le degré de la classe de dimension de degré minimum contenant C_0 , et le second est le minimum de l'ensemble des degrés des sous-classes de C_0 auxquelles il appartient.*

$$d^o(D) = (d^o(C_0), \text{Min}_{\{c, D \in c, c \text{ sous-classe de } C_0\}}(d^o(c)))$$

avec C_0 défini par $d^o(C_0) = \text{Min}_{\{C, D \in C\}}(d^o(C))$

Les degrés de pertinence définissent des classes d'équivalence sur les classes de pertinence selon le nombre de critères optionnels vérifiés.

Dans la perspective d'une visualisation des documents pour l'interaction, ces classes d'équivalence fournissent un guide de présentation selon deux dimensions : sur l'axe horizontal, sont présentées les classes de pertinence de même degré, et sur l'axe vertical s'échelonnent ces classes d'équivalence par ordre décroissant de degré de pertinence. L'axe horizontal n'est pas orienté, au contraire de l'axe vertical.

Ce principe pouvant être appliqué au niveau des classes de dimension et au niveau de chaque sous-classe, l'organisation des documents est finalement multi-dimensionnelle.

Un problème subsiste pour la visualisation : on souhaite présenter des classes de documents disjointes, sans perdre la sémantique des classes quant à la vérification des critères de pertinence.

Chapitre 3

Visualisation selon les classes de pertinence

Les classes de pertinence servent de principe organisateur pour la visualisation des documents retrouvés afin de rendre compte de la relation de pertinence entretenue par les documents avec le schéma de pertinence. Cependant une propriété des classes de pertinence n'est pas souhaitable : les classes de pertinence ne sont pas disjointes, c'est-à-dire qu'un même document peut apparaître dans plusieurs classes, même à un niveau donné (classe ou sous-classe).

Dans ce chapitre nous définissons une organisation des documents retrouvés plus adaptée, selon des *classes visuelles* de pertinence construites sur les classes de pertinence précédemment définies, puis nous décrivons ensuite la façon dont elles sont exploitées dans l'interaction.

3.1 Classes visuelles de pertinence

Après avoir défini les classes visuelles de pertinence en fonction des classes de pertinence, nous vérifions que les classes visuelles vérifient d'une part la propriété de disjonction souhaitée, et qu'elles sont *compatibles* avec les classes de pertinence, c'est-à-dire qu'elles reflètent, pour l'utilisateur, une sémantique analogue à celle traduite par les classes de pertinence.

3.1.1 Définition des classes visuelles de pertinence

Les propriétés souhaitables pour ces classes visuelles sont les suivantes :

Proposition 3.1 *À chaque niveau (classes et sous-classes d'une classe), les classes visuelles doivent être disjointes deux à deux et former ainsi une partition de l'ensemble des documents retrouvés.*

Proposition 3.2 *Les classes visuelles doivent être compatibles avec les classes de pertinence (propriété 3.3).*

Nous donnons une définition de l'ensemble des classes visuelles $\{\hat{C}_B\}_{B \in \mathcal{E}_B}$ en fonction des classes $\{C_B\}_{B \in \mathcal{E}_B}$.

Définition 3.1 (Classes visuelles)

$$\hat{C}_B = C_B \setminus \bigcup_{\{C, d^o(C) < d^o(C_B)\}} (C)$$

On ne visualise de C_B que les documents correspondant strictement aux critères de définition de cette classe. Ceux qui vérifient d'autres critères du schéma de pertinence que ceux-là se trouvent dans d'autres classes visuelles, dont le degré de pertinence est plus petit que le degré de C_B .

Cette définition générale fournit simultanément la définition pour les *sous-classes visuelles*, en substituant c à C .

Propriété 3.1 *Pour tout $B \in \mathcal{E}_B$, on a $\hat{C}_B \subseteq C_B$.*

Preuve 3.1 *Ce résultat est immédiat au vu de la définition 3.1. ■*

Définition 3.2 (Degré d'une classe visuelle) *On définit le degré d'une classe visuelle $d^o(\hat{C}_B)$ comme le degré de la classe C_B dont elle est issue (définitions 2.6 et 2.7).*

$$d^o(\hat{C}_{b_1 \dots b_n}) = d^o(C_{b_1 \dots b_n})$$

Les classes de pertinence, non visuelles, regroupent des documents partageant certaines caractéristiques quant à la relation de pertinence qu'ils entretiennent avec le schéma de pertinence. Une description du sens d'une classe de pertinence est donc donnée par un schéma de pertinence homologue (définitions 2.3 et 2.5). Pour ce qui est d'une classe visuelle, le schéma homologue de la classe dont elle est issue décrit logiquement un ensemble de documents *plus grand* que celui susceptible de se trouver dans cette classe visuelle, puisque cette dernière est privée des documents se trouvant dans les classes de degré inférieur pour ce niveau hiérarchique.

Pour définir les schémas homologues des classes visuelles, un nouveau statut est requis appelé *non-vérifié*. Nous adoptons pour $(K, \text{non-vérifié})$ la notation abrégée $\tilde{K}$.

Définition 3.3 (Schéma homologue d'une classe visuelle de dimension) *Le schéma homologue $\hat{S}_{b_1 \dots b_n}$ à une classe visuelle de dimension $\hat{C}_{b_1 \dots b_n}$ est le résultat de la substitution dans S des $\tilde{K}_i = (K_i, \text{optionnel})$ pour lesquels $b_i = 0$, par les $\tilde{K}_i = (K_i, \text{non-vérifié})$ et des $\tilde{K}_i = (K_i, \text{optionnel})$ pour lesquels $b_i = 1$, par K_i , le statut obligatoire étant sous-entendu.*

La figure IV.3.1 montre les schémas homologues des classes visuelles de dimension issues des classes de dimension représentées dans la figure IV.2.1.

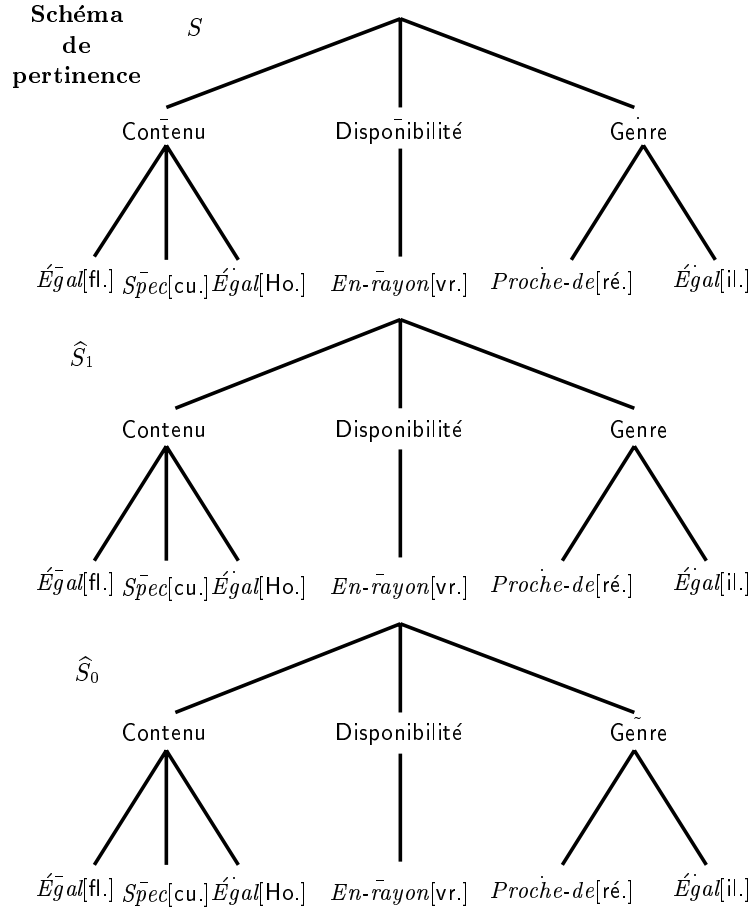


FIG. IV.3.1 – Schémas homologues de classes visuelles de dimension

La sémantique d'un schéma homologue d'une classe visuelle résulte d'une fonction d'évaluation binaire $\widehat{E}_{bin}$ qui définit l'ensemble des documents que la classe contient. La fonction d'évaluation est similaire en tout point à celle des schémas homologues, qui elle-même est identique à celle des schémas de pertinence (E_{bin} dans les définitions 1.5, 1.6 et 1.7), mais nous devons définir la sémantique du nouveau statut *non-vérifié*.

Définition 3.4 (Sémantique du statut non-vérifié pour les dimensions)

$$\widehat{E}_{bin}((K, \text{non-vérifié}))(D) = \neg(E_{bin}((K, \text{obligatoire}))(D))$$

Nous définissons de façon analogue les schémas homologues des sous-classes visuelles.

Définition 3.5 (Schéma homologue d'une sous-classe visuelle) *Le schéma homologue $\widehat{S}_{(b_1 \dots b_n, b'_1 \dots b'_n)}$ à une sous-classe visuelle $\widehat{c}_{b'_1 \dots b'_n}$, est le résultat de la substitution dans S des $\tilde{k}_i$ pour lesquels $b'_i = 0$ par les $\tilde{k}_i = \otimes(k_i, \text{non-vérifié})$ et des $\tilde{k}_i = \otimes(K_i, \text{optionnel})$ pour lesquels $b'_i = 1$, par k_i , le statut obligatoire étant sous-entendu.*

La figure IV.3.2 montre les schémas homologues des sous-classes visuelles de dimension issues des sous-classes représentées dans la figure IV.2.2.

La sémantique du statut *non-vérifié* quand il s'applique à un critère élémentaire est la même que pour un critère de dimension, comme le montre la définition qui suit.

Définition 3.6 (Sémantique du statut non-vérifié pour les crit. élém.)

$$\widehat{E}_{bin}((k, \text{non-vérifié}))(D) = \neg(E_{bin}((k, \text{obligatoire}))(D))$$

3.1.2 Propriétés des classes visuelles

Nous montrons que les classes visuelles de dimension sont disjointes deux à deux, et que les sous-classes visuelles d'une classe visuelle de dimension sont disjointes deux à deux.

Propriété 3.2 (Disjonction des classes visuelles) *Pour un niveau hiérarchique donné (ensemble des classes visuelles de dimension ou ensemble des sous-classes visuelles d'une classe visuelle donnée), les classes visuelles sont disjointes deux à deux.*

Preuve 3.2 En deux étapes :

1° Les unions des classes visuelles de même degré sont des ensembles disjoints deux à deux.

Par l'absurde :

Hypothèse : Supposons qu'il existe un document D tel que $D \in E_1 \cap E_2$, où $E_1 = \cup_{d^\circ(\widehat{C})=d_1}(\widehat{C})$ est l'union des classes visuelles de degré d_1 et $E_2 = \cup_{d^\circ(\widehat{C})=d_2}(\widehat{C})$ l'union des classes visuelles de degré d_2 .

Prenons pour fixer les idées

$$d_1 < d_2 \quad (i)$$

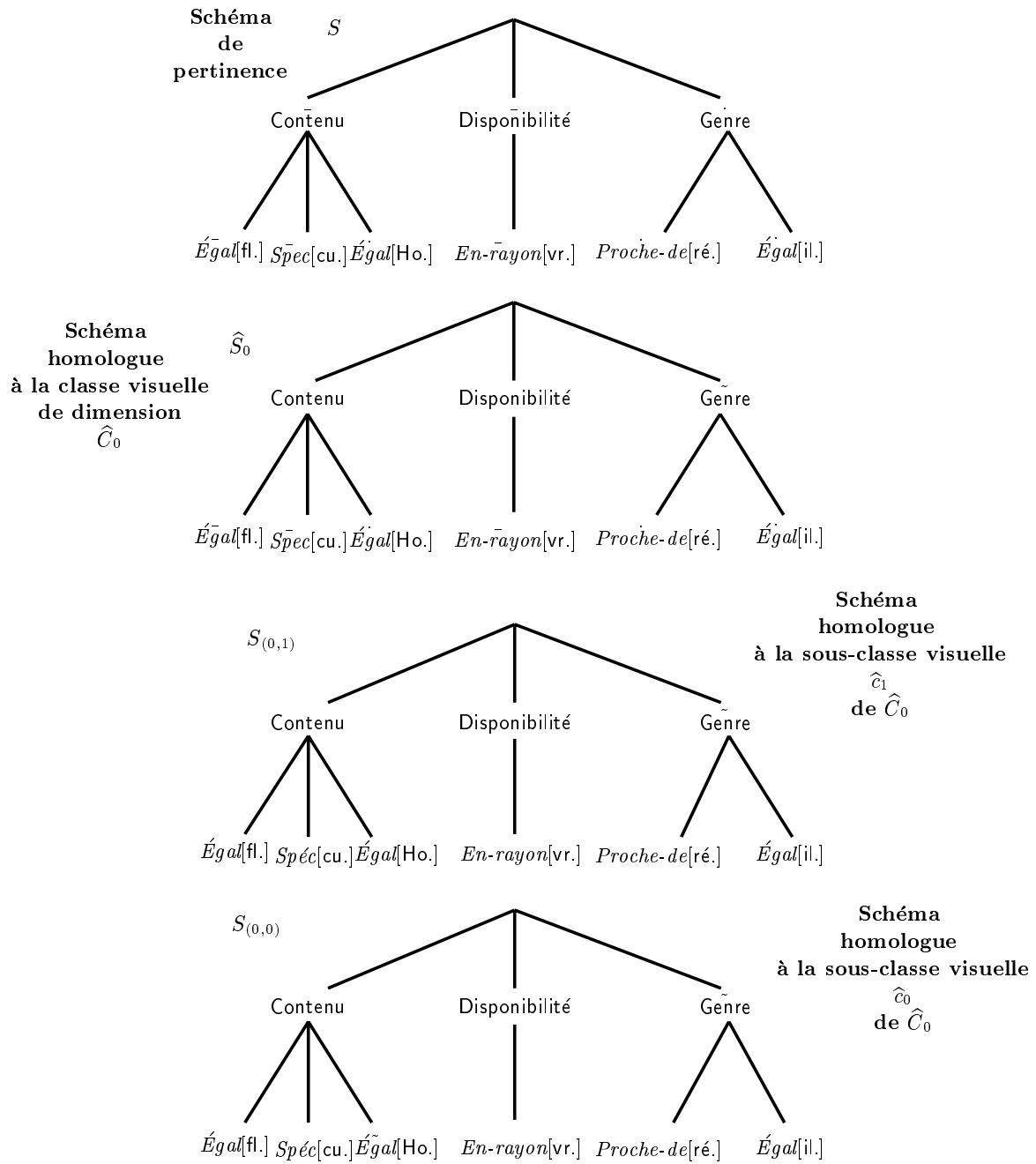


FIG. IV.3.2 – Schémas homologues de sous-classes visuelles

Comme $D \in E_2$, D appartient au moins à l'une des classes visuelles de degré d_2 , et on a

$$\exists \widehat{C}_2, d^o(\widehat{C}_2) = d_2 \wedge D \in \widehat{C}_2 \quad (ii)$$

De même, comme $D \in E_1$ on a

$$\exists \widehat{C}_1, d^o(\widehat{C}_1) = d_1 \wedge D \in \widehat{C}_1 \quad (iii)$$

Comme $d^o(\widehat{C}_1) = d^o(C_1)$ par la définition du degré d'une classe visuelle (définition 3.2), (i) implique $d^o(C_1) < d_2$, où C_1 et C_2 sont les classes de pertinence dont sont issues $\widehat{C}_1$ et $\widehat{C}_2$ respectivement.

Donc

$$C_1 \in \{C_k, d^o(C_k) < d_2\} \quad (iv)$$

Comme par définition (définition 3.1) $\widehat{C}_2 = C_2 \setminus \bigcup_{\{k, d^o(C_k) < d_2\}} (C_k)$, (iv) implique

$$\widehat{C}_2 \cap C_1 = \emptyset \quad (v)$$

Comme $\widehat{C}_1 \subseteq C_1$ (propriété 3.1) et $D \in \widehat{C}_1$ (d'après (iii)), on a

$$D \in C_1 \quad (vi)$$

Ainsi on a à la fois $D \in C_1 \cap \widehat{C}_2$ (d'après (vi) et (ii)), et $\widehat{C}_2 \cap C_1 = \emptyset$ (d'après (v)).

Contradiction.

2° Deux classes visuelles de même degré sont disjointes.

Par l'absurde et par récurrence décroissante sur le degré des classes :

Hypothèse: Supposons qu'il existe un document D et deux classes visuelles $\widehat{C}_1$ et $\widehat{C}_2$ tels que $d^o(\widehat{C}_1) = d^o(\widehat{C}_2) = d$ et

$$D \in \widehat{C}_1 \cap \widehat{C}_2 \quad (vii)$$

et montrons qu'il y a contradiction.

Cas de base :

Montrons que pour tout D vérifiant (vii) il existe C_0 de degré $d_0 < d$ tel que $D \in C_0$ et $D \notin \widehat{C}_0$.

Par la propriété 3.1, on a $\widehat{C}_1 \subseteq C_1$, et $\widehat{C}_2 \subseteq C_2$, donc

$$D \in C_1 \cap C_2 \quad (\text{viii})$$

Comme $d = d^o(C_1) = d^o(C_2)$, la propriété 2.6 s'applique et on a

$$(C_1 \cap C_2) \subseteq \cup_{d^o(C) < d} C \quad (\text{ix})$$

(viii) et (ix) impliquent que D appartient à au moins l'une des classes de degré strictement inférieur à d , c.-à-d. $\exists C, (d^o(C) < d_0) \wedge (D \in C)$.

Appelons C_0 cette classe, d_0 son degré et $\widehat{C}_0$ la classe visuelle associée. On a donc

$$D \in C_0 \quad (\text{x})$$

Par le 1° on sait que les unions des classes visuelles de même degré sont disjointes, donc comme $D \in \widehat{C}_1$ par hypothèse (vii), et comme $d^o(\widehat{C}_0) = d^o(C_0) < d_0$ par définition de C_0 (x), on a

$$D \notin \widehat{C}_0 \quad (\text{xi})$$

Cas général :

Montrons que pour tout D vérifiant (vii), s'il existe C_i de degré $d_i < d$ tel que $D \in C_i$ et $D \notin \widehat{C}_i$, alors il existe C_{i+1} de degré strictement inférieur à $d_{i+1} < d_i$ tel que $D \in C_{i+1}$ et $D \notin \widehat{C}_{i+1}$.

Supposons donc qu'il existe C_i de degré $d_i < d$ tel que

$$(D \in C_i) \wedge (D \notin \widehat{C}_i) \quad (\text{xii})$$

D'après la définition 3.1, on a $\widehat{C}_i = C_i \setminus \bigcup_{\{k, d^o(C_k) < d_i\}} (C_k)$. Par suite, (xii) implique $D \in \bigcup_{\{k, d^o(C_k) < d_i\}} (C_k)$.

Par suite il existe C_{i+1} de degré $d_{i+1} < d_i$ tel que $D \in C_{i+1}$.

Comme $D \in \widehat{C}_1$ (d'après (vii)), et comme $d^o(\widehat{C}_{i+1}) = d^o(C_{i+1}) = d_{i+1}$ et $d_{i+1} < d$ implique, par le 1°, que D ne peut appartenir à cette classe visuelle, et on a

$$D \notin \widehat{C}_{i+1} \quad (\text{xiii})$$

Conclusion de la récurrence :

Par récurrence décroissante sur le degré des classes, on obtient pour D vérifiant (vii) ,

$$\forall d' < d, \exists C_i, (D \in C_i) \wedge (D \notin \widehat{C}_i) \quad (xiv)$$

En particulier, pour $d' = 1$, une seule classe existe : c'est $C_{1\dots 1}$. Donc on a nécessairement $D \in C_{1\dots 1}$ et $D \notin \widehat{C}_{1\dots 1}$.

Or $C_{1\dots 1} = \widehat{C}_{1\dots 1}$.

Contradiction. ■

Ainsi, l'ensemble de toutes les sous-classes visuelles forme une partition de l'ensemble des documents retrouvés, et l'ensemble des classes visuelles de dimension forme une partition moins fine des documents en regroupant des sous-classes visuelles, ce qui satisfait la proposition 3.1.

Il faut encore nous assurer que les classes visuelles sont compatibles avec les classes de pertinence (proposition 3.2), c'est-à-dire que le degré du document obtenu à partir des classes visuelles est le même que celui obtenu à partir des classes de pertinence.

Propriété 3.3 (Compatibilité de la visualisation) *La visualisation des documents est compatible avec les classes de pertinence ssi*

$$\forall D, (D \in \widehat{C} \cap \widehat{c}) \implies d^o(D) = (d^o(\widehat{C}), d^o(\widehat{c}))$$

Preuve 3.3 *Soit D un document retrouvé de degré (d_1, d_2) .*

Comme les classes visuelles de dimension constituent une partition de l'ensemble des documents retrouvés (propriété 3.2), il existe un unique $\widehat{C}_0$ tel que $D \in \widehat{C}_0$. De même, l'ensemble des sous-classes visuelles de $\widehat{C}_0$ formant une partition de $\widehat{C}_0$, il existe un unique $\widehat{c}_0$ tel que $D \in \widehat{c}_0$.

Comme pour tout $\widehat{C}$, $\widehat{C} \subseteq C$ (propriété 3.1), on a aussi $D \in C_0 \cap c_0$, où C_0 et c_0 désigne les classe de dimension et sous-classe dont $\widehat{C}_0$ et $\widehat{c}_0$ sont issues, respectivement.

Par suite, $d_1 \leq d^o(C_0)$ et $d_2 \leq d^o(c_0)$, par définition du degré d'un document (définition 2.9).

Hypothèse : Supposons qu'il existe C' telle que $d^o(C') < d_1$ et $D \in C'$.

Comme $\widehat{C}_0 = C_0 \setminus \bigcup_{\{C_j, d^o(C_j) < d^o(C_0)\}} (C_j)$ par définition (définition 3.1), $D \in \widehat{C}_0$ implique $D \notin C'$.

Contradiction.

Par conséquent $d_1 = d^o(C_0) = d^o(\widehat{C}_0)$, par définition du degré d'une classe visuelle de dimension (définition 3.2).

Hypothèse: De même, supposons qu'il existe c' telle que $d^o(c') < d_2$ et $D \in c'$.

Comme $\widehat{c}_0 = c_0 \setminus \bigcup_{\{c_j, d^o(c_j) < d^o(c_0)\}} (c_j)$ par définition (définition 3.1), $D \in \widehat{c}_0$ implique $D \notin c'$.

Contradiction.

Par conséquent $d_2 = d^o(c_0) = d^o(\widehat{c}_0)$ par définition du degré d'une sous-classe visuelle (définition 3.2). ■

Nous donnons une définition récursive alternative pour les classes visuelles, qui met en évidence la façon dont les classes visuelles se construisent à partir des classes de pertinence. La classe visuelle de degré 1 est égale à la classe de pertinence de degré 1. Une classe visuelle donnée est égale à la classe de pertinence analogue dont on a supprimé les documents présents dans les classes visuelles de degré inférieur (que l'on a déjà définies).

Définition 3.7 (Classes visuelles, définition alternative)

$$\begin{cases} \widehat{C}_{1\dots 1} &= C_{1\dots 1} \\ \widehat{C}_B &= C_B \setminus \bigcup_{\{\widehat{C}, d^o(\widehat{C}) < d^o(\widehat{C}_B)\}} (\widehat{C}) \quad \text{pour } B \neq 1\dots 1 \end{cases}$$

Propriété 3.4 Cette définition est équivalente à la définition 3.1.

Preuve 3.4 Par récurrence sur le degré de $\widehat{C}'$

On note $\widehat{C}$ les classes visuelles selon la première définition et $\widehat{C}'$ les classes visuelles selon la définition alternative.

D'après la définition des degrés, on a pour tout B $d^o(\widehat{C}_B) = d^o(\widehat{C}'_B)$.

Pour $B = 1\dots 1$, on a bien

$$\widehat{C}_{1\dots 1} = C_{1\dots 1} = \widehat{C}'_{1\dots 1}$$

Pour $B \neq 1\dots 1$,

$$\widehat{C}_B = C_B \setminus \bigcup_{\{C, d^o(C) < d^o(C_B)\}} (C)$$

Montrons-le par récurrence sur le degré de $\widehat{C}'$.

Supposons que pour tout d et pour tout $\hat{C}'$ tel que $d^o(\hat{C}') = d$, on a $\hat{C}' = \hat{C}$.

Soit $\hat{C}'_B$ tel que $d^o(\hat{C}'_B) = d + 1$.

$$\hat{C}'_B = C_B \setminus \bigcup_{\{\hat{C}', d^o(\hat{C}') < d+1\}} (\hat{C}')$$

$$\hat{C}'_B = C_B \setminus \bigcup_{\{C, d^o(C) < d+1\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d^o(C)\}} (C_0))$$

Considérons le second terme $\bigcup_{\{C, d^o(C) < d+1\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d^o(C)\}} (C_0))$.

$$\begin{aligned} & \bigcup_{\{C, d^o(C) < d+1\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d^o(C)\}} (C_0)) = \\ & \bigcup_{\{C, d^o(C)=d\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d\}} (C_0)) \cup \bigcup_{\{C, d^o(C)=d-1\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d-1\}} (C_0)) \\ & \quad \cup \dots \cup \\ & \bigcup_{\{C, d^o(C)=2\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < 2\}} (C_0)) \cup \bigcup_{\{C, d^o(C)=1\}} (C) \end{aligned}$$

Notons $F_d = \bigcup_{\{C, d^o(C)=d\}} (C)$.

Pour tout x , on a

$$\bigcup_{\{C, d^o(C)=x\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < x\}} (C_0)) = \bigcup_{\{C, d^o(C)=x\}} (C) \setminus \bigcup_{\{C_0, d^o(C_0) < x\}} (C_0)$$

c.-à-d.

$$\bigcup_{\{C, d^o(C)=x\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < x\}} (C_0)) = F_x \setminus \bigcup_{i \in [1..x-1]} (F_i)$$

Le second terme s'écrit donc :

$$\begin{aligned} & \bigcup_{\{C, d^o(C) < d+1\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d^o(C)\}} (C_0)) = \\ & (F_d \setminus \bigcup_{i=1}^{d-1} (F_i)) \cup (F_{d-1} \setminus \bigcup_{i=1}^{d-2} (F_i)) \cup \dots \cup (F_2 \setminus \bigcup_{i=1}^1 (F_i)) \cup F_1 \end{aligned}$$

Or pour tout x , $F_x \setminus \bigcup_{i=1}^{x-1} (F_i) = \bigcup_{\{C, d^o(C)=x\}} (C)$.

Donc le second terme se simplifie en :

$$\begin{aligned} & \bigcup_{\{C, d^o(C) < d+1\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d^o(C)\}} (C_0)) = \\ & \bigcup_{\{C, d^o(C)=d\}} (C) \cup \bigcup_{\{C, d^o(C)=d-1\}} (C) \cup \dots \cup \bigcup_{\{C, d^o(C)=2\}} (C) \cup \bigcup_{\{C, d^o(C)=1\}} (C) \end{aligned}$$

c.-à-d.

$$\bigcup_{\{C, d^o(C) < d+1\}} (C \setminus \bigcup_{\{C_0, d^o(C_0) < d^o(C)\}} (C_0)) = \bigcup_{\{C, d^o(C) < d+1\}} (C)$$

Ainsi,

$$\hat{C}'_B = C_B \setminus \bigcup_{\{C, d^o(C) < d+1\}} (C)$$

et comme $d^o(C_B) = d + 1$, on a

$$\hat{C}'_B = C_B \setminus \bigcup_{\{C, d^o(C) < d^o(C_B)\}} (C) = \hat{C}_B$$

■

3.1.3 Conclusion

Ainsi nous avons défini pour toute requête et tout schéma de pertinence s'y appliquant, un ensemble de classes visuelles assorti d'un ordre partiel donné par la relation d'ordre sur les degrés des classes visuelles. Nous avons garanti la compatibilité de la sémantique de ces classes visuelles vis-à-vis des classes de pertinence, ainsi que la sémantique de leurs degrés vis-à-vis des degrés de pertinence. En particulier, la classe de degré $n + 1$ correspond à l'ensemble des documents non retrouvés.

La structure qui résulte des propriétés des classes visuelles est une partition de l'ensemble des documents retrouvés qui sont ordonnées partiellement selon les degrés des classes.

Les définitions et propriétés données dans cette section s'appliquent aussi aux sous-classes de pertinence, et permettent de définir les sous-classes visuelles.

3.2 Classes visuelles et interaction

3.2.1 Structure des classes visuelles

Étant donné un ensemble de classes visuelles $\mathcal{E}_{\hat{C}}$, correspondant à l'ensemble de classes visuelles de dimension ou à l'ensemble des sous-classes visuelles d'une classe visuelle de dimension donnée, les degrés définissent autant de classes d'équivalence sur cet ensemble qu'il y a de degrés. Pour le degré d , la classe d'équivalence associée est notée $\mathcal{E}_{\hat{C}}/d$ et est définie comme l'ensemble des classes visuelles de $\mathcal{E}_{\hat{C}}$ de degré d .

Définition 3.8 (Classe d'équivalence)

$$\mathcal{E}_{\hat{C}}/d = \{\hat{C} \in \mathcal{E}_{\hat{C}}, d^o(\hat{C}) = d\}$$

La visualisation des documents se fait au moyen des classes visuelles, en exploitant la structure que leur confèrent les deux niveaux hiérarchiques (classes et sous-classes) et les classes d'équivalence selon les degrés.

Une visualisation exhaustive des documents dans le plan implique une visualisation récursive, où la visualisation de l'ensemble des classes est analogue à la visualisation de l'ensemble des sous-classes d'une classe particulière.

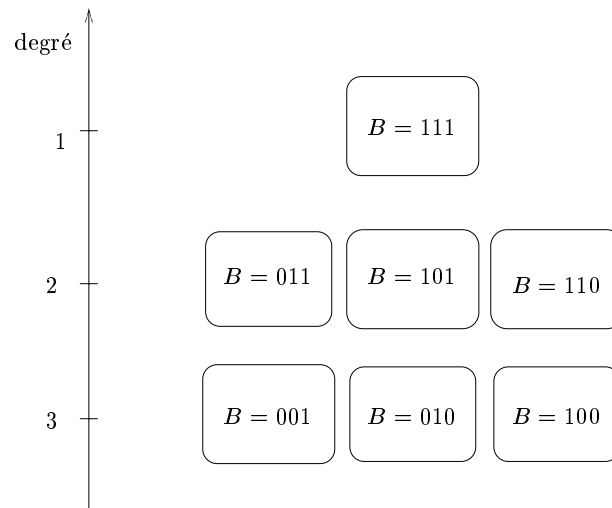
On appelle *plan de visualisation* la visualisation d'un niveau hiérarchique de classes, c.-à-d. l'ensemble des classes ou l'ensemble des sous-classes d'une classe, en omettant les classes de degré maximum, qui correspondent aux documents non retrouvés (section 2.4 page 116).

La figure IV.3.3 décrit un plan de visualisation avec les classes visuelles de dimension pour un schéma de pertinence avec trois critères de dimension optionnels. Les degrés sont représentés sur l'axe vertical selon une orientation sud-nord pour les degrés décroissant. L'axe horizontal n'est pas orienté.

Ce même schéma illustre un plan de visualisation des sous-classes d'une classe de pertinence particulière, pour laquelle on aurait aussi trois critères élémentaires optionnels.

Une visualisation exhaustive dans le plan se présente comme montré dans la figure IV.3.4.

Ce genre de visualisation n'est en général pas praticable, dans la mesure où l'espace de visualisation des documents sur l'écran serait trop réduite. De plus, l'utilité de présenter un si grand nombre d'informations dans leurs plus petits détails est douteuse, car un individu ne peut les appréhender. En revanche, une telle vue globale du résultat peut être utile pour permettre de prendre du recul. Pour la représenter, le concepteur du système doit avoir recours à des techniques de représentation symbolique du contenu des classes visuelles, pour lesquelles les schémas homologues aux classes visuelles constituent un bon support de conception.

FIG. IV.3.3 – *Exemple de plan de visualisation*

3.2.2 Modes de visualisation et manipulation

Les classes de pertinence sont définies de manière à associer directement le schéma de pertinence aux caractéristiques des documents contenus dans les classes, et cela par l'intermédiaire des schémas homologues qui mettent en évidence les caractéristiques de pertinence exprimées dans le schéma, qui sont partagées par les documents d'une même classe.

En fonction du contexte informatique dans lequel le système de recherche d'information est développé, la structure de visualisation décrite dans la section précédente peut être exploitée de différentes manières.

En plus de ce paramètre, d'autres paramètres liés à la problématique de la recherche d'information interviennent dans la conception de la visualisation à adopter. D'abord la nature des documents du corpus intervient. Selon le média, la taille des documents à visualiser, ou l'homogénéité du corpus, plus ou moins de documents pourront être présentés à la fois sur l'écran, ce qui contraint les possibilités de visualisation.

De manière plus intrinsèque à la problématique de la recherche d'information, en fonction de la nature de la sous-tâche que l'utilisateur cherche à accomplir, certains modes de visualisation sont plus appropriés que d'autres. Par exemple, si l'utilisateur cherche à évaluer sa satisfaction quant au résultat produit de manière globale, une vue d'ensemble des documents est requise. En revanche, pour raffiner un aspect du schéma de pertinence de manière à mieux le cibler, une visualisation focalisée sur les sous-classes d'une classe particulière est plus utile pour prendre conscience des

effets produits par les différents éléments de la requête, par exemple.

Ainsi, la démarche à adopter est la spécification de modes de visualisation adaptés pour certains à la production d'un jugement global de satisfaction et pour d'autres à la réalisation de diverses sous-tâches engendrées par les situations de non-satisfaction qui peuvent être rencontrées.

Dans la partie suivante, nous revenons sur ces notions et nous les appliquons à la conception d'un système de recherche d'information pour un corpus d'images.

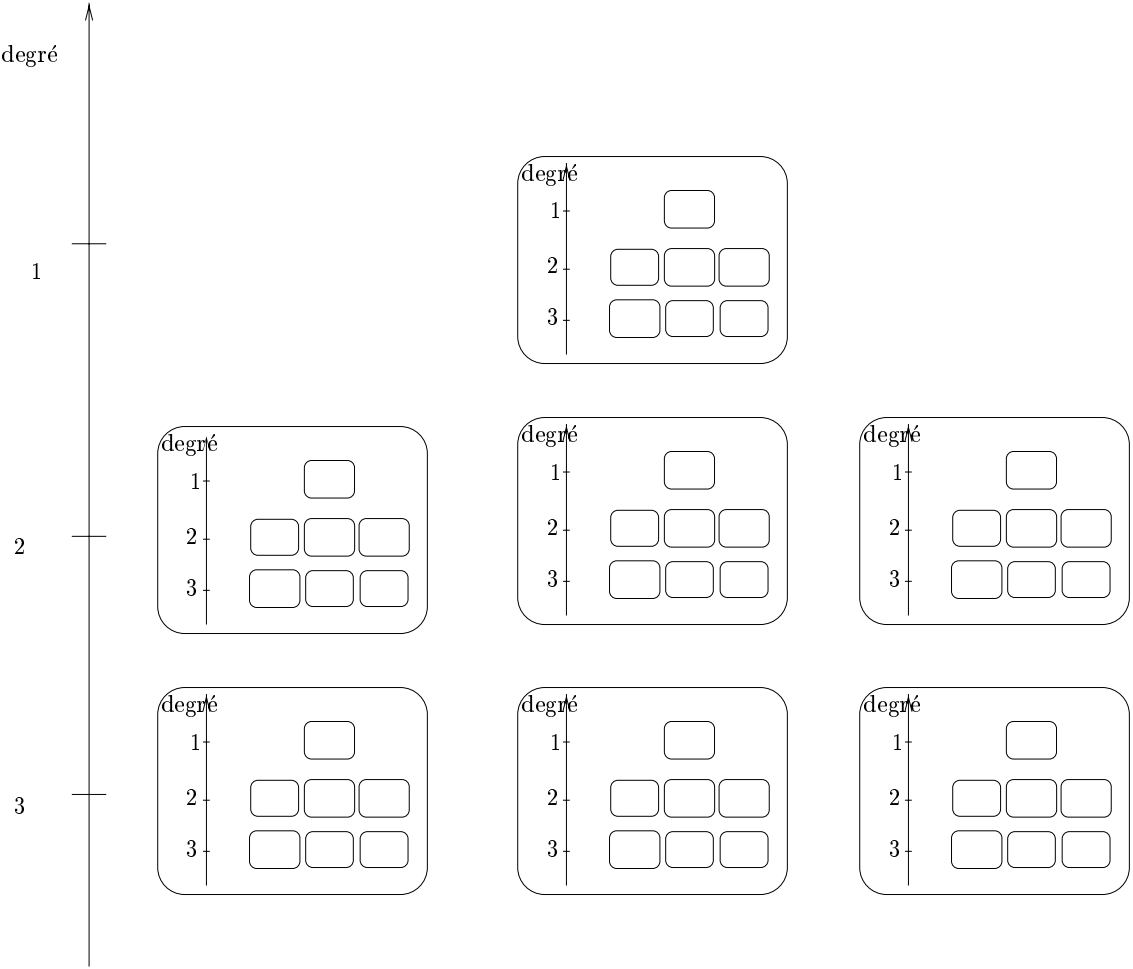


FIG. IV.3.4 – *Exemple de visualisation exhaustive récursive*

Chapitre 4

Utilisation interactive du système

Nous étudions ici les caractéristiques du processus itératif de recherche d'information dans la perspective offerte par notre modèle. Un processus itératif est une succession de *reformulations* du schéma de pertinence, chaque reformulation s'inscrivant dans une situation appelée *situation courante*.

La *reformulation manuelle* remet le *contrôle* du déroulement du processus entre les mains de l'utilisateur, qui évalue la situation courante pour déterminer la modification à apporter au schéma de pertinence et procéder manuellement à cette modification. Dans ce cadre, l'utilisateur élabore seul une stratégie d'utilisation du système pour résoudre son problème d'information.

L'éventail des possibilités de reformulation qui se présentent à un instant donné est grand, ce pourquoi il est souhaitable que le système assiste l'utilisateur dans l'élaboration de sa stratégie.

C'est dans ce but que nous étudions les possibilités d'automatisation de la reformulation selon les trois axes suivants :

- l'identification de caractéristiques système de la situation courante, relatives à la configuration de la réponse, qui donne au système un point de départ pour contrôler le déroulement du processus ; c'est ce que nous appelons la *vue système de la situation*, qui permet de définir un ensemble de situations stéréotypiques ;
- la définition d'*associations situation stéréotypique - action stéréotypique*, qui définit d'un point de vue logique le contrôle système du déroulement du processus ;
- l'automatisation de la réalisation concrète d'une reformulation du schéma de pertinence, qui se traduit par la définition d'un ensemble d'*opérations* sur le schéma de pertinence qui peuvent instancier les actions stéréotypiques.

Dans ce chapitre nous définissons d'abord l'objectif d'une utilisation du système (une session de recherche) et les critères d'évaluation de la qualité du déroulement d'une session. Une session s'organise selon une boucle principale de reformulation du schéma de pertinence dans laquelle se succèdent les tâches principales d'évaluation et de reformulation.

Nous étudions ensuite l'éventail des situations qui peuvent conduire à la réalisation de ces tâches, afin de définir un ensemble de situations stéréotypiques d'une part, et un ensemble d'actions répondant à ces situations d'autre part.

Dans ces deux premières sections, nous nous situons dans la perspective d'une utilisation d'un système de recherche d'information où la reformulation est faite de façon purement manuelle. Ainsi incombent à l'utilisateur toutes les tâches qui ne relèvent pas des fonctions internes au système (la correspondance et la mise en forme), et en particulier le contrôle de la correspondance et de la mise en forme, qui sont gérés par l'intermédiaire de modifications du schéma de pertinence. Nous faisons donc l'hypothèse que l'utilisateur est capable de développer une stratégie de recherche qui détermine le déroulement de la boucle principale, par la définition successive de sous-problèmes.

Dans la dernière section, nous discutons de la possibilité d'automatiser l'identification de la situation courante et la production de la reformulation adéquate du schéma de pertinence, afin de fournir à l'utilisateur un guide qui l'aide à développer sa stratégie. Nous comparons ces possibilités aux diverses techniques de reformulation automatique existantes.

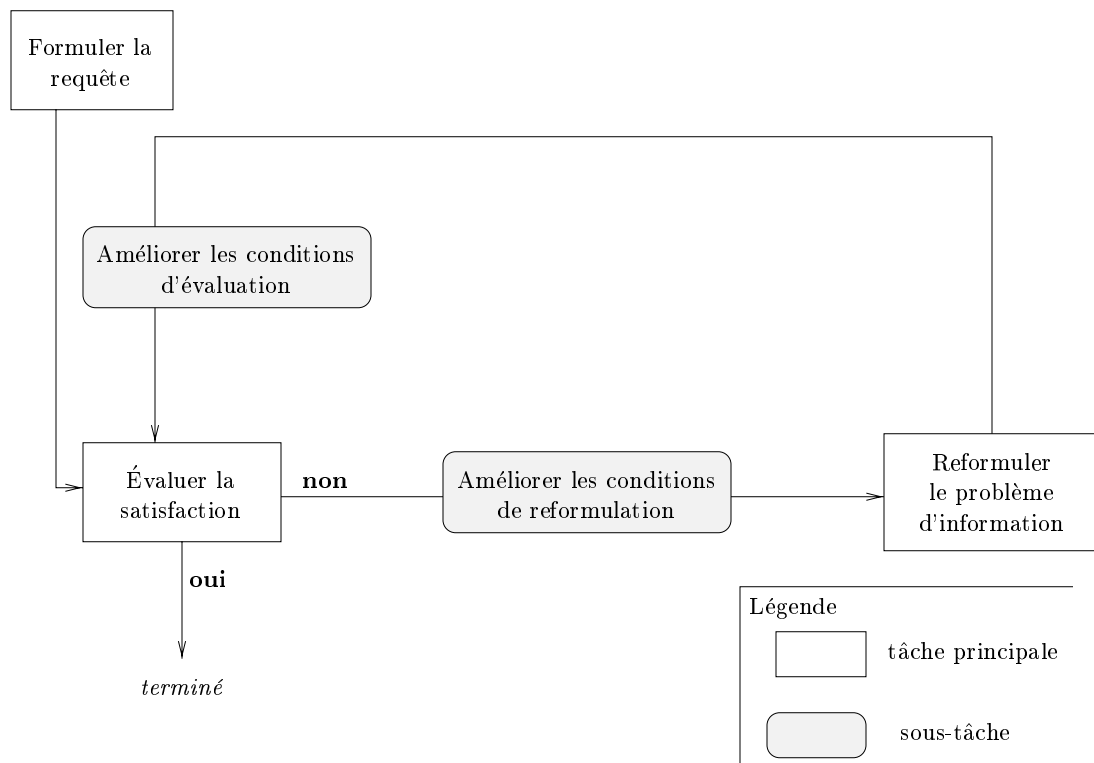
4.1 Objectif de l'interaction

Les tâches principales sont au nombre de trois : formuler la requête, évaluer la satisfaction, reformuler le problème d'information en cas de non-satisfaction.

Les conditions dans lesquelles ces tâches doivent être accomplies interviennent dans les performances du système. Ainsi, nous définissons deux sous-tâches dont l'objectif est de favoriser la réalisation des tâches principales : en amont de l'évaluation, intervient la sous-tâche «améliorer les conditions d'évaluation», et en amont de la reformulation, la sous-tâche «améliorer les conditions de reformulation» (figure IV.4.1). Chacune de ces sous-tâches induit une boucle imbriquée, comme le montre la figure IV.4.2.

Le système intervient dans chaque boucle : après chaque reformulation, il évalue le corpus selon le nouveau schéma de pertinence et produit un nouveau résultat, objet de la tâche d'évaluation suivante¹.

1. Comme nous le verrons plus loin, dans le cadre d'une automatisation de la reformulation, le système contribue, voire gère seul, les tâches d'évaluation et de reformulation des boucles imbriquées, ainsi que la tâche de reformulation de la boucle principale

FIG. IV.4.1 – *Tâches et sous-tâches*

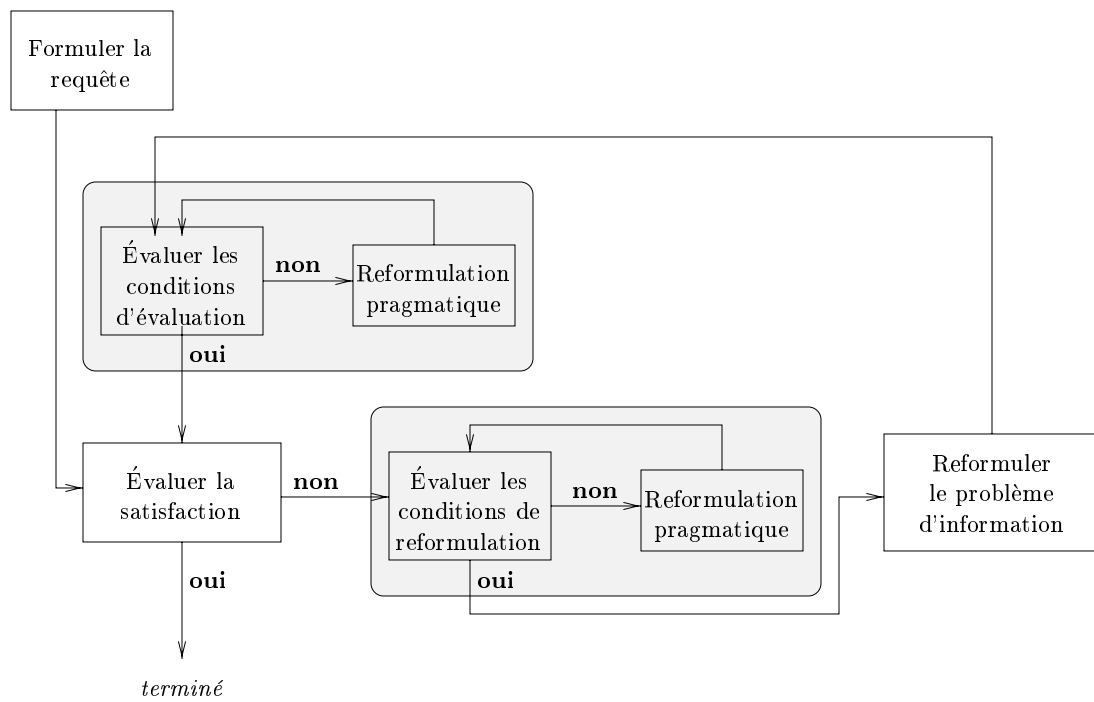


FIG. IV.4.2 – Boucles imbriquées des sous-tâches

Dans la suite, nous définissons l'objectif d'une session de recherche ainsi que notre critère général d'évaluation d'une session, afin de montrer comment notre modèle de pertinence système permet de satisfaire à ces critères. Pour cela, nous définissons la notion de situation problématique et la notion d'action associée en réponse à une telle situation, sous la forme d'une modification du schéma de pertinence qui permet de poursuivre le processus de recherche.

4.1.1 État de satisfaction optimale et évaluation d'une session

L'objectif final du processus de recherche est, pour l'utilisateur, d'atteindre un état de satisfaction optimal. Nous définissons un *état de satisfaction optimale* de manière opérationnelle : c'est un état où l'utilisateur considère avoir tiré le meilleur du corpus et termine l'interaction avec le système.

Cet objectif est différent de l'objectif «résoudre le problème d'information initial». Par exemple, dans le cas où le corpus ne contient aucun document pertinent, un utilisateur peut atteindre un état de satisfaction optimale sans avoir trouvé un seul document pertinent. Cependant il n'aura atteint cet état de satisfaction optimale que s'il est convaincu d'avoir tiré le meilleur du corpus. L'état de satisfaction optimale est donc assujetti au niveau de compréhension des réponses du système, au niveau de maîtrise de ses fonctionnalités, et au niveau de clarté de la vue sur le corpus que l'utilisateur a atteints.

Ainsi, l'objectif d'atteindre un état de satisfaction optimale met en jeu les critères classiques d'évaluation d'une session de recherche (retrouver des documents pertinents) *en conjonction* avec d'autres critères relatifs, orientés vers l'utilisateur. La qualité d'une session de recherche ne peut plus s'exprimer uniquement en termes de rappel et de précision, mais doit prendre en compte la qualité du déroulement du processus de recherche.

Considérant l'utilisation des systèmes comme un processus fondamentalement itératif, nous définissons une *bonne session* de recherche comme une session où à chaque étape, l'utilisateur dispose des éléments qui lui permettent de procéder à l'itération suivante, jusqu'à ce qu'il atteigne un état de satisfaction optimale ; une mauvaise session est par exemple une session où l'utilisateur atteint une impasse dans le déroulement du processus de recherche (il ne sait pas comment réagir à la situation dans laquelle il se trouve), sans avoir atteint un tel état de satisfaction.

Parmi les facteurs qui interviennent dans le déroulement du processus de recherche, nous nous intéressons exclusivement à ceux qui relèvent des fonctions du système liées à la pertinence système (fonction de correspondance et mise en forme). Nous laissons de côté les facteurs liés à la réalisation des interfaces des systèmes interactifs en général, et qui ne font pas directement référence à la fonction spécifique

de recherche d'information.

4.1.2 Situation de non-satisfaction et situation d'entrave

À une étape donnée d'une session, deux types de situations problématiques peuvent être rencontrés.

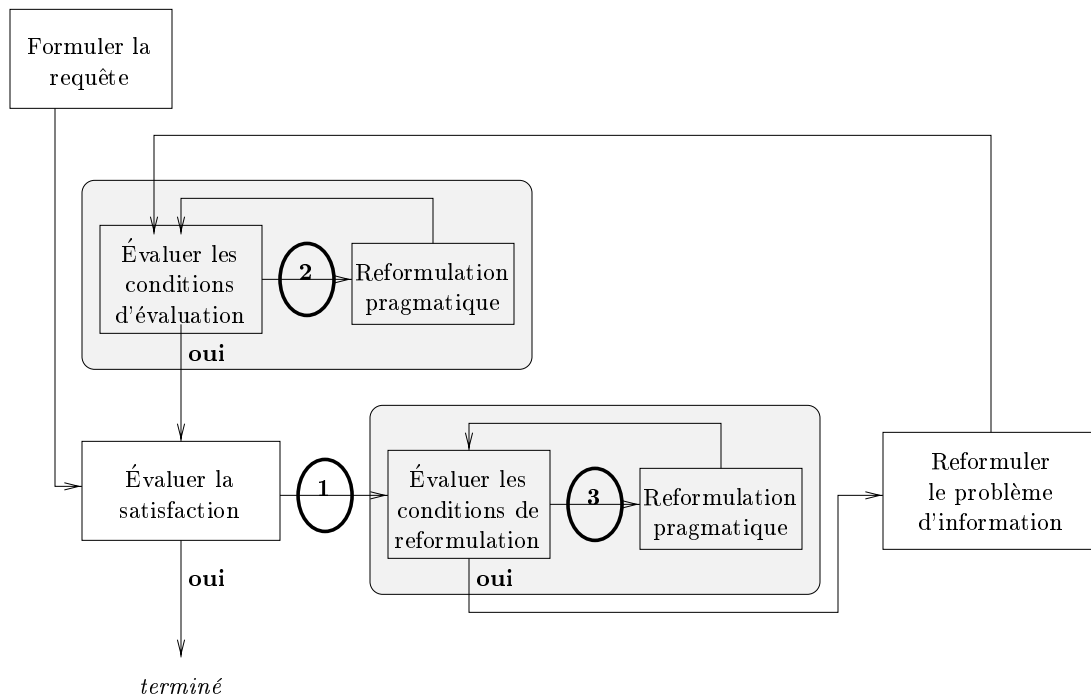


FIG. IV.4.3 – Boucles imbriquées des sous-tâches

Situation de non-satisfaction

Le premier type de situation, appelé *situation de non-satisfaction*, se rencontre après chaque tâche d'évaluation de la satisfaction, dans la boucle principale (identifiée par le cercle en gras numéro 1 dans la figure IV.4.3). Dans une telle situation, une reformulation du schéma de pertinence est requise, dans le but de résoudre un sous-problème qui permet à l'utilisateur de se rapprocher d'un état de satisfaction optimale.

Ce sous-problème d'information peut être de nature sémantique. Dans ce cas il relève soit d'une incompatibilité entre la sémantique que le système attribue à

un critère et la sémantique que lui attribue l'utilisateur dans son schéma individuel de pertinence, soit du manque de complétude de la formulation du problème d'information par rapport au contenu du corpus.

Le sous-problème d'information peut aussi être de nature pragmatique : il s'agit dans ce cas d'éclaircir le rapport entre les documents pertinents retrouvés jusque-là et le reste du corpus.

Ainsi, la reformulation du schéma de pertinence qui intervient dans ce premier type de situation est de nature à en modifier les *paramètres sémantiques* ou les *paramètres pragmatiques* selon les cas.

Situation d'entrave à la poursuite du processus itératif

Le second type de situation problématique, appelé *situation d'entrave à la poursuite du processus itératif*, se rencontre en amont de la réalisation de chacune des tâches principales (évaluation de la satisfaction et reformulation du problème d'information, identifiées par les cercles en gras numéro 2 et 3 dans la figure IV.4.3). Une situation d'entrave est une situation dans laquelle les conditions pour réaliser la tâche principale courante ne sont pas bonnes.

Dans un tel type de situation, une reformulation du schéma de pertinence est requise, dans le but d'améliorer ces conditions. Ainsi, la reformulation du schéma de pertinence qui intervient dans ce second type de situation est de nature à en modifier les *paramètres pragmatiques*, qui déterminent les conditions de visualisation du résultat du système.

4.1.3 Niveau de notre formalisation du déroulement d'une session de recherche

Vue système de la situation courante

Une *vue système de la situation* courante est un ensemble de mesures statistiques sur la réponse du système comme le nombre de documents retrouvés, le nombre de classes, le nombre de documents dans une classe donnée, le nombre de documents pertinents, etc.

Ces informations globales sur la réponse courante du système permettent de définir un ensemble de *situations stéréotypiques* identifiables par le système, comme un ensemble d'instanciations stéréotypiques des caractéristiques statistiques.

Démarche

Étant donné une vue système de la situation courante, incluant éventuellement un ensemble de jugements de pertinence, nous faisons l'hypothèse d'une ou plusieurs

situations de recherche (cette fois selon une vue utilisateur de la situation) où ce type de réponse pose problème. Nous proposons pour chaque situation de recherche une *intention* à court terme qui prend place dans la stratégie développée par l'utilisateur.

Pour chaque intention, nous décrivons l'*action* qui permet d'obtenir un meilleur résultat du système. L'action se traduit sous la forme d'une indication générale de reformulation du schéma de pertinence, qui doit modifier l'aspect sémantique ou l'aspect pragmatique de la réponse.

4.2 Situations stéréotypiques et actions

Nous avons déjà évoqué dans la partie II un certain nombre de situations problématiques, telles qu'on les rencontre dans les systèmes classiques. Nous définissons ici un ensemble de situations problématiques telles qu'elles peuvent être formalisées dans le cadre de notre modèle.

4.2.1 Définitions et notations

Dans la suite, nous utilisons les termes «classe» et «sous-classe» pour désigner les classes et sous-classes visuelles.

Un schéma de pertinence peut être noté :

$$S = \{\dot{\mathbf{K}}, \bar{\mathbf{K}}, \dot{\mathbf{k}}, \bar{\mathbf{k}}\}$$

où $\dot{\mathbf{K}}$ (respectivement $\bar{\mathbf{K}}$) est l'ensemble des critères de dimension optionnels (respectivement obligatoires), et $\dot{\mathbf{k}}$ (respectivement $\bar{\mathbf{k}}$) est l'ensemble des critères élémentaires optionnels (respectivement obligatoires).

Le schéma de pertinence résultant de la reformulation de S est noté $S' = \{\dot{\mathbf{K}}', \bar{\mathbf{K}}', \dot{\mathbf{k}}', \bar{\mathbf{k}}'\}$.

Le nombre de critères de dimension optionnels $Card(\dot{\mathbf{K}})$ est noté N_{OPT} , et le nombre de critères de dimension obligatoires $Card(\bar{\mathbf{K}})$ est noté N_{OBL} . Le nombre de critères élémentaires optionnels $Card(\dot{\mathbf{k}})$ est noté N_{opt} , et le nombre de critères élémentaires obligatoires $Card(\bar{\mathbf{k}})$ est noté N_{obl} . Le nombre total de critères élémentaires est noté N .

Vocabulaire pour définir une vue système de la situation

$N_{DR} = Card(S(\mathcal{E}_D))$	Nombre de Documents Retrouvés
N_{DP}	Nombre de Documents Pertinents
N_{DNP}	Nombre de Documents Non Pertinents
N_C	Nombre de Classes
N_{CP}	Nombre de Classes Pertinentes
N_{CNP}	Nombre de Classes Non Pertinentes
R_{pnp}	Rapport du nombre de documents Pertinents sur le nombre de documents Non Pertinents au sein d'une sous-classe donnée
N_{dsc}	Nombre de Documents dans une Sous-Classe donnée

Catégories d'actions possibles et effets produits

Nous définissons quatre catégories d'actions : **RelâcherP**, **RelâcherS**, **Détailler**, **Focaliser**. La première réalise une modification de nature exclusivement pragmatique sur le résultat, tandis que les trois autres actions réalisent une modification de nature sémantique ou pragmatique, selon l'intention de l'utilisateur, comme nous le montrons plus loin.

RelâcherP consiste à changer certains statuts obligatoires en statuts optionnels.

$$(\dot{\mathbf{k}} \cup \bar{\mathbf{k}} = \dot{\mathbf{k}}' \cup \bar{\mathbf{k}}') \text{ et } (\dot{\mathbf{K}} \subset \dot{\mathbf{K}}' \text{ ou } \dot{\mathbf{k}} \subset \dot{\mathbf{k}}')$$

Avec S' , plus de documents sont susceptibles d'être retrouvés, selon le même axe sémantique que celui suggéré par le schéma de pertinence S .

RelâcherS consiste à modifier un critère élémentaire fort, en un critère plus lâche s'il en existe un.

$$\begin{aligned} ((\dot{\mathbf{k}} \cup \bar{\mathbf{k}}) \setminus (\dot{\mathbf{k}}' \cup \bar{\mathbf{k}}')) &= \{\xi_i[t_i]\} \text{ et} \\ ((\dot{\mathbf{k}}' \cup \bar{\mathbf{k}}') \setminus (\dot{\mathbf{k}} \cup \bar{\mathbf{k}})) &= \{\xi'_i[t_i]\} \text{ et} \\ &(\forall i, \xi'_i \prec \xi_i) \end{aligned}$$

Avec S' d'autres documents sont susceptibles d'être retrouvés, selon un axe sémantique voisin de celui suggéré par le schéma de pertinence S .

Détailler consiste à ajouter un critère élémentaire en s'inspirant des caractéristiques des documents de la sous-classe. Un critère discriminant entre les documents jugés pertinents et les documents jugés non pertinents, doit être repéré.

$$(\dot{\mathbf{K}} = \dot{\mathbf{K}}') \text{ et } (\bar{\mathbf{K}} = \bar{\mathbf{K}}') \text{ et } (\dot{\mathbf{k}} \subset \dot{\mathbf{k}}' \text{ ou } \bar{\mathbf{k}} \subset \bar{\mathbf{k}}')$$

À l'ajout du critère, la sous-classe est séparée en fonction du nouveau critère ajouté.

Focaliser consiste à changer certains statuts de critères élémentaires en obligatoires.

$$(\dot{\mathbf{K}} = \dot{\mathbf{K}}') \text{ et } (\bar{\mathbf{K}} = \bar{\mathbf{K}}') \text{ et } (\dot{\mathbf{k}} \cup \bar{\mathbf{k}} = \dot{\mathbf{k}}' \cup \bar{\mathbf{k}}') \text{ et } (\bar{\mathbf{k}} \subset \bar{\mathbf{k}}')$$

En choisissant les statuts des critères vérifiés dans les classes pertinentes, cela permet de classer en tête des classes pertinentes.

4.2.2 Situations problématiques et actions

Nous présentons quatre grandes situations telles qu'elles peuvent être identifiées du point de vue du système : la pénurie, la non-pertinence, la non-discrimination, la non-focalisation. Pour chacune de ces vues système de situation, nous décrivons le type de problème qu'elle pose en fonction de sa position dans la boucle des tâches de la figure IV.4.2, nous développons un ensemble de vues utilisateur de ces situations, sous la forme d'une intention à court terme, et nous montrons quelle action permet de réaliser cette intention.

Pénurie

Dans l'hypothèse où l'utilisateur n'est pas satisfait, la situation où aucun document n'est retrouvé constitue une situation d'entrave à la poursuite du processus itératif.

Vue système de la situation	Type de situation problématique
$N_{DR} = 0$	Cas 1 : Entrave à l'évaluation Cas 2 : Entrave à la reformulation

Cette situation pose problème que ce soit pour évaluer la satisfaction (le corpus ne contient-il effectivement aucun document pertinent?) ou pour reformuler (sur quelle base reformuler le schéma de pertinence?).

L'objectif est de retrouver des documents malgré tout.

Dans le cas 1 (entrave à l'évaluation), ces documents pourraient devenir pertinents pour l'utilisateur, dans la mesure où le corpus ne comprend pas de documents aussi pertinents qu'espéré (1.1), ou au contraire convaincre l'utilisateur qu'il ne peut espérer tirer mieux du corpus (1.2).

Dans le cas 2, les documents obtenus permettraient à l'utilisateur d'examiner les caractéristiques des documents afin d'y puiser des idées pour reformuler son problème d'information (2).

	Intention de l'utilisateur	Action
1.1	Revoir à la baisse le schéma individuel de pertinence	RelâcherP
1.2	Se convaincre de l'absence de documents pertinents	RelâcherP
2	Trouver des idées de reformulation	RelâcherS

Dans les contextes 1.1 et 1.2, le résultat de **RelâcherP** donne une vue en perspective sur le contenu du corpus, dans l'axe du schéma de pertinence, grâce aux classes de degré supérieur, ce qui remédie au problème d'entrave à l'évaluation. Il permet aussi de focaliser sur chaque critère pour vérifier la sémantique que lui attribue le système, par l'exemple, et éventuellement la modifier (cas d'une entrave à la reformulation).

Dans le contexte 2, le résultat de **RelâcherS** donne une nouvelle perspective sur le corpus, pour un schéma de pertinence S' plus nuancé que S .

Non-pertinence

Dans l'hypothèse où l'utilisateur n'est pas satisfait, la situation où aucun des documents retrouvés n'est ou pertinent, ou non pertinent, constitue une entrave à reformulation.

Vue système de la situation	Type de situation problématique
$N_{DR} < > 0$ et ($N_{DP} = 0$ et/ou $N_{DNP} = 0$)	Entrave à la reformulation

En effet, cela signifie que l'utilisateur ne voit pas de lien entre le schéma de pertinence qu'il a formulé et les documents retrouvés (1), ou bien, dans le cas contraire, cela signifie qu'un ou plusieurs des critères abstraits qu'il a employés ne conviennent pas, quoique les termes de la requête conviennent (2).

	Intention de l'utilisateur	Action
1	Vérifier la sémantique des critères de S	RelâcherP
2	Identifier un critère abstrait plus approprié	RelâcherS

Dans le cas du contexte 1, l'utilisateur souhaite vérifier la sémantique des critères qu'il a employés. Pour cela il lui faudrait observer les effets produits par la vérification des critères élémentaires sur les documents.

Dans le cas du contexte 2, l'utilisateur a pu constater que les termes de sa requête sont corrects, mais les documents ne sont pas pertinents pour autant. Il cherche à identifier un critère abstrait plus approprié au sens qu'il veut exprimer pour un critère dont il a repéré l'inadéquation (éventuellement dans le cadre du contexte 1). Il lui faut donc observer l'effet de la modification d'un critère abstrait pour un

critère plus souple s'il en existe un.

Non-discrimination

Cette situation est relative à un ensemble de jugements de pertinence exprimés sur une sous-classe en particulier. Les documents de la sous-classe, s'ils sont nombreux, ne sont pas nécessairement tous examinés à ce stade. Une situation de non-discrimination est caractérisée par le fait que dans cette sous-classe donnée, on a à la fois des documents pertinents et non pertinents, sans que les uns dominent les autres en nombre.

Vue système de la situation	Type de situation problématique
$R_{\text{pnp}} \sim 1$	Non-satisfaction
$R_{\text{pnp}} \sim 1$ et N_{dsc} grand	Entrave à l'évaluation

Malgré la détection de documents pertinents, l'utilisateur peut ne pas être satisfait, dans au moins deux contextes.

S'il a l'intention de cerner de manière exhaustive les documents pertinents du corpus, il peut espérer préciser son problème d'information, qui n'est pas encore formulé de manière suffisamment discriminante, comme en témoigne le mélange de documents pertinents et non pertinents dans cette sous-classe (1).

Par ailleurs, si la sous-classe en question contient un grand nombre de documents, l'utilisateur veut éviter de consulter tous les documents de cette classe, sachant qu'il risque de rencontrer de nombreux documents non pertinents : cela entrave sa capacité à évaluer le résultat obtenu. Il cherche donc à séparer les documents pertinents des documents non pertinents sur la base des premiers exemples qu'il a jugés dans cette sous-classe (2).

	Intention de l'utilisateur	Action
1	Ajouter un nouveau critère pour préciser le problème d'information	Détailler
2	Séparer les documents pertinents des non pertinents	Détailler

Non-focalisation

La situation où les classes non pertinentes dominent en nombre les classes pertinentes est appelée situation de non-focalisation. Dans le contexte de notre modèle, une mauvaise précision globale du résultat ne pose pas nécessairement un problème étant donné le regroupement en classes des documents : il suffit d'ignorer certaines

classes non pertinentes. Cependant il existe des situations utilisateur où les situations de non-focalisation posent problème.

Vue système de la situation	Type de situation problématique
$N_{CP} < > 0$ et $\frac{N_{CNP}}{N_c} \sim 1$	Non-satisfaction

Lorsqu'il existe des classes pertinentes, mais que la plupart sont non pertinentes, l'utilisateur peut n'être pas satisfait dans la mesure où il s'interroge sur la position des documents retrouvés jusque-là dans l'ensemble du corpus. Les classes non pertinentes perturbent cette vue d'ensemble du corpus. Ainsi, l'utilisateur peut décider de régler le schéma de pertinence de façon à focaliser le plus possible sur les documents pertinents.

	Intention de l'utilisateur	Action
	Éliminer autant de classes non pertinentes que possible	Focaliser

4.2.3 Bilan

Cette analyse, qui ne prétend pas à l'exhaustivité, montre cependant qu'à une même vue du système sur une situation, peuvent être associés plusieurs types de problèmes. C'est en effet le cas pour la situation de pénurie. C'est l'intention de l'utilisateur, issue de sa situation de recherche, qui détermine quel problème pose la situation système et quelle action permet de le résoudre.

Nous constatons aussi que la spécification de certaines actions requièrent l'intervention de l'utilisateur (**Détailler**, **Focaliser**), alors que d'autres peuvent éventuellement s'en dispenser (**RelâcherP**, **RelâcherS**).

4.3 Possibilités d'automatisation de la reformulation

Nous discutons ici les possibilités d'automatisation de la reformulation permises par notre modèle.

Nous distinguons trois axes d'automatisation. D'abord l'automatisation d'opérations globales, qui correspondent aux techniques d'expansion de requête, telles que les techniques de reformulation automatique classique.

Puis nous définissons les opérations paramétrées, qui correspondent à une semi-automatisation de la reformulation, et pour lesquelles l'utilisateur contrôle les para-

mètres de la reformulation.

Ces opérations spécialisent les actions que nous avons décrites par association à des situations stéréotypiques dans la section précédente. Par exemple, l'action **RelâcherP** recouvre deux opérations globales : **RelâcherPDim**, **RelâcherPElem**.

De plus, ces opérations étendent le jeu de ces actions. Par exemple, nous définissons l'opération **ResserrerS** comme une action globale inverse de l'action **RelâcherS**.

4.3.1 Opérations globales

Les opérations globales définissent des actions qui s'appliquent de manière globale au schéma de pertinence S . Elles ne requièrent pas de l'utilisateur qu'il fournisse des paramètres.

$$S = [\dot{K}_1, \dots, \dot{K}_{\dot{n}}, \bar{K}_{\dot{n}+1}, \dots, \bar{K}_n]$$

avec pour tout $i \in [1..n]$,

$$K_i = (\text{Dim}_i, [\dot{k}_{i.1}, \dots, \dot{k}_{i.\dot{n}_i}, \bar{k}_{i.\dot{n}_i+1}, \dots, \bar{k}_{i.n_i}])$$

Relâcher les contraintes

Trois types d'opérations globales ont pour effet de relâcher les contraintes imposées par le schéma de pertinence : le passage des statuts de dimension obligatoires en statuts de dimension optionnels, le passage des statuts élémentaires obligatoires en statuts optionnels, et le passage des critères élémentaires stricts en critères élémentaires plus lâches.

Nous définissons trois opérations **RelâcherPDim**, **RelâcherPElem** et **RelâcherS** sur les schémas de pertinence qui réalisent chacune de ces actions.

RelâcherPDim associe à tous les critères de dimensions un statut optionnel.

$$\text{RelâcherPDim}(S) = [\dot{K}_1, \dots, \dot{K}_{\dot{n}}, \dot{K}_{\dot{n}+1}, \dots, \dot{K}_n]$$

RelâcherPElem associe à tous les critères élémentaires un statut optionnel.

$$\text{RelâcherPElem}(S) = [\dot{K}'_1, \dots, \dot{K}'_{\dot{n}}, \bar{K}'_{\dot{n}+1}, \dots, \bar{K}'_n]$$

avec pour tout $i \in [1..n]$,

$$K'_i = (\text{Dim}_i, [\bar{k}_{i.1}, \dots, \bar{k}_{i.\dot{n}_i}, \bar{k}_{i.\dot{n}_i+1}, \dots, \bar{k}_{i.n_i}])$$

RelâcherS remplace tous les critères élémentaires abstraits pour lesquels il existe un critère moins strict, par l'un d'eux. L'opération élémentaire qui remplace un

critère élémentaire abstrait par un critère moins strict s'il existe² est notée $\triangleleft(\xi)$ et est définie comme suit³ :

$$\triangleleft(\xi) = \begin{cases} \xi & \text{si } \nexists \xi' \in \mathcal{E}_\xi, \xi' \prec \xi \\ \xi' & \text{sinon, avec } \xi' \in \{\xi'' \in \mathcal{E}_\xi, \xi'' \prec \xi\} \end{cases}$$

Ainsi, **RelâcherS** est défini comme suit :

$$\text{RelâcherS}(S) = [\dot{K}'_1, \dots, \dot{K}'_{\dot{n}}, \bar{K}'_{\dot{n}+1}, \dots, \bar{K}'_n]$$

avec pour tout $i \in [1..n]$,

$$K'_i = (\text{Dim}_i, [\triangleleft(\bar{\xi}_{i.1})[t_{i.1}], \dots, \triangleleft(\bar{\xi}_{i.\dot{n}_i})[t_{i.\dot{n}_i}], \triangleleft(\bar{\xi}_{i.\dot{n}_i+1})[t_{i.\dot{n}_i+1}], \dots, \triangleleft(\bar{\xi}_{i.n_i})[t_{i.n_i}]])$$

Resserrer les contraintes

Trois types d'opérations globales ont pour effet de resserrer les contraintes imposées par le schéma de pertinence : le passage des statuts de dimension optionnels en statuts de dimension obligatoires, le passage des statuts élémentaires optionnels en statuts obligatoires, et le passage des critères élémentaires lâches en critères élémentaires plus stricts.

Nous définissons trois opérations **ResserrerPDim**, **ResserrerPElem** et **ResserrerS** sur les schémas de pertinence qui réalisent chacune de ces actions.

ResserrerPDim associe à tous les critères de dimensions un statut obligatoire.

$$\text{ResserrerPDim}(S) = [\dot{K}_1, \dots, \dot{K}_{\dot{n}}, \dot{K}_{\dot{n}+1}, \dots, \dot{K}_n]$$

ResserrerPElem associe à tous les critères élémentaires un statut obligatoire.

$$\text{ResserrerPElem}(S) = [\dot{K}'_1, \dots, \dot{K}'_{\dot{n}}, \bar{K}'_{\dot{n}+1}, \dots, \bar{K}'_n]$$

avec pour tout $i \in [1..n]$,

$$K'_i = (\text{Dim}_i, [\bar{k}_{i.1}, \dots, \bar{k}_{i.\dot{n}_i}, \bar{k}_{i.\dot{n}_i+1}, \dots, \bar{k}_{i.n_i}])$$

ResserrerS remplace tous les critères élémentaires abstraits pour lesquels il existe un critère plus strict, par l'un d'eux. L'opération élémentaire qui remplace un critère élémentaire abstrait par un critère plus strict s'il existe est notée $\triangleright(\xi)$ et est définie comme suit :

$$\triangleright(\xi) = \begin{cases} \xi & \text{si } \nexists \xi' \in \mathcal{E}_\xi, \xi \prec \xi' \\ \xi' & \text{sinon, avec } \xi' \in \{\xi'' \in \mathcal{E}_\xi, \xi \prec \xi''\} \end{cases}$$

Ainsi, **ResserrerS** est défini comme suit :

$$\text{ResserrerS}(S) = [\dot{K}'_1, \dots, \dot{K}'_{\dot{n}}, \bar{K}'_{\dot{n}+1}, \dots, \bar{K}'_n]$$

2. Le choix du critère si plusieurs conviennent fait partie des paramètres à régler dans le modèle opérationnel.

3. C'est une donnée du système qui est élaborée à partir du modèle d'indexation.

avec pour tout $i \in [1..\dot{n}]$,

$$K'_i = (\text{Dim}_i, [\triangleright(\bar{\xi}_{i.1})[t_{i.1}], \dots, \triangleright(\bar{\xi}_{i.\dot{n}_i})[t_{i.\dot{n}_i}], \triangleright(\bar{\xi}_{i.\dot{n}_i+1})[t_{i.\dot{n}_i+1}], \dots, \triangleright(\bar{\xi}_{i.n_i})[t_{i.n_i}]]])$$

Conclusion

Ces opérations qui appliquent des modifications systématiques à l'ensemble du schéma de pertinence ont les mêmes limites que les méthodes de reformulation automatique : elles peuvent trahir le besoin d'information de l'utilisateur. Elles ne sont applicables que dans le cas d'une action de reformulation peu précise comme dans les situations où le résultat est mauvais et un problème de structure du résultat doit être résolu (pénurie).

Pour régler les autres types de problèmes, l'intervention de l'utilisateur est requise pour sélectionner précisément la partie du schéma de pertinence à modifier et définir les termes précis de la modification.

4.3.2 Opérations paramétrées

Les opérations paramétrées sont les actions qui nécessitent que l'utilisateur fournisse les paramètres de la modification du schéma, dont voici la liste.

- suppression d'un critère élémentaire,
- modification d'un terme de la requête,
- modification d'un critère élémentaire s'appliquant à un terme de la requête,
- passage d'un statut de critère élémentaire optionnel en obligatoire,
- passage d'un statut de critère élémentaire obligatoire en optionnel.
- passage d'un statut de critère de dimension optionnel en obligatoire,
- passage d'un statut de critère de dimension obligatoire en optionnel.

Dans la partie V, nous présentons dans le cadre de notre expérimentation un ensemble d'exemples de processus de recherche où ces opérations interviennent. Ces exemples s'inscrivent dans le cadre d'une définition de la vue système de la situation étendue, qui prend en compte un ensemble de jugements de pertinence fournis par l'utilisateur. Ces paramètres s'expriment en fonction de la répartition des documents pertinents et non pertinents dans les classes. Cela permet de définir un nouvel éventail de situations typiques (situations lacunaire, tranchée, ambiguë). Malgré ce raffinement, ces exemples montrent la diversité des opérations qui peuvent être associées à une même vue système de la situation.

4.4 Résumé

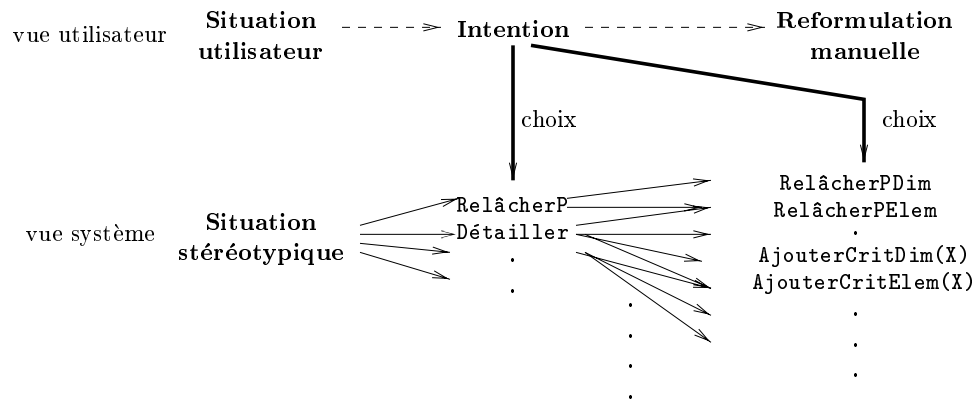


FIG. IV.4.4 – *Résumé de notre analyse de l'automatisation de la reformulation*

L'ensemble de notre analyse de l'automatisation de la reformulation est résumé dans la figure IV.4.4, qui met en parallèle la vue utilisateur et la vue système de la reformulation manuelle et notre approche pour son automatisation. Pour une même situation stéréotypique plusieurs actions peuvent convenir. C'est l'intention de l'utilisateur qui détermine l'action à réaliser. De plus une action peut se traduire par de multiples opérations. C'est l'utilisateur qui est le mieux placé pour faire le choix de la bonne opération en fonction de son intention. De plus un sous-ensemble de ces opérations requièrent une instanciation par des paramètres que l'utilisateur est seul à pouvoir déterminer.

Dans la partie V, nous utilisons cette analyse pour procéder à nos expérimentations dans le cas particulier de l'application à la recherche dans un corpus d'images [DBM97].

Chapitre 5

Bilan

Nous discutons ici des capacités d'adaptation de notre modèle, en les comparant à celles des techniques d'adaptation du système.

5.1 Niveau de granularité

Les techniques d'adaptation du système diffèrent d'abord par le niveau de granularité auquel l'adaptation a lieu. Pour les modèles fondés sur un ensemble prédéfini de «stratégies de recherche» (fonctions de correspondance), l'adaptation a lieu en amont du processus itératif de recherche. Pour les techniques de reformulation automatique, l'adaptation intervient à chaque itération de ce processus.

Notre modèle se situe au même niveau de granularité que ces techniques de reformulation automatique : à chaque itération la fonction de correspondance et la requête sont susceptibles d'être modifiées.

5.2 Portée de la modification

Le deuxième point est la portée de la modification engendrée par l'adaptation. Pour les modèles fondés sur les stratégies, c'est la fonction de correspondance qui est modifiée. Pour les techniques de reformulation, c'est la requête qui est modifiée.

Notre modèle intègre ces deux portées au sein du schéma de pertinence, où tant la requête que la fonction de correspondance sont paramétrées. La nouveauté est que la fonction de correspondance peut être modifiée *à chaque itération*, ce qui donne au système une capacité à évoluer de manière dynamique avec la situation de recherche d'information.

D'un point de vue statique, la paramétrisation de la fonction de correspondance dans notre modèle permet de l'adapter en fonction de chaque requête, ce qui n'est pas vrai des fonctions de correspondance correspondant à des stratégies de recherche

prédéfinies. C'est en cela que notre modèle permet de réduire la distance entre pertinence utilisateur et pertinence système, car le schéma de pertinence permet de refléter le schéma individuel de pertinence de l'utilisateur dans un contexte particulier, celui de la situation de recherche courante.

5.3 Degré et conditions d'intervention de l'utilisateur

Le troisième point par lequel les techniques diffèrent est le degré d'intervention de l'utilisateur dans la spécification de l'adaptation à réaliser, et les conditions dans lesquelles il spécifie cette adaptation.

Pour les modèles fondés sur les stratégies, l'utilisateur est seul maître du choix de la fonction de correspondance à appliquer, mais son choix s'effectue sans qu'il puisse établir un lien avec les effets que vont produire son choix, et sans disposer des moyens de contrôler l'adéquation de son choix *a posteriori* afin de l'améliorer si besoin est. Un problème de robustesse se pose, car nous avons montré que les cas d'inadéquation sont très probables.

Pour les techniques de reformulation automatique, l'utilisateur intervient à chaque itération pour par l'intermédiaire de jugements de pertinence. Pour le système, les jugements de pertinence spécifient complètement l'adaptation à réaliser. Pour cela, le système fait des hypothèses très fortes sur la sémantique des jugements de pertinence, en particulier celle de leur cohérence et de leur complétude. La clarté de la sémantique de la tâche de production des jugements de pertinence, ainsi que les moyens fournis par le système pour que l'utilisateur satisfasse au mieux ces hypothèses sont donc cruciaux. Pourtant, la sémantique que le système associe aux jugements de pertinence produits est orientée vers le système et difficilement transmissible à l'utilisateur. On qualifie d'incertitude les écarts dus à cette inadéquation. Quand il arrive que cette incertitude soit trop grande, un problème de robustesse se pose, ainsi qu'un problème d'intelligibilité du système, comme nous l'avons décrit auparavant.

Dans le cadre d'une utilisation manuelle d'un système fondé sur notre modèle, le degré d'intervention de l'utilisateur est du même ordre que pour les techniques de reformulation automatique, à cela près que la nature de l'intervention est plus élaborée. En effet, l'utilisateur doit spécifier la nature de l'adaptation. Cette complexité supplémentaire est compensée par la clarification de la sémantique des adaptations qu'apporte le schéma de pertinence et les classes de pertinence. De plus la paramétrisation de la visualisation des résultats par les statuts de critère fournit un nouvel axe d'interaction dédié à la mise en place de bonnes conditions pour la réalisation de ces tâches.

Partie V

Application

Notre modèle pose un ensemble de principes de conception. Il ne propose pas une fonction de correspondance particulière ni une technique de bouclage de pertinence particulière, mais un contexte dans lequel de telles techniques peuvent s'intégrer, et être promises à un plus grand profit du fait d'une meilleure qualité d'interaction avec l'utilisateur.

Les principes visant à établir cette qualité ont été établis selon une approche descendante, en partant des connaissances existant sur la pertinence pour l'individu, et en les déclinant dans le cadre de l'utilisation des systèmes de recherche d'information. Nous voulons montrer ici dans quels cas ces principes peuvent être profitables, et comment ils s'appliquent effectivement en pratique.

Dans le chapitre 1, nous développons un exemple d'application à un corpus d'images indexées manuellement, qui nous permet d'illustrer les associations situation-réactions que nous avons proposées dans la partie précédente. Puis, dans le chapitre 2, nous mettons en perspective la mise en place de divers types de contrôle de l'interaction. Enfin, nous cernons dans le chapitre 3 le champ d'application de notre modèle.

Chapitre 1

Exemple d'application

1.1 Le prototype VIZ

Le prototype VIZ implémente un système de recherche d'information conforme à notre modèle, à partir d'un corpus d'images indexées. Il implémente les évaluations binaire et symbolique de notre modèle de pertinence système, et les classes visuelles. L'interface d'interrogation permet de formuler un schéma de pertinence et de visualiser les résultats selon les classes et les sous-classes visuelles.

Ce prototype a pour but de concrétiser les concepts de notre modèle de pertinence système, et en aucun cas de proposer une interface destinée à un utilisateur final.

Nous présentons d'abord le passage de notre modèle de pertinence système à un modèle opérationnel qui prend en compte les spécificités du corpus, de l'indexation disponible, et de l'application visée.

Puis nous décrivons les principes informatiques de la réalisation du prototype.

1.1.1 Modèle opérationnel

Corpus, indexation et objectif de l'application

Corpus Le corpus est constitué de 658 photographies datant du début du siècle. Une première moitié environ de ces photographies illustre le vieux Paris, ses monuments, ses petits métiers, ses quartiers, des événements marquants de l'époque. L'autre moitié est plus hétéroclite, et représente des portraits d'hommes célèbres, des paysages, des vues d'autres villes que Paris, des objets d'art, etc. L'intérêt de ce corpus est donc d'ordre esthétique, historique, anecdotique, architectural, etc.

Le public potentiel est relativement varié, du journaliste qui cherche à illustrer un article, à l'étudiant en architecture qui se documente sur un type d'architecture particulier, en passant par l'amateur de vieilles photographies en général.

Indexation L'indexation recouvre plusieurs aspects des photographies, identifiés par onze attributs : Titre, Auteur, Genre de l'image, Commentaire, Contenu, Précisions, Connotation, Morphologie, Vue photographique, Lumière. Elle a été réalisée par des documentalistes selon un modèle attribut-valeur. L'attribut Numéro correspond simplement à l'identificateur de la photographie.

Les figures V.1.1 et V.1.2 donnent deux exemples de documents indexés. Les attributs Auteur, Genre de l'image, Contenu, Précisions, Connotation, Plan, Vue photographique, Lumière prennent leur valeur dans une liste de mots-clés au sens traditionnel du terme. Les attributs Titre et Commentaire prennent pour valeur de simples chaînes de caractères qui ne sont pas répertoriées dans une liste prédéfinie.

Les attributs Titre, Auteur et Genre de l'image relèvent des attributs externes du document. L'attribut Contenu fait référence à la description du contenu sémantique de l'image. Les attributs Commentaire et Précisions fournissent des informations contextuelles qui complètent la description du contenu sémantique de l'image. L'attribut Connotation donne un point de vue subjectif sur l'image, selon la subjectivité de l'indexeur. Enfin les attributs Plan, Vue photographique et Lumière décrivent des caractéristiques techniques de la photographie.

Application visée Comme le montre la figure V.1.2, le niveau de précision de l'indexation du contenu en matière d'architecture permet d'envisager une interrogation du corpus relativement technique, de même que les précisions qui fournissent un contexte historique ou géographique, permettent d'envisager une interrogation du point de vue de la culture générale. L'attribut Connotation permet aussi d'envisager une interrogation plus qualitative et dans un but esthétique.

Ainsi, la nature et le niveau de détail et de précision de l'indexation permettent d'envisager un grand éventail de possibilités d'utilisations et de type d'utilisateurs de ce corpus.

Notre prototype a pour objectif de ne pas restreindre son utilisation à une catégorie d'utilisations ou d'utilisateurs, afin de mettre à profit les possibilités offertes par l'indexation. Ainsi des situations de recherche nombreuses et variées sont susceptibles d'intervenir, ce qui constitue un contexte d'expérimentation intéressant pour notre modèle de pertinence système qui doit s'adapter aux diverses situations.

Arbre de relations entre termes Pour l'attribut Contenu, l'indexation a donné lieu à la construction d'un arbre qui organise les divers termes utilisés selon une structure arborescente. Les feuilles de l'arbre sont les termes d'indexation, et les termes d'indexation relevant d'un même terme générique T sont groupés comme les fils de ce terme. Ce procédé est réappliqué récursivement pour former un arbre. La table V.1.1 donne un extrait de cet arbre. L'extension -D ajoutée à certains mots-clés signifie «divers», et regroupe tous les termes spécifiques du père de ce terme,



Champs	Valeurs
Numéro	4699
Titre	Félix Nadar posant dans une nacelle tenant ses jumelles.
Auteur	Nadar Atelier
Genre de l'image	autre portrait
Commentaire	Félix Nadar (1820–1910) Il fit de véritables ascensions dans son ballon «le Géant» dès 1863.
Contenu	autre instrument d'optique, autre costume, moustache, foulard, ciel, montgolfière, autre photographe, toile de fond, partie de, chapeau, autre maturité, autre accessoire de pose, homme
Précisions	Nadar Félix Tournachon Pseudo. Félix
Connotation	pittoresque, anecdotique
Morphologie	plan d'ensemble, vue de face, photographie en atelier

FIG. V.1.1 – Une photographie et son indexation



Champs	Valeurs
Numéro	1480
Titre	Partie gothique d'une église reconstruite au XVIIème siècle
Auteur	Atget Jean-Eugène
Genre de l'image	
Commentaire	La voûte du bas-côté nord n'a jamais été remaniée.
Contenu	portillon, bas-côté, bois-d, ronde-bosse-d, voûte d'ogive, pilier, confessionnal, pilastre, nord
Précisions	Église Saint Julien le Pauvre
Connotation	vide, austère
Morphologie	vue perspective, vue d'intérieur

FIG. V.1.2 – Autre exemple d'image indexée

et qui ne sont pas répertoriés explicitement. Ainsi, ARTISAN DU CUIR-D dénote tous les artisans du cuir qui ne sont ni cordonniers, ni tanneurs, ni bottiers.

Il en est de même pour l'attribut **Genre**, dont l'arbre ne comprend que deux niveaux sous la racine (table V.1.2).

Les termes effectivement employés comme termes d'indexation sont donc les feuilles de l'arbre, soit les termes les plus spécifiques, dans le but de vérifier au mieux l'hypothèse d'indépendance des termes d'indexation. Les relations de généralité et spécificité sont destinées à être utilisées dans le cadre de l'interrogation et la correspondance.

Nous notons *Fils* la relation de spécificité immédiate d'un terme du thésaurus pour un autre. Soit t_1 et t_2 deux termes quelconques du thésaurus ; $Fils(t_1, t_2)$ signifie que t_1 est l'ascendant immédiat de t_2 .

La relation de spécificité généralisée est notée *Spéc*. Soit t_1 et t_2 deux termes quelconques du thésaurus ; $Spéc(t_1, t_2)$ signifie que t_1 est un ascendant quelconque de t_2 .

Les autres attributs d'indexation n'ont pas donné lieu à la construction d'un arbre traduisant des relations de généralité et spécificité.

Cependant l'attribut **Morphologie** présente une structure sur les termes, qui traduit le regroupement de certaines caractéristiques selon des sous-attributs (table V.1.3).

Pour les autres attributs (à l'exception de **Titre** et **Commentaire**) les termes sont simplement stockés dans une liste. Les divers ensembles de termes d'indexation par attribut sont disjoints. L'attribut **Précisions** comprend en général des noms propres qui sont associés à des objets du contenu, typiquement le nom d'un personnage ou d'un lieu apparaissant sur la photographie.

Les attributs **Titre** et **Commentaire** ont pour valeur une chaîne de caractère en texte libre.

Nous revenons plus loin sur l'indexation, pour analyser la façon dont nous pouvons l'utiliser pour l'interrogation.

Interrogation

Pour l'interrogation, nous identifions sept dimensions d'interrogation correspondant aux attributs d'indexation disponibles (table V.1.4). Les attributs **Titre** et **Commentaire** ne participent pas à l'interrogation, car nous n'avons pas implémenté de critère élémentaire abstrait qui réalise des traitements sur la langue naturelle.

Requête et schéma de pertinence

La structure à deux niveaux du schéma de pertinence de notre modèle formel se traduit en utilisant les dimensions d'interrogation. L'interrogation se fait en langage

```

o ARTISANAT
  PROFESSION DE L'ARTISANAT
    ARTISAN
      ARTISAN DU CUIR
        CORDONNIER
        ARTISAN DU CUIR-D
        TANNEUR
        BOTTEUR
      ARTISAN DU BÂTIMENT
        RAMONEUR
        MARBRIER
        ÉLECTRICIEN
        MAÇON
        PEINTRE EN BÂTIMENT
        COUVREUR
        PLOMBIER
        CHARPENTIER
        PLÂTRIER
        ARTISAN DU BÂTIMENT-D
        TAILLEUR DE PIERRE
        VITRIER
        MENUISIER
        CHAUFFAGISTE
        SERRURIER
      VANNIER
      BARBIER
      ARTISAN DU TEXTILE
        TISSERAND
        DENTELLIÈRE
        CARDEUR
      COMPAGNON
      ARTISAN DU MÉTAL
        MARÉCHAL-FERRANT
        CHARRON
        ÉTAMEUR
        FERRONNIER
        FORGERON
        FONDEUR
        REMOULEUR
        CHAUDRONNIER
        ORFÈVRE
        FERBLANTIER
      LAVANDIÈRE
      MEUNIER
      VERRIER
      HORLOGER
      TONNELIER
      PALUDIER
      REMPAILLEUR
      POTIER
      COIFFEUR
    CHEF D'OEUVRE
    OUTILLAGE ARTISANAL
      OUTIL DE L'ARTISANAT
        RABOT
        ROUET
        PINCE
        CLÉ À MOLETTE
        LIME
        TENAILLE
        TOURNEVIS
        ÉTAU
        QUENOUILLE
        FUSEAU
        BATTE
        BOÎTE À ONGLET
        GOUGE
        ENCLUME
        OUTIL DE L'ARTISANAT-D
        CREUSET
        BURIN
        SCIE
        VRILLE
        MARTEAU
        TRUELLE
        SPATULE
        CISAILLE
        MOULE
      FOUR
      MÉTIER À TISSER
        MÉTIER À TISSER-D
        NAVETTE
      ÉTABLI
      OUTILLAGE ARTISANAL-D
      MEULE

```

TAB. V.1.1 – *Extrait de l'arbre des relations pour l'attribut Contenu*

- GENRE DE L'IMAGE
 - PAYSAGE
 - PAYSAGE RURAL
 - PAYSAGE-D
 - PAYSAGE MARITIME
 - PAYSAGE URBAIN
 - PORTRAIT
 - PORTRAIT « À LA MANIÈRE DE »
 - PORTRAIT COLLECTIF
 - PORTRAIT-D
 - PETIT MÉTIER
 - SCÈNE DE RUE
 - SCÈNE D'INTÉRIEUR
 - PHOTOGRAPHIE ÉVÉNEMENTIELLE

TAB. V.1.2 – *Arbre des relations pour l'attribut Genre*

- ANALYSE MORPHOLOGIQUE DU DOCUMENT
 - VUE PHOTOGRAPHIQUE
 - VUE DE TROIS-QUARTS
 - VUE DE PROFIL
 - VUE DE DOS
 - VUE D'ENSEMBLE
 - VUE DE FACE
 - VUE OBLIQUE
 - VUE FRONTALE
 - VUE PERSPECTIVE
 - VUE D'EXTÉRIEUR
 - VUE D'INTÉRIEUR
 - LUMIÈRE
 - EFFET DE LUMIÈRE
 - CONTRASTE
 - CLAIR
 - LUMIÈRE FRISANTE
 - LUMIÈRE-D
 - CONTRE-JOUR
 - REFLET
 - SOMBRE
 - PHOTOGRAPHIE DE NUIT
 - PLAN
 - GROS PLAN
 - PLAN RAPPROCHE
 - CONTRE PLONGÉE
 - PLONGÉE
 - PLAN AMÉRICAIN
 - PLAN D'ENSEMBLE
 - PLAN MOYEN
 - PROFONDEUR
 - PHOTOGRAPHIE COULEUR
 - OPTIQUE
 - VIGNETTAGE
 - EFFET DE LIGNE
 - PHOTOGRAPHIE EN ATELIER
 - TEMPS DE POSE
 - FLOU

TAB. V.1.3 – *Arbre des termes pour l'attribut Morphologie*

Attribut d'indexation	Dimension d'interrogation
Numéro	Numéro
Auteur	Auteur
Genre de l'image	Genre de l'image
Contenu	Contenu
Connotation	Connotation
Morphologie	Morphologie
Précisions	Précisions
Commentaire	
Titre	

TAB. V.1.4 – *Les sept dimensions d'interrogation*

contrôlé sur les éléments de l'arbre ou la liste des mots-clés, dimension par dimension.

L'arbre associé à la dimension **Contenu** permet de définir les critères élémentaires abstraits suivants :

1° *Égal*, égalité ou spécificité d'un terme du document pour t ;

2° *Vois*, voisinage d'un terme du document avec t ou les spécifiques de t .

Nous notons $t \in D$ le fait que le document indexé par D contient dans la dimension appropriée le terme t .

Le critère élémentaire d'égalité est défini par son évaluation binaire comme suit :

$$\acute{E}gal[t](D) = \begin{cases} \top & \text{si } t \in D \text{ ou } \exists t' \in D, t' = Spéc(t) \\ \perp & \text{sinon} \end{cases}$$

Le critère élémentaire de voisinage est défini par son évaluation binaire comme suit :

$$Vois[t](D) = \begin{cases} \top & \text{si } \exists t' \in D, \exists t'', t' = Spéc(t'') \text{ et } t = Spéc(t'') \text{ et } t'' \neq \top \\ \perp & \text{sinon} \end{cases}$$

On notera la précaution qui consiste à exclure de la relation de voisinage le cas où c'est la racine de l'arbre qui constitue l'ascendant commun au terme du document et au terme de la requête.

Une relation d'ordre total peut être définie sur ces deux critères élémentaires abstraits. Nous la définissons de manière opérationnelle en donnant les opérations $\triangleright$ et $\triangleleft$.

$$\begin{aligned} \triangleright(\acute{E}gal) &= \acute{E}gal \\ \triangleright(Vois) &= \acute{E}gal \end{aligned}$$

$$\begin{aligned}\triangleleft(\acute{E}gal) &= Vois \\ \triangleleft(Vois) &= Vois\end{aligned}$$

Pour la dimension **Genre**, le critère *Vois* n'a pas de sens car *Vois[t]* désigne tout l'arbre (figure V.1.2). Ainsi seul le critère *Égal* est applicable.

Pour les autres dimensions, seul le critère élémentaire d'égalité est applicable.

1.1.2 Description technique et interface

Moyens informatiques

Les images sont stockées au moyen du SGBD O2, système de gestion des bases de données orienté objet. Les calculs de correspondance sont effectués par l'intermédiaire du langage de requête de O2.

L'interface est implémentée en JAVA, ce qui permet d'offrir les fonctionnalités hypertextes de HTML pour la visualisation des documents retrouvés¹.

Interface

Nous décrivons l'interface à partir de l'exemple de schéma de pertinence donné dans la figure V.1.3.

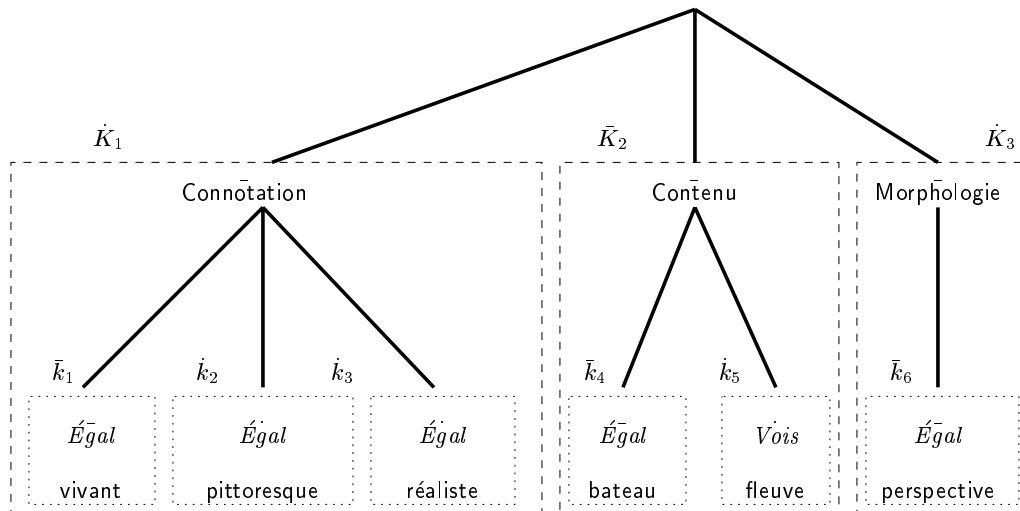


FIG. V.1.3 – Exemple de schéma de pertinence pour la description de l'interface

1. Le prototype est disponible sur le WWW, à l'adresse suivante : http://www-clips.imag.fr/mrim/nathalie.denos/index_perso.html. Il a été implémenté par Mourad Mechkour.

La fenêtre principale de l'interface du prototype (figure V.1.4) comprend 6 régions : les commandes (région 1), le schéma de pertinence (région 2), la vue d'ensemble du schéma de pertinence (région 3), les classes (région 4), les sous-classes (région 5) et les images (région 6).

Schéma de pertinence (interface en entrée) La région 1 comprend, de gauche à droite, les boutons Lancer la recherche, Ajouter un terme, Initialiser le schéma de pertinence, Initialiser la dimension, Supprimer le terme.

La région 2 permet l'édition directe du schéma de pertinence, mais elle ne donne qu'une vue partielle du schéma, car un seul terme par dimension est visible à la fois. Une dimension correspond à une ligne horizontale, et comprend, de gauche à droite, le statut de la dimension, le nom de la dimension, le terme du critère élémentaire actuellement affiché, son statut, et le critère abstrait qui s'y applique.

Classes et sous-classes (interface en sortie) La région 3 donne une vue d'ensemble du schéma de pertinence, où les critères élémentaires ou de dimension sont en vert s'ils sont optionnels, et en noir s'ils sont obligatoires.

La région 4 montre les classes de pertinence retrouvées, où une classe est sélectionnée. La région 5 montre les sous-classes de la classe sélectionnée. À chaque classe ou sous-classe est associé le code qui l'identifie et le nombre de documents qu'elle contient entre crochets. Les classes et les sous-classes sont ordonnées verticalement en fonction de leur degré de pertinence, les plus pertinentes en tête (plus petit de degré).

Dans la région 6 les documents contenus dans la classe ou sous-classe sélectionnée sont affichés. Ils peuvent être affichés sous la forme de la liste des titres des photographies, sous la forme d'une mosaïque des photographies, ou encore une par une en gros plan. Dans cette dernière présentation, l'indexation associée à l'image peut être visualisée. Les termes d'indexation qui ont participé à la pertinence système du document sont en rouge.

Les relations entre les régions 4, 5, et 6 sont gérées par des liens hypertextes. Au sein de la région 6, les diverses représentations des images sont aussi gérées de cette façon, et la présentation image par image permet de feuilleter les images en passant directement d'une image à la suivante.

Déroulement d'une session

La valeur par défaut des critères abstraits est *Égal*, et le statut par défaut d'une dimension est obligatoire.

Dans un premier temps, l'utilisateur élabore un premier schéma de pertinence en ajoutant un ou plusieurs termes et en ajustant les critères abstraits et les statuts au moyens des menus «pop-up». Puis il lance la recherche. Enfin il parcourt les classes

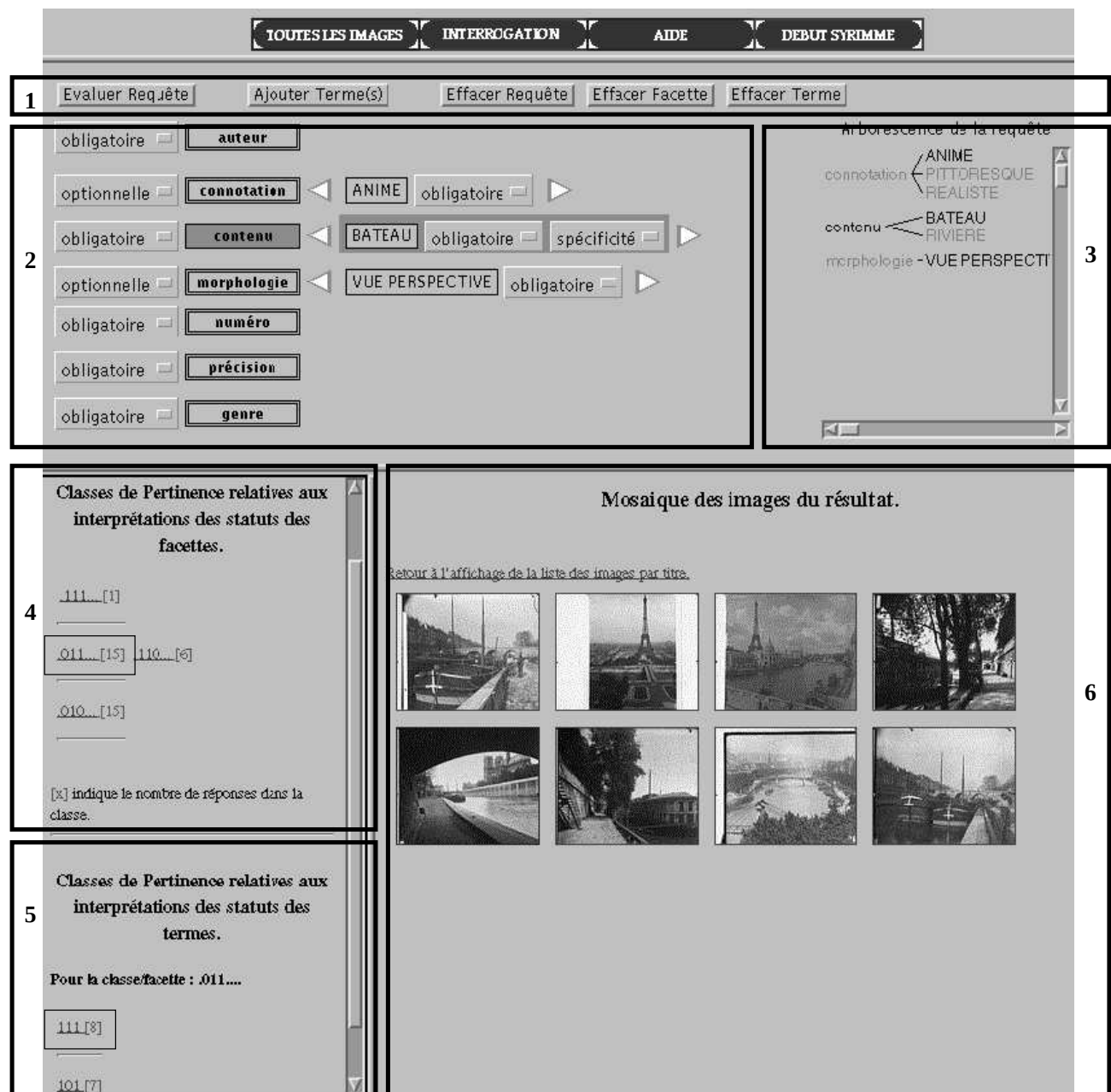


FIG. V.1.4 – Interface du prototype VIZ

et les sous-classes des documents retrouvés dans le but d'évaluer sa satisfaction, et éventuellement de procéder à une reformulation manuelle.

1.2 Propriétés des dimensions

Pour effectuer la tâche de formulation, l'utilisateur met en place une stratégie d'utilisation du système. Dans une certaine mesure, le système peut jouer un rôle pour guider l'élaboration de cette stratégie. En particulier, le système peut

- contraindre la formulation de façon à traduire un certain nombre de propriétés de l'indexation ;
- définir un ensemble de types de recherche indicatifs à partir de la notion de priorités entre critères, et proposer des profils de schéma de pertinence associés.

Nous discutons d'abord des particularités de chaque dimension en soi et en rapport avec les propriétés de l'indexation dont nous disposons, et des stratégies d'utilisation que ces particularités suggèrent. Puis, nous décrivons les principes de formulation des statuts de critères de dimension (critère principal et critère secondaire). Enfin nous discutons des possibilités de contrôle de la formulation du schéma de pertinence.

1.2.1 Particularités des dimensions

Dimension Connotation

Pour la dimension Connotation, l'indexation est très subjective. Typiquement, le terme *agréable* n'indexe qu'une image (figure V.1.5), et le fait que cette image ait une connotation agréable et pas l'image de la figure V.1.6 (indexée par *accueillant*) relève de la subjectivité de l'indexeur. Le besoin d'une structure organisant les termes d'indexation de la dimension Connotation se fait sentir.

Notre démarche consistant à définir un système s'accommodant des contraintes imposées par le corpus et son indexation, nous examinons la façon de tirer parti de cette situation, en développant une stratégie d'utilisation de la dimension Connotation.

Le schéma de pertinence

S_1	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	agréable
-------	--------------------	-------------	--------------------	----------

ramène un unique document (figure V.1.5)².

2. Désormais les schémas de pertinence seront notés sous cette forme de tableau, avec en colonne, de gauche à droite, les statuts de dimension, les noms des dimensions, les statuts des critères élémentaires, les critères abstraits et enfin les termes.



Champs	Valeurs
Numéro	2793
Titre	Une vasque dans le jardin du Luxembourg devant la fontaine Médicis
Auteur	Atget Jean-Eugène
Genre de l'image	
Commentaire	
Contenu	bassin, grille, végétation-D, marche, socle, fontaine-D, ornement géométrique-D, vasque, été
Précisions	jardin du Luxembourg, fontaine de Médicis
Connotation	agréable
Morphologie	plan moyen, vue d'extérieur, reflet

FIG. V.1.5 – *Vasque dans le jardin du Luxembourg*



Champs	Valeurs
Numéro	2350
Titre	Maison de Victor Hugo et Adèle Foucher à Gentilly
Auteur	Atget Jean-Eugène
Genre de l'image	
Commentaire	Maison des Foucher où Victor Hugo séjourna en 1822 avant son mariage avec Adèle.
Contenu	tabouret, végétation-D, porte-D, communs-D, maison-D, table-D, jardin-D, façade-D, cour-D, auvent, pierre-D, escalier-D
Précisions	Adèle Foucher, Victor Hugo
Connotation	accueillant, poétique
Morphologie	plan moyen, vue d'extérieur, contraste

FIG. V.1.6 – *Maison de Victor Hugo et Adèle Foucher à Gentilly*

En consultant les termes d'indexation pour la dimension **Connotation**, il apparaît que plusieurs d'entre eux ont un sens proche de **agréable**, notamment **accueillant** (1 image), **apaisant** (50 images), **charmant** (1 image), **ensoleillé** (5 images), **gai** (8 images), etc. D'autres images pertinentes pourraient donc être retrouvées en augmentant le critère de dimension, quitte à rendre son expression redondante.

Ainsi il est recommandé de formuler le critère de **Connotation** de manière redondante afin de couvrir un ensemble de sens proches.

Pour cela, on peut ajouter des termes proches de **agréable** et associer à ces critères élémentaires le statut optionnel. Cela permet de visualiser la répartition des images en fonction de ces différents termes, et de comprendre la logique de l'indexation connotative. Dans tous les cas les propriétés de l'indexation connotative incitent à utiliser cette dimension avec précaution, par exemple en lui associant toujours un statut optionnel comme nous le discutons plus loin.

Le schéma de pertinence

S_2	<i>obligatoire</i>	Connotation	<i>optionnel</i>	agréable
			<i>optionnel</i>	accueillant
			<i>optionnel</i>	apaisant
			<i>optionnel</i>	charmant
			<i>optionnel</i>	ensoleillé
			<i>optionnel</i>	gai

ramène une unique classe de dimension qui contient

1° deux sous-classes de degré 5 correspondant à la conjonction de **apaisant** et **gai** pour une image et à la conjonction de **ensoleillé** et **gai** pour une autre image ;

2° six sous-classes de degré 6 correspondant à l'intervention de chacun des termes de la dimension **Connotation** ;

3° la sous-classe de degré 7 comprenant les images non pertinentes.

L'ensemble des sous-classes de degré strictement inférieur à 7 correspond ainsi de manière plus réaliste à l'intention associée au schéma de pertinence S_1 .

Dimension **Auteur**

La dimension **Auteur** présente elle aussi des particularités intrinsèques, confirmées par les propriétés de l'indexation : chaque image a au plus un auteur.

Par conséquent, un schéma de pertinence qui demande une conjonction de plusieurs auteurs n'a pas de sens. Un schéma de pertinence comprenant

S_3	<i>obligatoire</i>	Auteur	<i>obligatoire</i>	X
			<i>obligatoire</i>	Y

ne ramène aucun document, quels que soient X et Y distincts³.

En revanche, un schéma de pertinence comprenant

S_4	$\langle statut \rangle$	Auteur	<i>optionnel</i>	X
			<i>optionnel</i>	Y

permet de visualiser les images dont l'auteur est X et celles dont l'auteur est Y dans des sous-classes séparées.

Dimension Numéro

De même, la dimension **Numéro** est une clé, unique, pour l'identification des images. La seule utilisation concevable de cette dimension pour l'interrogation est la désignation d'un ensemble d'images. Pour cela, la formulation requise est similaire à celle relative à la dimension **Auteur**.

Dimension Morphologie

Cette dimension correspond à un niveau de description technique du document. Elle regroupe plusieurs rubriques («Vue photographique», «Lumière», «Plan») auxquelles s'ajoutent divers attributs. Ces rubriques ne sont pas utilisables en tant que termes pour l'interrogation (au contraire du terme générique **portrait**, pour la dimension **Genre**).

Au sein d'une même sous-dimension, certains termes s'excluent mutuellement : par exemple on n'a jamais à la fois une plongée et une contre-plongée, ni un gros plan et un plan d'ensemble.

Cependant on peut avoir à la fois **plan d'ensemble** et **profondeur**, ou bien **vue de profil** et **vue perspective**, et même **vue de profil** et **vue de face** (figure V.1.7). Ce dernier cas montre que l'indexation ne peut pas refléter la structure du document, qui dans ce cas comprend deux photographies (deux portraits) du même sujet selon deux angles différents.

Dimensions Contenu et Précisions

Ces deux dimensions jouent des rôles similaires dans la formulation d'un problème d'information (rôle thématique).

Les termes de la dimension **Précisions** ne sont pas organisés hiérarchiquement, au contraire des termes relevant de la dimension **Contenu**. Pour permettre d'exploiter les différents critères élémentaires de pertinence pour l'indexation du contenu, nous ne groupons pas ces deux axes d'indexation en une même dimension d'interrogation.

3. Pour ce même schéma, si la dimension **Auteur** est affectée d'un statut *optionnel*, le critère n'est jamais vérifié et ne présente par conséquent aucun intérêt.



Champs	Valeurs
Numéro	4702
Titre	Portrait d'Aristide Bruant
Auteur	Nadar Atelier
Genre de l'image	
Commentaire	
Contenu	canne, poète, portrait-D, costume-D, botte-D, cha- peau, pied, homme, chanteur, table-D, velours
Précisions	Bruant Aristide
Connotation	
Morphologie	vue d'ensemble, photographie en atelier, vue de profil, vue de face

FIG. V.1.7 – *Portrait d'Aristide Bruant*

Dimension Genre

L'arborescence qui organise les termes relatifs à cette dimension comprend deux termes génériques : **Portrait** et **Paysage**, et un ensemble réduit de termes divers. De même que pour la dimension **Morphologie**, il est fréquent (mais pas systématique) que deux termes voisins s'excluent mutuellement.

Ainsi, quand une image est indexée par plusieurs termes relatifs à la dimension **Genre**, on a rarement deux termes relevant du même terme générique. Par exemple, une photographie de paysage est rarement à la fois un **Paysage rural** et un **Paysage urbain**.

L'interrogation peut avoir lieu sur les termes génériques portrait et paysage (alors que cela n'a pas de sens pour la dimension **Morphologie**), car ces termes désignent l'union de leurs termes spécifiques.

Nous reprenons plus loin ces stratégies d'utilisation des dimensions pour montrer comment elles permettent de définir un contrôle sur la formulation.

1.2.2 Statuts de dimensions et types de recherche

Les dimensions utilisées dans la formulation de la requête et leur statut peuvent caractériser des types de recherche, au sens où l'on peut leur associer des stratégies de recherche d'information pour l'utilisateur.

Un problème d'information peut par exemple se formuler de manière thématique : le schéma de pertinence comprend alors un critère de **Contenu** qui joue le rôle de critère principal. En fonction des documents retrouvés, d'autres critères de dimension peuvent s'ajouter pour sélectionner un sous-ensemble des documents retrouvés selon d'autres points de vue, par exemple la morphologie. Le schéma de pertinence a alors la forme suivante⁴ :

$$S_5$$

<i>obligatoire</i>	Contenu	
<i>optionnel</i>	Morphologie	

On peut ainsi définir le *type de requête thématique* comme l'ensemble des requêtes dont la formulation initiale est de la forme :

$$S_6$$

<i>obligatoire</i>	Contenu	
<i>optionnel</i>	<Dim>	
$\vdots$		
<i>optionnel</i>	<Dim>	

4. Les profils de schéma de pertinence sont notés de cette façon pour signifier que certains éléments (ici les critères élémentaires et leurs statuts) restent à spécifier pour former un schéma de pertinence complet.

Un problème d'information peut aussi se formuler avant tout en termes connotatifs : le schéma de pertinence comprend alors un critère de **Connotation** qui joue le rôle de critère principal. En fonction des documents retrouvés, d'autres critères de dimension peuvent s'ajouter pour sélectionner un sous-ensemble des documents retrouvés selon d'autres points de vue, par exemple le contenu, et la morphologie. Le schéma de pertinence a la forme générale suivante :

$$S_7$$

<i>obligatoire</i>	Connotation	
<i>optionnel</i>	Contenu	
<i>optionnel</i>	Morphologie	

On peut ainsi définir le *type de requête connotatif* comme l'ensemble des requêtes dont la formulation initiale est de la forme :

$$S_8$$

<i>obligatoire</i>	Connotation	
<i>optionnel</i>	$\langle \text{Dim} \rangle$	
$\vdots$		
<i>optionnel</i>	$\langle \text{Dim} \rangle$	

D'une manière générale, on parle de *critère de dimension principal* pour désigner le critère de dimension formulé initialement, auquel est associé un statut obligatoire, et de *critères de dimension secondaires*, pour les critères qui s'ajoutent au critère de dimension principal en tant que critères optionnels.

De façon analogue, on peut définir un *type de requête exploratoire*, où la dimension utilisée ($\langle \text{Dim} \rangle$) correspond à une direction d'exploration du corpus. Le schéma de pertinence initial a la forme :

$$S_9$$

<i>optionnel</i>	$\langle \text{Dim} \rangle$	
------------------	------------------------------	--

Pour ce type de schéma, la réponse contient la totalité des documents du corpus, séparés en deux classes : ceux qui vérifient le critère de dimension, et ceux qui ne le vérifient pas. Ce résultat constitue un point de départ pour l'élaboration de nouveaux critères permettant de focaliser la recherche en fonction du contenu du corpus.

1.2.3 Contrôle pour la formulation

Contrôles possibles pour la formulation d'un critère de **Connotation**

En raison des propriétés de l'indexation, pour ne pas pénaliser le rappel, il est recommandé d'utiliser dans un critère de connotation autant de termes proches que possible, en les assortissant du statut optionnel.

On peut ainsi définir une formulation typique d'un critère de connotation, comme suit

S_{10}	$\langle statut \rangle$	Connotation	<i>optionnel</i>	t_1
			<i>optionnel</i>	t_2
			$\vdots$	$\vdots$
			<i>optionnel</i>	t_n

Dans la perspective de soulager l'utilisateur d'une partie du contrôle tout en le guidant dans sa formulation, cette dimension se prête à l'application de techniques d'expansion de requête. Il faut pour cela définir une relation de proximité sur les termes de connotation, et procéder à une expansion automatique du critère de dimension.

Cependant, en fonction de la situation courante, et selon le rôle que joue le critère de connotation, l'expansion automatique de requête est plus ou moins opportune.

Prenons le cas d'une situation où de nombreux documents sont retrouvés, et où l'objectif final de l'utilisateur est d'en sélectionner un ou deux, sans souci d'exhaustivité. Le critère de connotation peut servir à réaliser une sélection grossière qui oriente la consultation des documents retrouvés, afin d'éviter d'en parcourir un trop grand nombre. Dans ce cas, le critère de connotation exprime une préférence, et l'objectif est que ce critère sélectionne un sous-ensemble des documents retrouvés de taille raisonnable (autour de 10, par exemple). L'expansion de requête ne s'avère alors utile que si elle respecte cette contrainte. On pourrait définir dans ce but une méthode d'expansion de requête contrôlée par la sélectivité des termes sur l'ensemble des documents retrouvés.

Par ailleurs, la dimension **Connotation** gagne à être utilisée pour formuler un critère secondaire. En effet, le manque de précision et d'objectivité de l'indexation sur laquelle elle repose rend difficile une recherche précise sur la connotation. Cela peut être traduit par un statut optionnel imposé en permanence dans le schéma de pertinence.

S_{11}	<i>optionnel</i>	Connotation	
----------	------------------	-------------	--

Contrôles possibles pour la formulation d'un critère d'Auteur

Les documents du corpus ont toujours au plus un auteur. On peut répercuter explicitement ce phénomène sur la forme générale des critères d'auteur, en imposant qu'ils aient l'une des deux formes suivantes.

S_{12}	$\langle statut \rangle$	Auteur	<i>obligatoire</i>	t
----------	--------------------------	--------	--------------------	---

S_{13}	$\langle statut \rangle$	Auteur	<i>optionnel</i>	t_1
			<i>optionnel</i>	t_2
			$\vdots$	$\vdots$
			<i>optionnel</i>	t_n

Par exemple, sitôt que plus d'un terme est entré dans la dimension **Auteur**, les statuts associés aux critères élémentaires sont automatiquement changés en *optionnel*.

Contrôles possibles pour la formulation d'un critère de Numéro

Il en est de même pour les critères de numéro : on peut restreindre leur expression à ces deux formes sans perdre de généralité.

S_{14}	$\langle statut \rangle$	Numéro	<i>obligatoire</i>	t
----------	--------------------------	--------	--------------------	---

S_{15}	$\langle statut \rangle$	Numéro	<i>optionnel</i>	t_1
			<i>optionnel</i>	t_2
			$\vdots$	$\vdots$
			<i>optionnel</i>	t_n

Contrôles possibles pour la formulation d'un critère de Morphologie

La formulation d'un critère de morphologie peut se réaliser à plat, sans tenir compte des diverses rubriques que l'indexation recouvre (Plan, Vue, Lumière, Divers). Cependant, si la morphologie constitue la dimension principale d'interrogation, il peut être utile de structurer à nouveau cette dimension en sous-dimensions, afin de permettre une formulation plus fine. Il suffit pour cela de généraliser la définition du schéma de pertinence à une structure à trois niveaux hiérarchiques au lieu de deux.

<i>obligatoire</i>	Morphologie	$\langle statut \rangle$	Vue	$\langle statut \rangle$	t
		$\langle statut \rangle$	Lumière	$\langle statut \rangle$	t
		$\langle statut \rangle$	Plan	$\langle statut \rangle$	t
		$\langle statut \rangle$	Divers	$\langle statut \rangle$	t

1.3 Exemples de situations et réactions

Dans le chapitre 4, nous avons défini quatre types de situations problématiques (pénurie, non-pertinence, non discrimination et non-focalisation) auxquelles nous avons associé des types d'actions possibles en fonction de paramètres liés à la situation. Nous illustrons ici un à un ces types de situations autour d'exemples tirés du prototype VIZ. Pour chacun de ces types, nous donnons un schéma de pertinence et

son résultat ; nous décrivons la situation problématique dans laquelle se trouve l'utilisateur ; nous appliquons de diverses manières les actions associées à cette situation et analysons le résultat obtenu.

Les exemples sont tous construits autour du schéma de pertinence donné dans la figure V.1.8, qui traduit la recherche de documents où des bateaux et des fleuves apparaissent, dans le contexte d'un paysage, et qui produisent un effet apaisant.

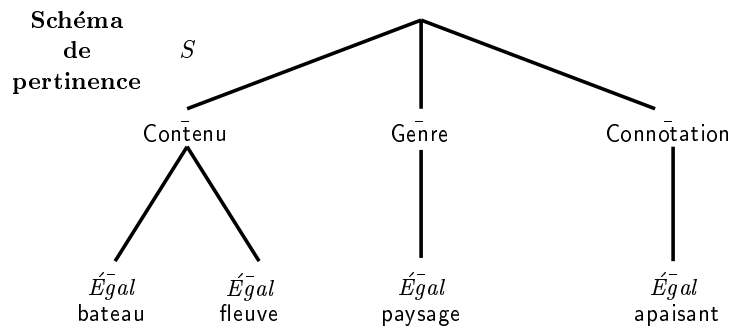


FIG. V.1.8 – Exemple de schéma de pertinence pour VIZ

Nous faisons l'hypothèse que le critère de **Contenu** constitue le critère de recherche principal, et que parmi ses critères élémentaires, **bateau** est le critère élémentaire principal. Par la suite, nous ajoutons dans certains cas une hypothèse supplémentaire ou une autre, afin de préciser la nature du problème d'information quand cela s'avère nécessaire pour la situation à illustrer.

1.3.1 Pénurie

La situation de pénurie est définie comme une situation où aucun document n'est retrouvé et où l'utilisateur n'est pas satisfait.

Le schéma de pertinence S_1 suivant, établit une situation où $N_{DR} = 0$, c'est-à-dire qu'aucun document n'est retrouvé.

S_1	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	paysage	
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	apaisant	

Nous développons ci-dessous 5 exemples d'actions de type **RelâcherP**, et deux exemples d'actions de type **RelâcherS** et discutons de leur capacité à permettre de sortir de cette situation de pénurie. Nous complétons les exemples d'action de type **RelâcherS** sur un autre exemple (S_8), qui nous permet d'illustrer un meilleur

contexte d'application des actions de type **RelâcherS** en cas de pénurie. Enfin nous montrons comment la technique d'expansion de requête appliquée à la dimension **Connotation** permet aussi de résoudre une situation de pénurie, par l'intermédiaire d'actions de type **Détailler**. Il s'agit d'un cas marginal, car l'adéquation de cette action à la situation tient aux propriétés particulières de la dimension **Connotation** que nous avons discutées plus haut.

Actions de type RelâcherP

Quatre actions de type **RelâcherP** peuvent être appliquées aux dimensions de ce schéma, pour donner S_2 (on relâche le statut de la dimension **Connotation**), S_3 (on relâche le statut de la dimension **Genre**), et S_4 (on relâche les statuts des dimensions **Connotation** et **Genre**).

Le schéma S_2 ci-dessous produit une classe vide et une classe comprenant quatre images, parmi lesquelles aucun bateau n'apparaît en gros plan (tableau V.1.5⁵ et figure V.1.9).

S_2	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	paysage	
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	apaisant	

Cl.	S-Cl.	Tot.Cl.
0	4578 4387 0162 0163	4

TAB. V.1.5 – Réponse pour le schéma de pertinence S_2

Le schéma S_3 ci-dessous produit une classe vide et une classe comprenant trois images, où des bateaux apparaissent en assez gros plan, dans un contexte urbain (grands ponts et constructions) (tableau V.1.6 et figure V.1.10).

S_3	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i>	paysage	
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	apaisant	

5. Les tableaux présentent la réponse du système en les organisant en classes et sous-classes. Les lignes horizontales doubles séparent les classes de degrés différents; les lignes horizontales simples séparent les sous-classes de degré différents. Une classe est décrite de gauche à droite par son numéro, son contenu, et le nombre total de documents qu'elle contient; récursivement, une sous-classe est décrite de gauche à droite par son numéro, son contenu et le nombre de documents qu'elle contient.

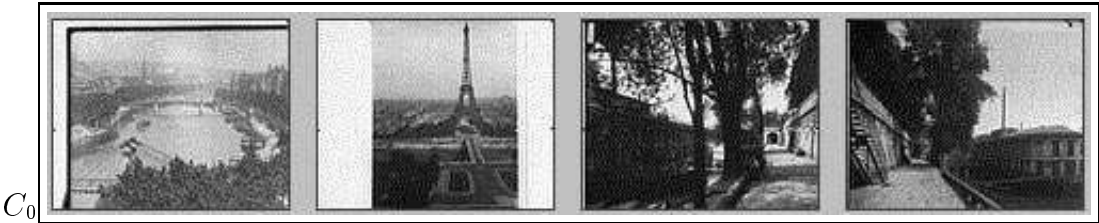


FIG. V.1.9 – Réponse du schéma de pertinence S_2

Cl.	S-Cl.	Tot.Cl.
0	0078 1289 1290	3

TAB. V.1.6 – Réponse pour le schéma de pertinence S_3



FIG. V.1.10 – Réponse du schéma de pertinence S_3

Le schéma S_4 ci-dessous produit quatre classes, la première étant vide (conjonction des trois critères de dimension). Les trois autres classes correspondent respectivement à la classe non vide du schéma S_2 , la classe non vide du schéma S_3 , et une classe de degré de pertinence moindre comprenant 5 images (tableau V.1.7 et figure V.1.11).

S_4	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i>	paysage	
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	apaisant	

Cl.	S-Cl.	Tot.Cl.
10	0162 4387 0163 4578	4
01	1290 1289 0078	3
00	4507 4508 0018 1293 0010	5

TAB. V.1.7 – Réponse pour le schéma de pertinence S_4

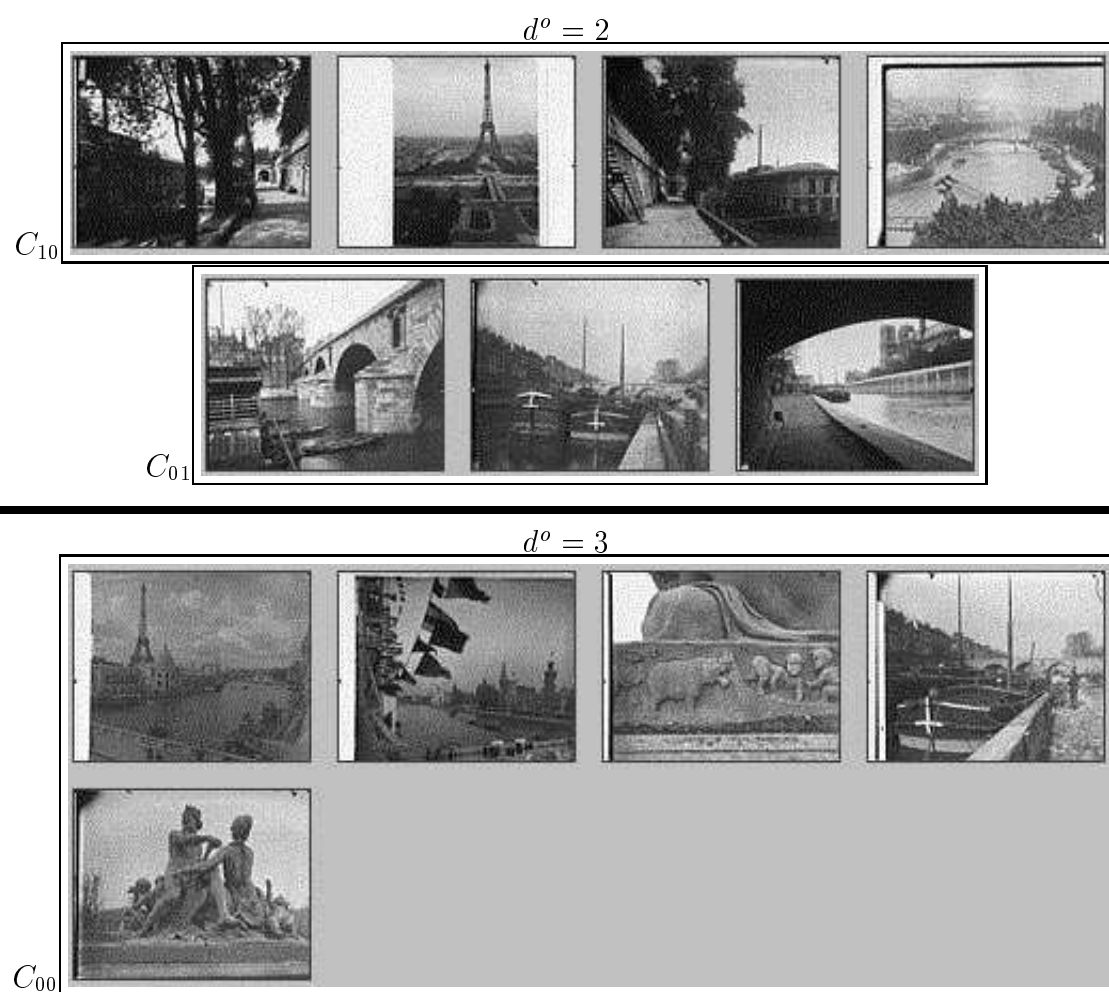
bateau étant le critère élémentaire principal de la dimension **Contenu**, une action de type **RelâcherP** peut être appliquée au critère relatif au **fleuve**, pour s'assurer que le terme **fleuve** ne nuit pas au rappel, ce qui donne le schéma de pertinence S_5 ci-dessous. Ce schéma produit une sous-classe comprenant 4 images représentant des bateaux dans un contexte maritime exclusivement (tableau V.1.8 et figure V.1.12).

S_5	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>optionnel</i>	fleuve	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	paysage	
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	apaisant	

Cl.	S-Cl.	Tot.Cl.
	0 4741 4562 4486 4600	4
		4

TAB. V.1.8 – Réponse pour le schéma de pertinence S_5

Ces actions de type **RelâcherP** permettent d'obtenir une réponse non vide. Si l'on applique toutes les actions simultanément (schéma de pertinence S_6), on obtient

FIG. V.1.11 – Réponse du schéma de pertinence S_4 FIG. V.1.12 – Réponse du schéma de pertinence S_5

le même éventail de réponses, organisé hiérarchiquement selon des classes et sous-classes (tableau V.1.9 et figure V.1.13).

S_6	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	paysage	
	<i>optionnel</i>		<i>obligatoire</i>	apaisant	

Cl.	S-Cl.					Tot.Cl.
11	0	4486 4600 4741 4562 (la classe non vide de S_5)	4			4
10	1	4387 4578 0163 0162 (la classe non vide de S_2)	4			7
	0	4590 4579 4564 (nouveaux documents où fleuve n'apparaît pas)	3			
01	1	0078 1290 1289 (la classe non vide de S_3)	3			3
00	1	4507 4508 1293 0010 0018 (la dernière classe de S_4)	5			23
	0	4606 4743 4567 4561 4568 4638 4782 4776 4767 4768 4570 4655 4511 4516 0173 0140 2024 2495 (nouveaux documents où fleuve n'apparaît pas)	18			

TAB. V.1.9 – Réponse pour le schéma de pertinence S_6

Cette réponse comprend à la fois des documents pertinents et des documents non pertinents et constitue un point de départ pour l'évaluation, ce qui résout le problème d'entrave à l'évaluation. De plus, elle fournit une explication des rôles joués par les différents critères en matière de sélectivité sur le corpus.

Actions de type RelâcherS

Une action de type **RelâcherS** peut être appliquée au critère élémentaire de contenu **fleuve** (S_7), afin d'explorer les autres possibilités de contexte aquatique pour les bateaux.

S_7	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	fleuve	<i>Vois</i>
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	paysage	
	<i>obligatoire</i>		<i>obligatoire</i>	apaisant	

On n'obtient aucune image à nouveau. Cette reformulation a donc échoué. Cet échec n'est pas anecdotique : lorsque le schéma de pertinence est très contraint (de nombreux critères obligatoires), les effets que peuvent produire les modifications des critères élémentaires abstraits sont souvent négligeables.

En revanche considérons le schéma de pertinence S_8 , dont la réponse est vide elle aussi (pénurie), mais qui est moins contraint.

S_8	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>obligatoire</i>		<i>obligatoire</i>	rivière	<i>Égal</i>

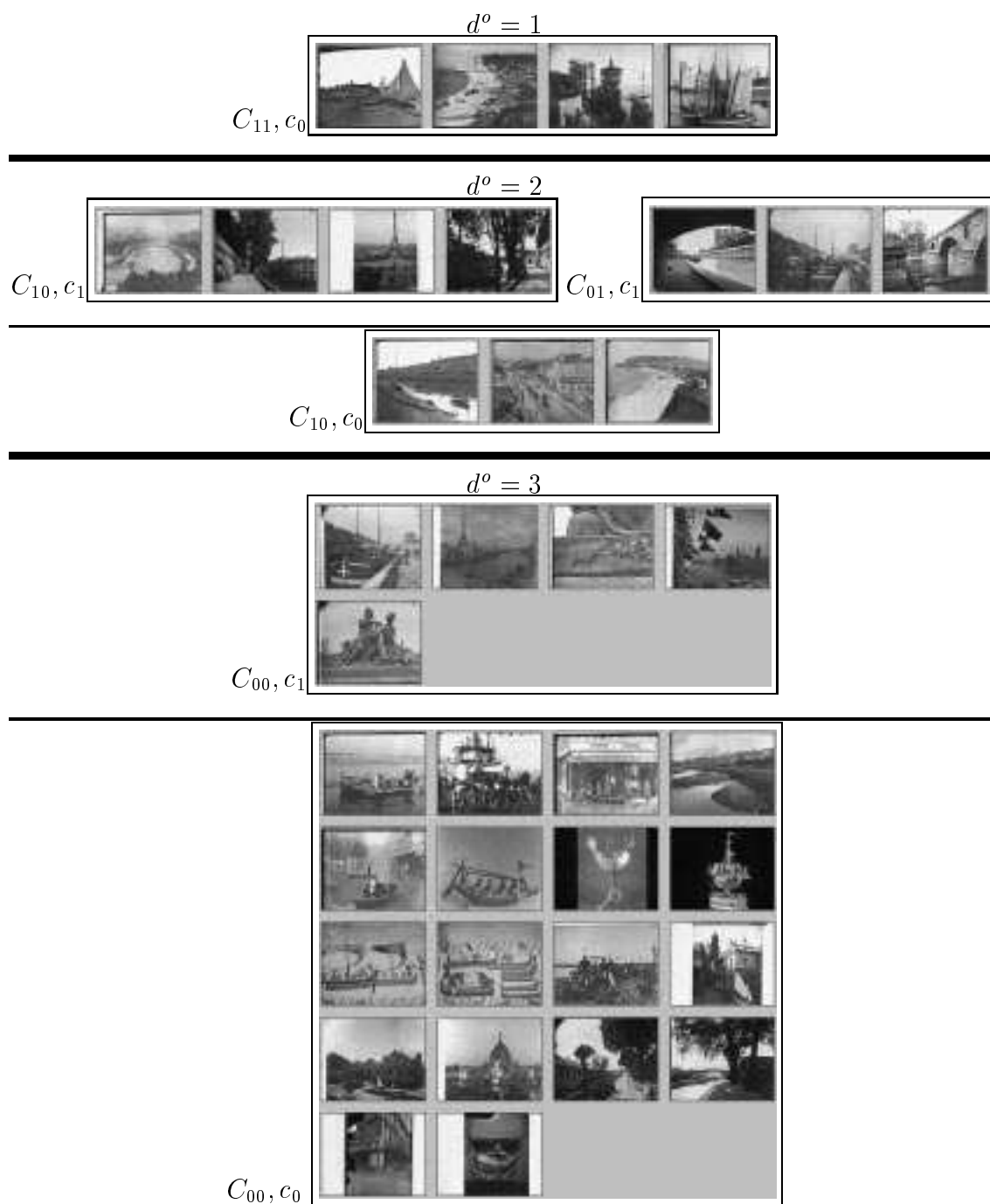


FIG. V.1.13 – Réponse du schéma de pertinence S_6 (voir un agrandissement pages iii et iv de l'annexe)

Si l'on relâche le critère abstrait de rivière en le passant à *Vois* (S_9), on obtient 12 images présentant des fleuves ou d'autres types de cours d'eau, avec des bateaux (tableau V.1.10 et figure V.1.14).

S_9	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>obligatoire</i>	rivière	<i>Vois</i>

Cl.	S-Cl.	Tot.Cl.
	4578 4387 4507 4508 0018 0078 1293 1289 1290 0010 0162 0163	12

TAB. V.1.10 – Réponse pour le schéma de pertinence S_9

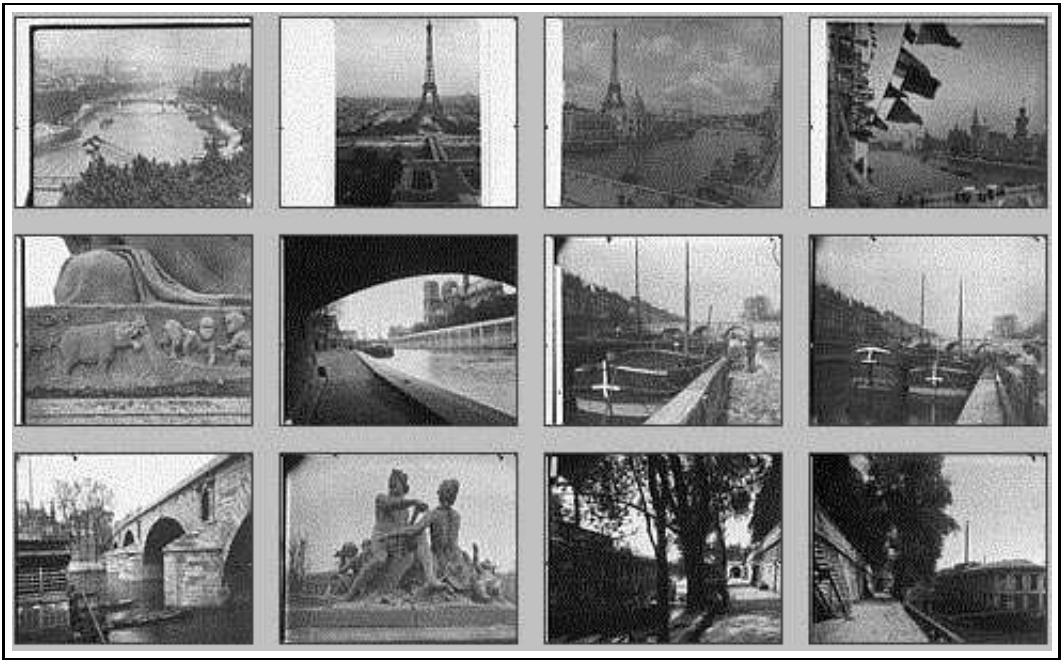


FIG. V.1.14 – Réponse du schéma de pertinence S_9

Ainsi la forme de la requête fournit des indices quant à la reformulation la plus utile. Des connaissances statistiques sur le lien entre ces indices et l'utilité de la reformulation peut être utilisée pour l'automatisation de la reformulation. Par exemple nous pouvons proposer que si le nombre de critères obligatoires d'un schéma de pertinence est grand, les actions de type **RelâcherS** ont peu de chances d'augmenter le résultat, et donc de résoudre une situation de pénurie.

Expansion de requête pour la dimension **Connotation** (Détailler)

Enfin, les remarques relatives à la dimension **Connotation** (section 1.2) peuvent s'appliquer ici pour que le critère de connotation reflète mieux le sens voulu. Pour cela on ajoute au schéma S_1 des critères élémentaires de connotation proches de **apaisant**, avec un statut optionnel (S_{10}).

S_{10}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	paysage	
	<i>obligatoire</i>	Connotation	<i>optionnel</i>	apaisant	
			<i>optionnel</i>	paisible	
			<i>optionnel</i>	harmonieux	
			<i>optionnel</i>	poétique	

Cl.	S-Cl.				Tot.Cl.
	0010	0163	0162	2	4
	0000	4578	4387	2	

TAB. V.1.11 – Réponse pour le schéma de pertinence S_{10}

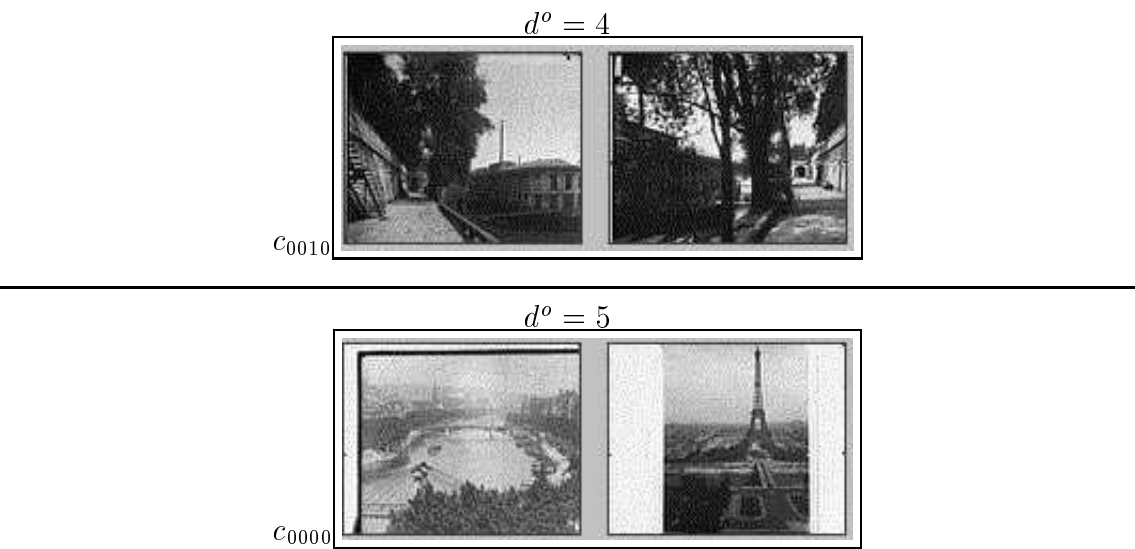


FIG. V.1.15 – Réponse du schéma de pertinence S_{10}

La réponse comprend 4 documents, dont deux vérifient le critère **poétique** et deux autres aucun des critères de connotation (tableau V.1.11 et figure V.1.15).

1.3.2 Non-pertinence

La situation où aucun des documents retrouvés n'est ni pertinent, ni non pertinent est une situation de non-pertinence.

Le schéma de pertinence S_{11} suivant établit une telle situation, car il produit une classe vide et une classe comprenant 4 images, sur lesquelles aucun bateau n'apparaît en suffisamment gros plan pour satisfaire l'utilisateur (il s'agit du même schéma que S_2 , tableau V.1.12 et figure V.1.9).

S_{11}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	paysage	
			<i>obligatoire</i>	apaisant	

Cl.	S-Cl.	Tot.Cl.
0	4578 4387 0162 0163	4

TAB. V.1.12 – Réponse pour le schéma de pertinence S_{11}

Nous donnons deux exemples d'actions de type **RelâcherP** et un exemple d'action de type **RelâcherS** qui s'appliquent à ce schéma pour résoudre le problème d'entrave à la reformulation que pose ce résultat. Nous ajoutons un nouvel exemple de schéma de pertinence (S_{14}) auquel nous appliquons une action de type **RelâcherS**, qui complète l'illustration des informations que permet d'obtenir ce type d'action dans une situation de non-pertinence.

Actions de type RelâcherP

On peut procéder à une action du type **RelâcherP** sur le critère de genre afin de vérifier la sémantique de ce critère, ce qui donne le schéma de pertinence S_{12} (il s'agit du même schéma que S_4 , tableau V.1.13 et figure V.1.11).

S_{12}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	paysage	
			<i>obligatoire</i>	apaisant	

On obtient alors à nouveau une classe vide, une classe comprenant les 4 images déjà retrouvées avec le schéma précédent, une classe (pour la conjonction des critères de contenu et de connotation, et la négation du critère de genre) comprenant 3 images, où l'une d'elles présente des péniches en gros plan et une dernière classe

Cl.	S-Cl.	Tot.Cl.
10	0162 4387 0163 4578	4
01	1290 1289 0078	3
00	4507 4508 0018 1293 0010	5

TAB. V.1.13 – Réponse pour le schéma de pertinence S_{12}

(pour le critère de contenu et la négation des deux autres) comprenant 5 images, où deux d'entre elles présentent des bateaux fluviaux en gros plan (tableau V.1.13 et figure V.1.11).

Cela tend à prouver que le critère de genre **paysage** est mal choisi, car les images où l'on peut voir des bateaux en gros plan ne sont pas indexées comme des paysages. On peut déduire que le sens attribué par les indexeurs au terme **paysage** est lié au fait qu'il n'existe pas de premier plan marquant qui vole la vedette au paysage lui-même, ou bien que le corpus ne comprend pas d'image avec un bateau fluvial en gros plan dans le contexte d'un paysage.

Ce constat conduit à une seconde étape de reformulation. Dans le cas où l'on déduit qu'il s'agit d'un défaut de formulation, on supprime ce critère de genre (**Supprimer**) ou bien on établit le statut optionnel de ce critère comme définitif, afin de traduire cette incertitude pour la suite de la recherche.

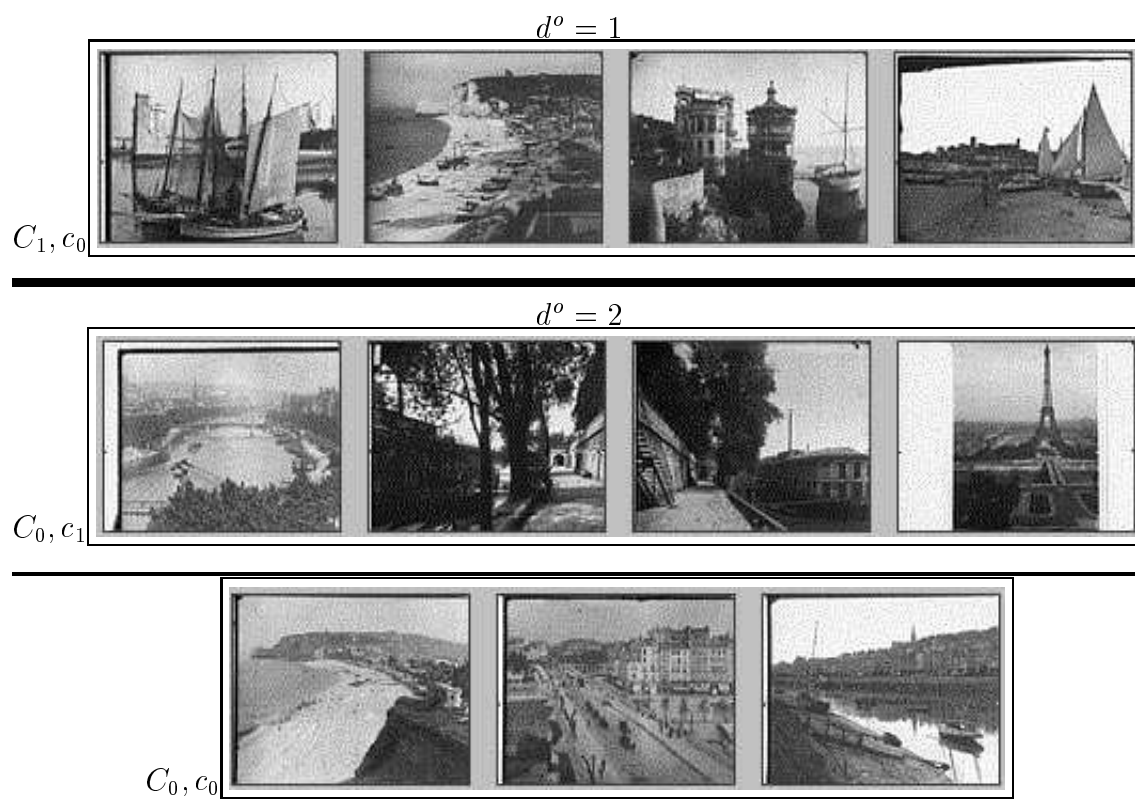
On peut aussi procéder à une action du type **RelâcherP** sur le critère élémentaire **fleuve** afin de vérifier la sémantique de ce terme, ce qui donne le schéma de pertinence S_{13} .

S_{13}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
			<i>optionnel</i>	fleuve	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	paysage	
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	apaisant	

Cl.	S-Cl.	Tot.Cl.
1	0 4486 4562 4741 4600	4
0	1 4578 0162 0163 4387	4
	0 4564 4590 4579	3
		7

TAB. V.1.14 – Réponse pour le schéma de pertinence S_{13}

On obtient alors à nouveau une classe vide, et une classe comprenant 11 images, organisées en 2 sous-classes : celle de degré 1 contient les images précédemment

FIG. V.1.16 – Réponse du schéma de pertinence S_{13}

retrouvées (contexte fluvial), et celle de degré 2 comprend 7 images (contexte maritime), ce qui confirme la capacité du terme **fleuve** à produire le sens voulu (tableau V.1.14 et figure V.1.16).

Actions de type RelâcherS

Étant donné le problème posé par la réponse du schéma S_{12} évoqué précédemment, on peut chercher à modifier la sémantique du critère **paysage** pour qu'il traduise le sens que l'on cherche à exprimer. Dans le contexte de cette application, on ne dispose pas d'une structure sur l'ensemble des termes d'indexation relatifs au genre, et par conséquent de critères abstraits permettant de relâcher les critères élémentaires du point de vue de leur sémantique. Si c'était le cas, on pourrait relâcher le critère abstrait associé à **paysage** (action de type **RelâcherS**) afin de lui donner un sens moins fort pour traduire l'incertitude de l'adéquation de ce terme à la formulation du problème d'information.

Prenons un autre exemple. Le schéma de pertinence S_{14} produit une sous-classe comprenant 4 images représentant des bateaux dans un contexte maritime exclusivement, ce qui constitue une situation de non-pertinence (il s'agit du même schéma que S_5 , tableau V.1.15 et figure V.1.12).

S_{14}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>optionnel</i>		<i>obligatoire</i>	fleuve	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	paysage	
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	apaisant	

Cl.	S-Cl.	Tot.Cl.
0	4741 4562 4486 4600	4
		4

TAB. V.1.15 – Réponse pour le schéma de pertinence S_{14}

On peut appliquer une action de type **RelâcherS** en relâchant le critère abstrait associé à **fleuve**, afin de voir si cela permet de retrouver des documents suffisamment pertinents (S_{15}).

S_{15}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>optionnel</i>		<i>obligatoire</i>	fleuve	<i>Vois</i>
	<i>obligatoire</i>	Genre	<i>obligatoire</i>	paysage	
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	apaisant	

La réponse reste la même (il s'agit du même schéma que S_5 , tableau V.1.16 et figure V.1.12). La raison en est que les documents indexés par les voisins (**torrent**,

Cl.	S-Cl.	Tot.Cl.
	0 4741 4562 4486 4600 4	4

TAB. V.1.16 – Réponse pour le schéma de pertinence S_{15}

berge, rivière, autre cours d'eau) ne présentent pas de bateau. Cette particularité du corpus fait que cette action ne permet pas de sortir de la situation de non-pertinence. Cependant cette étape fournit l'information supplémentaire que ce n'est pas le critère fleuve qui compromet le rappel.

1.3.3 Non-discrimination

Le schéma de pertinence S_{16} établit une situation de non-discrimination. En effet, 49 images sont retrouvées dans une unique classe, et en consultant simplement les 12 premières images, on constate que certaines mettent en scène des cours d'eau dans un contexte urbain, alors que l'on cherche des cours d'eau dans un contexte rural de préférence (hypothèse supplémentaire sur le problème d'information) (tableau V.1.17 et figure V.1.17).

S_{16}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	cours d'eau	<i>Égal</i>
----------	--------------------	---------	--------------------	-------------	-------------

Cl.	S-Cl.	Tot.Cl.
	4572 4377 4591 4592 4578 4581 4471 4752 4481 4727 4721 4656 4387 4485 4732 4671 4505 4507 4508 4497 4751 4731 0080 0018 3240 2351 0078 0079 0486 1293 1339 1340 0447 1289 1290 1215 0235 0011 0010 0013 0014 0015 0101 0145 0162 0163 0016 0017 2073	49

TAB. V.1.17 – Réponse pour le schéma de pertinence S_{16}

Nous donnons deux exemples d'actions de type **Détailler** qui, en s'appliquant successivement, permettent de séparer les documents pertinents des documents non pertinents. Une fois le résultat obtenu, nous montrons en quoi cette modification profite à la suite de la recherche, en introduisant une action de type **RelâcherS** (S_{19}). Nous comparons les résultats obtenus pour ce schéma avec le résultat que l'on aurait obtenu si l'on n'avait pas effectué la seconde action **Détailler**.

Actions de type Détailler

On identifie un critère discriminant entre les images pertinentes et les autres, sur la base de quelques exemples tirés des douze premières images : le critère de genre



FIG. V.1.17 – Réponse du schéma de pertinence S_{16} (voir un agrandissement pages *v* et *vi* de l'annexe)

paysage. On procède à une action de type **Détailler** en ajoutant ce critère, d'où le schéma de pertinence S_{17} . On l'ajoute comme optionnel, car rien ne prouve que ce critère est effectivement approprié.

S_{17}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	cours d'eau	<i>Égal</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i>	paysage	

Cl.	S-Cl.	Tot.Cl.
1	0163 0162 0145 4721 4731 4377 4578 4656 4581 4752 4387 4485 4751 0101	14
0	4572 4591 4592 4471 4481 4727 4732 4671 4505 4507 4508 4497 0080 0018 3240 2351 0078 0079 0486 1293 1339 1340 0447 1289 1290 1215 0235 0011 0010 0013 0014 0015 0016 0017 2073	35

TAB. V.1.18 – Réponse pour le schéma de pertinence S_{17}

On obtient une classe avec 14 images où l'on trouve à la fois des paysages ruraux et des paysages urbains, et une classe avec 35 images qui correspondent à première vue (les 12 premières) à celles que l'on cherchait à éliminer (tableau V.1.18 et figure V.1.18).

L'intention d'ajouter un critère discriminant est donc réalisée par cette action de type **Détailler**, mais pas de manière définitive, puisque dans la première classe se mêlent encore des documents non pertinents aux documents pertinents.

Si l'on suppose de plus que l'on n'est pas intéressé par les paysages urbains, on peut procéder à une nouvelle action de type **Détailler** qui vise cette fois à séparer les documents non pertinents (paysages urbains) des autres. On introduit en optionnel le critère **paysage urbain** (S_{18}).

S_{18}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	cours d'eau	<i>Égal</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i>	paysage	
			<i>optionnel</i>	paysage urbain	

Cl.	S-Cl.	Tot.Cl.
1	1 4731 4578 4387 4656 0101 0163 0162 0145 8	14
	0 4721 4377 4581 4752 4485 4751 6	
0	4572 4591 4592 4471 4481 4727 4732 4671 4505 4507 4508 4497 0080 0018 3240 2351 0078 0079 0486 1293 1339 1340 0447 1289 1290 1215 0235 0011 0010 0013 0014 0015 0016 0017 2073	35

TAB. V.1.19 – Réponse pour le schéma de pertinence S_{18}

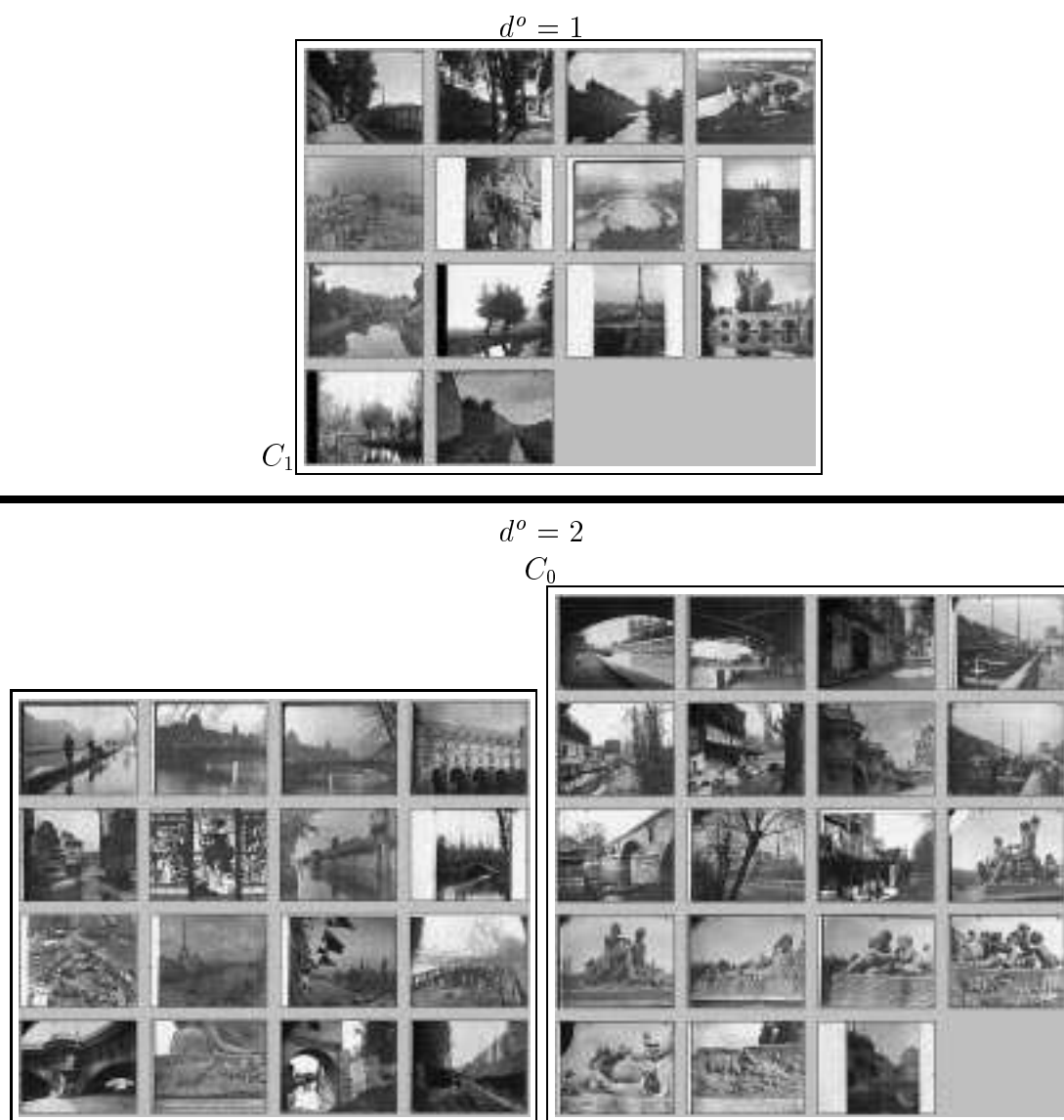


FIG. V.1.18 – Réponse du schéma de pertinence S_{17} (voir un agrandissement pages vii, viii et ix de l'annexe)

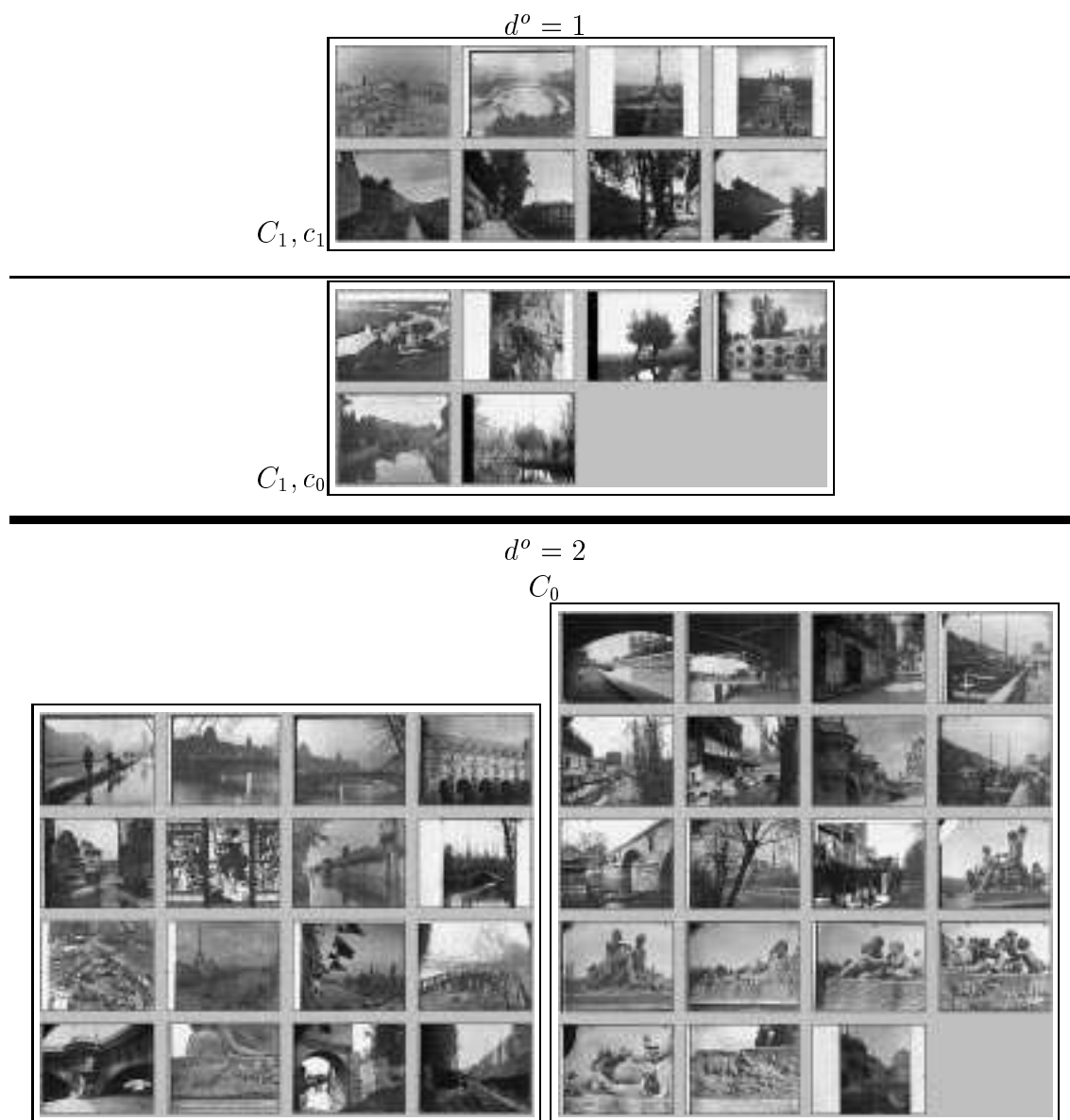


FIG. V.1.19 – Réponse du schéma de pertinence S_{18} (voir un agrandissement pages *x*, *xi* et *xii* de l'annexe)

La classe correspondant à la conjonction des critères de contenu et de genre est organisée selon deux classes, la première comprenant 8 documents non pertinents (contexte urbain), et la seconde 6 documents pertinents (autres contextes) (tableau V.1.19 et figure V.1.19).

L'intention de séparer les documents pertinents des autres est donc réalisée.

Bénéfice à moyen terme pour la suite de la recherche

Cette démarche peut paraître superfétatoire car il était aisé d'examiner les 14 documents et de séparer par l'esprit les pertinents des non pertinents. Mais si l'on veut poursuivre la recherche en élargissant le critère de contenu, il est appréciable que le critère de genre soit formulé de manière précise, afin d'organiser les nouveaux documents à venir, qui pourraient être nombreux, comme nous le montrons ci-dessous.

Dans le but de connaître les limites du corpus en matière de cours d'eau, on peut procéder à une action de type **RelâcherS** en étendant le critère de contenu **cours d'eau** à son voisinage sémantique (S_{19}).

S_{19}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	cours d'eau	<i>Vois</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i> <i>optionnel</i>	paysage paysage urbain	

Cl.	S-Cl.																Tot.Cl.
1	1	4656	4579	4596	4578	4600	4387	4731	0163	0101	0162	0145	11				25
	0	4581	4565	4564	4377	4563	4562	4741	4485	4721	4752	4594	4605	4751	3645	14	
0	4606 4382 4572 4561 4591 4592 4631 4471 4713 4473 4481 4727 4570 4732 4671 4505 4507 4508 4497 4601 0080 0018 3240 2351 0078 0079 0486 1293 1339 1340 0447 1289 1290 1215 0235 0011 0010 0013 0014 0015 0016 0017 2073 2495																44

TAB. V.1.20 – Réponse pour le schéma de pertinence S_{19}

On obtient une classe comprenant 25 documents où les deux critères de contenu et de genre sont vérifiés, et une classe comprenant 44 documents où seul le critère de contenu est vérifié. La première classe comprend deux sous-classes, la première (11 images) étant associée, comme pour le résultat de S_{18} , aux documents dont le genre ne convient pas (paysage urbain), et la seconde (14 images) à ceux dont le genre convient. Les images supplémentaires font figurer la mer, ou un plan d'eau (tableau V.1.20 et figure V.1.20).

Si l'on n'avait pas effectué l'action de type **Détailler** pour séparer ces deux sous-classes, on aurait obtenu 25 documents où documents pertinents et non pertinents seraient mêlés, ce qui pose un problème d'entrave à l'évaluation.

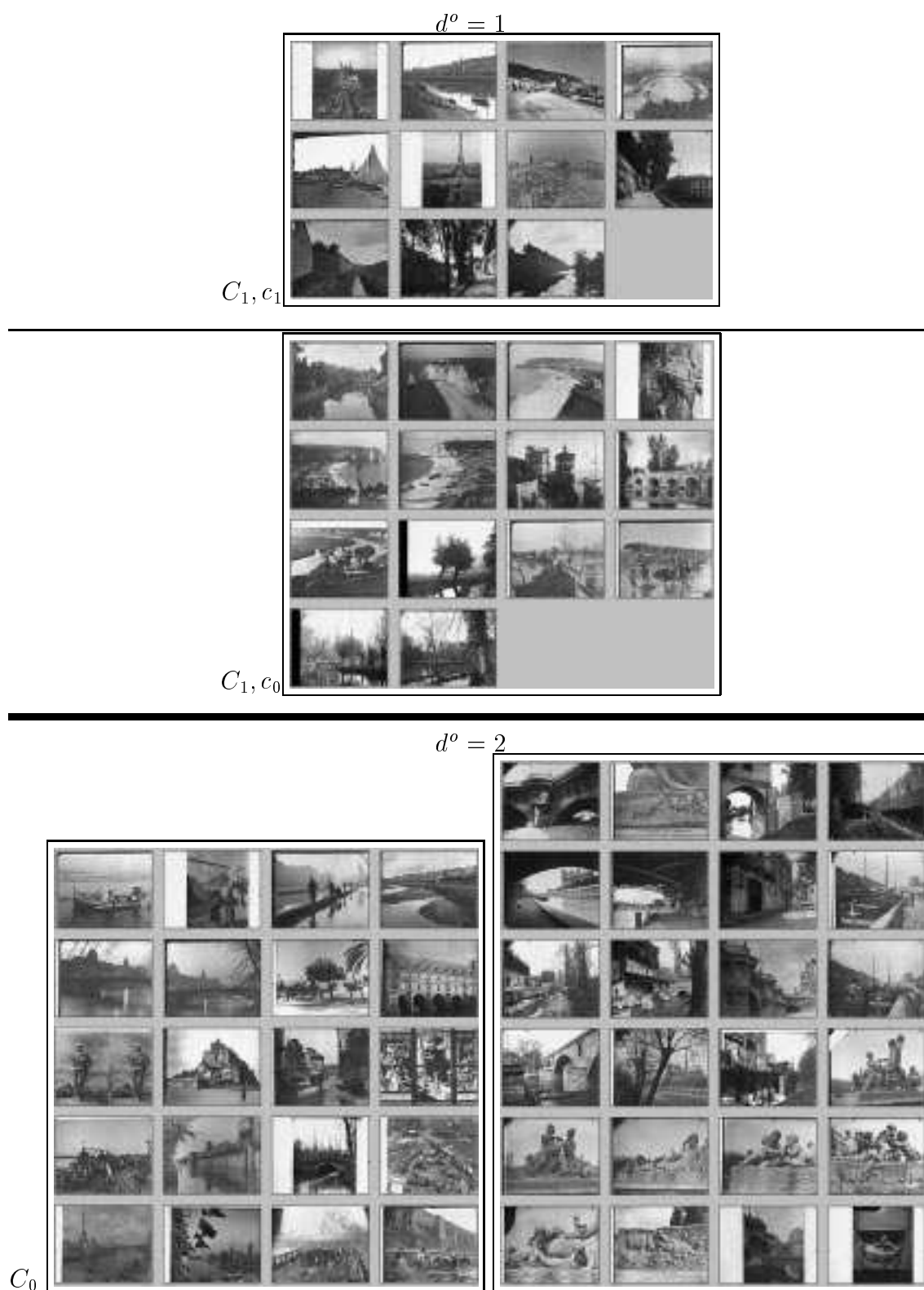


FIG. V.1.20 – Réponse du schéma de pertinence S_{19} (voir un agrandissement pages *xiii*, *xiv* et *xv* de l'annexe)

1.3.4 Non-focalisation

Actions de type ResserrerS

Le schéma de pertinence S_{20} établit une situation de non-focalisation : les images où figurent la mer ou des plans d'eau ne sont pas pertinentes (il s'agit du même schéma que S_{19} , tableau V.1.21 et figure V.1.20).

S_{20}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	cours d'eau	<i>Vois</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i> <i>optionnel</i>	paysage paysage urbain	

Cl.	S-Cl.	Tot.Cl.
1	1 4656 4579 4596 4578 4600 4387 4731 0163 0101 0162 0145 11	25
	0 4581 4565 4564 4377 4563 4562 4741 4485 4721 4752 4594 4605 4751 3645 14	
0	4606 4382 4572 4561 4591 4592 4631 4471 4713 4473 4481 4727 4570 4732 4671 4505 4507 4508 4497 4601 0080 0018 3240 2351 0078 0079 0486 1293 1339 1340 0447 1289 1290 1215 0235 0011 0010 0013 0014 0015 0016 0017 2073 2495	44

TAB. V.1.21 – Réponse pour le schéma de pertinence S_{20}

On supprime les sous-classes comprenant ces images en passant le critère abstrait de cours d'eau à *Égal*. On retrouve le schéma S_{21} déjà rencontré plus haut (S_{18}), car il s'agit de l'action inverse de la précédente (tableau V.1.22 et figure V.1.19).

S_{21}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	cours d'eau	<i>Égal</i>
	<i>optionnel</i>	Genre	<i>obligatoire</i> <i>optionnel</i>	paysage paysage urbain	

Cl.	S-Cl.	Tot.Cl.
1	1 4731 4578 4387 4656 0101 0163 0162 0145 8	14
	0 4721 4377 4581 4752 4485 4751 6	
0	4572 4591 4592 4471 4481 4727 4732 4671 4505 4507 4508 4497 0080 0018 3240 2351 0078 0079 0486 1293 1339 1340 0447 1289 1290 1215 0235 0011 0010 0013 0014 0015 0016 0017 2073	35

TAB. V.1.22 – Réponse pour le schéma de pertinence S_{21}

Actions de type ResserrerP

Le schéma de pertinence S_{22} établit une situation de non-focalisation car les deux classes où le critère de genre n'est pas vérifié ne sont pas pertinentes (il s'agit du même schéma que S_6 , tableau V.1.23 et figure V.1.13).

S_{22}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>optionnel</i>	Genre	<i>optionnel</i>	fleuve	<i>Égal</i>
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	paysage	
	<i>optionnel</i>		<i>obligatoire</i>	apaisant	

Cl.	S-Cl.					Tot.Cl.
11	0	4600 4562 4741 4486			4	4
10	1	4578 0163 4387 0162			4	7
	0	4579 4590 4564			3	
01	1	0078 1289 1290			3	3
00	1	1293 4507 0018 4508 0010			5	23
	0	4606 4743 4567 4561 4568 4638 4782 4776 4767 4768 4570 4655 4511 4516 0173 0140 2024 2495			18	

TAB. V.1.23 – Réponse pour le schéma de pertinence S_{22}

En passant le critère de genre en obligatoire (action de type ResserrerP), on obtient le schéma S_{23} , où ces classes disparaissent (il s'agit du même schéma que S_{13} , tableau V.1.24 et figure V.1.16).

S_{23}	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>obligatoire</i>	Genre	<i>optionnel</i>	fleuve	<i>Égal</i>
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	paysage	
	<i>optionnel</i>		<i>obligatoire</i>	apaisant	

Cl.	S-Cl.					Tot.Cl.
1	0	4600 4562 4741 4486			4	4
0	1	4387 4578 0163 0162			4	7
	0	4590 4579 4564			3	

TAB. V.1.24 – Réponse pour le schéma de pertinence S_{23}

1.4 Exemple de recherche

Nous développons ici un exemple d'utilisation du prototype VIZ. Nous commençons par définir un problème d'information, et sa formulation initiale. Au fur et à mesure du déroulement de la session, la formulation de ce problème d'information évolue dans une certaine direction déterminée par les modifications de la situation et des préférences de l'individu qui utilise le système.

Pour éviter les confusions, les éléments de la formulation du problème d'information sont notés comme suit «élément». Cette notation désigne le sens que l'utilisateur attribue à ce terme, tandis que la notation *élément* désigne le sens que le système attribue à ce terme.

Le processus d'utilisation que nous développons se déroule selon les phases suivantes :

- Nous décrivons la réponse courante du système.
- Nous décrivons les jugements de pertinence émis avec leurs raisons sous-jacentes.
- Nous synthétisons l'évaluation selon 1) les nouveautés que cette étape a apporté en matière d'idées de formulation, et 2) la nature des problèmes identifiés.
- Nous décrivons les modifications du schéma de pertinence suscitées par l'évaluation.

De plus, nous supposons que lorsqu'une classe comprend plus de 16 documents, seuls les 16 premiers sont consultés et jugés, de façon à faire intervenir le paramètre de la difficulté à détecter la pertinence de documents mal classés.

1.4.1 Problème d'information

L'objectif est d'illustrer un article qui traite de la navigation pour les vacanciers, de façon à suggérer le caractère apaisant de vacances au bord de l'eau. L'article développe plus particulièrement la navigation sur les fleuves, les lacs et les canaux, qui présente moins de dangers pour les amateurs que la navigation en mer.

1.4.2 Étape 1

S_1	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>obligatoire</i>	Connotation	<i>obligatoire</i>	apaisant	
	<i>obligatoire</i>	Morphologie	<i>obligatoire</i>	gros plan	

Ce schéma de pertinence produit une situation de pénurie (aucun document n'est retrouvé), qui constitue un problème de nature pragmatique. La modification à apporter est de type *RelâcherP*.

On procède à une action **RelâcherP** sous la forme de deux opérations qui donnent aux dimensions secondaires (**Connotation** et **Morphologie**) un statut optionnel.

1.4.3 Étape 2

S_2	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	apaisant	
	<i>optionnel</i>	Morphologie	<i>obligatoire</i>	gros plan	

Le résultat comprend au total 37 documents. Le tableau V.1.25 donne les titres des photographies classe par classe, tandis que la figure V.1.21 montre les photos elles-mêmes.

Comme on pouvait le prévoir, la classe de degré 1 C_{11} est vide, puisque son schéma homologue correspond au schéma de pertinence précédent S_1 .

Deux classes de degré 2 apparaissent: C_{10} et C_{01} .

C_{10} comprend les documents retrouvés vérifiant le critère de **Connotation** et pas le critère de **Morphologie**. Les documents de cette classe représentent des bateaux (réels cette fois) dans un contexte maritime ou fluvial, dont certains sont bien visibles et en assez gros plan, tandis que d'autres sont vus de loin. Les bateaux apparaissant sur des fleuves y apparaissent dans un contexte urbain (immeubles, ponts).

C_{01} comprend les documents retrouvés vérifiant le critère de **Morphologie** et pas le critère de **Connotation**. Un unique document est retrouvé, qui représente effectivement un bateau, mais ce bateau apparaît dans la photographie en tant que représentation sculptée d'un bateau.

Enfin la classe C_{00} de degré 3 comprend 29 documents. Seuls les 16 premiers vont être consultés, conformément au procédé choisi.

C_{00} comprend les documents retrouvés ne vérifiant ni le critère de **Connotation** ni le critère de **Morphologie**. D'abord certaines photos montrent des représentations de bateaux (au même titre que la classe C_{01}). L'ensemble des autres photos sont variées tant en matière de morphologie (plan moyen, plongée, etc.) qu'en matière d'animation (les personnages posent ou sont actifs). Certaines sont des paysages maritimes variés, d'autres des photos prises dans les villes. Enfin, une image («CAVALIERS DÉBARQUANT D'UN BAC AU TURKESTAN») représente un paysage non construit et sauvage, mais le bateau (un bac sur un lac) est très peu visible.

Étant donné le problème d'information que nous prenons pour exemple, voici les jugements de pertinence associées à cette réponse, avec leurs raisons sous-jacentes.

Classe C_{10} : Parmi les 7 documents retrouvés de cette classe, où seul le critère de **Connotation** est vérifié, les bateaux apparaissant dans un contexte maritime ne sont pas pertinents, car l'utilisateur ne veut pas de paysage maritime, qui évoquent un certain danger pour des amateurs. Par ailleurs, les autres images (contexte fluvial)

Cl.	S-Cl.	Tot.Cl.
10	LE RESTAURANT « LA RÉSERVE » PRÈS DE NICE VUE DE LA CATHÉDRALE NOTRE-DAME PRISE SOUS L'ARCHE DU PONT DE L'ARCHEVÊCHÉ VOILIERS SUR LA GRÈVE DÉTAIL DU PONT MARIE ET D'UN BAIN FLOTTANT BATEAUX SUR LA PLAGE PORT D'ÉTRETAT VOILIER ET TOUR SAINT NICOLAS DANS L'AVANT PORT DE LA ROCHELLE DEUX PÉNICHES À QUAI DEVANT LE PONT MARIE	7
01	HIPPOPOTAME ATTAQUANT DEUX PYGMÉES EN BARQUE SUR LE NIL	1
00	HOMMES AU TRAVAIL DANS UN BOUCHOT MATELOTS AU TRAVAIL GARÇONS DE CAFÉ OFFICIAIENT SUR DES BARQUES SOUS L'OEIL VIGILANT ET DISCRET DE L'ARMÉE MARÉE BASSE À TROUVILLE AU DÉBUT DU SIÈCLE GROUPE POSANT AU DESSUS DE LA PLAGE D'ÉTRETAT EN REGARDANT LES PROMENEURS VUE DE PARIS AU DÉBUT DU SIÈCLE DÉPUTÉS VAQUANT GAIEMENT À LEURS AFFAIRES EN BARQUE LE PORT DE TROUVILLE À MARÉE BASSE LA SAÔNE À LYON AU DÉBUT DU SIÈCLE SCULPTURE DE BARQUE ÉGYPTIENNE OU LES RAMEURS DIRIGÉS PAR LEUR CAPITAINE SONT EN PLEINE ACTION COQUILLAGE MONTÉ EN NAVETTE À ENCENS NEF RELIQUAIRE DE SAINTE-URSULE BRODERIE DE BAYEUX : LES NORMANDS TRAVERSENT LA MANCHE DÉTAIL DE LA TAPISSERIE DE BAYEUX : CONSTRUCTION DES BATEAUX NORMANDS CAVALIERS DÉBARQUANT D'UN BAC AU TURKESTAN PAVILLON D'INDOCHINE À L'EXPOSITION UNIVERSELLE DE 1900 TOUR EIFFEL DEPUIS L'ANCIEN PALAIS DU TROCADÉRO PAVILLON ANNAMITE À L'EXPOSITION UNIVERSELLE DE 1900 DÔME CENTRAL ET FONTAINE DU PROGRÈS À L'EXPOSITION DE 1889 PAVILLON DES SECTIONS ÉTRANGÈRES À L'EXPOSITION UNIVERSELLE DE 1900 SUR LA RIVE GAUCHE PAVILLON DES SECTIONS ÉTRANGÈRES LORS DE L'EXPOSITION UNIVERSELLE DE 1900 TRAFIC DE PÉNICHES PRÈS DU PONT MARIE BAS QUAI ET BAIN GROUPE SCULPTÉ : LA SEINE ET LA MARNE BAIN VIGIER SUR UN QUAI DE LA SEINE BAINS VIGIER SUR UN QUAI DE LA SEINE CANAL DE L'OURCQ FONTAINE À UN ANGLE DE RUE MISÉRICORDE DE STALLE FIGURANT UN BATELIER	29

TAB. V.1.25 – Réponse pour le schéma de pertinence S_2

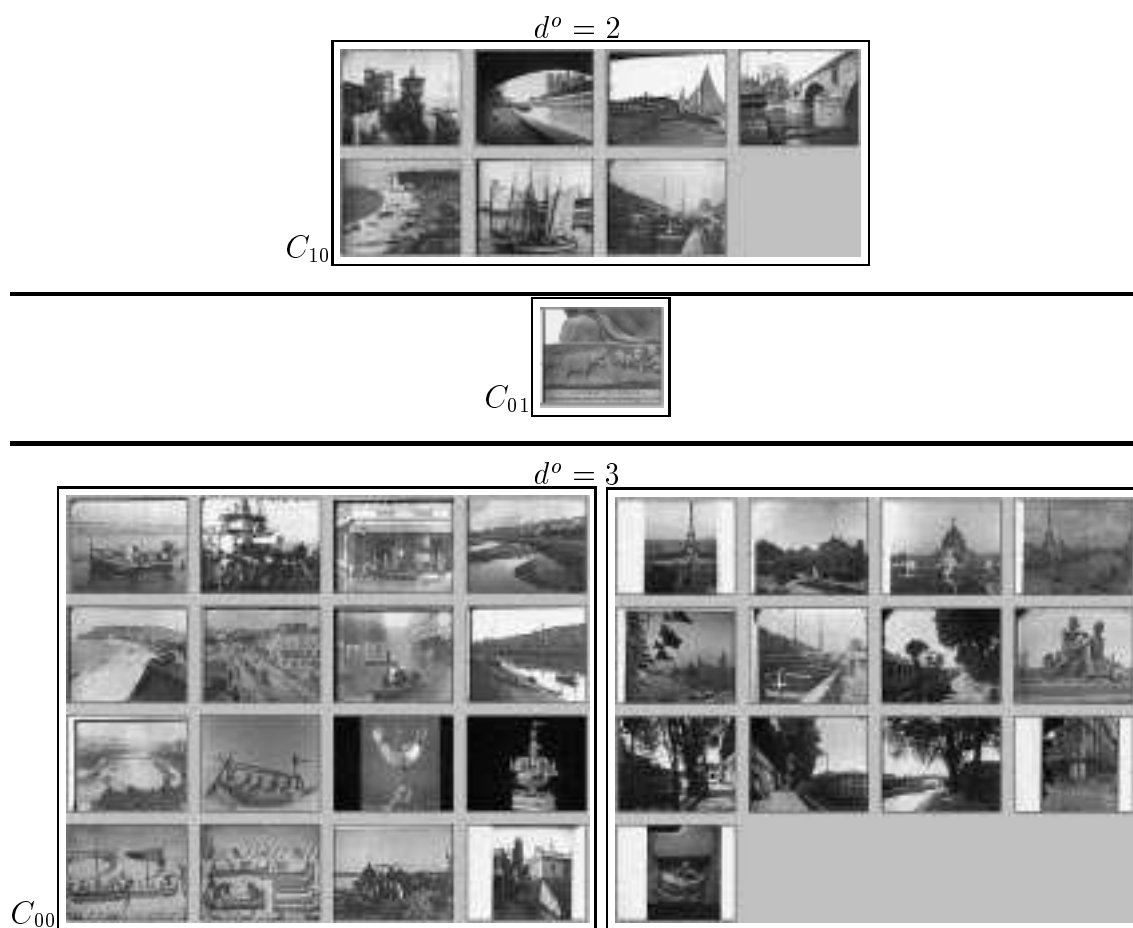


FIG. V.1.21 – Résultat du schéma de pertinence S_2 (voir agrandissement pages xvii et xviii de l'annexe)

ne sont pas pertinentes non plus, car le contexte urbain ne coïncide pas avec la connotation apaisante que l'utilisateur cherche à produire.

Ceci montre que le critère de **Connotation** n'exprime pas ce que l'utilisateur signifie.

Classe C_{01} : L'unique document de cette classe (ou seul le critère de **Morphologie** est vérifié) ne convient pas, car l'utilisateur ne veut pas d'une photo avec un bateau représenté, mais un bateau réel.

Cela montre que le critère de **Morphologie** peut prendre un sens contraire au sens attendu par l'utilisateur : il s'agit ici d'un gros plan sur un détail d'une sculpture, et non le fait que le bateau apparaisse en gros plan.

Classe C_{00} : Parmi les 16 documents de cette classe effectivement visualisés, où aucun des critères de dimension optionnels n'est vérifié, on retrouve des photos non pertinentes du même type que celle de la classe C_{01} (représentation de bateau), ainsi que de la classe C_{10} (paysage maritime, contexte urbain pour un fleuve). Les bateaux apparaissant dans un contexte maritime ne sont pas pertinents, car l'utilisateur ne veut pas de paysage maritime. Par ailleurs, les autres images (contexte fluvial) ne sont pas pertinentes non plus, car le contexte urbain ne coïncide pas avec l'idée que l'utilisateur se fait de la connotation apaisante.

Parmi ces images, cependant, certaines présentent un bateau visible et en gros plan, qui satisferait l'utilisateur si le contexte était approprié. Or ces documents, qui vérifient le critère de «**Morphologie**» dans le sens que l'utilisateur cherche à exprimer, devraient apparaître dans la classe C_{01} , si l'expression de ce critère était conforme à l'interprétation qu'en fait le système. Comme ce n'est pas le cas, on peut déduire que le critère de **Morphologie** traduit mal le sens que l'utilisateur cherche à exprimer.

L'image qui se distingue de celles-ci par son contexte à la fois non maritime et non construit, n'est pas pertinente car le bateau est trop peu visible.

De plus, l'utilisateur découvre à cette occasion qu'il préférerait une photo prise en France ou dans un pays occidental. Cette précision lui apparaît nécessaire seulement depuis qu'il a pris conscience de l'étendue couverte par le corpus en matière de lieu de prise des photographies.

La dernière image évoquée est pertinente dans une certaine mesure, car elle semble bien vérifier certains aspects du contexte dans lequel le bateau doit apparaître (non maritime, non construit), bien qu'elle ne soit pas pertinente du tout en matière de lieu de prise de la photo, ni en matière de visibilité du bateau.

Synthèse de l'évaluation : Globalement nous nous trouvons dans une situation de non-pertinence, c'est-à-dire qu'aucun des documents retrouvés n'est pertinent.

Les nouveautés en matière de formulation à l'issue de cette évaluation sont les suivantes.

- «non maritime» (ou bien fleuve ou lac ou canal);
- «non construit» ou «sauvage» (qui précise le sens de «apaisant»);
- «localisé en France ou dans un pays occidental» ou «non exotique».

Les modifications à apporter au schéma de pertinence sont les suivantes.

- «bateau visible» n'est pas correctement exprimé par **Morphologie(gros plan): Substituer**;
- «non maritime» (ou bien fleuve ou lac): **Détailler**;
- «non construit» ou «sauvage» (qui adapte le sens de «apaisant»): **Détailler**;
- «localisé en France ou dans un pays occidental» ou «non exotique»: **Détailler**.

Le terme **bateau** semble avoir une sémantique appropriée. Par conséquent on n'a pas de raison dans la boucle principale de modifier fondamentalement ce critère ou de le passer en optionnel. Par ailleurs, l'ensemble des documents de la classe C_{00} (qui n'a pas encore été entièrement visualisée) a des chances de contenir des documents pertinents, si le corpus en comprend. Il faut donc travailler sur la pertinence et la non-pertinence partielle des documents de cette classe, selon les critères qui ne sont pas encore bien exprimés.

Nous allons essayer de séparer les documents pertinents des non pertinents vis-à-vis de la connotation, dans la classe C_{00} (S'). Nous choisissons de simuler une action de bouclage de pertinence sur la dimension de **Contenu**, car c'est en termes de contenu que se formule le plus simplement la différence entre les documents pertinents des non pertinents : la présence d'immeubles, de rues, de quais, de voitures, de ponts, d'une cathédrale, reflète un contexte urbain non désiré. Les termes associés à ces éléments indésirables peuvent être sélectionnés manuellement à partir de l'ensemble des termes apparaissant dans les documents non pertinents et n'apparaissant pas dans les documents pertinents.

1.4.4 Étape 3

On essaye de séparer les documents pertinents des non pertinents dans la classe C_{00} (S'), par une action **Détailler**, qui ajoute les critères indésirables comme op-

tionnels.

S_3	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	immeuble	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	rue	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	quai	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	véhicule hippomobile	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	pont	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	cathédrale	<i>Égal</i>
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	apaisant	
	<i>optionnel</i>	Morphologie	<i>obligatoire</i>	gros plan	

C'est la sous-classe $c_{(00,000000)}$ qui est susceptible de contenir des documents pertinents. Elle comprend 14 documents. Les 29-14=15 documents qui se trouvent dans les autres sous-classes sont effectivement non pertinents.

Parmi les 14 documents retrouvés dans la sous-classe $c_{(00,000000)}$, qui sont cette fois tous examinés (14 est inférieur au seuil 16), aucun n'est pertinent. À nouveau ces documents présentent des caractéristiques très variées, mais comme il y en a peu, ce n'est pas gênant pour évaluer cet ensemble.

On peut conclure à l'absence de documents pertinents dans le corpus, sauf si l'on envisage la possibilité que le critère relatif à **bateau** exprime mal le besoin, contrairement à l'impression première.

Nous trouvant dans une situation de non-satisfaction, l'objectif est de s'assurer qu'effectivement aucun document du corpus ne convient. On procède à une action de type **RelâcherS** sur ce critère de **Contenu** (S_4). On consulte les documents de la classe C_{10} pour sélectionner un ensemble de termes pouvant être tenus pour critères de non-pertinence vis-à-vis du **Contenu**.

1.4.5 Étape 4

S_4	<i>obligatoire</i>	Contenu	<i>obligatoire</i>	bateau	<i>Vois</i>
	<i>optionnel</i>		<i>optionnel</i>	immeuble	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	rue	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	quai	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	véhicule hippomobile	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	pont	<i>Égal</i>
	<i>optionnel</i>		<i>optionnel</i>	cathédrale	<i>Égal</i>
	<i>optionnel</i>	Connotation	<i>obligatoire</i>	apaisant	
	<i>optionnel</i>	Morphologie	<i>obligatoire</i>	gros plan	

La réponse est beaucoup plus grande. En effet, dans l'arbre des relations entre termes, **bateau** a pour père le terme **transport**, qui a pour fils tous les types de

transport possibles (urbains, marins, aériens, animal, etc.). La classe C_{10} comprend 10 documents et la classe C_{00} en comprend 110. La sous-classe de C_{00} qui nous intéresse, $c_{(00,000000)}$, en comprend 45.

Si l'on regarde les nouveaux documents de C_{10} (figure V.1.22), on constate qu'effectivement d'autres types de véhicules que les bateaux y apparaissent. Aucun n'est donc pertinent.



FIG. V.1.22 – S_4 : classe C_{10} toutes sous-classes confondues

Il en est de même pour les 16 premiers documents de la sous-classe $c_{(00,000000)}$ (figure V.1.23), où apparaissent des automobiles, des montgolfières, des voitures à cheval, des chevaux montés, etc. La visualisation de la relation de pertinence entre ces documents et le schéma de pertinence confirme cette explication.

Ainsi, nous parvenons à la conclusion que le corpus ne comprend pas de document pertinent, pour ce problème d'information et dans cette situation.

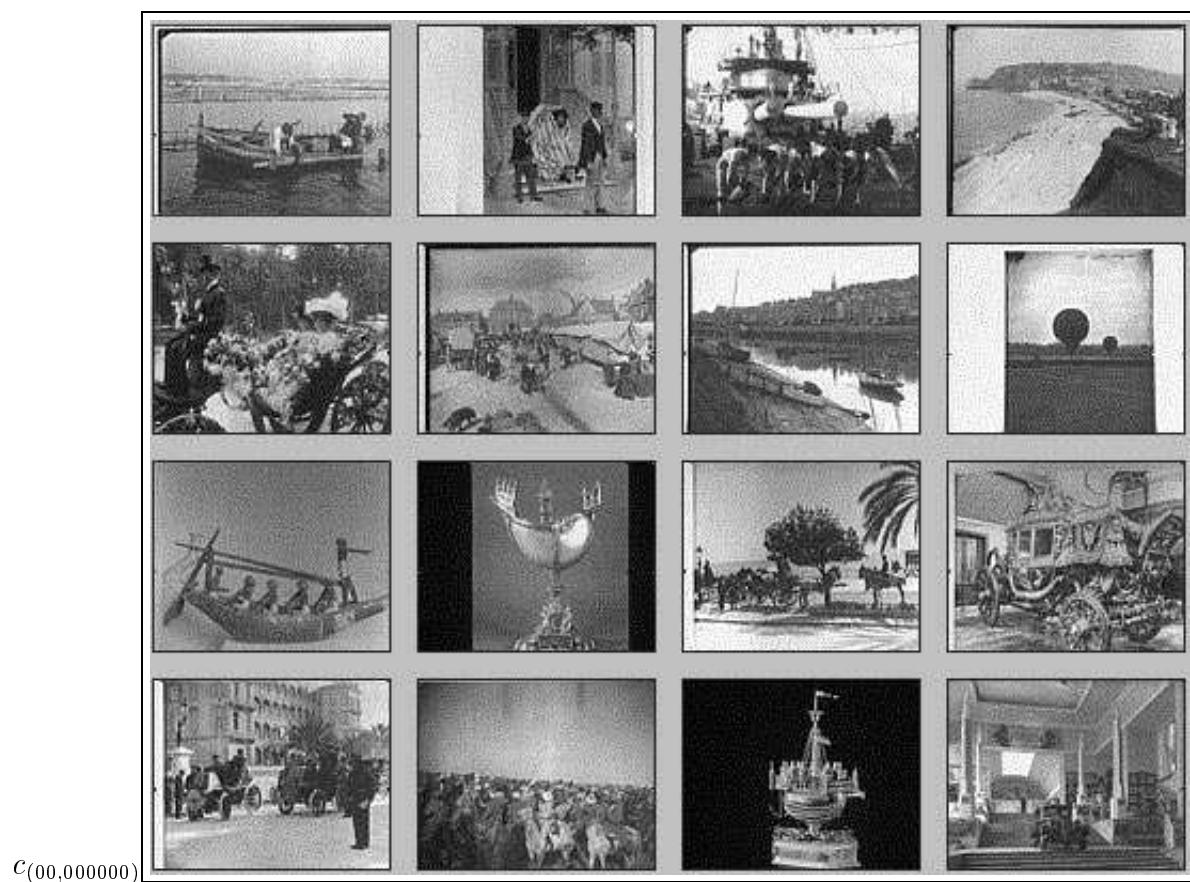


FIG. V.1.23 – S_4 : 16 premiers documents de la sous-classe $c_{(00,000000)}$

Chapitre 2

Perspectives de contrôle de l'interaction

Nous avons développé jusqu'ici l'éventail des possibilités d'adaptation de la pertinence système en relation avec la situation de recherche. La combinatoire qu'engendre cet éventail de variations fait du contrôle du déroulement de l'interaction un point critique de l'utilisabilité du système.

Nous exposons dans ce chapitre trois façons de contrôler l'interaction. Nous donnons d'abord une description d'un mode manuel où seule la formulation est contrôlée par le système. Puis nous décrivons, à l'autre extrême, un mode automatique où la reformulation est réalisée automatiquement.

Pour chacun de ces modes d'interaction, nous ne présentons pas l'intégralité des caractéristiques du système, mais simplement

- la façon dont le système intervient dans la boucle d'interaction décrite dans le chapitre 4 (figure V.2.1), ainsi que
- la nature des informations sur lesquelles ces interventions du système s'appuient.

2.1 Mode manuel avec formulation contrôlée

Nous définissons le mode manuel comme un mode où les seules connaissances que le système fait intervenir pour contrôler l'interaction sont des connaissances statiques, relatives aux bonnes propriétés de formulation du schéma de pertinence définies précédemment.

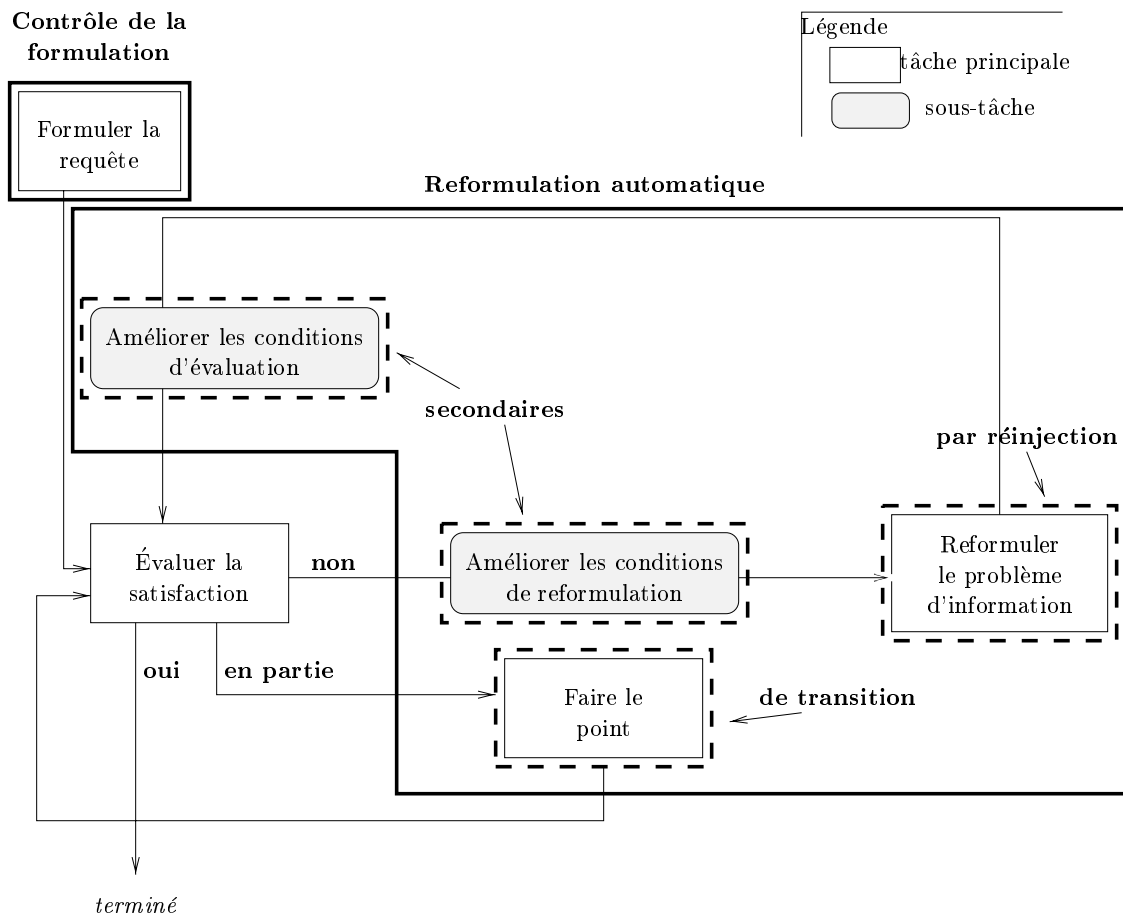


FIG. V.2.1 – Localisation du contrôle dans la boucle d'interaction

2.1.1 Correction du schéma de pertinence

Le premier type de contrôle réalise la correction du schéma de pertinence. Il consiste à garantir que le schéma de pertinence formulé possède un sens au regard des propriétés générales de l'indexation.

Nous avons identifié deux dimensions qui se prêtent à ce type de contrôle : **Auteur** et **Numéro**. Ces deux dimensions n'ont de sens que si elles présentent l'une des deux formes suivantes :

Forme 1

$\langle statut \rangle$	$\langle Dim \rangle$	<i>obligatoire</i>	<i>t</i>
--------------------------	-----------------------	--------------------	----------

Forme 2

$\langle statut \rangle$	$\langle Dim \rangle$	<i>optionnel</i>	<i>t₁</i>
		<i>optionnel</i>	<i>t₂</i>
		$\vdots$	$\vdots$
		<i>optionnel</i>	<i>t_n</i>

Par conséquent les statuts associés aux termes ne sont pas des paramètres réglables par l'utilisateur. Il suffit donc de les faire disparaître de l'interface pour empêcher une formulation incorrecte. Une interprétation implicite est associée à ces critères de dimension, selon la règle suivante : si un seul critère élémentaire est présent, son statut est *obligatoire* ; s'il y en a plus d'un, leur statut est *optionnel* .

La question suivante se pose : la suppression de la mention explicite des statuts associés aux critères élémentaires de cette dimension nuit-elle à l'intelligibilité du système pour l'utilisateur ? La réponse à cette question dépend de l'ensemble de l'interface, et notamment de la façon dont le lien entre le schéma de pertinence et les classes de pertinence y est rendu sensible.

Si ce lien fait jouer un rôle explicite aux statuts (comme c'est le cas dans l'interface rudimentaire notre prototype), il est probable que la disparition des statuts pour les dimensions **Auteur** et **Numéro** nuise à l'interaction. Pour éviter cela, les statuts sont laissés visibles, mais leur valeur est fixée automatiquement par le système en fonction du nombre de critères élémentaires présents.

2.1.2 Types de recherche

Le deuxième type de contrôle s'apparente aux stratégies fondées sur une typologie des problèmes d'information. Ce type de contrôle permet de réduire la surcharge cognitive pour l'utilisateur.

Nous avons donné plus haut trois exemples de types de recherche, auxquels nous avons associé des profils de schéma de pertinence. Ces profils traduisent la nature des

critères de dimension utilisés, et les priorités entre ces critères (recherche thématique, connotative, exploratoire, etc.).

Profil thématique	<i>obligatoire</i>	Contenu	
	<i>optionnel</i>	⟨Dim⟩	
	⋮		
	<i>optionnel</i>	⟨Dim⟩	

Profil connotatif	<i>obligatoire</i>	Connotation	
	<i>optionnel</i>	⟨Dim⟩	
	⋮		
	<i>optionnel</i>	⟨Dim⟩	

Profil exploratoire	<i>optionnel</i>	⟨Dim⟩	
----------------------------	------------------	-------	--

Pour traduire ce contrôle dans l'interface, une première solution consiste à présenter en préambule à la tâche de formulation, un ensemble de profils, en donnant une description qualitative du type de recherche (en termes de situation de recherche, par exemple, ou avec un exemple). L'utilisateur peut choisir l'un de ces profils et y inscrire la suite de sa formulation. Cette solution comporte le risque de mettre d'emblée l'utilisateur devant un choix difficile. Il faut donc ménager un ou deux profils très neutres : le profil de recherche par le contenu (en tant que profil par défaut), et le profil exploratoire (approprié pour un utilisateur qui n'a qu'une idée vague de ce qu'il cherche).

Une seconde solution repose sur une interface en entrée évolutive : le système fait apparaître progressivement les diverses dimensions, en commençant par la dimension **Contenu**, qui est la plus fréquemment employée pour exprimer le critère de dimension principal. Cette solution présente l'avantage de simplifier les premières étapes de l'interaction. Elle est donc appropriée pour une approche «tutorée» du système. En revanche, elle fait courir le risque d'une utilisation étriquée du système. Il faut donc veiller à ce que le système propose régulièrement d'introduire de nouvelles dimensions d'interrogation.

2.1.3 Heuristiques d'utilisation

Le troisième type de contrôle relève des caractéristiques du schéma de pertinence particulier qui est formulé, au niveau des critères élémentaires. La dimension **Connotation** en donne un exemple caractéristique dans le prototype VIZ.

Il est fréquent qu'un critère de connotation réduit à un terme, ou comprenant plusieurs termes avec un statut obligatoire, compromette le rappel.

L'utilisation de la dimension **Connotation** peut être contrainte de la même façon que les dimensions **Auteur** et **Numéro**, en imposant le profil suivant sitôt que plus d'un critère élémentaire est présent, mais cette contrainte est trop forte, du fait qu'elle interdit une recherche de type connotatif détaillée. Il est préférable d'introduire la notion de profil recommandé (ou profil par défaut), tout en ménageant la possibilité de modifier les statuts.

Profil recommandé pour la connotation	$\langle statut \rangle$	Connotation	<i>optionnel</i>	t_1
			<i>optionnel</i>	t_2
			$\vdots$	$\vdots$
			<i>optionnel</i>	t_n

Par ailleurs, l'introduction d'un processus d'expansion de requête interactif permet de résoudre le problème de rappel que peut poser un unique critère élémentaire. Le déclenchement de l'expansion doit être inhibé dans le cas d'une recherche connotative, ou bien contrôlée par l'utilisateur.

Une autre solution consiste à intégrer le contrôle de la formulation de cette dimension dans le cadre de la reformulation automatique, en ajoutant les critères de connotation rencontrés dans les documents jugés pertinents.

2.2 Mode automatique avec reformulation automatique

À l'opposé du mode manuel, où toutes les reformulations incombent à l'utilisateur à l'exception de celles liées aux propriétés intrinsèques du système, nous définissons le mode automatique comme un mode où toutes les reformulations sont traitées de manière automatique.

Pour satisfaire au principe de compréhension, la reformulation doit se traduire de manière visible sur le schéma de pertinence, et doit être validée de manière interactive.

2.2.1 Trois types de reformulation

Nous définissons trois types de reformulation, en fonction de leur rôle dans le déroulement de la recherche.

La *reformulation par réinjection* se situe au niveau de la boucle d'interaction principale. Elle utilise les jugements de pertinence recueillis sur les documents retrouvés pour proposer un ensemble de modifications sémantiques à apporter au schéma de pertinence (ajout de termes, modification de critères abstraits). Il n'est

pas nécessaire que la totalité de la réponse ait été examinée et évaluée par l'utilisateur pour déclencher ce type de reformulation.

La *reformulation secondaire* se situe au niveau des boucles d'interaction secondaires. Elle a pour but de résoudre les situations problématiques de type pénurie et non-pertinence afin de permettre de poursuivre la boucle principale. Elle s'appuie sur les caractéristiques de la réponse courante pour identifier une situation problématique de type pénurie ou non-pertinence, et proposer une modification pragmatique à apporter au schéma de pertinence (modification des statuts). Par conséquent, la reformulation secondaire n'utilise pas de jugements de pertinence.

La *reformulation de transition* se situe au niveau de la boucle principale, dans la situation particulière où l'utilisateur a réalisé une évaluation complète de la réponse, et veut faire le point avant de poursuivre la recherche (*situation de transition*). Elle utilise un nouveau type de jugement de pertinence, les jugements de pertinence sur les classes et sous-classes, pour proposer une modification pragmatique qui permet de focaliser la réponse sur les documents pertinents.

2.2.2 Principe de la réinjection

À partir des documents jugés pertinents, le système peut établir

1. une liste de termes candidats pour augmenter le schéma de pertinence ;
2. une liste de critères abstraits à modifier.

Pour établir une liste des termes candidats, le principe des techniques classiques de relevance feedback pour le modèle vectoriel doit être adapté pour tenir compte, d'une part de l'absence de l'hypothèse de complétude des jugements de pertinence recueillis, et d'autre part de l'absence d'une pondération des termes d'indexation et des termes de requête. Le système doit utiliser les documents jugés pertinents pour établir une liste des termes utilisés dans ces documents, en les rangeant par ordre de plus grande utilité pour la reformulation.

L'*utilité d'un terme pour la reformulation* est liée non seulement à la fréquence d'apparition de ce terme dans l'ensemble des documents jugés pertinents (hypothèse du relevance feedback classique), mais aussi à la rareté de ce terme dans l'ensemble des documents de la classe où il apparaît.

Le facteur d'utilité lié à la fréquence d'apparition permet de retrouver des documents semblables au plus grand nombre de ceux qui ont été déjà retrouvés. Le facteur d'utilité lié à la rareté dans la classe permet de retrouver des documents relativement dissemblables, et néanmoins potentiellement pertinents. Ces deux facteurs antagonistes traitent des situations différentes : le premier convient lorsque le rappel est déjà suffisamment bon, alors que le second est utile en début de session, lorsque la formulation du problème d'information est encore incomplète et que des classes

entières de documents pertinents n'ont pas encore été retrouvés. N'ayant que peu de moyen de savoir dans quel cas on se trouve à un instant donné, les deux facteurs doivent être tous deux pris en compte, en combinaison.

Pour établir une liste des critères abstraits à modifier, le système doit induire à partir des documents jugés pertinents, le critère abstrait qui permet de retrouver ces documents avec les termes déjà présents dans le schéma de pertinence. Par exemple, si les documents jugés pertinents sont indexés par des termes voisins d'un terme du schéma courant, le système doit proposer le passage d'un critère abstrait *Égal* à un critère abstrait *Vois*. Ce processus d'induction n'est pas toujours réalisable en fonction de la nature formelle du calcul impliqué par les critères abstraits disponibles. Dans le cas de VIZ, l'induction est réalisable pour un coût relativement limité. En revanche, pour un critère abstrait de nature statistique, ce processus peut ne pas être implémentable.

2.2.3 Trois modes de reformulation par réinjection

Notre modèle permet de définir trois modes de reformulation par réinjection : la réinjection classique, la réinjection limitée à une dimension, la réinjection au sein d'une classe. Les modifications associées relèvent des types d'action *Détailler* (ajout d'un terme), et *RelâcherS* (modification d'un critère abstrait).

La *réinjection classique* consiste à établir une liste des modifications candidates (ajouts de termes et modifications de critères abstraits) à partir de tous les documents jugés pertinents. La liste comprend des modifications qui relèvent à la fois des dimensions déjà utilisées dans le schéma de pertinence courant, et des dimensions encore non utilisées.

La *réinjection limitée à une dimension* consiste à établir une liste des modifications candidates à la réinjection pour une dimension donnée. Tous les documents jugés pertinents sont pris en compte, mais on fait intervenir le facteur d'utilité en tenant compte de la classe à laquelle chaque document appartient. Par exemple, les termes issus de documents pertinents qui appartiennent aux classes où le critère de dimension concerné est vérifié, auront plus de poids que les autres.

La *réinjection limitée à une classe* consiste à établir une liste des modifications candidates à la réinjection en ne considérant que les jugements de pertinence effectués dans une classe particulière. Cela permet à l'utilisateur de raisonner dans le cadre limité d'une classe de pertinence, ce qui peut être utile dans une situation de non-discrimination.

2.2.4 Reformulation secondaire

Dans une situation du type pénurie ou non-pertinence, on déclenche une reformulation dont le but est de supprimer l'entrave à l'évaluation ou à la reformulation

afin de débloquent la situation problématique. Les modifications engendrées par ce type de reformulation relèvent des types d'action **RelâcherP** et **ResserrerP**. Il ne requiert pas de jugement de pertinence pour être réalisé.

Comme nous l'avons montré dans le chapitre 4, d'une part, le système peut détecter les situations de pénurie et de non-pertinence, et d'autre part il existe des opérations globales de modification du schéma de pertinence associées aux actions **RelâcherP** et **ResserrerP**, ce qui permet d'automatiser complètement ce processus de reformulation. Cependant, l'initiative peut être laissée à l'utilisateur de limiter l'action à certaines dimensions seulement, en particulier dans le cas où l'utilisateur considère que l'une des dimensions est déjà correctement formulée.

2.2.5 Reformulation de transition

Un utilisateur peut marquer une classe ou une sous-classe de pertinence entière comme pertinente, ou comme non pertinente. C'est ce que nous appelons les *jugements de pertinence étendus*. Ils permettent d'effectuer un autre type de reformulation concernant la structure du schéma de pertinence.

En effet, une classe ou une sous-classe est associée à un schéma homologue qui décrit une combinaison booléenne des critères de dimension ou des critères élémentaires. Étant donné un ensemble de jugements de pertinence sur les classes ou les sous-classes, le système peut calculer un ensemble de contraintes sur les critères de dimension ou les critères élémentaires.

Par exemple, si toutes les classes où un critère donné n'est pas vérifié, sont jugées pertinentes, le système propose de supprimer ce critère. Si toutes les classes où un critère donné est vérifié sont jugées pertinentes, et toutes les classes où ce même critère n'est pas vérifié sont jugées non pertinentes, le système propose de rendre ce critère obligatoire.

Les cas tranchés que nous venons d'évoquer peuvent être rares. Une première solution pour rendre cette reformulation plus souvent applicable, consiste à définir un calcul des contraintes qui intègre la gestion des cas incertains. Une autre solution consiste à réclamer de l'utilisateur le marquage des documents qui sont certainement pertinents, afin de limiter les modifications du schéma à celles qui n'éliminent pas ces documents marqués.

2.2.6 Boucle de validation de la reformulation

Nous proposons d'instaurer deux niveaux de validation des reformulations, comme décrit dans la figure V.2.2.

D'abord l'utilisateur valide a priori une reformulation qui lui est proposée, et il confirme cette validation le cas échéant, au vu de la nouvelle réponse obtenue. Cela permet de préserver une boucle imbriquée dans le schéma d'interaction, dont le rôle

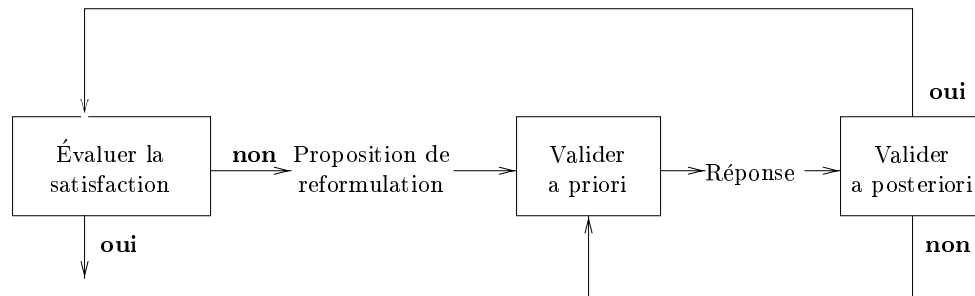


FIG. V.2.2 – *Boucle de reformulation automatique avec boucle de validation imbriquée*

est de permettre à l'utilisateur de faire des essais afin d'affiner sa compréhension du système. Cette boucle doit aussi permettre d'éviter des erreurs grossières de la reformulation automatique.

Chapitre 3

Champ d'application

Notre modèle de pertinence système a pour but de permettre à un utilisateur d'interagir avec le système pour exprimer son problème d'information de façon plus complète et plus précise que dans le cadre de systèmes où la sémantique de la pertinence système est implicite. Les systèmes de recherche d'information où la pertinence système se présente à l'utilisateur comme une boîte noire, tant pour l'évaluation de la requête que pour la prise en compte des jugements de pertinence, atteignent un certain niveau de performance auquel nous voyons une limite théorique : les hypothèses qui sous-tendent la pertinence système figent le cadre d'évaluation des requêtes, et l'absence d'un dispositif interactif permettant à l'utilisateur de comprendre ce cadre et de s'y adapter, limite la qualité des réponses quand la situation de l'utilisateur le positionne hors de ce cadre. Ce cadre d'évaluation est celui de la correspondance thématique.

Nous présentons ici les types de contexte d'application dans lesquels notre modèle de pertinence système est susceptible d'améliorer les performances en principe. Nous discutons en particulier des possibilités d'intégrer aux techniques existantes les principes de notre modèle et du profit qui peut en être tiré. Nous insistons sur la possibilité d'extraire automatiquement des caractéristiques d'indexation qui permettent de définir divers critères abstraits.

3.1 Classification des documents retrouvés

Dans le cadre d'un corpus de documents techniques, les termes de l'interrogation sont généralement peu ambigus en raison de la spécificité des termes employés, et de l'homogénéité thématique du corpus qui se cantonne à un domaine particulier. C'est ce qui fait, entre autres, le succès du modèle vectoriel, où l'hypothèse d'une base orthogonale de termes d'indexation est relativement réaliste, et où l'ambiguïté des termes pose rarement problème dans l'interrogation.

Cependant, les termes d'indexation extraits des documents peuvent jouer des rôles variés dans le contenu sémantique du document. Par exemple, un thème peut être le sujet principal d'un article, ou bien y figurer simplement à titre d'exemple, ou encore être rapidement évoqué pour donner une référence à un autre article dont c'est le sujet principal.

Prenons l'exemple d'un corpus d'articles scientifiques relatifs au domaine de la médecine, et considérons un utilisateur qui recherche «une étude comparative des symptômes de l'hypothyroïdie», avec la requête suivante :

étude comparative	p_1
symptôme	p_2
hypothyroïdie	p_3

Supposons que le corpus ne contienne aucun document dont le sujet principal constitue une étude comparative des symptômes de l'hypothyroïdie, mais comprenne des études comparatives d'autres symptômes que ceux de l'hypothyroïdie (ensemble de documents noté $\mathcal{E}_{D_1}$ indexés par **étude comparative**, **symptôme**, **X**), et aussi des articles traitant de l'hypothyroïdie en général (ensemble des documents $\mathcal{E}_{D_2}$ indexés par **hypothyroïdie**), certains d'entre eux (ensemble $\mathcal{E}_{D_3} \subseteq \mathcal{E}_{D_2}$) citant d'autres travaux où une étude comparative des symptômes de l'hypothyroïdie est réalisée, et cela par l'intermédiaire d'une référence à un autre article non présent dans le corpus. Dans ces derniers documents, les poids de **étude comparative** et **symptôme** sont faibles en raison de l'apparition rare de ces termes dans l'article (tf), et leur fréquence relativement haute dans l'ensemble du corpus (idf).

Si tous les termes de la requête sont assortis du même poids, le modèle vectoriel classera en tête les documents traitant d'études comparatives et de symptômes ($\mathcal{E}_{D_1}$), puis viendront les documents traitant de l'hypothyroïdie ($\mathcal{E}_{D_2}$) avec en tête ceux qui font allusion à une étude comparative de symptômes ($\mathcal{E}_{D_3}$). Si l'ensemble $\mathcal{E}_{D_1}$ est grand, l'ensemble des documents pertinents ($\mathcal{E}_{D_3}$) sera difficile à repérer dans un classement linéaire.

Quels poids p_1, p_2, p_3 associer aux termes de la requête pour placer les documents de $\mathcal{E}_{D_3}$ en tête de classement ? Si l'on réduit le poids des termes **étude comparative** et **symptôme**, les premiers documents retrouvés traiteront de l'hypothyroïdie avant tout, en tant que sujet principal, mais à nouveau, si l'ensemble $\mathcal{E}_{D_2}$ comprend un grand nombre de documents, les documents pertinents où apparaissent les citations d'articles relatifs à l'hypothyroïdie seront probablement noyés dans la masse des documents en raison du faible poids des termes **étude comparative** et **symptôme**, tant dans la requête que dans le document.

Le bouclage de pertinence automatique a pour but d'ajuster les poids des termes de la requête ainsi que l'ensemble des termes lui-même, par l'intermédiaire de jugements de pertinence. Mais dans la situation où les documents pertinents ($\mathcal{E}_{D_3}$)

ne sont pas détectés en raison de leur faible classement, le bouclage de pertinence ne peut pas être mis à profit (entrave à la reformulation). Par ailleurs, il est difficile d'évaluer à partir de ce résultat, la possibilité pour qu'il existe néanmoins des documents pertinents dans le corpus (entrave à l'évaluation).

Ainsi les situations problématiques que nous avons définies peuvent se présenter dans le cadre du modèle vectoriel. L'intégration d'une organisation des documents retrouvés selon une classification qui ne tient pas compte des poids, comme nous le proposons dans notre modèle, peut apporter une solution intéressante.

Pour cela, nous simplifions notre modèle en réduisant le schéma de pertinence à un seul niveau hiérarchique en raison de l'absence de dimensions d'interrogation identifiables. Voici le schéma de pertinence obtenu.

<i>obligatoire</i>	étude comparative
<i>obligatoire</i>	symptôme
<i>obligatoire</i>	hypothyroïdie

Les documents de $\mathcal{E}_{D_3}$ apparaîtront dans la première classe et les autres documents dans les suivantes.

Dans le cas d'un corpus ne comprenant aucun document pertinent ($\mathcal{E}_{D_3} = \emptyset$), les classes de pertinence qui peuvent être produites en relâchant les statuts de **étude comparative** et de **symptôme** permettent de faire apparaître les documents du corpus où seulement une partie des critères est vérifiée, ce qui permet à l'utilisateur de se convaincre que le corpus ne comprend pas de document pertinent.

<i>optionnel</i>	étude comparative
<i>optionnel</i>	symptôme
<i>obligatoire</i>	hypothyroïdie

Ainsi, même dans le cadre d'une indexation vectorielle, nous pouvons gagner en facilité de détection de la pertinence : les documents pertinents sont plus facilement repérés car visibles en tête d'une classe. Cette détection de la pertinence facilitée permet de résoudre dans certains cas une situation d'entrave à la reformulation. De plus, nous pouvons gagner en qualité d'évaluation de la satisfaction, car la classification du résultat fournit des indices pour évaluer la correction de la formulation de la requête, et des informations quant à la présence de documents répondant à la requête.

Le modèle vectoriel peut néanmoins être appliqué pour ordonner les documents au sein d'une classe. Cela pose des problèmes de complexité de calcul, du fait que la mesure de similarité doit être normalisée non plus vis-à-vis de l'ensemble du corpus, mais vis-à-vis de la classe. Si l'on procède naïvement, les coefficients de normalisation doivent être recalculés pour chaque classe, à chaque évaluation. Il serait intéressant

d'étudier la façon dont ces calculs peuvent être optimisés tout en conservant l'intérêt du modèle vectoriel et de son indexation automatique sur les textes.

3.2 Clarification du sens des termes employés dans la requête

Dans le cas de corpus dont la thématique est ciblée autour d'un domaine peu spécialisé, ou encore dans le cas de corpus dont le contenu thématique est très hétérogène, notre approche peut contribuer à améliorer l'utilisation des systèmes par l'intermédiaire du dispositif de désambiguisation qui peut être développé au moyen des classes de pertinence.

En effet, dans ce type d'application, un même terme peut avoir plusieurs sens en fonction du contexte, et il est important de permettre à l'utilisateur de clarifier le sens que prend un terme de sa requête pour le système. Les sous-classes de degré $n - 1$ et de degré 2 permettent cette clarification, comme nous l'avons décrit en détail dans le cadre du prototype VIZ.

3.3 Ensemble de critères abstraits

Notre modèle permet d'introduire des nuances dans le rôle que joue un thème pour contribuer à la pertinence d'un document, par l'intermédiaire des critères abstraits. La définition d'un ensemble de critères abstraits requiert un certain raffinement de l'indexation, qui doit véhiculer les informations suffisantes pour évaluer ces nuances. Il est intéressant d'examiner la possibilité d'extraire automatiquement des textes de telles informations.

C'est ce que propose Paradis [Par96] dans son modèle d'indexation de documents textuels. Il montre que l'extraction des thèmes d'un document peut être raffinée si l'on utilise la structure du discours. Cette structure est ou bien disponible directement sous la forme d'un marquage du texte, ou bien dérivée en reconnaissant certaines structures syntaxiques qui caractérisent les contextes d'apparition de thèmes jouant un rôle particulier dans le discours.

Par exemple, les structures «Cet article étudie X», «Dans cette partie, nous nous intéressons à X» caractérisent le contexte d'apparition d'un thème principal ou d'un thème secondaire d'un article scientifique. Cela produit l'élément d'indexation (X,ThèmePrincipal) ou (X,ThèmeSecondaire). De même, la structure «Pour des détails sur X, voir l'article [...]» caractérise le contexte d'apparition d'un thème pas nécessairement abordé, mais pour lequel une référence est donnée, ce qui produit l'élément d'indexation (X,ThèmeRéféréncé).

Le modèle d'indexation de Paradis consiste à accompagner les thèmes indexant un document, d'un élément décrivant le rôle qu'ils jouent dans le discours. Dans le cadre de notre modèle, ces caractéristiques d'indexation qui accompagnent les thèmes permettent de définir des critères abstraits distincts. Notamment, avec les exemples précédants nous pouvons définir un critère *ConstitueLeThèmePrincipal*, *ConstitueUnThèmeSecondaire*, *DonneUneRéférencePour*. Ces divers critères permettent de faire varier le sens de la requête de manière fine, et en particulier de réagir à une situation problématique de pénurie ou de non-pertinence.

Reprenons l'exemple que nous avons cité plus haut, concernant l'étude comparative des symptômes de l'hypothyroïdie.

<i>obligatoire</i>	<i>ConstitueLeThèmePrincipal</i> [étude comparative]
<i>obligatoire</i>	<i>ConstitueLeThèmePrincipal</i> [symptôme]
<i>obligatoire</i>	<i>ConstitueLeThèmePrincipal</i> [hypothyroïdie]

Devant une réponse où aucun document pertinent n'est retrouvé, certains critères abstraits peuvent être changés pour *DonneUneRéférencePour*, afin de retrouver les documents qui fournissent des références au sujet en question.

<i>obligatoire</i>	<i>DonneUneRéférencePour</i> [étude comparative]
<i>obligatoire</i>	<i>DonneUneRéférencePour</i> [symptôme]
<i>obligatoire</i>	<i>ConstitueLeThèmePrincipal</i> [hypothyroïdie]

Dans le cas de grand corpus, le nombre de caractéristiques d'indexation extraites des documents constitue un point critique pour la faisabilité de l'indexation automatique. Ainsi, le nombre de caractéristiques extraites au-delà des mots-clés eux-mêmes doit rester petit, et le nombre des critères abstraits qu'il est possible d'implémenter est limité en conséquence.

Cependant divers critères abstraits peuvent être définis à partir d'une indexation automatique mono-dimensionnelle. Nous développons cette idée dans la suite.

3.4 Dimensions d'interrogation

Dans le cas de grands corpus homogènes vis-à-vis de la thématique ou du type de document, des connaissances sur le domaine ou sur le type de document peuvent être utilisées pour définir, à partir d'une indexation statistique par mots-clés, diverses dimensions d'interrogation et divers critères abstraits.

Reprenons notre exemple de corpus d'articles relevant du domaine médical. Un réseau sémantique peut regrouper les termes ayant trait aux types d'études et aux méthodologies, un autre aux objets des études (noms de maladies, de symptômes,...),

etc. Ainsi, nous pouvons définir diverses dimensions d'interrogation au sein desquelles les termes de l'interrogation seront traités de façon spécifique à la dimension dont ils relèvent.

Par exemple, un document indexé par le terme *étude comparative*, le sera souvent avec un faible poids en raison de la fréquence des études comparatives dans le corpus, et du petit nombre d'apparitions de ce terme dans le document : typiquement, ce terme n'apparaît que dans le titre, au contraire du thème principal qui se retrouve de façon récurrente dans le document. Ainsi, les termes indiquant le type de sujet (étude comparative, étude analytique, etc.), seront intégrés différemment dans la fonction de correspondance, de telle façon que le poids n'intervienne pas. En revanche, la pertinence d'un document pour les termes décrivant l'objet de l'étude pourra être traité de la manière classique, ou celle proposée dans la section 3.1.

De plus, divers critères abstraits peuvent être définis grâce au réseau sémantique. En permettant d'appliquer explicitement des techniques d'expansion de requêtes, divers critères abstraits peuvent être définis pour une même dimension d'interrogation. Par exemple, l'hypothyroïdie a pour symptômes le crétinisme et le goitre, en particulier. Étant donné un réseau sémantique spécialisé dans lequel les maladies sont associées à leurs divers symptômes, la requête évoquée précédemment peut se traduire de la façon suivante :

<i>obl</i>	Intention	<i>obl</i>	<i>EstDuType</i>	étude comparative
<i>obl</i>	Contenu	<i>obl</i>	<i>ConstitueLeThèmePrincipal</i>	hypothyroïdie
		<i>opt</i>	<i>TraiteDesSymptômesDe</i>	hypothyroïdie

3.5 Bilan et perspectives d'application

3.5.1 Bilan

Globalement, il apparaît que le principe de classification des documents retrouvés de notre modèle peut profiter à des corpus pour lesquels seule une indexation statistique est réaliste. Nous avons montré par un exemple que cela permet une amélioration dans les cas où des situations problématiques sont rencontrées. De plus, si le domaine, le corpus et/ou le contexte d'utilisation sont assez bien connus, il est possible de développer des dimensions d'interrogation et un ensemble de critères abstraits qui permettent d'enrichir la sémantique de l'interaction.

En revanche, dans le cadre du World Wide Web par exemple, il est difficile de concevoir une interrogation structurée qui vaille pour tous les documents du réseau, et pour tous les types de besoins d'interrogation qu'un tel corpus peut susciter, étant donné le manque de structure de ce corpus qui évolue sans cesse.

3.5.2 Perspectives d'application sur le corpus du World Wide Web

Cependant, si nous nous projetons dans l'avenir, nous pouvons espérer que des outils pour les auteurs de documents du World Wide Web seront développés, pour permettre à l'information qui est disponible sur le réseau, d'être effectivement et efficacement accessible. Ces outils pourront définir des types de documents en fonction des destinataires souhaités de l'information qu'ils véhiculent au moyen d'un ensemble de marqueurs permettant d'y repérer des caractéristiques d'indexation intéressantes pour la recherche d'information (voir les recommandations de la TEI, *Text Encoding Initiative*). Les auteurs des documents se conformeront à ce schéma pour rédiger les documents qu'ils souhaitent rendre accessibles aux utilisateurs du réseau.

Cette perspective optimiste, si elle peut sembler lointaine, n'est pas pour autant irréaliste. En effet, l'auteur d'un document destiné à être lu par d'autres, a en général en tête un type de destinataire, ou un ensemble de types de destinataires. Partant de cela, il est possible de définir les besoins d'interrogation des destinataires potentiels du document. L'auteur a tout intérêt à faire l'effort de se conformer aux standards de rédaction relatifs à ce type de destinataire, s'il en existe, s'il veut que son document soit rendu accessible.

Dans cette optique, notre modèle de pertinence système peut être vu comme un moteur de recherche paramétré selon les types de recherche. Pour chaque type de recherche, un ensemble de dimensions et/ou de critères abstraits adaptés aux informations de marquage disponibles, peuvent être associés pour tirer le meilleur du sous-corpus constitué des documents destinés au même type de destinataire.

3.5.3 Perspectives d'application à des corpus de documents non textuels

Nous avons essentiellement traité des documents textuels, pour lesquels l'auteur est identifiable. En revanche, pour des données non textuelles, le marquage des documents en cours de «rédaction» n'est plus possible. L'effort requis pour permettre une indexation automatique sémantique des images ou des documents audio est donc un investissement qui ne peut être demandé à l'auteur.

Cependant pour un corpus constitué de documents homogènes (une collection de tableau d'un musée, ou bien l'ensemble des disques CompactDisc de musique actuellement édités), on peut envisager une indexation progressive des documents, comme le suggère la démarche adoptée par le site «All Music Guide» du World Wide Web. Ce site fournit un accès à une base de données de 230,000 disques de musique disponibles sur le marché. L'interface organise des documents hypertextes qui décrivent les disques, les musiciens qui interviennent dans ces disques, et les

mouvements musicaux dont relèvent ces disques. La recherche peut s'effectuer sur un musicien, un titre de disque, ou un titre de chanson apparaissant dans le disque, ce qui constitue une recherche du type base de données.

Cependant, chaque fois qu'une page relative à un disque est consultée, un formulaire est présenté, où les utilisateurs qui connaissent bien le disque en question sont invités à porter un jugement de nature qualitative sur le disque. Ainsi, avec le temps, se constitue une indexation qui a trait à des caractéristiques qualitatives sur les documents, qui ne peuvent être extraits automatiquement et qui constitue un travail trop important et difficile pour un groupe d'indexeurs réduit.

Mettre à contribution les utilisateurs de l'information pour valoriser cette information nous semble être une voie intéressante pour gérer l'accès à une quantité d'information grandissante.

Chapitre 4

Bilan

4.1 Bilan

Du point de vue théorique, notre modèle fournit un outil qui permet d'étudier la dépendance de la pertinence envers la situation. Cela constitue une approche orthogonale aux études qui cherchent à capturer les régularités de la pertinence (propriétés valables quelque soit l'utilisateur et la situation de recherche). Il s'agit ici au contraire de gérer ses variations en fonction de la situation, de façon à obtenir un système robuste dans les conditions réelles d'utilisation.

La solution que nous proposons pour atteindre cette robustesse, consiste à mettre en œuvre les trois principes «Comprendre», «Détecter» et «Reformuler». Nous faisons l'hypothèse que les performances des systèmes peuvent être améliorées si l'utilisateur peut résoudre les problèmes qu'il rencontre dans l'interaction avec le système :

- «Comprendre» : problème de l'accord sur le sens du langage de requête (par exemple, «le terme que j'emploie a-t-il bien le même sens pour le système»),
- «Détecter» : problème de la facilité à détecter la pertinence (par exemple, «ce document est pertinent seulement si le corpus ne comprend pas d'autres documents plus pertinents» ou encore «je ne peux évaluer la pertinence d'un si grand nombre de documents»),
- «Reformuler» : problème de la mise en évidence d'éléments pour améliorer la formulation du problème d'information («comment modifier la formulation de mon problème d'information alors que la réponse ne me fournit aucun indice pour le faire»).

Nous avons montré que ces divers *problèmes* s'inscrivent dans des *situations problématiques*. En structurant le langage de requête (critères abstraits, statuts) et la réponse du système (classes de pertinence), la sémantique de l'interaction entre

le système et l'utilisateur est enrichie. Notamment, la notion de situation fait partie du vocabulaire commun au système et à l'utilisateur.

Ainsi, le système peut identifier au moins partiellement la situation problématique courante (par exemple une situation de non-pertinence), à laquelle il associe un ensemble de problèmes possibles ; dans une situation de non-pertinence, le problème à résoudre peut être clarifier la sémantique d'un terme «Comprendre», ou obtenir un point de vue en perspective sur le corpus pour confirmer l'absence de documents plus pertinents que ceux déjà retrouvés «Détecter», voire trouver dans les nouveaux documents des éléments de reformulation «Reformuler». À chaque problème le système associe une reformulation (ou un profil de reformulation que l'utilisateur doit compléter) permettant de le résoudre.

De son côté, l'utilisateur doit valider l'identification de la situation, et préciser le problème particulier qu'il veut résoudre. Il obtient alors une suggestion de reformulation qui lui permet de sortir de sa situation problématique, que ce soit au niveau de la boucle principale ou des boucles secondaires.

Ainsi le système établit une interaction qui permet de réaliser une reformulation adaptée à la situation.

De plus, la structure du schéma de pertinence, et le lien entre cette structure et la réponse, permettent de reformuler le problème d'information de manière plus fine, notamment en attaquant les divers aspects du problème d'information séparément, dimension par dimension.

4.2 Intégration des techniques existantes

Notre modèle permet d'intégrer plusieurs techniques d'adaptation du système dans le même contexte. Il permet d'utiliser chaque technique dans le champ d'application où elle est le plus approprié.

La notion de *stratégie de formulation* (section 1.2) reprend la notion de stratégie par typologie de façon souple. Elle fournit un guide qui allège la charge cognitive de l'utilisateur lors de la première étape de formulation du problème d'information. Elle n'impose pas pour autant une contrainte formelle sur l'ensemble de la session, puisque les profils de recherche sont des instances particulières du schéma de pertinence général.

Plus généralement, la notion de *situation* (section 1.3) constitue un point de départ pour définir et étudier expérimentalement la notion de *stratégie d'utilisation du système pour l'utilisateur*.

La notion de *schéma homologue* associé aux classes de pertinence constitue une formulation intuitive de requêtes booléennes. La sémantique des opérateurs booléens qui pose problème au niveau de la formulation, est ici exploitée comme un outil pratique de visualisation. Il constitue une définition opérationnelle des notions de

rappel, précision, focalisation, qui donnent à l'utilisateur les moyens de manipuler la réponse.

Notre modèle fournit un contexte de définition de paradigmes d'interaction pour la recherche d'information, en ouvrant l'éventail des possibilités de répartition du contrôle entre l'utilisateur et le système. Les solutions intermédiaires entre les solutions de reformulation purement manuelle et les solutions de reformulation complètement automatique peuvent être discutées dans ce contexte.

Les concepts de critère de pertinence et de statut sont suffisamment généraux pour être appliqués dans un grand nombre d'applications de recherche d'information, comme nous l'avons discuté précédemment, et ils peuvent se traduire dans l'interface de manières variées.

4.3 Extensions souhaitables du modèle

L'ensemble des classes de pertinence est défini comme les diverses combinaisons de vérification des statuts optionnels : deux classes diffèrent par les valeurs associées aux statuts optionnels. Il serait intéressant d'intégrer dans l'ensemble des classes de pertinence les diverses valeurs que peuvent prendre les critères abstraits. Cela permettrait de visualiser sur les documents l'effet produit par les divers critères abstraits, afin d'en éclairer le sens pour l'utilisateur. Cette extension ne peut se faire simplement en raison de la combinatoire qu'elle engendre : un contrôle des classes effectivement visualisées doit être instauré pour guider la consultation des classes.

Au sein d'une classe, les documents ne sont pas présentés de manière ordonnée. Il serait intéressant de les ordonner de façon à optimiser la détection de la pertinence. Nous avons évoqué ce sujet en parlant d'utilité pour la reformulation. Nous proposons de considérer qu'un utilisateur consulte au plus les 12 premiers documents d'une classe qui en contient un plus grand nombre¹. Pour rendre cette consultation la plus profitable possible, il faudrait faire apparaître en tête un échantillon représentatif de l'ensemble des documents de la classe, notamment en faisant valoir la diversité des documents compris dans la classe.

1. Ce seuil peut être fixé en fonction du type de document du corpus : un plus grand nombre pour des images examinées d'un coup d'oeil, un plus petit nombre pour des documents textuels dont il faut lire le résumé pour en évaluer la pertinence.

Partie VI

Conclusion

Chapitre 1

Bilan

L'objectif de cette thèse est de concevoir la fonction d'interrogation d'un système de recherche d'information de façon à prendre en compte les facteurs de performance des systèmes qui relèvent de l'interaction.

1.1 Contexte

L'originalité de ce travail consiste dans la prise en compte de la nature interactive du processus de recherche d'information dans la conception des fonctions internes du système. Il se présente comme un trait d'union entre les trois axes de recherche concernant la fonction d'interrogation :

- améliorer la fonction de correspondance d'un point de vue statique, selon le paradigme d'évaluation de Cranfield (améliorer rappel et précision par rapport à un ensemble de collections test) ;
- définir de nouveaux paradigmes d'interrogation en termes d'interaction entre le système et l'utilisateur (assister l'utilisateur dans l'exploration d'un corpus par exemple) ;
- concevoir la fonction de correspondance comme une fonction dynamique, qui doit s'adapter à la situation de recherche de l'utilisateur.

L'approche de conception des systèmes de recherche d'information qui ressort de notre modèle de pertinence système se situe à la croisée de ces trois axes de recherche, et constitue un trait d'union entre des avancées souvent menées indépendamment les unes des autres. Elle permet d'aborder la problématique de recherche d'information dans son ensemble, en établissant un lien formel étroit entre la fonction de correspondance (premier axe) et la dimension interactive de l'interrogation (deuxième axe). Ce lien permet d'envisager d'un point de vue formel la dynamique

du processus de recherche d'information qui intervient lors de l'utilisation effective des systèmes, ce qui fournit un cadre pour le développement formel et opérationnel de l'adaptation dynamique de la pertinence système à la situation de recherche (troisième axe).

La fonction de correspondance modélise traditionnellement la pertinence à partir des *régularités* que présente la pertinence d'un individu à l'autre. De cette manière, les performances des systèmes peuvent être améliorées dans le cadre d'applications où les utilisateurs et les utilisations qui sont faites des informations recherchées sont relativement homogènes. Dans le cas d'applications de recherche d'information destinées à une utilisation moins ciblée, l'hypothèse de travail classique qui consiste à définir la correspondance de façon unique doit être abandonnée, car la variété des utilisateurs et des utilisations des systèmes ne permet plus de négliger les variations de la pertinence liées à la situation particulière de l'utilisateur.

Les travaux de recherche qui consistent à définir un système dont la fonction de correspondance est adaptable se heurtent au problème du contrôle de l'adaptation du système. Quand le système dispose d'un éventail d'adaptation assez large, il ne dispose pas d'assez d'informations lui permettant de choisir l'adaptation appropriée. C'est pourquoi ces recherches s'orientent essentiellement dans deux directions : soit l'éventail des possibilités d'adaptation demeure réduit, ce qui limite la portée de l'amélioration, soit l'adaptation s'applique à des paramètres auxquels on peut appliquer des techniques probabilistes ou d'apprentissage, ce qui revient à nouveau d'une certaine façon à modéliser les régularités des phénomènes de pertinence.

1.2 Apports

Nous avons choisi une direction alternative qui consiste à enrichir l'interaction avec l'utilisateur, de façon à permettre au système de recueillir un plus grand éventail d'informations permettant le contrôle de l'adaptation. La difficulté que présente cette approche consiste dans la nécessité de prendre en compte conjointement des paramètres relevant de la fonction interne de correspondance et des paramètres relevant de l'utilisation interactive des systèmes.

Cela nous a amené à définir les principes souhaitables du fonctionnement interactif des systèmes en relation avec les caractéristiques souhaitables de la fonction de correspondance. Cette démarche fait de notre modèle de pertinence système un modèle orienté vers l'utilisateur et l'utilisation, qui tient pourtant compte des contraintes imposées par les conditions de modélisation de la fonction de correspondance.

Notamment, la sémantique booléenne de la combinaison des éléments du schéma de pertinence apparaît à l'utilisateur sous une forme simple dans l'interface en entrée, puisqu'elle se réduit à la conjonction, ce qui évite les dangers de formulation erro-

née fréquents dans la formulation booléenne des requêtes. La disjonction n'apparaît qu'au travers de l'interface en sortie, de manière intuitive, comme un outil pratique permettant de régler la focalisation de la réponse grâce aux classes de pertinence.

Les deux sous-fonctions de l'interrogation que sont la fonction de correspondance et la fonction de mise en forme de la réponse du système jouent des rôles d'égale importance dans notre modèle (alors que la seconde est souvent négligée dans les modèles de pertinence système), et sont fortement couplées pour permettre à l'utilisateur de connecter les données en entrée (requête et jugements de pertinence entrés par l'utilisateur) aux données en sortie du système (réponse du système). Ce couplage est réalisé à travers la visualisation des réponses selon les classes de pertinence dont la sémantique se traduit de manière immédiate sur le schéma de pertinence (équivalent de la requête). La notion de degré de pertinence est ainsi renouvelée : elle prend un sens qualitatif à la fois explicite et intuitif, grâce aux classes de pertinence définies par les statuts.

Notre modèle est conçu de façon à favoriser la réalisation des activités qui incombent à l'utilisateur lors du processus itératif de recherche, c'est-à-dire détecter la pertinence des documents, évaluer sa propre satisfaction vis-à-vis de la réponse courante, et reformuler son problème d'information s'il n'est pas satisfait.

Cette conception permet d'évaluer la capacité d'un utilisateur à exprimer des informations plus complètes sur son problème d'information (plus que des termes ou des jugements de pertinence binaires). Ces informations supplémentaires fournissent au système les éléments de contrôle de l'adaptation du système qui lui manquaient.

Nous avons limité les hypothèses relatives à l'indexation en définissant les critères élémentaires abstraits de pertinence comme une interface entre les fonctions d'indexation et d'interrogation, ce qui fait de notre modèle un modèle assez général pouvant servir de plate-forme pour l'étude de la dynamique de l'utilisation des systèmes en termes de pertinence.

1.3 Problèmes soulevés

Notre modèle soulève des problèmes en matière d'interface pour la visualisation et l'interaction avec un grand nombre d'informations coordonnées entre elles, qui requièrent un travail important de conception de l'interface. En particulier le couplage entre le schéma de pertinence et les classes de pertinence qu'il permet de définir n'est pas rendu de façon probante dans l'interface de notre prototype. D'une manière générale, le lien entre l'interface et le noyau fonctionnel des systèmes de recherche d'information interactifs constitue un sujet de recherche à part entière [MN96].

Par ailleurs, la nature qualitative du degré de pertinence que nous avons défini ne dispense pas d'ordonner les documents au sein d'une classe ou d'une sous-classe.

En effet, une classe ou une sous-classe peut très bien contenir un grand nombre de documents dans certains cas. Les fonctions de correspondance qui ordonnent les documents peuvent être utilisées telles quelles. Elles peuvent aussi être adaptées pour normaliser les mesures non pas sur l'ensemble du corpus mais sur l'ensemble des documents de la classe, ce qui resitue l'ordre dans la perspective du schéma de pertinence.

Elles pourraient aussi être conçues de façon à identifier les documents les plus dissemblables d'une même sous-classe, afin de présenter en tête un échantillon des documents de la classe qui rende compte de la diversité des documents qu'elle contient. Une tel ordre sur les documents d'une classe faciliterait la détection de la pertinence de la classe, ainsi que la tâche de reformulation en suggérant de possibles précisions à apporter à la requête. Les techniques de clustering appliquées à une classe peuvent constituer un point de départ pour définir un tel type de classement.

Chapitre 2

Perspectives

2.1 Systèmes de recherche d'information interactifs et adaptatifs

Notre modèle fournit un contexte formel pour l'étude de problèmes que soulève aujourd'hui la généralisation de l'utilisation des systèmes de recherche d'information.

La taille des corpus et leur hétérogénéité rend les performances des systèmes plus sensibles aux ambiguïtés sémantiques introduites dans la formulation du problème d'information. De plus, le public d'utilisateurs non spécialistes réclame non seulement une grande robustesse des systèmes, mais aussi une interaction à la fois simple et intelligible.

Selon notre point de vue, pour être robustes les systèmes doivent être raffinés et permettre une adaptation. Pour être intelligibles, leur adaptation doit pouvoir être expliquée. Une adaptation paramétrée du système permet de donner une explication de l'adaptation, comme c'est le cas dans notre modèle, mais cela accroît la complexité de l'interaction. Ainsi, pour satisfaire à l'exigence de simplicité, l'utilisation du système doit être guidée. Ce guide peut se présenter sous la forme d'une automatisation partielle de la reformulation, ce qui ouvre un axe de recherche pour lequel notre modèle peut servir de point de départ formel.

En effet, de la même façon que des expérimentations ont été et sont encore menées sur la pertinence indépendamment de l'utilisation d'un système pour identifier les caractéristiques de la pertinence *en soi*, notre modèle permet d'envisager des expérimentations sur la pertinence *dans le cadre de l'utilisation d'un système* de recherche d'information, dans le but d'identifier les stratégies que développent les utilisateurs.

L'objectif est d'étudier systématiquement les régularités qui interviennent dans la dynamique du processus de recherche d'information, afin d'explorer la possibilité de définir une typologie des utilisations des systèmes dans ces termes dynamiques.

C'est sur la base d'une telle typologie qu'un guide pour la reformulation peut être conçu, rejoignant ainsi la problématique d'exploration assistée d'un corpus.

Dans cette même veine, notre modèle, en fournissant des éléments formels pour la définition d'une vue système de la situation de recherche, constitue un premier pas pour une contribution à la problématique relativement neuve de l'évaluation des systèmes de recherche d'information interactifs.

2.2 Lien avec les modèles d'indexation avancés

Les critères élémentaires abstraits de pertinence dépendent fortement de l'indexation. Considérant que l'indexation d'un corpus est réalisée dans le but d'être utilisée par la fonction de correspondance d'un système de recherche d'information, un langage d'indexation doit s'accompagner de son «mode d'emploi» pour l'interrogation, c'est-à-dire de sa sémantique en matière de pertinence. L'ensemble du langage d'indexation et de sa sémantique en matière de pertinence constitue ce que l'on appelle un modèle d'indexation [Par96]. Dans cette perspective, la définition des critères élémentaires abstraits de pertinence fait partie intégrante du modèle d'indexation. L'ensemble des critères élémentaires abstraits de pertinence constitue ainsi l'interface entre la fonction d'indexation et la fonction d'interrogation.

L'interrogation ne peut s'adapter que dans la mesure où l'indexation est suffisamment riche pour permettre la définition d'un éventail assez large de critères de pertinence.

Le travail de Paradis [Par96] est représentatif du courant des progrès actuels dans le domaine des modèles d'indexation. Il définit un modèle d'indexation qui permet d'exploiter les données textuelles pour dériver les thèmes des documents avec une plus grande précision : les thèmes d'indexation sont associés à des règles de dérivation indiquant le rôle qualitatif de chaque thème dans le document. Dans ce contexte, indexation automatique n'est plus synonyme de description statistique des documents : elle inclut des informations qualitatives lui permettant d'adapter la description des documents à la perception qu'est susceptible d'en avoir un individu particulier dans sa situation.

Ainsi les modèles d'indexation évoluent vers une conception duale de la nôtre, que sous-tend l'idée d'indexation adaptable à la situation. Ce courant promet des développements qui pourraient apporter une solution au problème de l'utilisation des indexations qui comprennent des informations numériques statistiques, dans le cadre d'une interrogation adaptative. Nous rencontrons ce problème par exemple dans notre modèle, pour attribuer un ordre aux documents au sein d'une classe de pertinence.

Finalement, ces modèles d'indexation avancés, qui incluent à la fois des processus automatiques d'indexation et une sémantique riche en informations qualitatives, per-

mettent d'envisager la généralisation de la mise en œuvre du principe d'adaptation de la pertinence système à un éventail de plus en plus large d'applications.

Bibliographie

- [AK93] Avrahami (J.) et Kareev (Y.). – What do you expect when you ask for “a cup of coffee and a muffin or a croissant” ? on the interpretation of sentences containing multiple connectives. *International Journal on Man-machine Studies*, vol. 38, 1993, pp. 429–434.
- [Bar94] Barry (C. L.). – User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, vol. 45, n° 3, 1994, pp. 135–141.
- [Bel80] Belkin (N. J.). – The problem of ‘matching’ in information retrieval. *In: Theory and application of information research*, éd. par Harbo (O.) et Kajberg (L.), pp. 187–197. – London, Mansell, 1980.
- [BH94] Bruza (P. D.) et Huibers (T. W. C.). – Investigating aboutness axioms using information fields. *In: Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, éd. par Croft (W. B.) et van Rijsbergen (C. J.), pp. 112–121. – Dublin, Ireland, juillet 1994.
- [BH95] Bruza (P. D.) et Huibers (T. W. C.). – *How nonmonotonic is aboutness?* – Technical report UU-CS-1995-09, Department of Computer Science, Utrecht University, The Netherlands, mars 1995.
- [Bor84] Borgman (C. L.). – Psychological research in human-computer interaction. *In: Annual Review of Information Science and Technology*, pp. 33–64. – White Plains, NY, Knowledge Industry Publications, 1984.
- [Bru94] Bruce (H. W.). – A cognitive view of the situational dynamism of user-centered relevance estimation. *Journal of the American Society for Information Science*, vol. 45, n° 3, 1994, pp. 142–148.
- [CDBK86] Chiaramella (Y.), Defude (B.), Bruandet (M.-F.) et Kerkouba (D.). – IOTA: a full text information retrieval system. *In: Proceedings of the*

- Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 207–213. – Pisa, Italia, 1986.
- [Che92] Chevallet (J.-P.). – *Un modèle logique de recherche d'informations appliqué au formalisme des graphes conceptuels. Le prototype ELEN et son expérimentation sur un corpus de composants logiciels*. – Thèse de PhD, Université Joseph Fourier Grenoble I, 1992.
- [CK67] Cuadra (C. A.) et Katter (R. V.). – *Experimental studies of relevance judgments: Final report: Volume I: Project summary*. – Rapport technique, Santa Monica, CA, System Development Corp., 1967.
- [Coo71] Cooper (W. S.). – A definition of relevance for information retrieval. *Information Storage and Retrieval*, vol. 7, 1971, pp. 19–37.
- [Coo73] Cooper (W. S.). – On selecting a measure of retrieval effectiveness, part 1. The subjective philosophy of evaluation. *Journal of the American Society for Information Science*, vol. 24, n° 2, 1973, pp. 87–100.
- [Cro95] Croft (W. B.). – What do people want from information retrieval? – the top 10 research issues for companies that use and sell IR systems. *D-Lib Magazine*, novembre 1995. – <http://www.dlib.org/>.
- [CT84] Croft (W. B.) et Thompson (R. H.). – The use of adaptive mechanisms for selection of search strategies in document retrieval systems. *In: Third Joint BCS-ACM Symposium*. – Cambridge, 1984.
- [CvR96] Crestani (F.) et van Rijsbergen (C. J.). – *A study of probability kinematics in information retrieval*. – Deliverable n° 3, ESPRIT BRA Project No. 8134 - FERMI, janvier 1996.
- [Dan86] Daniels (P. J.). – Cognitive models in information retrieval – an evaluative review. *Journal of Documentation*, vol. 42, n° 4, 1986, pp. 272–304.
- [Dan94] Daniłowicz (C.). – Modelling of user preferences and needs in boolean retrieval systems. *Information Processing and Management*, vol. 30, n° 3, 1994, pp. 363–378.
- [DBM97] Denos (N.), Berrut (C.) et Mechkour (M.). – An image retrieval system based on the visualization of system relevance via documents. *In: Database and Expert Systems Applications, 8th International Conference, DEXA '97*, éd. par Hameurlain (A.) et Tjoa (A. M.). pp. 214–224. – Toulouse, France, septembre 1997.

- [DN86] Dervin (B.) et Nilan (M.). – Information needs and uses. *In: Annual Review of Information Science and Technology*, éd. par Williams (M. E.), pp. 3–33. – White Plains, N. Y., Knowledge Industry Publications, 1986.
- [Fuh92] Fuhr (N.). – Probabilistic models in information retrieval. *The Computer Journal*, vol. 35, n° 3, 1992, pp. 243–255.
- [GB95] Green (R.) et Bean (C. A.). – Topical relevance relationships – ii. an exploratory study and preliminary typology. *Journal of the American Society for Information Science*, vol. 46, n° 9, 1995, pp. 654–662.
- [GDCG90] Greene (S. L.), Devlin (S. J.), Cannata (P. E.) et Gomez (L. M.). – No ifs, ands or ors: a study of database querying. *International Journal on Man-Machine Studies*, vol. 32, 1990, pp. 303–326.
- [Gre95] Green (R.). – Topical relevance relationships – i. why topic matching fails. *Journal of the American Society for Information Science*, vol. 46, n° 9, 1995, pp. 646–653.
- [HBW92] Hancock-Beaulieu (M.) et Walker (S.). – An evaluation of automatic query expansion in an online library catalogue. *Journal of Documentation*, vol. 48, n° 4, 1992, pp. 406–21.
- [How94] Howard (D. L.). – Pertinence as reflected in personal constructs. *Journal of the American Society for Information Science*, vol. 45, n° 3, 1994, pp. 172–185.
- [Hui96] Huibers (T. W. C.). – *An Axiomatic Theory for Information Retrieval*. – Thèse de PhD, Department of Computer Science, Utrecht university, The Netherlands, novembre 1996.
- [Kel55] Kelly (G. A.). – *The psychology of personal constructs – Vol. 1: A theory of personality*. – New York, W. W. Norton, 1955.
- [Mar77] Maron (M. E.). – On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, janvier 1977, pp. 38–43.
- [Mar85] Maron (M. E.). – Probabilistic retrieval methods. *Progress in Communication Sciences*, vol. 5, 1985, pp. 145–176.
- [MK95] Müller (A.) et Kutschekmanesch (S.). – Using abductive inference and dynamic indexing to retrieve multimedia sgml documents. *In: Proceedings of the Final Workshop on Multimedia Information Retrieval (MI-*

- RO'95*), éd. par Ruthven (I.). – Glasgow, Scotland, septembre 1995. eWiC, electronic Workshops in Computing series.
- [MN96] Mulhem (Philippe) et Nigay (Laurence). – Interactive information retrieval systems: From user centred interface design to software design. *In: Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 326–334. – Zürich, Suisse, August 18–22 1996.
- [MT94] Müller (A.) et Thiel (U.). – Query expansion in an abductive information retrieval system. *In: RIAO 94 conference proceedings: Intelligent multimedia information retrieval systems and management*, pp. 461–480. – Rockefeller University, New York, octobre 1994.
- [MW86] MacCall (F.) et Willet (P.). – Criteria for the selection of search strategies in best-match document retrieval. *International Journal on Man-Machine Studies*, vol. 25, 1986, pp. 317–326.
- [Nie90] Nie (J.). – *Un modèle logique général pour les systèmes de recherche d'informations - Application au prototype RIME*. – Thèse de PhD, Université Joseph Fourier, 1990.
- [Par93] Park (T. K.). – The nature of relevance in information retrieval: an empirical study. *Library Quarterly*, vol. 63, n° 3, 1993, pp. 318–351.
- [Par96] Paradis (F.). – *Modèle d'indexation pour les documents textuels structurés*. – Thèse de PhD, Université Joseph Fourier – Grenoble I, Groupe MRIM - CLIPS-IMAG, 1996.
- [PD91] Prieto-Díaz (R.). – Implementing faceted classification for software reuse. *Communications of the ACM*, vol. 34, n° 5, mai 1991, pp. 88–97.
- [QF93] Qiu (Y.) et Frei (H. P.). – Concept based query expansion. *In: Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 160–169. – Pittsburgh, PA, juin 1993.
- [RJ71] Rocchio Jr. (J. J.). – Relevance feedback in information retrieval. *In: The SMART retrieval System-Experiments in Automatic Document Processing*, éd. par Salton (G.), chap. 14, pp. 313–323. – Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1971.

- [RS67] Rees (A. M.) et Schultz (D. G.). – *A field experimental approach to the study of relevance assessments in relation to document searching: Final report: Volume I*. – Rapport technique, Cleveland, OH, Case Western Reserve University, School of Library Science, Center of Documentation and Communication Research, 1967.
- [Rut96] Ruthven (I.). – *Context generation through selection of pertinent material*. – Deliverable n° 3, ESPRIT BRA Project No. 8134 - FERMI, janvier 1996.
- [Sar75] Saracevic (T.). – Relevance: a review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, vol. 39, 1975, pp. 321–343.
- [SB95] Smeaton (A.) et Berrut (C.). – *Running TREC-4 Experiments: A Chronological Report of Query Expansion Experiments Carried out as Part of TREC-4*. – Rapport technique n° CA-2095, Ireland, Dublin City University–School of Computer Applications, 1995.
- [Sch91] Schamber (L.). – Users' criteria for evaluation in multimedia information seeking and use situations. – 1991. Unpublished doctoral dissertation, Syracuse University, Syracuse, N. Y.
- [Sch94] Schamber (L.). – Relevance and information behavior. In: *Annual Review of Information Science and Technology*, éd. par Williams (M. E.), chap. 1. – Medford, N. J., Learned Information, 1994.
- [SEN90] Schamber (L.), Eisenberg (M. B.) et Nilan (M. S.). – A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management*, vol. 26, n° 6, 1990, pp. 755–776.
- [SJ81] Spark Jones (K.). – The Cranfield tests. In: *Information retrieval experiment*, éd. par Sparck Jones (K.), chap. 13, pp. 256–284. – London, Boston, Sydney, Butterworths, 1981.
- [SM83] Salton (Gerard) et McGill (M. J.). – *Introduction to modern Information Retrieval*. – New York, McGraw Hill Book Company, 1983.
- [Sma94] Smaïl (M.). – *Raisonnement à base de cas pour une recherche évolutive d'informations; prototype Cabri-n – Vers la définition d'un cadre d'acquisition des connaissances*. – Nancy, France, Thèse de PhD, Université Henri Poincaré, 1994.

- [Tay86] Taylor (R. S.). – *Value-added processes in information systems*. – Norwood, NJ, Ablex Publishing Corp., 1986.
- [TGMS95] Thiel (U.), Gulla (J. A.), Müller (A.) et Sein (A.). – Dialogue strategies for multimedia retrieval: intertwining abductive reasoning and dialogue planning. *In: Proceedings of the Final Workshop on Multimedia Information Retrieval (MIRO'95)*, éd. par Ruthven (I.). – Glasgow, Scotland, septembre 1995. eWiC, electronic Workshops in Computing series.
- [TM96] Thiel (U.) et Müller (A.). – «Why was this item retrieved?»: new ways to explore retrieval results. *In: Information retrieval and hypertext*, éd. par Agosti (M.) et Smeaton (A.), chap. 8, pp. 181–201. – Boston, London, Dordrecht, Kluwer Academic Publishers, 1996.
- [VB96] Veerasamy (A.) et Belkin (N. J.). – Evaluation of a tool for visualization of information retrieval results. *In: Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 85–92. – Zürich, Suisse, août 1996.
- [Vic59a] Vickery (B. C.). – The structure of information retrieval systems. *In: Proceedings of the International Conference on Scientific Information*, pp. 1275–1289. – Washington D.C, 1959.
- [Vic59b] Vickery (B. C.). – Subject analysis for information retrieval. *In: Proceedings of the International Conference on Scientific Information*, pp. 855–865. – Washington D.C, 1959.
- [vR92] van Rijsbergen (C. J.). – Probabilistic retrieval revisited. *The Computer Journal*, vol. 35, n° 3, 1992, pp. 291–298.
- [WB95] Willie (S.) et Bruza (P.). – User's models of the information space. *In: Proceedings of the Eighteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, éd. par Fox (E.), Ingwersen (P.) et Fidel (R.). pp. 205–209. – Seattle, WA, juillet 1995.
- [Wil73] Wilson (P.). – Situational relevance. *Information Storage and Retrieval*, vol. 9, n° 8, 1973, pp. 457–471.
- [WY89] Wong (S. K. M.) et Yao (Y. Y.). – *Information retrieval based on axiomatic decision theory*. – Rapport technique n° TR-89-985, Ithaca, Cornell University, avril 1989.

- [Yao95] Yao (Y. Y.). – Measuring retrieval effectiveness based on user preference of documents. *Journal of the American Society for Information Science*, vol. 46, n° 2, 1995, pp. 133–145.

Partie VII

Annexe

Annexe A

Agrandissements de résultats

A.1 Agrandissements – 1

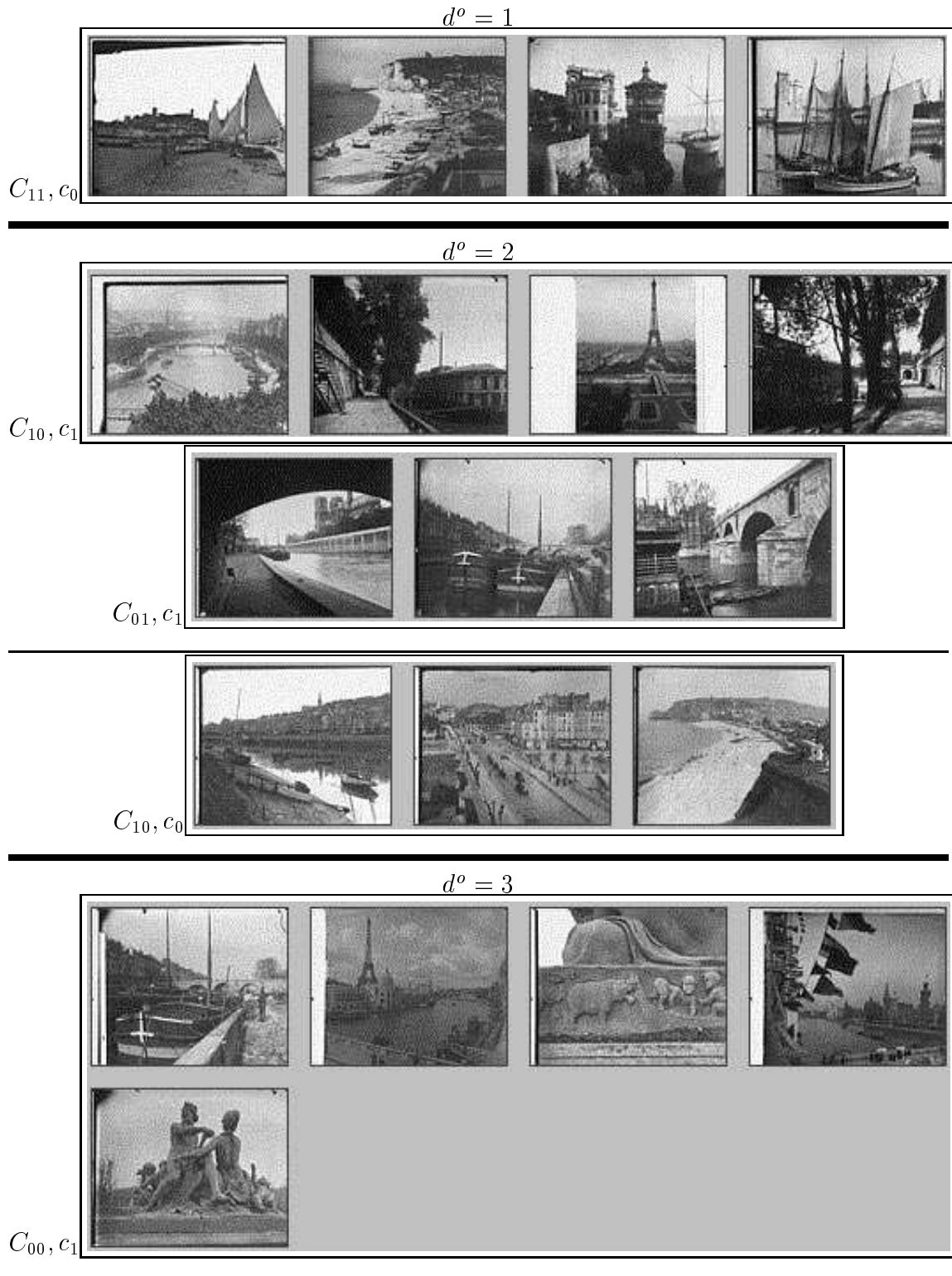


FIG. VII.A.1 – Réponse des schémas de pertinence S_6 et S_{22} (à suivre, figure VII.A.2)

($d^o = 3$, suite)

C_{00}, c_0

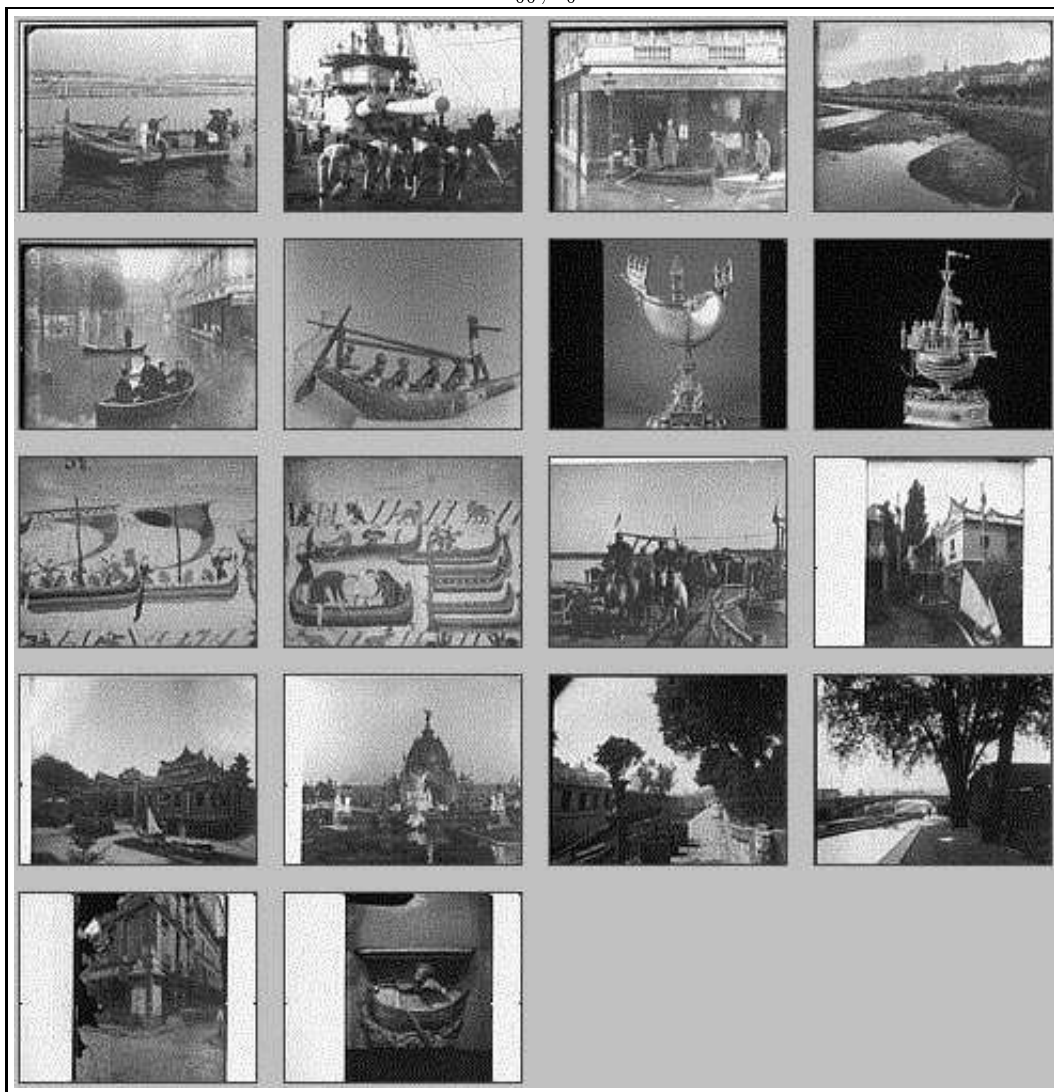


FIG. VII.A.2 – Réponse des schémas de pertinence S_6 et S_{22} (suite et fin)

(à suivre)

FIG. VII.A.3 – Réponse du schéma de pertinence S_{16} (à suivre, figure VII.A.4)

(suite et fin)

FIG. VII.A.4 – Réponse du schéma de pertinence S_{16} (suite et fin)



FIG. VII.A.5 – Réponse du schéma de pertinence S_{17} (à suivre, figure VII.A.6)

$d^o = 2$
 C_0 (à suivre)



FIG. VII.A.6 – Réponse du schéma de pertinence S_{17} (suite, à suivre figure VII.A.7)

$(d^o = 2)$
(C_0 , suite)

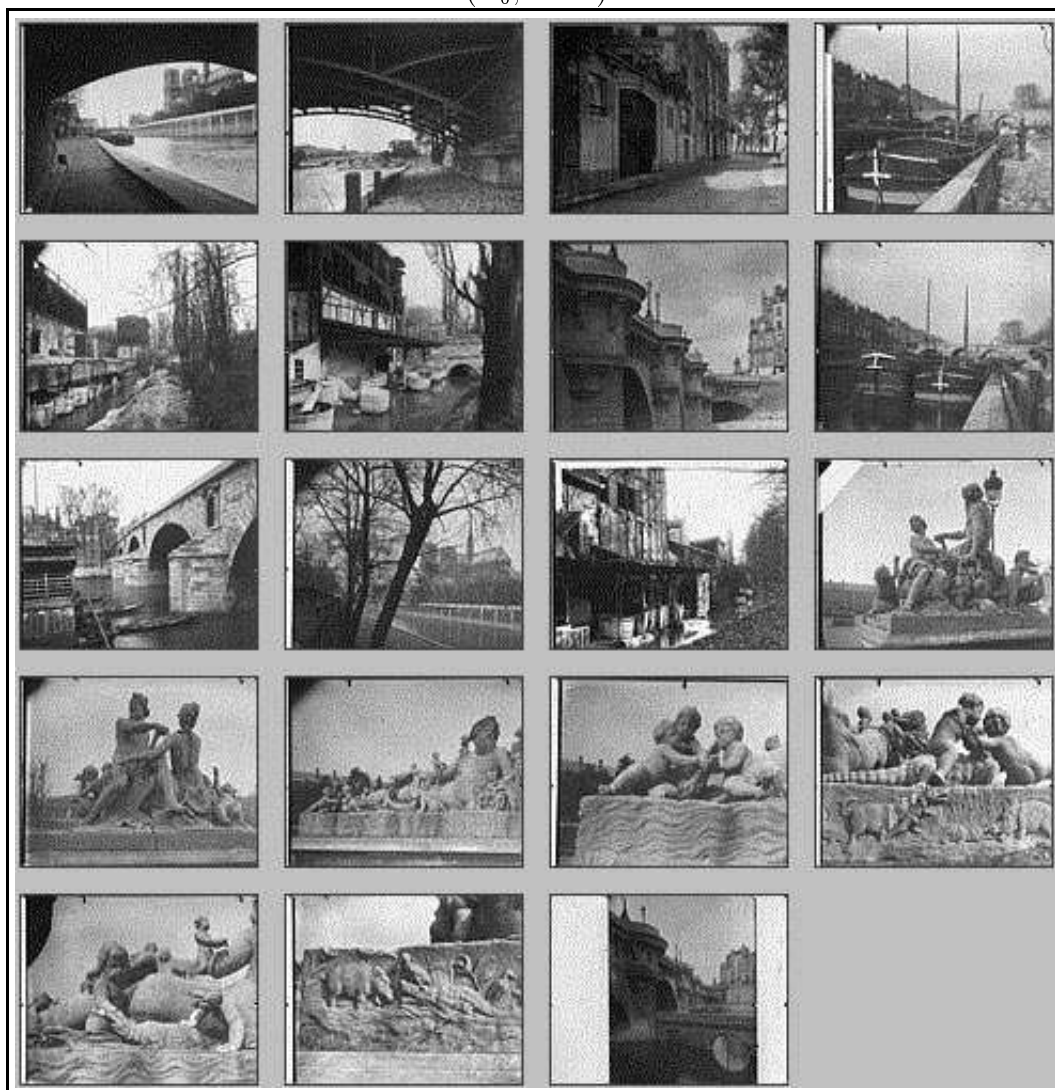


FIG. VII.A.7 – Réponse du schéma de pertinence S_{17} (suite et fin)



FIG. VII.A.8 – Réponse des schémas de pertinence S_{18} et S_{21} (à suivre, figure VII.A.9)

$d^o = 2$
 C_0 (à suivre)



FIG. VII.A.9 – Réponse des schémas de pertinence S_{18} et S_{21} (suite, à suivre figure VII.A.10)

($d^o = 2$)
(C_0 , suite et fin)



FIG. VII.A.10 – Réponse des schémas de pertinence S_{18} et S_{21} (suite et fin)

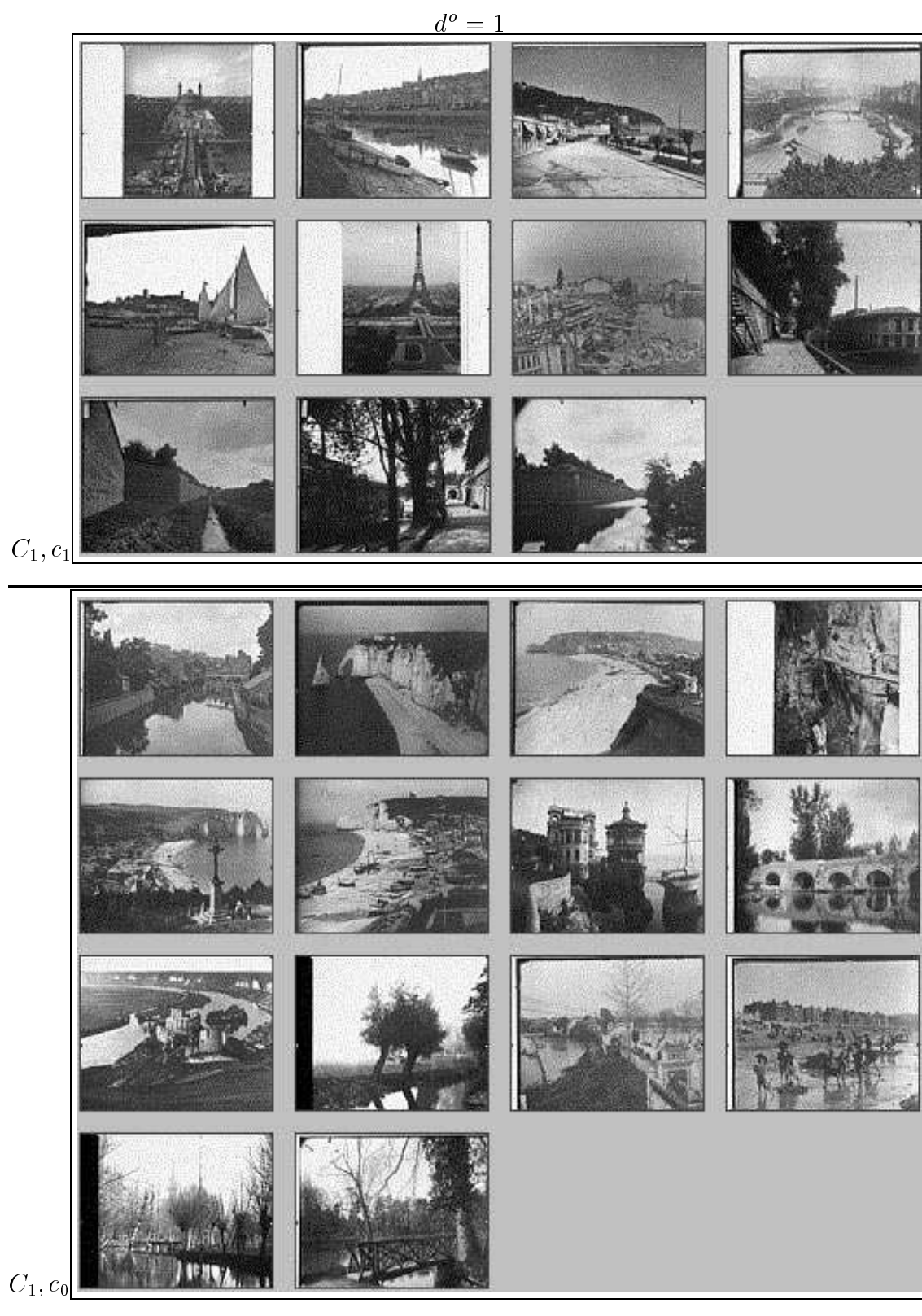


FIG. VII.A.11 – Réponse des schémas de pertinence S_{19} et S_{20} (à suivre, figure VII.A.12)

$d^o = 2$
 C_0 (à suivre)



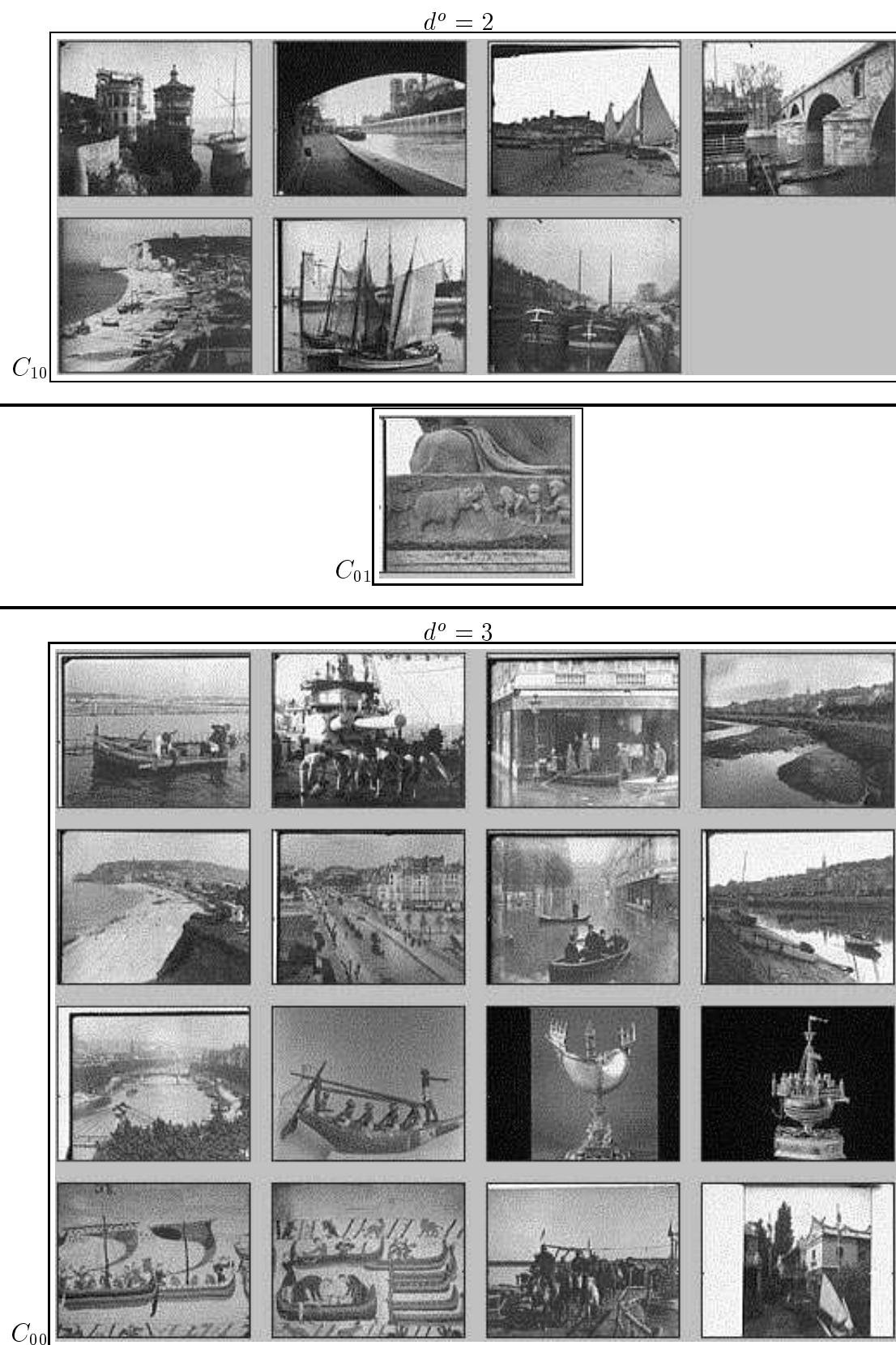
FIG. VII.A.12 – Réponse des schémas de pertinence S_{19} et S_{20} (suite, à suivre figure VII.A.13)

($d^o = 2$)
(C_0 , suite et fin)



FIG. VII.A.13 – Réponse des schémas de pertinence S_{19} et S_{20} (suite et fin)

A.2 Agrandissement – 2

FIG. VII.A.14 – Résultat du schéma de pertinence S_2 (à suivre, figure VII.A.15)

($d^o = 3$)
 (C_{00} , suite et fin)

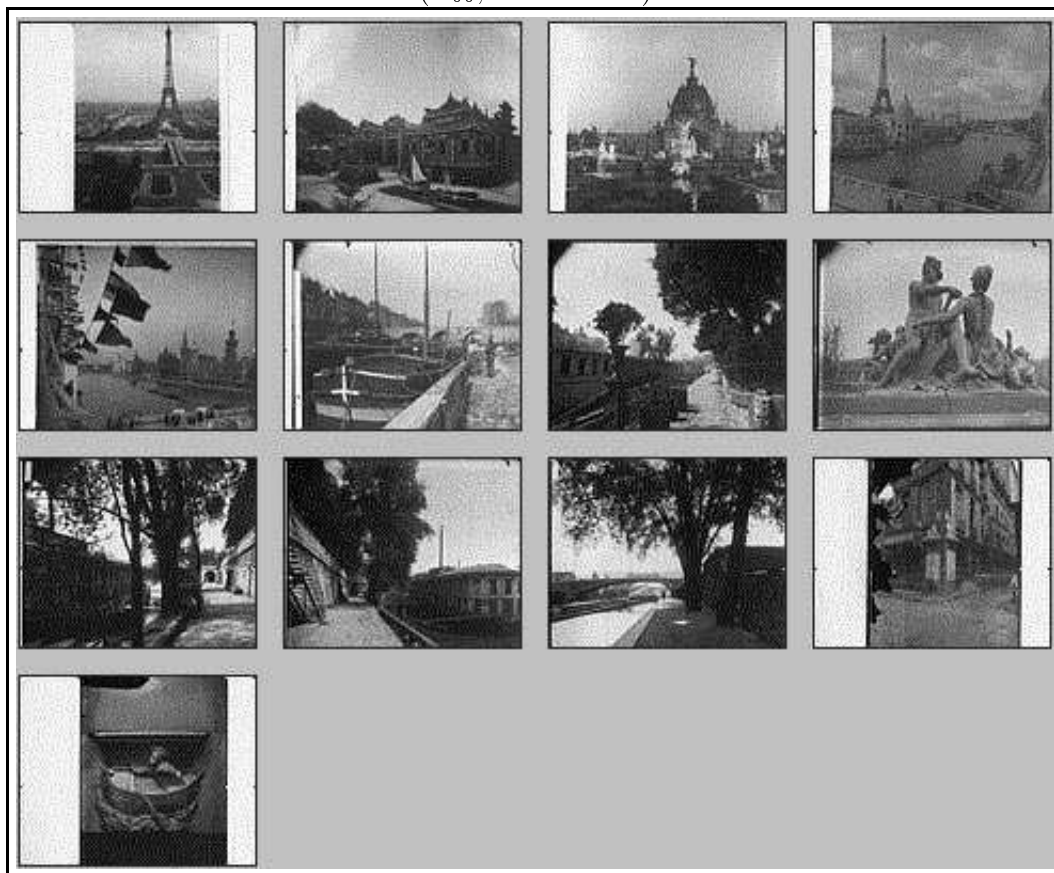


FIG. VII.A.15 – *Résultat du schéma de pertinence S_2 (suite et fin)*